申报	系列：专任教师
	专业：农业电气化与自动化
	职称：副教授

业绩成果材料

（申报人的业绩成果材料包括论文、科研项目、获奖以及其他成果等）

单　位（二级单位）_华南农业大学数信、软件学院

姓　名　涂淑琴

材料核对人：　　单位盖章：

核对时间：

华南农业大学制

目　录

五、其他业绩

【佐证材料切记与目录页所列页码对应，不要用图片格式的材料进行打印。】

1.2关于“2021百个专任教师党支部思政精品课-数据库应用” 立项通知（合同）及有关佐证材料

附件 2

华南农业大学百个专任教师党支部
课程思政精品示范课建设项目验收登记表

课程类别： **公共基础课程**

课程名称： **数据库应用**

支部名称： **教工计算机科学党支部**

项目负责人： **涂淑琴**

项目参与人：

（不含项目负责人） 徐锐，万华，方凤美，张春玲，杨磊，张丽霞

立项时间： **2021 年 5 月 7 日**

填表时间： **2022 年 5 月 23 日**

华南农业大学党委组织部

华南农业大学本科生院 制

二 O 二二年

一、项目既定建设举措执行情况

1.已完成的既定建设举措：以项目申报表为参照，分条列举已经执行和落实的主要建设（改革）举措(步骤、计划、措施等)（800 字以内），需提供相关证明材料。

数据库技术是计算机信息系统与应用的核心技术和重要基础。《数据库应用》课程的教学目标是使非计算机专业学生系统地掌握数据库系统的基本原理和工作机制，掌握数据库设计方法和步骤，具备设计数据库模式及开发数据库应用系统的基本能力。该门课程的主要特点是理论与实践联系非常密切，需要培养学生具有较强的计算思维能力及运用数据库专业知识解决实际工程问题能力。通过课程的学习，使学生理解不忘初心，砥砺前行，勇挑重担，才能成为大国工匠。

项目以习近平新时代中国特色社会主义思想为指导，以党的十九大精神为指引，深入学习贯彻全国全省高校思想政治工作会议精神，深入挖掘专业核心课程的育人功能，落实立德树人根本任务，把思想政治工作贯穿《数据库应用》教学全过程。该课程教学中的思政实施步骤主要根据项目申报书分为三步：

（1） 已经执行融合课程内容挖掘思政元素。例如在讲解数据库技术发展历程中，讲述我国数据库 40 年的发展历程，探讨科技的发展，传递科技强国的思想，培养学生的家国情怀和使命担当；在关系数据库查询和 SQL 语句时，讲述查询的多种实现方法；百度、谷歌等搜索引擎的查询效率，培养学生养成认真负责的工作态度、仔细的工匠精神和求真务实的科学精神；在讲解数据库设计，分析数据库设计步骤及方法，讨论好的数据库的设计标准，培养学生的创新能力、抗压能力和团队合作能力。已经设计 5 个学时课程教案，包含了这些课程内容及对应的思政元素。

（2） 已经按照不同时代的学生特点进行科学引导。针对 00 后学生，思想活跃，讲课以幽默诙谐的语言、润物细无声的方式自然和谐地向学生传递社会主义核心价值观，给学生带来真实的体验感与获得感，关心学生，与学生多沟通，加强学生政治信仰、社会责任、价值取向的教育。

（3） 已经结合课程教学创新方法进行思政元素的灌输。已经采用线上线下、雨课堂和腾讯会议新型教学方式，培养学生爱国主义、职业道德规范、精益求精的工匠精

神和科学思维和自主创新精神。

项目已经在校内开展了半年的实践应用（2022.2-2022），运行效果良好，附件已提供相关证明材料。

2.未完成的既定建设举措：项目申报表已设定的，但目前尚未实施或者未完全实施的建设(改革)举措，分条列举，并说明未执行相应建设举措的原因。（300 字以内）

无

二、项目预期成果达成情况

1.已完成的预期成果清单：以项目申报表为参照，分条列举已经完成的与本项目密切相关的主要建设成果清单（500 字以内），并附成果证明材料。

以项目申报表为参照，已经完成的与本项目密切相关的主要建设成果清单如下:

（1） 修订的数据库应用的课程教案

根据本门课本每章的课程内容，插入对应的思政元素，新建一份修订的课程教案。

（2） 1 次课程思政公开课佐证

在 2022 年 6 月 7 号上午，本人在院楼 618 开展关于数据库融合思政的公开课，由数信学院组织。

（3） 设计了 4 个典型教学案例

第一个为：在讲解数据管理技术发展阶段，探讨科技的发展，传递科技强国的思想，培养学生的家国情怀和使命担当；

第二个为：讲解数据库关系查询时，通过引入百度、谷歌等搜索引擎的查询效率，传递严谨的工作学习态度，一丝不苟的工匠精神；

第三个为：讲解数据库完整性，通过从学校的学风建设、制度管理探讨制度约束的重要性，进而讨论数据库的完整性约束，引导学生认识制度约束的重要性，强化学风建设；

第四个为：讲解数据库设计，通过分析数据库设计步骤及方法，讨论好的数据库的设计标准，培养学生的创新能力、抗压能力和团队合作能力。

2.未完成或超期完成的成果清单：项目申报时设定，但尚未完成的建设成果，分条列举并说明原因。项目申报时未设定，但目前超出预期完成的建设成果（必须与项目建设直接相关），分条列举（300 字以内），并附成果证明材料。

无

三、项目建设成果应用、推广、示范

项目主要建设成果在校内外的实践应用情况、推广共享情况和对教学改革的示范促进作用(600 字以内)，需附实证或证明材料。

项目主要建设成果在校内开展了半年的实践应用，时间为 2022.2-2022.6。开课学生为 2021 级艺术学院广电编导和环境设计专业共计 170 名。具体地，其推广共享情况和对教学改革的示范促进作用如下：

（1）在课程讲解中，通过课程内容引入具体的实例，同时结合当前热点问题和身边发生的事件，用学生熟悉的语言、感兴趣的方式将思政教育无缝衔接到教学中，让学生在认知、情感、行为上产生共鸣，更有说服力。之后，发现学生们更加认真上课，对于知识点的吸收也更好。

（2）在理论教学过程中，运用多样化的教学方法和手段，如在课堂上通过腾讯视频、案例讲解、情境教学、雨课堂弹幕投稿、课堂讨论、回答问题等多种教学方法手段，以思政内容融入并牵引专业知识，运用多媒体和课堂互动讨论，鼓励自由发言，提出创新问题，增加课堂参与度与趣味性，把政治性过强的内容转为现实案例研讨，在知识传授中弘扬民族精神，做到以思政促知识、促能力，促进整体教学效果的提升。

（3）在课程实践与实验中，针对学生不同层次、不同阶段的学习，采用了多层次、递进式的数据库实验教学体系，精心设计教学案例。在实践中，让学生组队完成数据库综合创新应用实验，该部分内容要求学生以小组为单位完成，成员之间分工协作，扮演不同的角色，合作完成系统的开发。在这个过程中，学生之间相互学习讨论，培养了学生探索研究能力及团队作能力，形成独立自主思考，探索解决方案。

材料包括：2021 年 9 月以副主编出版了《数据库技术及应用教程》和《数据库技术与应用教程上机指导与习题》（第 3 版）

四、其他需要说明的问题及后续建设规划

分析目前项目建设仍然存在的主要未解决的问题及对策，填写后续建设设想或应用推广计划等（500 字以内）

在目前项目建设中，仍然存在的主要未解决的问题就是本门课程内容较多，学时较短，在讲解思政内容时，会不知不觉进行思政元素拓展，导致专业内容没有正常讲解完。因此，在后续课程建设设想中，思索如何平衡专业知识点和思政内容的讲解，让学生在思政内容中更好地吸收专业知识点。

五、项目经费使用情况

（请具体列出项目经费收入细目和项目支出细目）
项目支出明细：
（1） 打印资料：1000 元

项目负责人签章：涂淑琴
2022年 6 月 27 日

六、项目结题初评专家及意见

	序号	姓名	职称/职务	所在单位	联系方式
结题评审专家信息（专家至少3人，含1名校外专家）	1	曾署才	教授/副院长	本科生院	13427633688
	2	金华	教授	华南师范大学	13570103286
	3	肖政宏	教授	广东技术师范大学	13726897598
	4	陈洁	副教授	马克思主义学院	13512704126
专家组意见（300字以内）	该项目以习近平新时代中国特色社会主义思想为指导，以党的十九大精神为指引，深入学习贯彻全国全省高校思想政治工作会议精神，深入挖掘专业核心课程的育人功能，落实立德树人根本任务，已经把思想政治工作贯穿《数据库应用》教学全过程。项目完成数据库应用课程的新教案，已开展一次公开思政课讲授，并设计了4个典型教学案例。在校内开展了半年的实践应用，效果良好。项目负责人以副主编在2021年9月出版了《数据库技术及应用教程》和《数据库技术与应用教程上机指导与习题》两本教材，取得了不错的成果。项目存在的主要问题就是该门课程内容较多，学时较短，在讲解思政内容时，会不知不觉进行思政元素拓展，导致专业内容没有正常讲解完。 总体来看，本项目研究按计划执行，研究成果达到了规定目标。建议通过验收。 2022 年 6 月 28 日				

— 6 —

项目结题初评专家及意见

项目名称：《数据库应用》课程思政精品示范课建设

	序号	姓名	职称/职务	所在单位	联系方式	签名
结题评审专家信息（专家至少3人，含1名校外专家）	1	曾曙才	教授/副院长	本科生院	13427633688	
	2	金华	教授	华南师范大学	13570103286	
	3	肖政宏	教授	广东技术师范大学	13726897598	
	4	陈洁	副教授	马克思主义学院	13512704126	
专家组意见（300字以内）	（需将项目建设任务执行情况、成果完成情况、成果实践应用情况、项目创新点、项目建设存在的主要问题、改进建议等具体说明，并给出总体评价，请附专家结题时签名原始材料） 该项目以习近平新时代中国特色社会主义思想为指导，以党的十九大精神为指引，深入学习贯彻全国全省高校思想政治工作会议精神，深入挖掘专业核心课程的育人功能，落实立德树人根本任务，已经把思想政治工作贯穿《数据库应用》教学全过程。项目完成数据库应用课程的新教案，已开展一次公开思政课讲授，并设计了4个典型教学案例。在校内开展了半年的实践应用，效果良好。项目负责人以副主编在2021年9月出版了《数据库技术及应用教程》和《数据库技术与应用教程上机指导与习题》两本教材，取得了不错的成果。项目存在的主要问题就是该门课程内容较多，学时较短，在讲解思政内容时，会不知不觉进行思政元素拓展，导致专业内容没有正常讲解完。 总体来看，本项目研究按计划执行，研究成果达到了规定目标。建议通过验收。 2022年 6月 28日					

七、项目所在单位审核意见

该项目完成了课程思政精品课的预期建设目标，并在学院开展了一次思政课的示范讲课，达到了结题要求，同意结题。

单位负责人签章：

2021年 6 月 29 日

八、学校主管部门审核意见

负责人签章：

公章：

年　　月　　日

1.3关于“2022年度华南农业大学课程思政-示范课堂” 立项通知（合同）及有关佐证材料

华南农业大学
课程思政示范课堂申报书

学院名称（盖章）数学与信息学院

所属课程名称 数据库应用

课堂所在章节 第 7 章 7.3 节 SQL 语言查询

课堂主讲人 涂淑琴

申报日期 2022.9.15

本科生院制

二〇二二年

附件：

2022年省级课程思政改革示范拟推荐项目和校级课程思政示范项目认定结果名单

序号	推荐类型	所属学院	项目负责人	备注
1	课程思政试点学院	资源环境学院	周艳华	拟认定校级
2	课程思政试点学院	经济管理学院	谭莹	拟认定校级
3	课程思政试点学院	兽医学院	鲍金勇	拟认定校级

序号	推荐类型	项目名称	所属学院	项目负责人	备注
1	课程思政示范团队	动物生物化学与饲料生物技术教学团队	动物科学学院	习欠云	拟认定校级，同时拟推荐省级
2	课程思政示范团队	数学建模课程思政示范团队	数学与信息学院、软件学院	房少梅	拟认定校级，同时拟推荐省级
3	课程思政示范团队	电子信息工程专业基础课程思政教学团队	电子工程学院 （人工智能学院）	宋淑然	拟认定校级，同时拟推荐省级
4	课程思政示范团队	食品微生物学课程思政示范团队	食品学院	钟青萍	拟认定校级
5	课程思政示范团队	无机及分析化学课程思政示范团队	材料与能源学院	刘英菊	拟认定校级
6	课程思政示范团队	压花艺术教学团队	园艺学院	陈国菊	拟认定校级
7	课程思政示范团队	刑事法实践教学团队	人文与法学学院	杜国明	拟认定校级
8	课程思政示范团队	水产专业基础课程思政教学团队	海洋学院	王梅芳	拟认定校级

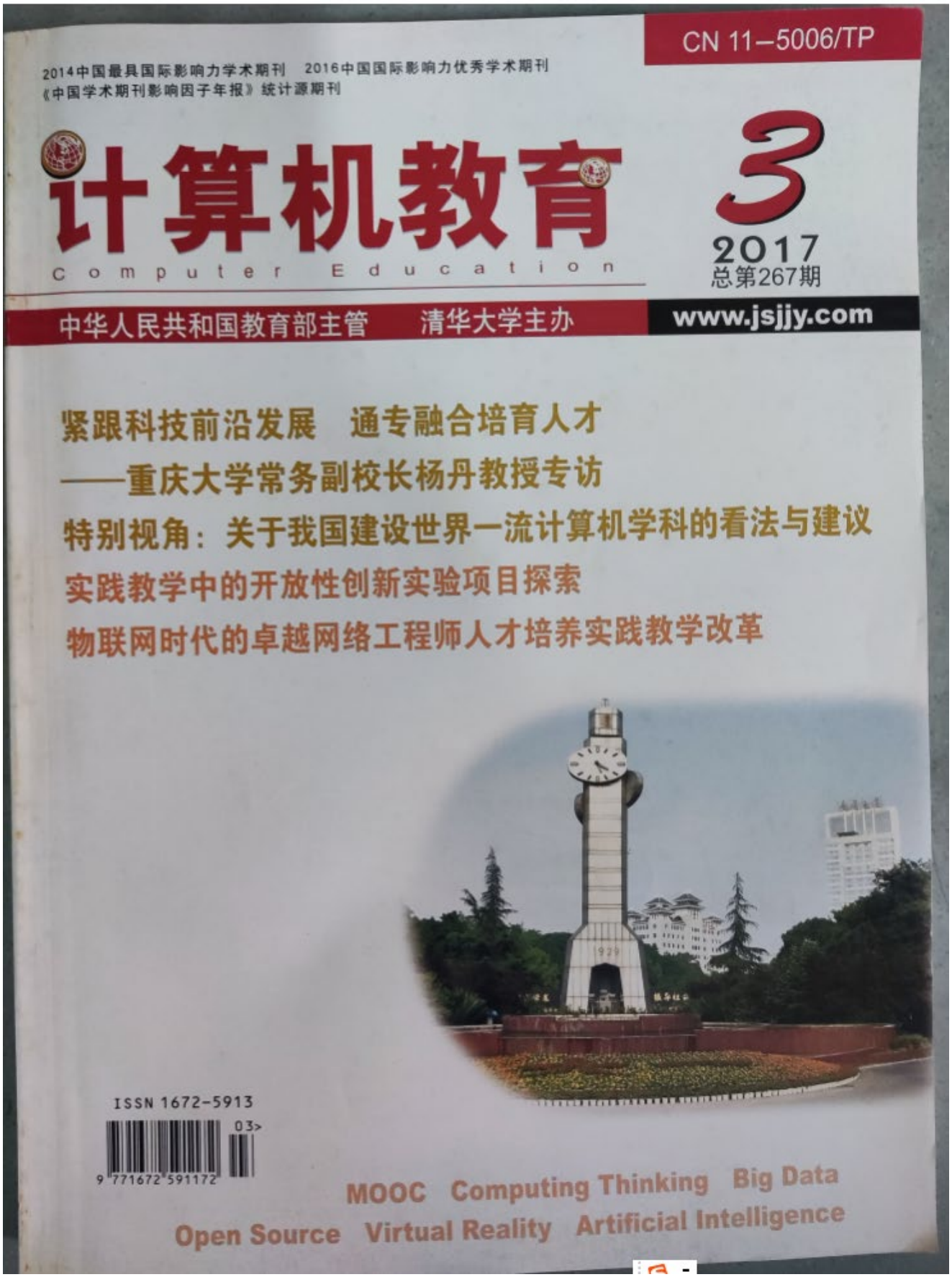
40	课程思政示范课堂	国兰鉴赏	基础实验与实践训练中心	羊海军	拟认定校级
41	课程思政示范课堂	刑事诉讼的基本原则	人文与法学学院	王琳	拟认定校级
42	课程思政示范课堂	《化学生态学》课程思政的建设与实践	资源环境学院	黎平	拟认定校级
43	课程思政示范课堂	茶树栽培学	园艺学院	晏嫦妤	拟认定校级
44	课程思政示范课堂	《水产动物组织胚胎学》课程示范课堂	海洋学院	范兰芬	拟认定校级
45	课程思政示范课堂	《大学数学 I》	数学与信息学院、软件学院	李凤	拟认定校级
46	课程思政示范课堂	现代汉语课程示范课堂	人文与法学学院	李颖	拟认定校级
47	课程思政示范课堂	兽医外科学	兽医学院	周沛	拟认定校级
48	课程思政示范课堂	数据库应用	数学与信息学院、软件学院	涂淑琴	拟认定校级
49	课程思政示范课堂	“文明互鉴·国际视野·文化自信”--《流行文化中的世界英语》的课程思政教学课堂	外国语学院	吴章欣	拟认定校级
50	课程思政示范课堂	电路	电子工程学院 （人工智能学院）	张 霞	拟认定校级
51	课程思政示范课堂	太平天国运动第 1 节课	人文与法学学院	杨品优	拟认定校级
52	课程思政示范课堂	结构力学思政课堂	水利与土木工程学院	黄欢	拟认定校级
53	课程思政示范课堂	《设计材料与加工工艺》课程思政示范课堂	工程学院	金鸿	拟认定校级
54	课程思政示范课堂	土地利用规划学	公共管理学院	刘光盛	拟认定校级

54	课程思政典型案例	兽医外科学课程思政课程思政典型案例-疝	兽医学院	贾坤	拟认定校级
55	课程思政典型案例	融入思政元素的《动物育种学》绪论	动物科学学院	袁晓龙	拟认定校级
56	课程思政典型案例	家蚕的养殖与抽丝	动物科学学院	黄志君	拟认定校级
57	课程思政典型案例	将思政元素融入《蚕桑概论》课程教学中的探索	动物科学学院	杨婉莹	拟认定校级
58	课程思政典型案例	物理光学	电子工程学院 （人工智能学院）	杨初平	拟认定校级

2.教改论文：

（1）普刊1：全国计算机等级考试二级Access的要点研究与分析

（涂淑琴，万华，张春玲,全国计算机等级考试二级Access的要点研究与分析[J], 计算机教育, 2017(3), 133-135.）



计算机教育
Jisuanji Jiaoyu

2017年3月10日 第3期 总第267期

2003年创刊

主管 中华人民共和国教育部
主办 清华大学

编辑出版发行：《计算机教育》杂志社
社址：北京市海淀区双清路学研大厦B座606室
邮编：100084 传真：(010) 62770175-3405
编辑部电话：(010) 62770175-3402—3406
广告营销：(010) 62770175-3418
杂志社邮箱：jsjjy@vip.163.com
网址/投稿平台：http://www.jsjjy.com
刊号：ISSN 1672-5913 CN 11-5006/TP
邮发代号：80-171

广告经营许可证号：京海工商广字第0368号
印 制：北京地大彩印有限公司
定 价：30.00元

目 次

扫一扫
或搜索jsjjyzz
微信也精彩

本刊为
中国知网数据库(CNKI)全文收录期刊
中国期刊全文数据库(GJFD)全文收录期刊
中国学术期刊综合评价数据库(CAJCED)统计源期刊
中国重要会议论文全文数据库(CPCD)收录期刊
万方数据—数字化期刊群全文收录期刊

计算机教育

2017年度协办单位

清华大学计算机科学与技术系	书 记	孙茂松
北京航空航天大学软件学院	院 长	吕卫锋
天津市大学软件学院	院 长	蒋秀明
北京林业大学信息学院	院 长	陈志泊
哈尔滨工业大学计算机科学与技术学院/软件学院	院 长	王亚东
复旦大学软件学院	副院长	赵一鸣
国防科学技术大学计算机学院	副院长	张春元
南海东软信息技术职业学院	院 长	杨 利
河南理工大学计算机学院	院 长	贾宗璞
西安邮电大学计算机学院	院 长	王忠民
大连理工大学软件学院	院 长	罗钟铉
北京邮电大学软件学院	执行院长	邝 坚
北京交通大学软件学院	院 长	卢 苇
苏州大学计算机科学与技术学院	院 长	杨季文
北京交通大学计算机学院	副院长	于双元
北京理工大学计算机学院	院 长	黄河燕
北京工业大学计算机学院	副院长	王 丹
上海交通大学软件学院	常务副院长	胡 飞
黄淮学院信息工程学院	院 长	耿红琴
华南理工大学软件学院	院 长	王振宇
武汉科技大学计算机科学与技术学院	副院长	符海东
湖北工业大学计算机学院	院 长	王春枝
哈尔滨理工大学计算机科学与技术学院	院 长	陈德运
深圳大学计算机与软件学院	常务副院长	明 仲
郑州轻工业学院计算机与通信工程学院	院 长	甘 勇

特别支持单位

清华大学出版社计算机与信息分社	卢先和
中国铁道出版社教材研究开发中心	严晓舟
Intel Cooperation	Elizabeth Eb
北京永信至诚科技股份有限公司	蔡晶晶
机械工业出版社计算机分社	胡毓坚
北京华章图文信息有限公司	周中华
合肥智圣系统集成有限公司	王 浩


清華大學出版社
开发利器
大数据时代的
THE DEVELOPMENT OF BIG DATA ERA
Scala
开发快速入门
从零开始学
Storm
Spark MLlib
机器学习实践
每天5分钟
玩转OpenStack
Hadoop+Spark
Python
网络爬虫实战
基于Apache Kylin
视觉机器学习
20讲

文章编号：1672-5913(2017)03-0133-03
中图分类号：G642

全国计算机等级考试二级 Access 的要点研究与分析

涂淑琴，万　华，张春玲
（华南农业大学 数学与信息学院，广东 广州 510642 ）

摘　要：二级 Access 作为全国计算机等级考试科目之一，现已成为华南农业大学非计算机专业学生参与人数最多的计算机水平认证考试。为提高 Access 课程的教学质量，在分析二级 Access 考试题型分布基础上，总结客观选择题的要点及分值分布，分析 3 个层次上机操作题的重点与难点内容，并通过实例研究对应的实用策略。

关键词：全国计算机等级考试；客观选择题；上机考试

1　背　景

全国计算机等级考试 (National Computer Rank Examination，简称 NCRE) 是经原国家教育委员会 (现教育部) 批准，由教育部考试中心主办，用于考查应试人员计算机应用知识与能力的全国性计算机水平考试体系，其合格证书全国通用，是持有人计算机应用能力的证明，供用人单位参考 [1]。教育部考试中心将 Access 数据库列为全国计算机等级考试二级考试的内容之一，作为数据库应用方向的考核科目，该科目也是华南农业大学非计算机专业学生参与人数最多的考试之一，引起广大师生的广泛关注及重视。然而，面向非计算机专业开设的计算机类课程数量少、学时短，如何在有限的教学时间内提高该考试通过率，是非计算机专业数据库人才培养的新挑战。

2　二级Access考试题型分布

二级 Access 科目考试时间为 120 分钟，考试形式为机试，考试环境为 Windows 7 + Access 2010，考试题型及分值分配见表 1。

表 1　二级 Access 的考试题型及分值分配

试题类型	内容	分值
客观选择题	公共基础知识	10
	Access 基础知识	30
操作题	基本操作题	18
	简单应用题	24
	综合应用题	18

在二级 Access 数据库科目的考核中，客观选择题占 40 分。其中，公共基础知识占 10 分，主要内容包括基本数据结构与算法、程序设计基础、软件工程基础；另外 30 分是 Access 的基础知识，包括数据库基础理论、关系运算、SQL 查询语句以及 Access 的七大对象（表、查询、窗体、报表、页、宏和 VBA）概念和知识 [2-3]。

操作题占 60 分，分为 3 类题型：基本操作题、简单应用题和综合应用题。基本操作题主要考查数据库中创建表的操作，通常占 18 分；简单应用题主要考查 SQL 查询与窗体的结合等操

基金项目：华南农业大学数学与信息学院教育教学改革与研究项目（SXJG201503，SXJG201506）；华南农业大学教育教学改革与研究一般项目（JG15074）。
第一作者简介：涂淑琴，女，讲师，研究方向为数据库技术、图像处理和智能算法，tushuqin@163.com。

作，通常细分为 4 小题，占 24 分；综合应用题难度最大，主要考查窗体、报表、宏和 VBA 四大对象之间的综合应用，在该类题中，难点是结合窗体中的按钮与宏和 VBA 操作实现具体功能，占 18 分。

3 客观选择题

客观选择题包含公共基础知识和 Access 基础知识，可分成以下两类进行研究。

3.1 公共基础知识要点

该部分内容是数据结构、程序设计基础和软件工程的基本知识。其中，数据结构主要考算法、栈、队列、二叉树和排序的基本概念，要点包含算法复杂度求解、栈和队列的特点、线性结构与非线性结构、二叉树的结点个数及遍历排序和二分查找。队列和栈的特点经常作为考点，例如栈操作遵循后进先出的特点，队列具有先进先出的特点。二叉树方面主要考察求解二叉树的深度。程序设计基础主要考核程序设计方法、对象、属性、继承、多态等概念。软件工程基础包括软件生命周期、软件设计、程序测试与调试方法。公共基础课程共占 10 分，分值少且内容分布广泛，不能投入大量时间研究所有知识点，可以研究近 5 年的真题命题规律，挖掘其重要知识点及常考点进行高效复习。

3.2 Access 基础知识

Access 基础知识包括数据库基础理论，主要有数据库概念设计、关系运算、SQL 查询语句以及 Access 的 7 种对象的概念和知识。数据库概念设计考核内容主要是 E-R 图模型，包括实体、联系、属性等。关系运算考核内容主要包括选择、投影、自然连接和除法操作。Access 基础知识内容多且庞杂，笔者统计了 5 年共 50 套题的数据并进行分析，Access 基础知识的考题分布见表 2。考点集中在 SQL 查询和 VBA 知识点上，其中 SQL 查询约占 6 分，VBA 内容约占 11 分，这两个知识点在 Access 基础知识方面占据 50% 以上分值，因此要重点对待。

表 2 Access 基础知识分值分布

知识点	分值分布 / 分
表	4
查询及 SQL 语句	6
窗体	3
报表	3
宏	3
VBA	11

4 操作题的要点研究与总结

二级 Access 数据库应用的操作题主要考核学生能否正确处理数据，要求学生掌握数据库的基本操作技能，掌握窗体、报表、宏和 VBA 的操作技能，并熟练应用这些技术实现数据库综合应用 [4]。根据操作题的 3 个层次，笔者把操作要点分成以下 3 类进行研究。

4.1 基本操作题

这部分内容考核数据库中表的操作，包括数据表视图和设计视图。数据表视图的要点是数据格式设置，包括字体大小、行高、隐藏、冻结列、导入数据等。表设计视图的操作包括表的数据类型、主键、关联关系、格式、输入掩码、有效性规则、增删修改表字段等。

这部分操作相对简单，通过随机统计近两年 50 套题的数据，考题分布见图 1，其中主键设置、格式设置、输入掩码和有效规则是基本操作的常考内容。

4.2 简单应用题

该部分考核查询的操作，查询类型有 5 种：选择查询、交叉表查询、参数查询、操作查询和 SQL 查询。选择查询是最常考的查询，根据指定条件，从一个或者多个数据源中获取数据并显示结果。交叉表查询主要将查询的字段分成两组，一组以行标题的方式显示在表格的左边，一组以列标题的方式显示在表格的顶端，在行和列交叉的地方对数据进行综合、平均计数或者其他类型

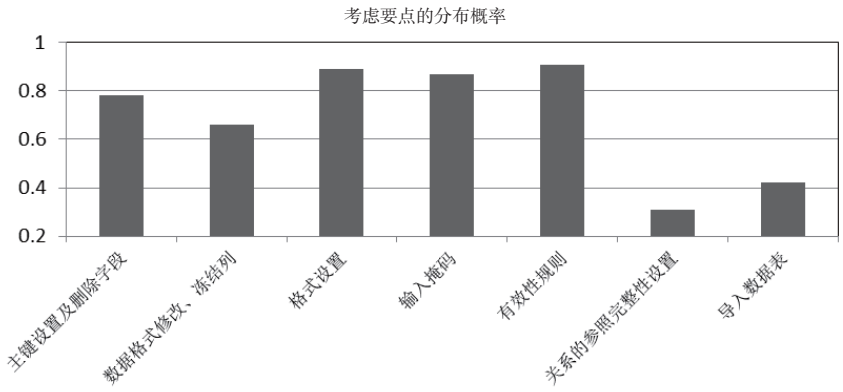


图 1　基本操作题要点分布概率

的计算，实现简单数据统计功能 [5]。参数查询在执行时会弹出提示框输入参数信息，可查询与该参数相关的所有记录。操作查询是需要运行才能执行的查询，可以根据指定条件生产新表、更新数据表、追加数据表、删除数据表操作。SQL 查询是用 SQL 语句创建的查询，上面所有查询功能都可以用 SQL 语句实现。

查询的考查包括查询的建立和查询设计器的设置。查询的建立最好使用设计视图创建。查询设计器包括表的字段名称、表名称、排序、是否显示字段和条件。其中条件的设置比较复杂，包括单个条件表达式、模糊查找条件、配合逻辑运算符进行组合条件查找、在查询的条件中使用函数等。

4.3　综合应用题

综合应用题主要考核窗体、报表、宏和 VBA 操作的融合。窗体与报表的控件设置是必考内容，控件分为绑定控件、未绑定控件和计算控件 3 类。常考的几个控件为：窗体与报表，包括修改窗体标题、记录选择器、浏览按钮和分隔线等设置；在窗体与报表页眉区添加标签控件，进行修改其名称、标题、字体、字大小、加粗等操作；在窗体报表页脚区添加按钮控件，修改名称、标题、按钮上文字的颜色和文字加粗，设置单击按钮事件属性为某个宏、设置某单选按钮默认值属于 0。

另外，报表常考点为页脚添加函数与计算控件，显示时间和页码。当设置计算控件时要注意设置控件格式，例如，在报表页脚添加一个计算控件，输出页码，其显示格式为“当前页 / 总页数”，应添加一个文本框，在文本框中或者其控件来源输入公式：=[page]&”/”&[pages]。宏操作主要考察将窗体或命令按钮的单击事件属性设定为给定的宏对象，例如已有宏对象“m1”，将窗体中名为“Command1”的命令按钮的单击事件属性设定为宏对象“m1”。

5　结　语

随着计算机等级考试的不断发展和变化，在教学中教师要不断研究知识点并总结新教学方法，深入研究教学知识点和考试内容的联系和区别，在提高教学水平的同时提高二级 Access 等级考试的通过率。最后，为提高等级考试通过率，考生在备考 Access 时应以考试大纲与实际应用为中心，多练习历年真题和开发实际应用系统。

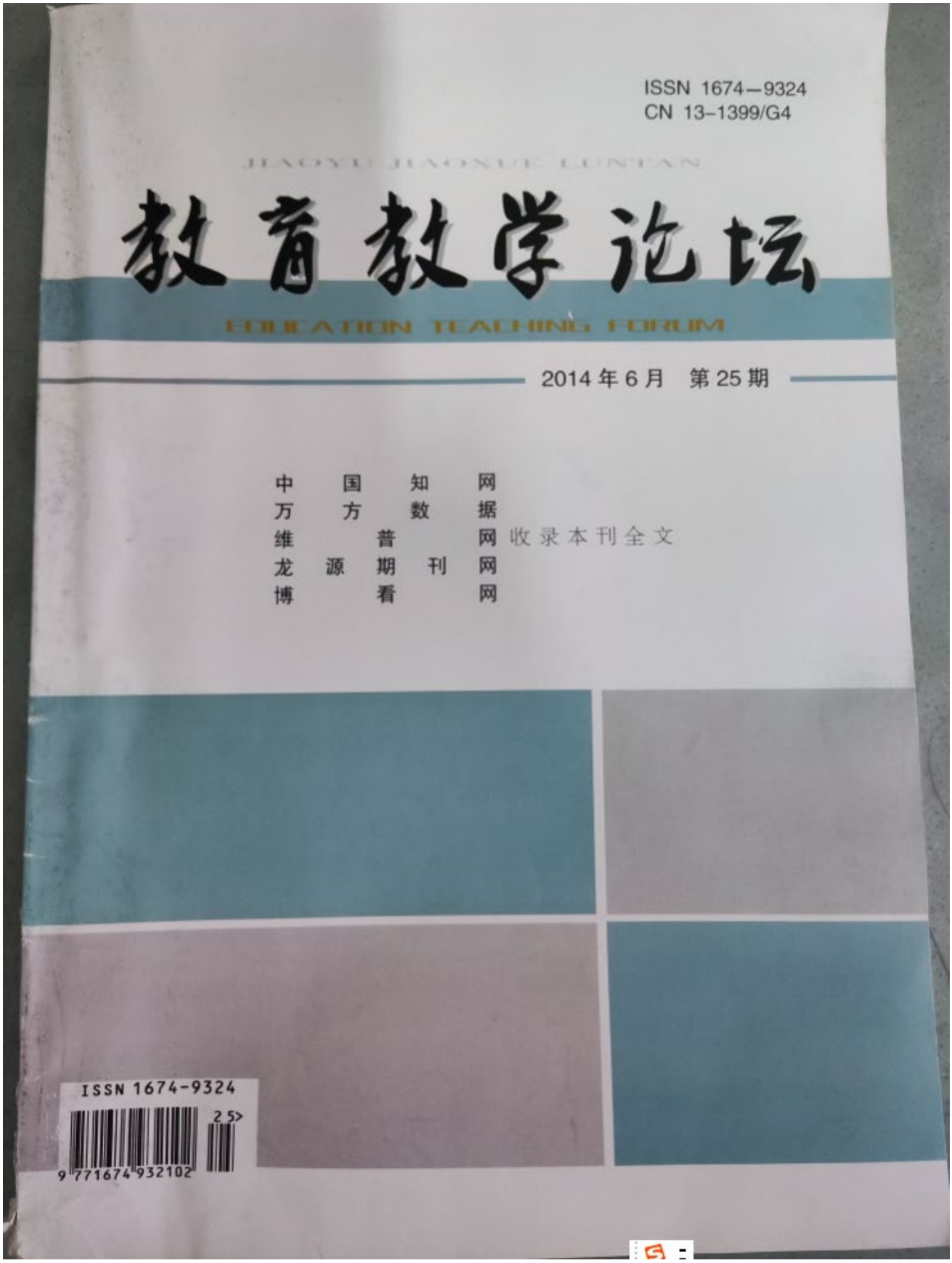
参考文献：

[1] 赵洪帅. 浅谈我校计算机等级考试(二级Access)教学改革[J]. 中央民族大学学报(自然科学版), 2010, 19(4): 89-93.
[2] 潘京. 对全国计算机等级考试(Access)的思考[J]. 计算机教育, 2011(7): 43-45.
[3] 康伟民. 对高校ACESS教学与计算机等级考试的几点思考[J]. 中国校外教育, 2013(8): 165.
[4] 刘娟. 计算机等级考试下的“Access数据库程序设计”教学探讨[J]. 安庆师范学院学报(自然科学版), 2011, 17(4): 120-122.
[5] 陶颖. 基于ACCESS科目的全国计算机等级考试操作要点的研究与总结[J]. 农村经济与科技, 2016, 27(2): 164-165.

（编辑：孙怡铭）

（2）普刊2：《数据库技术》课程在艺术类专业的教学改革

（涂淑琴,薛月菊,黄晓琳,等.《数据库技术》课程在艺术类专业的教学改革[J].教育教学论坛,2014(25):49-50.）



目录 contents

特别关注

科学管理

教师建设

教改创新

教育教学论坛

JIAOYU JIAOXUE LUNTAN

2014年6月 第25期 总第158期

期刊名称 教育教学论坛
主 管 河北出版传媒集团有限责任公司
主 办 河北教育出版社有限责任公司
花山文艺出版社有限责任公司
出 版 教育教学论坛杂志社
印 刷 石家庄市飞达印刷有限公司
发 行 河北省报刊发行局
出版日期 2014年6月18日
主 编 韩新保
发行范围 国内外公开
定 价 40.00元
国内统一连续出版物号 CN 13-1399/G4
国际标准连续出版物号 ISSN 1674-9324
广告许可证号 1301024000093
地 址 石家庄市联盟路705号
邮 编 050061
订 购 处 全国各地邮局
邮发代号 18-219
网 址 http://www.jyjxltzzs.com/
电子信箱 jyjxlt@jyjxltzzs.com
联系电话 0311-87760976

教法研究

学法指导

教师观点

专题研讨

教学设计

民主教学

职业教育

学生教育

探索与实践

学校与社会

阅读与写作

课程建设

信息技术

实验平台

快乐课堂

学前教育

德育美育

心理疏导

检测评价

学术前沿

《数据库技术》课程在艺术类专业的教学改革

涂淑琴,薛月菊,黄晓琳,张 晓
(华南农业大学 信息学院,广东 广州 510640)

摘要:针对本校艺术类专业的特点和学习要求,以《数据库技术》课程教学为主题,结合艺术类专业学生的特点,提出艺术类专业学生在《数据库技术》课程教学内容、教学方式、课程考核上的一些改进设想。

关键词:《数据库技术》;艺术类专业;课程设计
中图分类号:G642.0 文献标志码:A 文章编号:1674-9324(2014)25-0049-02

目前,几乎所有高校都有开设数据库课程。华南农业大学从2005年开始,《数据库技术》作为艺术类专业学生的计算机公共基础课。数据库课程具有理论基础强、实践应用性突出等特点。在本人的教学中,发现艺术类专业学生由于该课程理论太难而学习积极性不高,普遍厌烦这门课程。因此,结合艺术类专业学生的特点以及他们专业的要求,必须进行实践和理论整合,提高教学质量。如何改革、设计数据库课程,使之适应艺术类专业教育的培养目标,是我们这些数据库教师需要认真研究的问题。艺术类专业学生的自身特点:艺术类学生普遍形象思维能力发达,有自我的特点,在学习、生活中具有较强的自主自由性,易受情绪支配。心情好,学习积极性就高 如果状态不好,就完全放弃。他们情感丰富,富于想象力和创造性,对新事物和新技术有较强的接受能力和学习能力,但同时会出现容易冲动、欠缺理性的行为,而且他们凭兴趣学习,对于没有兴趣的课程则消极学习或者直接放弃,缺乏计划性和控制力。同时,艺术专业学生在文化基础教育方面比较薄弱,因此在计算机基础知识方面,存在理论知识薄弱的特点,少部分学生实验操作能力强,大部分学生比较弱。针对艺术类学生的以上特点,以本人5年艺术专业数据库课程的教学效果来说,主要有以下问题。

一、学生学习兴趣不浓,缺乏主动学习的热情

《数据库技术》课程概念比较抽象,基本理论知识比较多[1],这部分内容又涉及关系数学和数据模型创建,学生不是很容易掌握,而目前这部分内容对艺术学生的专业没有任何直接有效的帮助,因此学生缺乏学习的热情,也不会去主动学习。

二、教材内容不恰当和滞后

目前高校《数据库技术》教学采用的教材大多是依据全国计算机等级考试大纲进行编写的,里面包含数据库原理

三、课堂教学原则创新

课堂教学原则创新体现在以下四个方面。

1.主体性原则。让学生成为课堂教学的主人,让课堂教学焕发出生命的活力,把学习主动权交给学生,充分调动其学习的主动性和创造性。

2.开放性原则。创新性课堂教学应该是个开放的教学空间。学生的心智应是开放的、自由的、不受压抑的;教学的内容不是单一的、封闭的,不拘泥于教材和教师的知识视野;教学结果不满足于课本和教师的所谓标准答案。

3.挑战性原则。在课堂教学过程中,提倡学生向教师挑战,鼓励学生质疑问难,鼓励学生提出与课本不同的看法,鼓励学生通过自己的探索,否定权威的结论。

4.差异性原则。创新性课堂教学不实行“一刀切”、“齐步走”,要通过分层教学最大限度地满足学生因智力、个性特征、学习习惯个体差异发展的需要。

四、课堂教学内容创新

以培养财经类学生学习能力、创新能力、解决问题能力为目标,创新现有的教学内容体系,就必须改革现有的教学内容体系,构成新的教学内容体系,把有利于能力的形成作为筛选的对象,删除陈旧知识,增加训练能力的知识和前沿新知识,减少记忆性知识分量。减轻学生的课业负担,使学生有时间、有条件接触自然,参加社会实践,以适应新形势下的学生就业素质的培养。对教学内容创新可从以下几方面着手:⑴ 从教学内容的背景入手。注意教材背景的适当拓宽,开阔学生的视野。⑵ 从教学内容的新颖入手。改变教学内容僵化,课堂气氛死气沉沉的局面,增加贴近学生现实生活、反映现代科技发展的最新成果的内容,最大限度地激发学生的创新意识。

五、教学方法创新

废除不利于培养能力的传统教学方法,运用能培养学生自学能力、分析处理信息能力的启发式教学、讨论式教学、模拟实验方法和社会实践方法。启发式教学其精髓在于启发学生积极思考,培养学生的科学思维方式,鼓励学生大胆质疑,激发学生的学习积极性和创新精神。讨论式教学是在自学的基础上进行的,并在自学之后对自学成果进行检验使学生在不断发表见解中得到锻炼,这样可以培养学生的自学能力、表达能力、应变及创新能力。模拟实验方法是培养学生专业技能的重要途径,通过模拟实验,将学生置于现实工作中,使所学的理论知识转化为技能,为毕业后尽快就业打下基础。模拟实验应在教师指导和示范下,有目的、有计划地进行,并做到单项练习与综合练习相结合,集中练习与分散练习相结合。通过教师的指导和示范,既能让学生获得规范化的技能动作,又能培养学生的观察能力,还可以采用社会实践的方法进行教学,教师应紧密结合教学内容,组织学生开展有目的的社会实践活动,使学生认识到掌握理论知识的重要性和必要性,巩固学习成果,培养学生的动手能力和运用知识的能力。

六、考试制度创新,推动教学改革

创新考试制度,推动教学改革,要从考试内容、考试方式、评分和评判方式四个方面进行。考试内容要做到知识、能力和素质考核都有。根据学科特点,采用闭卷、开卷、笔试、口试、论文等考试方式。评分方式可采用百分制、五分制。评判成绩要把终极考核与过程考核相结合,注重过程考核,把创新作为评判的重要依据。

总之,在财经类教学中只有实施创新,才能提高学生的理论知识与实践能力,增强其在社会上的竞争能力,使他们早日成为经济领域中的重要人才。

参考文献:

[1]刘世清,刘家勋.教育信息技术实用教程(全国教育教学法“十五”规划教育部重点科研成果)[M].北京:电子工业出版社,2003.

和计算机编程的理论基本知识。对艺术类专业学生，数据库中VBA部分和计算机编程一般不要求掌握；另一方面，越来越多的艺术领域与计算机技术相结合而产生的新技术手段，必须让艺术类专业学生掌握。艺术类专业的《数据库技术》教学在教材选择上面临一个两难问题[2]：一方面非计算机专业数据库及计算机编程的内容比较难；另一方面艺术专业与计算机数据库相关的内容过少，同时教材更新时间长，教材内容明显滞后。

三、实验课程和实际应用脱节

传统的实践教学很多都是作为理论课程的补充，理论知识学习后，通过上机实践验证，完成对相应理论知识的理解。上机实践内容相互孤立，没有体现出前后知识点之间的联系和知识的综合应用，没有完成整个工程项目的开发过程。很多学生完成实验册子上的每个小实验，但不知道融合各个分知识点，也不知道如何完成整个工程项目。学生很少有机会实践一个数据库应用系统的设计和完整开发过程，更谈不上创新能力的培养。

四、《数据库技术》在艺术类专业的教学改革方案

1.根据艺术类专业学生特点，采用合理的教学方式。针对艺术类专业学生，我们不能采用对理工专业学生的理性教学方式，否则，很难达到预计的教学效果。我们可以采用感性教学方式，引导学生在理论与实际应用中理解知识，培养学生的学习兴趣。例如，在讲解"数据库设计E-R模型理论"知识时，以服装设计信息管理系统为例，将设计者与服装作为实体，研究它们之间的关系，提高了学生的学习兴趣。在讲解窗体与宏应用中，以讲解如何制作大家熟悉的QQ登陆界面，讲解控件工具箱中文本、标签、图像和按钮各个工具的作用，这样使学生觉得很有乐趣。课堂上，采用课堂讲解、回答问题和课堂互动等[3]多种方式学习，鼓励学生课前预习，课后复习，培养学生主动学习和创新能力。教学过程中注重示例讲解。例如，讲解参数查询操作时，以学校学生一卡通充值为例子，将学号设成参数，根据学号参数迅速找到学生的个人信息并充值，使学生的思维活跃起来，积极地参与讨论问题。针对艺术类专业学生自主自由性的特性，在布置实验操作练习的要求上，主张自主性和创新性。只要求他们实现系统中的某些功能，方法不限制，细节不严格要求，充分发挥学生的主观能动性，让他们把数据库操作实验当成自己设计的作品。

2.设计合理和适当的教学内容。《数据库技术》课程主要帮助学生完成两大任务：一是学习设计数据库（系统软件开发），二是学习使用数据库（应用软件开发）。教材与课件是学生学习数据库的主要方式，对教学的作用不言而喻。因此，要综合考虑教材内容的完整性、易用性以及对新概念、新技术的反映。针对艺术类专业学生的特点，在保证内容完整性的前提下[4]，所选教材应实例丰富，课后有满足学生复习、巩固所学内容要求的适量习题和实验、实训，内容清晰。而对示例数据库选择，要保证示例数据库内容完整，容易理解与接受，并要符合他们感性的理解。为此，本人首先根据学生的接受能力，将所学的难点与实际应用结合阐述，让学生觉得难点不难。例如对数据模型理论的讲解，本人针对数据建模的抽象性与难理解性，将其与汽车、轮船等模型做比较讲解，使学生印象深刻，充分理解数据模型的重要性。对实验教学内容的安排一般设置上机题和验证题，这些必须与学生的认知规律相一致，针对不同专业的学生因势利导，让学生有所收获。以下是本人针对艺术学院学生设计的Access教学工作执行计划表：

周次	讲授内容	理论课时	上机课时
第1~2周	课程简介，数据库概述，数据模型和关系模型的概念	4	
第3~4周	关系代数运算、创建数据库、表结构设计和格式设置	4	2
第5~6周	表之间联系与查询建立、建立参数查询与操作查询	4	4
第7~8周	数据库设计，建立E-R模型，规范化基础，3R范式	4	
第9~10周	SQL命令建立查询、动作查询	4	2
第11~12周	窗体建立，控件工具箱的使用和属性、方法的使用	4	2
第13周	报表的建立方法，图表报表和分组报表的使用	2	2
第14~15周	宏、宏组与条件宏的建立、运行，宏与窗体结合使用	4	4
第16周	小型数据库（服装设计信息管理系统）的开发与复习	2	
总计		32	16

3.设计有效可行的课程考核方法，充分体现应用能力的培养。制定一套切实可行的评分标准，对于考核学生理论掌握与应用创新非常重要。针对艺术专业学生，理论知识掌握较薄弱、动手创新能力较强，并且其学习目标与侧重点不同，考核方式也有所区别。该门课程本人采用平时考核占主导地位（70%），期末考试占30%。平时考核采用课后作业、上课点名、随堂测试和课内上机实验操作技能考核相结合的方法（作业15%；上课点名5%；随堂测试20%；上机实验30%），这部分主要考查学生平时的学习能力与操作技能，成绩占70%；以闭卷考试方式进行期末考试，成绩占30%，考查学生平时操作与基本知识的理解。增加平时成绩的分值，注重日常的学习和积累，通过多次考核综合评定成绩[5]。对于点名的考核，采取以提问方式代替直接点名的方法，增强了学生听课的专注程度以及教学中的互动，效果较好。上机实验成绩提升至30%，同时增加实验中的随堂上机考核环节，能提高学生解决实际应用问题的能力。注重加强学生平时学习情况的考核，课程中相对独立的学习内容结束后增加单元测试环节，以此作为随堂测验20%的依据。将期末考试成绩占总成绩的30%，这样既可以避免艺术专业学生死记硬背理论知识的缺点，又可以防止部分学生靠最后几天复习就过关的情况。探索数据库课程考试内容与方法的变革，重视理论与实践技能知识的考核，对促进学生主动学习具有良好的效果。

《数据库技术》课程是一门理论和实践密切结合的学科。改革《数据库技术》课程的教学方法、教学内容和考核模式，灵活地并创造性地掌握好教学过程和兴趣切入点，为艺术类专业学生营造良好的学习热情和氛围，对提高艺术类专业学生整体的数据库应用水平和综合素质起到了重要的作用。

参考文献：

[1]王珊，萨师煊.数据库系统概论[M].北京：高等教育出版社，2006.

[2]章雁.艺术类专业计算机基础教学的研究[J].海峡科学，2012，(07)：63-65.

[3]申玉静，谭业武.数据库案例教学在计算机专业专科教育职业化改革中的应用与研究[J].福建电脑，2010，(1)：22-25.

[4]李萍，张译丹.艺术类专业"数据库技术"课程设计探讨[J].中国有线电视，2011，(07)：825-827.

[5]钟辉.数据库课程教学改革探索与实践[J].沈阳建筑大学学报，2010，12(3)：379-381.

作者简介：涂淑琴（1978-），女，江西南昌人，在读博士，讲师，主要从事数据库，数据挖掘；薛月菊（1969-），女，云南昆明人，博导，教授，主要从事数据挖掘，智能算法。

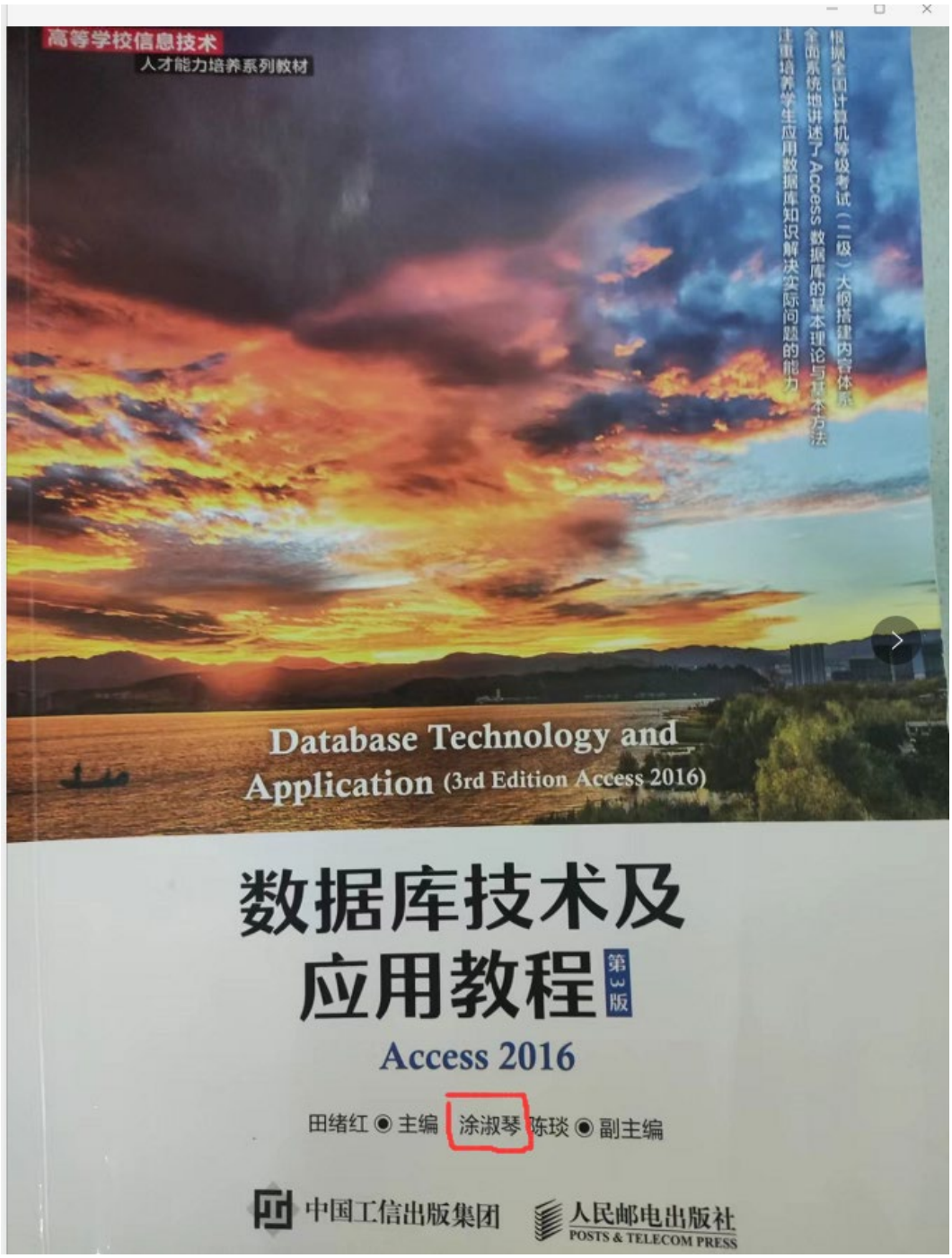
JIAOYU JIAOXUE LUNTAN

教育教学论坛

网址:http://www.jyjxltzzs.com/ 电子信箱:jyjxlt@jyjxltzzs.com

CN 13-1399/G4 邮发代号:18-219 定价:40.00元

5. （1）编写教材：一般教材《数据技术及应用教程》
《数据库技术与应用教程》.主编：田绪红，副主编：涂淑琴，陈琰,2021.人民邮电出版社



目 录

数据库技术及应用教程（第二版）

Access 2016

本书主要介绍数据库的基本理论与基本方法，并结合 Access 详细介绍数据库的具体操作。全书共 12 章，主要内容包括数据库系统概述、关系数据库、Access 数据库管理系统、表的操作、数据库设计、数据查询、关系数据库标准语言 SQL、数据库应用开发技术、VBA 程序设计、网上书城信息管理系统综合实例、数据库安全与管理、数据库技术新发展等。本书强调基本概念的准确性、基本原理的正确性，并通过接近实际数据库应用的例子，详细介绍数据库的设计原理与步骤。

扫此二维码下载
本书配套资源

人邮教育
www.ryjiaoyu.com
教材服务热线：010-81055256
反馈 / 投稿 / 推荐信箱：315@ptpress.com.cn
人邮教育服务与资源下载社区：www.ryjiaoyu.com

ISBN 978-7-115-57142-7

定价：52.00元

封面设计：董志桢

图书在版编目（CIP）数据

数据库技术及应用教程 : Access 2016 / 田绪红主编. — 3版. — 北京 : 人民邮电出版社, 2021.9
高等学校信息技术人才能力培养系列教材
ISBN 978-7-115-57142-7

Ⅰ. ①数… Ⅱ. ①田… Ⅲ. ①关系数据库系统—高等学校—教材 Ⅳ. ①TP311.138

中国版本图书馆CIP数据核字(2021)第165726号

内 容 提 要

本书主要介绍数据库的基本理论与基本方法，并结合 Access 详细介绍数据库的具体操作。全书共 12 章，主要内容包括数据库系统概述、关系数据库、Access 数据库管理系统、表的操作、数据库设计、数据查询、关系数据库标准语言 SQL、数据库应用开发技术、VBA 程序设计、网上书城信息管理系统综合实例、数据库安全与管理、数据库技术新发展等。本书强调基本概念的准确性、基本原理的正确性，并通过接近实际数据库应用的例子，详细介绍数据库的设计原理与步骤。

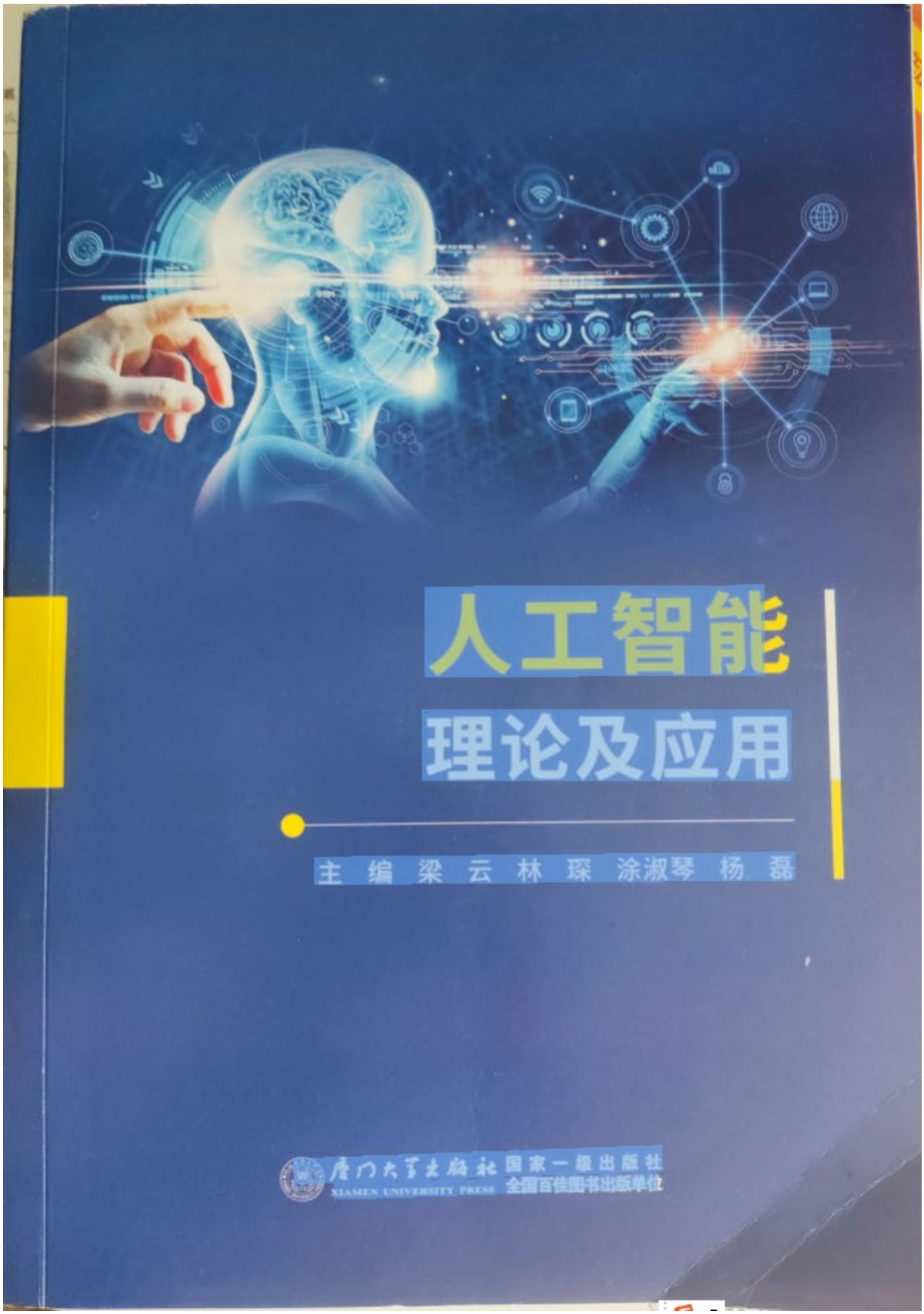
本书既可作为高等学校非计算机专业数据库技术课程的教材，也可作为计算机爱好者的自学参考书。

◆ 主　　编　田绪红
副 主 编　涂淑琴　陈　琰
责任编辑　许金霞
责任印制　王　郁　马振武
◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　https://www.ptpress.com.cn
三河市中晟雅豪印务有限公司印刷
◆ 开本：787×1092　1/16
印张：14.5
字数：386 千字
2021 年 9 月第 3 版
2021 年 9 月河北第 1 次印刷

定价：52.00 元

读者服务热线：(010)81055256　印装质量热线：(010)81055316
反盗版热线：(010)81055315
广告经营许可证：京东市监广登字 20170147 号

(2) 编写教材：一般教材《人工智能理论及应用》
《人工智能理论及应用》. 主编: 梁云，林琛，涂淑琴 杨磊.2024.厦门大学出版社.



图书在版编目（CIP）数据

人工智能理论及应用 / 梁云等主编. -- 厦门 : 厦门大学出版社, 2024. 8. -- ISBN 978-7-5615-9475-9

Ⅰ. TP18

中国国家版本馆 CIP 数据核字第 2024B6J949 号

责任编辑　郑　丹
美术编辑　李嘉彬
技术编辑　许克华

出版发行　厦门大学出版社
社　　址　厦门市软件园二期望海路 39 号
邮政编码　361008
总　　机　0592-2181111　0592-2181406(传真)
营销中心　0592-2184458　0592-2181365
网　　址　http://www.xmupress.com
邮　　箱　xmup@xmupress.com
印　　刷　厦门市竞成印刷有限公司

开本　787 mm×1 092 mm　1/16
印张　18.5
字数　440 千字
版次　2024 年 8 月第 1 版
印次　2024 年 8 月第 1 次印刷
定价　45.00 元

目　录

人工智能
理论及应用

责任编辑 郑 丹
美术编辑 李嘉彬
封面供图 VEER www.veer.com

T0475-1-1
ISBN 978-7-5615-9475-9
扫码了解更多
定价:45.00元

二、 科研项目

1.主持：关于基于多目标跟踪的群养生猪异常行为研发项目的立项通知（合同）及有关佐证材料

表9 科研课题情况涂淑琴 主持的课题

序号	项目名称	评审等级	项目来源	合同经费/实到经费	立项时间	结题时间	课题组总人数	本人排名	是否结题	备注
1	基于RGB-D传感器的百香果成熟度判别系统应用示范	B	广东省科技厅	5.0	2015-07-31	2018-01-01	4	1	是	
2	尿液细胞实例分割算法研发		横向	7.0	2022-06-16	2024-06-01	5	1	是	
3	尿液细胞多目标跟踪算法		横向	8.0	2022-10-08	2024-10-20	6	1	是	
4	基于多目标跟踪的群养生猪异常行为研发	A	横向	150.0	2024-07-12	2029-07-12	15	1	否	
5	基于Mask R-CNN和Soft-NMS融合的群养粘连猪实例分割方法专利转让		横向	1.0	2024-07-18	2024-10-26	2	1	是	

科技处审核人及盖章：

年　　月　　日

合同编号：

技术开发（委托）合同

项目名称：基于多目标跟踪的群养生猪异常行为研发

委 托 方（甲 方）：广州创纽人工智能科技有限公司

受 托 方（乙 方）：华南农业大学

签订时间：2024 年 7 月 12 日

签订地点：广州

有效期限：合同签定之日起 5 年

中华人民共和国国家科学技术部印制

填 写 说 明

一、本合同为中华人民共和国科学技术部印制的技术开发(委托)合同示范文本，各技术合同登记机构可推荐技术合同当事人参照使用。

二、本合同书适用于一方当事人委托另一方当事人进行新技术、新产品、新工艺、新材料或者新品种及其系统的研究开发所订立的技术开发合同。

三、签约一方为多个当事人的，可按各自在合同关系中的作用等，在“委托方”、“受托方”项下（增页）分别排列为共同委托人或共同受托人。

四、本合同书未尽事项，可由当事人附页另行约定，并作为本合同的组成部分。

五、当事人使用本合同书时约定无须填写的条款，应在该条款处注明“无”等字样。

技术开发（委托）合同

委托方（甲方）：广州创纽人工智能科技有限公司

住所地：广州市增城区新塘镇光华路 19 号 2730 号

法定代表人：马书杰

通讯地址：广州市增城区新塘镇光华路 19 号 2730 号

电　　话：15889957352　　传真：

电子信箱：836547465@qq.com

受托方（乙方）：华南农业大学

住所地：广东广州天河区五山路 483 号

法定代表人：薛红卫

通讯地址：广东广州天河区五山路 493 号华南农业大学数学与信息学院计算机科学与工程系

电话：13751838469　　传真：020-85285393

电子信箱：tushuqin@163.com

本合同甲方委托乙方就基于多目标跟踪的群养生猪异常行为研发项目，并支付研究开发经费和报酬，乙方接受委托并进行此项研究开发工作。双方经过平等协商，在真实、充分地表达各自意愿的基础上，根据《中华人民共和国合同法》的规定，达成如下协议，并由双方共同恪守。

第一条：甲方委托乙方进行技术服务的内容如下：

1．技术服务的目标：实现基于多目标跟踪的群养生猪异常行

为研发

2．技术服务的内容：（1）支持部署在 windows 操作系统和 MSSQL 数据库环境的群养生猪异常行为识别接口服务，提供基于 Http 的 API 接口服务（RestFul 方法，HTTP POST 方式，数据格式为 json 格式）。（2）对群养生猪进行数据集的采集与标注，完成不少于 30 个 1 分钟视频数据集的预处理。（3）建立群养生猪的多目标跟踪及异常行为识别的模型，其模型准确率在 92%以上。（4）开发群养生猪多目标跟踪与异常行为识别的软硬件系统。包括：猪舍多源信息采集模块、通信模块、智能控制模块及服务器。（5）辅助甲方完成与此项目相关的用户培训及技术咨询服务。

3．技术服务的方式：（1）算法研发期间组织项目会议，讨论需求、研究及辅助甲方组织验收等相关工作；（2）算法研发完成后，提供不少于一年期的咨询服务；（3）辅助甲方提供用户培训及技术咨询服务。

第二条：乙方应按下列要求完成技术服务工作：

1．技术服务地点：华南农业大学、广州创纽人工智能技术有限公司。

2．技术服务期限：自合同签定之日 5 年。

3．技术服务进度：（1）合同签定之日起一年内提供群养生猪多目标跟踪算法的技术方案并组织初验。（2）合同签定之日起 48 个月内完成异常行为算法研发任务提交甲方验收。（3）验收通过后开始算法运行维护、协助甲方进行用户培训等工作，工作期一年。

第三条：为保证乙方有效进行技术服务工作，甲方应当向乙方提供下列工作条件和协作事项：

1．提供技术资料：

（1）明确自然场景下养殖场中群养生猪的视频要求，要求自然场景下群养生猪的头数在[10,20]以内。（2）确定群养生猪的多目标跟踪的时间长短，异常行为判别，白天及黑夜的异常行为准则。3）每种异常行为类视频文件放在一个文件夹下。4）视频等资源不侵犯第三方专利权、著作权或其他技术秘密。

2．提供工作条件：

（1）算法的运行环境，包括服务器、网络等必要的软硬件资源；

（2）运维、用户培训若需现场办公，提供必要的工作环境。

第四条：甲方向乙方支付技术服务报酬及支付方式为：

1．技术服务费总额为：人民币壹佰伍佰万元整(¥1500,000.00)

2．技术服务费由甲方 分期 （一次或分期）支付乙方。

具体支付方式和时间如下：

（1） 合同签定后三十个工作日内，支付合同总金额的 30%，450,000.00 元。合同签定后六十个工作日内支付合同总金额 30%，45,000.00 元。合同签定后九十个工作日内支付项目尾款 40%，60000.00 元。每次支付完费用后 10 个工作日内，提供等额的增值税普通发票。

乙方开户银行名称、地址和帐号为：

开户银行：广州工行五山支行

地址：广州市天河区五山路483号

帐户名：华南农业大学

帐号：3602002609000310520

第五条：双方确定因履行本合同应遵守的保密义务如下：

甲方：

1.保密内容（包括技术信息和经营信息）：乙方单位及人员信息、乙方所提供研发技术方案、源代码、技术文档等。

2．涉密人员范围：甲方与此项目相关的工作人员。

3．保密期限：合同签定之日起，至项目结束后一年。

乙方：

1.保密内容（包括技术信息和经营信息）：甲方单位信息、本合同信息。

2．涉密人员范围：乙方与此项目相关的工作人员。

3．保密期限：合同签定之日起，至项目结束后一年。

第六条：本合同的变更必须由双方协商一致，并以书面或邮件形式确定。但有下列情形之一的，一方可以向另一方提出变更合同权利与义务的请求，另一方应当在30日内予以答复；逾期未予答复的，视为同意。

第七条：双方确定以下列标准和方式对乙方的技术服务工作成果进行验收：

1．乙方完成技术服务工作的形式：远程及现场交流。按项目时间安排组织现场或视频会议交流。

2．技术服务工作成果的验收标准：（1）算法以可调用软件包方式提供，可正常运行，群养生猪异常识别准确率不低于92%，且不会随识别图片增加出现额外延迟。（2）基于多目标跟踪的群养生猪异常行为算法的源代码、技术文档（包括算法说明书和用户手册等）完整。（3）算法及技术文档不侵犯第三方专利权、著作权或其他技术秘密。

3．技术服务工作成果的验收方法：由甲方组织进行成果评价。

4．验收的时间和地点：2029 年 7 月，华南农业大学或广州创纽人工智能技术有限公司。

第八条：双方确定：

1．在本合同有效期内，甲方利用乙方提交的技术服务工作成果所完成的新的技术成果，归 甲 方所有。

2．在本合同有效期内，乙方利用甲方提供的技术资料和工作条件所完成的新的科研论文和成果，归 乙 方所有。

3．在本合同有效期内，若双方需要共同申请软件著作权及发明专利，由双方协调规定署名、所获得利益及费用分配比例。

第九条：双方确定，在本合同有效期内，甲方指定 于会明 为甲方项目联系人，乙方指定 涂淑琴 为乙方项目联系人。一方变更项目联系人的，应当及时以书面形式通知另一方，未及时通知并影响本合同履行或造成损失的，应承担相应的责任。

第十一条：双方确定，出现下列情形，致使本合同的履行成为不

必要或不可能的，可以解除本合同：

1. 发生不可抗力；

2. 一方有违反本合同的，另一方有权解除合同；

3. 合同期满；

4. 双方同意终止合同的；

第十二条：双方因履行本合同而发生的争议，应协商、调解解决。协商、调解不成的，确定按以下第 1 种方式处理：

1. 提交 广州市 仲裁委员会仲裁；

2. 依法向人民法院起诉。

第十三条：本合同一式 3 份，具有同等法律效力。

第十四条：本合同经双方签字盖章后生效。

甲方：广州创纽人工智能科技有限公司（盖章）

法定代表人／委托代理人：（签名）

2024 年 7 月 12 日

乙方：华南农业大学（盖章）

法定代表人／委托代理人：（签名）

2024 年 7 月 12 日

印花税票粘贴处：

（以下由技术合同登记机构填写）

合同登记编号：

1．申请登记人：_____

2．登记材料：（1）_____

（2）_____

（3）_____

3．合同类型：_____

4．合同交易额：_____

5．技术交易额：_____

技术合同登记机构（印章）

经办人：

年　　月　　日

2.主持：基于RGB-D传感器的百香果成熟度判别系统应用示范项目的立项通知（合同）及有关佐证材料

广东省科学技术厅文件

粤科规财字〔2015〕151号

广东省科学技术厅关于下达2015年度省协同创新与平台环境建设专项资金项目计划的通知

各地级以上市科技局（委）、顺德区经济和科技促进局，有关单位：

2015年度省协同创新与平台环境建设专项资金项目已经公示无异议，现按规定下达给你们，并就有关事项通知如下：

一、本次下达的科技计划项目共585项，经费47988万元。

二、各级主管部门和项目承担单位收到本通知后，须尽快按照《广东省科学技术厅关于省科技计划项目合同书管理的实施细则（试行）》（粤科函规划字〔2013〕1097号）有关规定与省科技厅签订项目合同书，并协助下达财政资金（资金计划由省财政厅另文下达）。

三、各级主管部门应履行项目的日常监管职责，督促项目承

担单位做好项目的组织实施，并配合省有关部门组织开展的监督检查、绩效评价、验收结题、项目审计等相关工作。

四、各项目承担单位要抓紧项目的组织实施，严格按照科技经费的使用范围和有关规定管好用好财政资金，确保按期完成科研任务，提升创新能力。项目在研过程中每自然年度第 1 个月内须在省科技业务管理阳光政务平台（网址：http://pro.gdstc.gov.cn）填报上年度执行情况报告。项目完成后，要按照《广东省科学技术厅关于省科技计划项目结题管理的实施细则（试行）》（粤科监审字〔2014〕121 号）有关规定进行结题。

附件：2015 年度省协同创新与平台环境建设专项资金项目计划安排表

广东省科学技术厅

2015 年 9 月 28 日

公开方式：依申请公开

抄送：省财政厅。

广东省科学技术厅办公室　　2015 年 9 月 25 日印发

附件：

2015年省协同创新与平台环境建设专项资金项目计划安排表

单位：万元

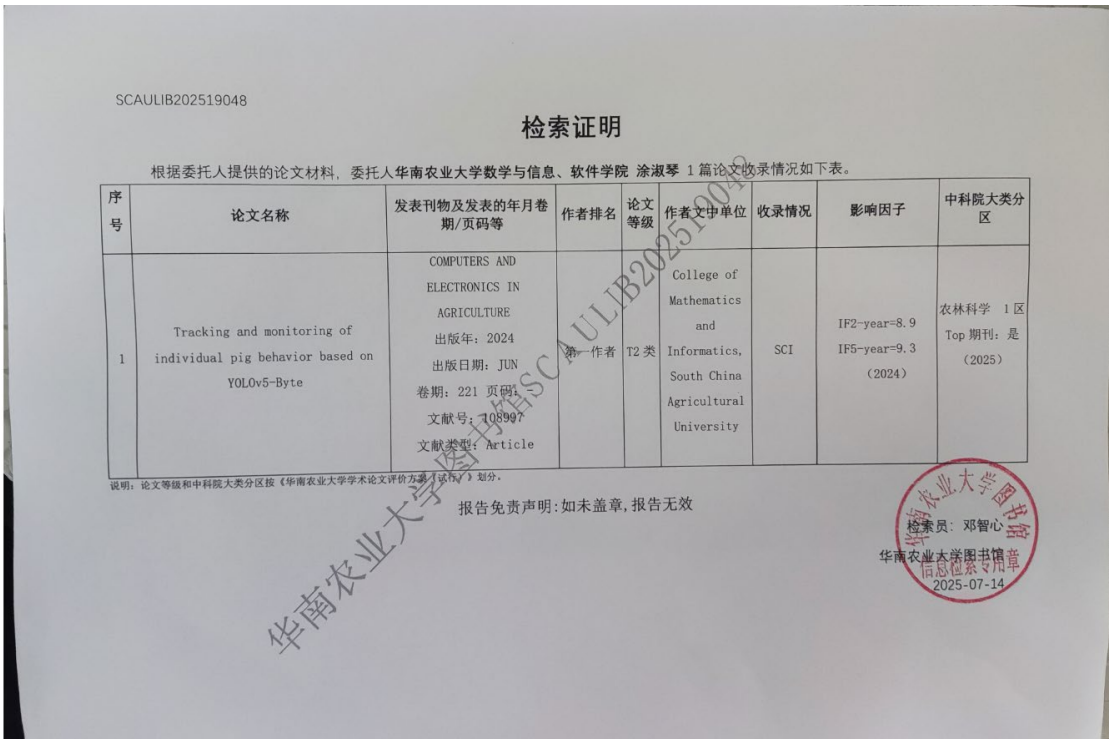
序号	财政预算编码	承担单位	项目立项编号	项目名称	负责人	立项金额
2	156003	华南农业大学				930
1		华南农业大学	2015A050502044	细胞内质网应激调控自噬在猪瘟病毒感染致病中的机制研究	陈金顶	50
2		华南农业大学	2015B090903076	广东省昆虫行为调控工程技术研究中心	何晓芳	100
3		华南农业大学	2015B090903077	广东省现代生态农业与循环农业工程技术研究中心建设	章家恩	100
4		华南农业大学	2015B090903074	广东省光学农业工程技术研究中心建设	刘应亮	100
5		华南农业大学	2015A020224036	梅县农作物种植中水肥滴灌自动化控制装置应用示范	岳学军	5
6		华南农业大学	2015A020224032	基于物联网的茶叶质量追溯系统应用示范	潘春华	5
7		华南农业大学	2015B090903075	广东省土地信息工程技术研究中心建设	胡月明	100
8		华南农业大学	2015A020224038	基于RGB-D传感器的百香果成熟度判别系统应用示范	涂淑琴	5
9		华南农业大学	2015A020224033	连平县丰隆种植专业合作社特色农产品信息网络平台建设与应用	殷建军	5
10		华南农业大学	2015A020224029	阳山县桑树病害虫害防控技术应用示范	黄志君	5
11		华南农业大学	2015A020224030	以HRQOL为导向的慢性病综合防治技术在龙川县的推广应用	童峰	5
12		华南农业大学	2015A020224031	杂交兰新品种产业化生产与示范	郭和蓉	5
13		华南农业大学	2015A020224037	冬种马铃薯高产优质栽培技术在龙川县示范	全锋	5
14		华南农业大学	2015A050502045	基于秸秆细胞壁成分分析的水稻种植资源筛选与秸秆利用	吴蔼民	50
15		华南农业大学	2015A050502043	土壤厌氧消毒法（ASD）控制番茄青枯病的关键技术与示范	蔡昆争	50
16		华南农业大学	2015B050501009	国际农业航空施药技术联合实验室建设	兰玉彬	150
17		华南农业大学	2015A040406002	农业高校科技成果入股转化机制研究	严会超	50
18		华南农业大学	2015A020224039	新丰县牛羊健康养殖科技人才培训	贾坤	5
19		华南农业大学	2015A070709014	高校低碳校园建设的探索	吕辉雄	10
20		华南农业大学	2015A070710018	《走近农用无人机》科普专题片策划与制作	胡年春	10
21		华南农业大学	2015A020224035	无抗与保健功能鸡蛋生产技术的应用示范	石达友	5
22		华南农业大学	2015A020224040	鸭稻共作生产有机稻米关键技术在和平县的应用示范	陈志鸿	5
23		华南农业大学	2015A020224034	梅州金柚种植过程中基于图像识别基础上的病虫害信息采集分析及产量预估系统研发	郭艾侠	5
24		华南农业大学	2015B090901059	新型广谱抗菌药土拉霉素原料和制剂的产业化研究	方炳虎	100

2.主参B类：基于视频监控大数据分析的奶牛行为研究与应用

1纵向项目：基于视频监控大数据分析的奶牛行为研究与应用；
编号 2016A050502050；
广东省科技厅；50；
排名2

三、论文、著作等

1.检索证明



SCAULIB202519048

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息、软件学院 涂淑琴** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Tracking and monitoring of individual pig behavior based on YOLOv5-Byte	COMPUTERS AND ELECTRONICS IN AGRICULTURE 出版年：2024 出版日期：JUN 卷期：221 页码： 文献号：108997 文献类型：Article	第一作者	T2 类	College of Mathematics and Informatics, South China Agricultural University	SCI	IF2-year=8.9 IF5-year=9.3 （2024）	农林科学 1 区 Top 期刊：是 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：邓智心

华南农业大学图书馆

2025-07-14

SCAULIB202519050

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息、软件学院 涂淑琴** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images	PRECISION AGRICULTURE 出版年：2020 出版日期：OCT 卷期：21 5 页码：1072-1091 文献类型：Article	第一作者	T2 类	College of Mathematics and Informatics, South China Agricultural University	SCI	IF2-year=5.385 IF5-year=5.004 （2020）	农林科学 1 区 Top 期刊：是 （2020）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：邓智心

华南农业大学图书馆

2025-07-14

SCAULIB202519056

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数信学院 涂淑琴** 6篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Behavior Tracking and Analyses of Group-Housed Pigs Based on Improved ByteTrack	ANIMALS 出版年：2024 出版日期：NOV 卷期：14 22 页码：- 文献号：3299 文献类型：Article	第一作者	A类	华南农业大学数信学院	SCI	IF2-year=2.7 IF5-year=3.2 （2024）	农林科学 2区 Top期刊：否 （2025）
2	The urine formed element instance segmentation based on YOLOv5n	SCIENTIFIC REPORTS 出版年：2024 出版日期：NOV 19 卷期：14 1 页码：- 文献号：28658 文献类型：Article	第一作者	B类	华南农业大学数信学院	SCI	IF2-year=3.9 IF5-year=4.3 （2024）	综合性期刊 3区 Top期刊：否 （2025）
3	Tracking and Behavior Analysis of Group-Housed Pigs Based on a Multi-Object Tracking Approach	ANIMALS 出版年：2024 出版日期：OCT 卷期：14 19 页码：- 文献号：2828 文献类型：Article	第一作者	A类	华南农业大学数信学院	SCI	IF2-year=2.7 IF5-year=3.2 （2024）	农林科学 2区 Top期刊：否 （2025）
4	RpTrack: Robust Pig Tracking with Irregular Movement Processing and Behavioral Statistics	AGRICULTURE-BASEL 出版年：2024 出版日期：JUL 卷期：14 7 页码：- 文献号：1158 文献类型：Article	第一作者	A类	华南农业大学数信学院	SCI	IF2-year=3.6 IF5-year=3.8 （2024）	农林科学 2区 Top期刊：否 （2025）
5	A passion fruit counting method based on the lightweight YOLOv5s and improved DeepSORT	PRECISION AGRICULTURE 出版年：2024 出版日期：JUN 卷期：25 3 页码：1731-1750 文献类型：Article	第一作者	A类	华南农业大学数信学院	SCI	IF2-year=6.6 IF5-year=7.4 （2024）	农林科学 2区 Top期刊：否 （2025）
6	MaskDis R-CNN: An instance segmentation algorithm with adversarial network for herd pigs	IET IMAGE PROCESSING 出版年：2023 出版日期：OCT 卷期：17 12 页码：3488-3499 文献类型：Article	第一作者	B类	华南农业大学数信学院	SCI	IF2-year=2.0 IF5-year=2.0 （2023）	计算机科学 4区 Top期刊：否 （2023）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明：如未盖章，报告无效

检索员：欧群

华南农业大学图书馆

2025-07-14

Engineering Village

EI 收录

1.

Accession number: 20224513059600

Title: **Behavior Recognition and Tracking Method of Group-housed Pigs Based on Improved DeepSORT Algorithm**

Title of translation: 基于改进 DeepSORT 的群养生猪行为识别与跟踪方法

Authors: Tu, Shuqin ; Liu, Xiaolong ; Liang, Yun ; Zhang, Yu ; Huang, Lei ; Tang, Yinjie

Author affiliation: College of Mathematics and Informatics, South China Agricultural University, Guangzhou; 510642, China
College of Electronic Engineering, South China Agricultural University, Guangzhou; 510642, China

Corresponding author: Zhang, Yu (zhangyu@scau.edu.cn)

Source title: Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery

Abbreviated source title: Nongye Jixie Xuebao

Volume: 53

Issue: 8

Issue date: August 2022

Publication year: 2022

Pages: 345-352

Language: Chinese

ISSN: 10001298

CODEN: NUYCA3

Document type: Journal article (JA)

Publisher: Chinese Society of Agricultural Machinery

Abstract: Behavior recognition and tracking of group-housed pigs are an effective aid to monitor pigs' health status in smart farming. In real farming scenarios, it is still challenging to automatically track the behavior of group-housed pigs by using computer vision techniques due to the pigs' overlapping occlusion and illumination change, which cause the identity (ID) of pig to switch wrongly. To improve the situation, an improved DeepSORT algorithm of behavior tracking based on YOLO v5s was proposed. The improvement of the algorithm included two parts. One was that the trajectory processing and data association were improved in the scene where there was a fixed number of pigs. This reduced ID switch and enhanced tracking stability. The other was that the behavior information from YOLO v5s detection algorithm was introduced into the tracking algorithm, thereby achieving behavior recognition of pigs in tracking. The experimental results showed that YOLO v5s algorithm had a mAP of 99.3% and an F1 of 98.7% in object detection. In terms of re-identification, the Top – 1 accuracy of the experiment was 99.88% . In terms of tracking, the method achieved a favorable performance with a MOTA of 91.9%, an IDF1 of 89.2% and an IDS of 33. Compared with the original DeepSORT algorithm, the proposed method improved 1.0 percentage points and 16.9 percentage points in MOTA and IDF1 respectively, and decreased 83.8% in IDS. This showed that the improved DeepSORT algorithm was able to achieve behavior tracking of group-housed pigs with stable ID. The method can provide technical support for no-contact automatic monitoring of pigs.

Number of references: 30

Main heading: Object detection

Controlled terms: Behavioral research - Data handling - Mammals - Object recognition - Tracking (position)

Uncontrolled terms: Behavior tracking - Behaviour recognition - DeepSORT - Group-housed pig - Health status - Multi-object tracking - Objects detection - Percentage points - Recognition methods - Tracking method

Classification code: 461.4 Ergonomics and Human Factors Engineering - 723.2 Data Processing and Image Processing - 971 Social Sciences

Numerical data indexing: Percentage 8.38E+01%, Percentage 8.92E+01%, Percentage 9.19E+01%, Percentage 9.87E+01%, Percentage 9.93E+01%, Percentage 9.988E+01%

DOI: 10.6041/j.issn.1000-1298.2022.08.037

Database: Compendex

查证员：王庆涛

Engineering Village

EI 收录

1.

Accession number: 20225113283169

Title: **Missing Seedling Localization Method for Sandalwood Trees in Complex Environment Based on YOLOv4 and Double Regression Strategy**

Title of translation: **基于 YOLOv4 和双重回归的复杂环境檀香树缺苗定位方法**

Authors: Zhang, Yu ; Xu, Haoran ; Niu, Jiajun ; Tu, Shuqin ; Zhao, Wenfeng

Author affiliation: College of Electronic Engineering, College of Artificial Intelligence, South China Agricultural University, Guangzhou; 510642, China

College of Mathematics and Informatics, South China Agricultural University, Guangzhou; 510642, China

Corresponding author: Tu, Shuqin (tushuqin@163.com)

Source title: Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery

Abbreviated source title: Nongye Jixie Xuebao

Volume: 53

Issue: 11

Issue date: November 2022

Publication year: 2022

Pages: 299-305+340

Language: Chinese

ISSN: 10001298

CODEN: NUYCA3

Document type: Journal article (JA)

Publisher: Chinese Society of Agricultural Machinery

Abstract: In the process of planting sandalwood trees on a large scale, there are problems such as low efficiency, high cost, and difficulty in the supervision of manual planting of missing seedlings, and the necessary companion plants for each sandalwood tree and other crops interspersed between the trees, further deepening the difficulty of checking and replenishing. For these problems, a seedling deficiency detection and precise localization method in complex environment was proposed based on YOLOv4 algorithm and double regression strategy. Firstly, the YOLOv4 target detection model was used to achieve sandalwood plant detection from remote sensing images collected by UAV. Then the missing seedling localization algorithm (MSL) was constructed based on the double linear regression and extended column line fixing strategy: arbitrary sandalwood trees were selected as the benchmark, column regions were divided according to the pixel coordinates, and column lines were fitted to the sandalwood trees in each column region by using linear regression; for the omitted sandalwood trees that were not classified into columns after fitting, the attribution was judged again with the extended regression line strategy, and the column lines were optimized by linear regression again. Finally, the missing seedlings were calculated and localized according to the spacing at the time of planting. The results showed that the precision was 86.82%, the recall was 82.25%, the F1-score was 84.47%, and the running time was 8.19 s, respectively. In summary, this method combined the rapidity of DJI UAV remote sensing image acquisition system, the accuracy of YOLOv4 algorithm and double regression strategy, which can be used to achieve real-time intelligent seedling deficiency detection and accurate localization of sandalwood trees under complex growth conditions.

Number of references: 29

Main heading: Object detection

Controlled terms: Aircraft detection - Linear regression - Remote sensing - Seed - Unmanned aerial vehicles (UAV)

Uncontrolled terms: Column line - Complex environments - Double linear regression - Localisation - Localization method - Missing seedling localization - Objects detection - Plantings - Sandalwood tree - YOLOv4

Classification code: 652.1 Aircraft, General - 716.2 Radar Systems and Equipment - 723.2 Data Processing and Image Processing - 821.4 Agricultural Products - 922.2 Mathematical Statistics

Numerical data indexing: Percentage 8.225E+01%, Percentage 8.447E+01%, Percentage 8.682E+01%, Time 8.19E+00s

DOI: 10.6041/j.issn.1000-1298.2022.11.030

Database: Compendex

查证员：王庆涛

Computers and Electronics in Agriculture 239 (2025) 110935

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

The OKByte-AR: a multi-stage MOT framework for identifying aggressive interactions in pigs

Shuqin Tu [a], Yuefei Cao [a], Liang Mao [b,*], Yun Liang [a], Hairan Yang [a], Baiyang Tang [a], Fang Yuan [c]

[a] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China
[b] Undergraduate School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen 518055, China
[c] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

ARTICLE INFO

Keywords:
pig aggressive interaction identification
multi-stage MOT framework
new evaluation metric
novel dataset

ABSTRACT

Aggressive interactions among pigs play a crucial role in establishing social hierarchies, and severe aggressive behaviors also lead to injuries and economic losses in commercial pig farming. Recent advances in deep learning-based methods, such as object detection and multi-object tracking (MOT), have enabled accurate pig aggressive behavior detection and tracking. However, these methods have difficulty in identifying the attacker and the attacked in an aggressive interaction. To address these challenges, we proposed OKByte-AR, a multi-stage MOT framework that integrated oriented bounding box (OBB) and keypoint detection, MOT, and post-tracking processing for pig aggression relationship identification. Additionally, we created a new evaluation metric, Aggression Relationship Evaluation Score (ARES), to fill the gap in existing metrics for accurately evaluating pig aggression relationship recognition. We constructed a novel dataset annotated with OBBs, keypoints, and aggression relationship links. Experimental results demonstrated that our method outperformed the baseline YOLO11 in detection performance, with mAP50 and mAP50-95 of 90.90 % and 82.70 %, respectively, which are 5.10 % and 15.30 % higher compared to YOLO11. For tracking performance, our method achieved the highest scores across four key MOT metrics compared to ByteTrack, OC-SORT, Deep OC-SORT, and BoT-SORT, with a Higher Order Tracking Accuracy (HOTA) of 85.78 %, Multiple Object Tracking Accuracy (MOTA) of 98.50 %, Identification F1 Score (IDF1) of 99.25 %, and only 2 Identity Switch (IDSW). In aggression relationship identification, our approach achieved an average ARES score of 66.92 % across the validation set. These results highlight the robustness and practicality of our approach for precision livestock farming and automated aggression monitoring. We shared our code at https://github.com/Ultran-cyf/OKByte-AR.

1. Introduction

Aggressive interactions among pigs are vital for establishing social hierarchies, as they determine dominant-submissive relationships (Meese and Ewbank 1973). Although such interactions are necessary for forming the group hierarchy, they often result in serious injuries, including skin trauma, infection, and even fatal injuries. The injured pigs have more difficulty in feeding, leading to lower growth rates (Stookey and Gonyou 1994). In stable social groups, aggressive interactions are much lower, and the risk of injury is lower (Jensen 1982). However, in commercial settings, it is common practice for pigs to be regrouped, which results in an unstable social group. Each time pigs are regrouped, a new hierarchy must be formed, which triggers a large number of aggressive interactions (Fraser et al., 2023). Moreover, intensive feeding environments expose pig groups to intensified competition, further exacerbating injuries (Stukenborg et al., 2011). Therefore, timely intervention during aggressive events is essential to reducing injuries and the risk of infection, accelerating hierarchy stabilization, and minimizing economic losses in pig farming.

With the rapid advancement of computer vision, many researchers have applied object detection techniques to study pig aggression. For example, Xia et al. (2025) proposed a hybrid model, PAB-Mamba-YOLO, integrating Mamba and YOLO architectures to detect aggressive behaviors such as mounting, nose hitting, tail biting, and ear biting in weaned piglets. Yan et al. (2024) developed an adaptive dual-modal deep neural network to detect the onset and offset of aggressive behaviors in group-housed pigs. Gao et al. (2023) introduced a CNN-GRU hybrid model and Ji et al. (2023) designed a Temporal Shift Module (TSM), respectively, to differentiate between four aggressive and five non-aggressive behaviors. Similarly, Fraser et al., 2023 employed a CNN

* Corresponding author.

https://doi.org/10.1016/j.compag.2025.110935
Received 14 May 2025; Received in revised form 3 August 2025; Accepted 24 August 2025
Available online 30 August 2025

with image differentials to recognize six aggressive and three non-aggressive behaviors. Chen et al. (2020) tested a method based on CNN and Long Short-Term Memory (LSTM) to recognize aggression.

While these studies have made significant contributions to the detection of pig aggression, their use of horizontal bounding box (HBB) leads to excessive background noise, reducing the precision of aggression target identification. To address this issue, some researchers (e.g., Fu et al., 2025, Li et al., 2025, Lu et al., 2024) have introduced oriented

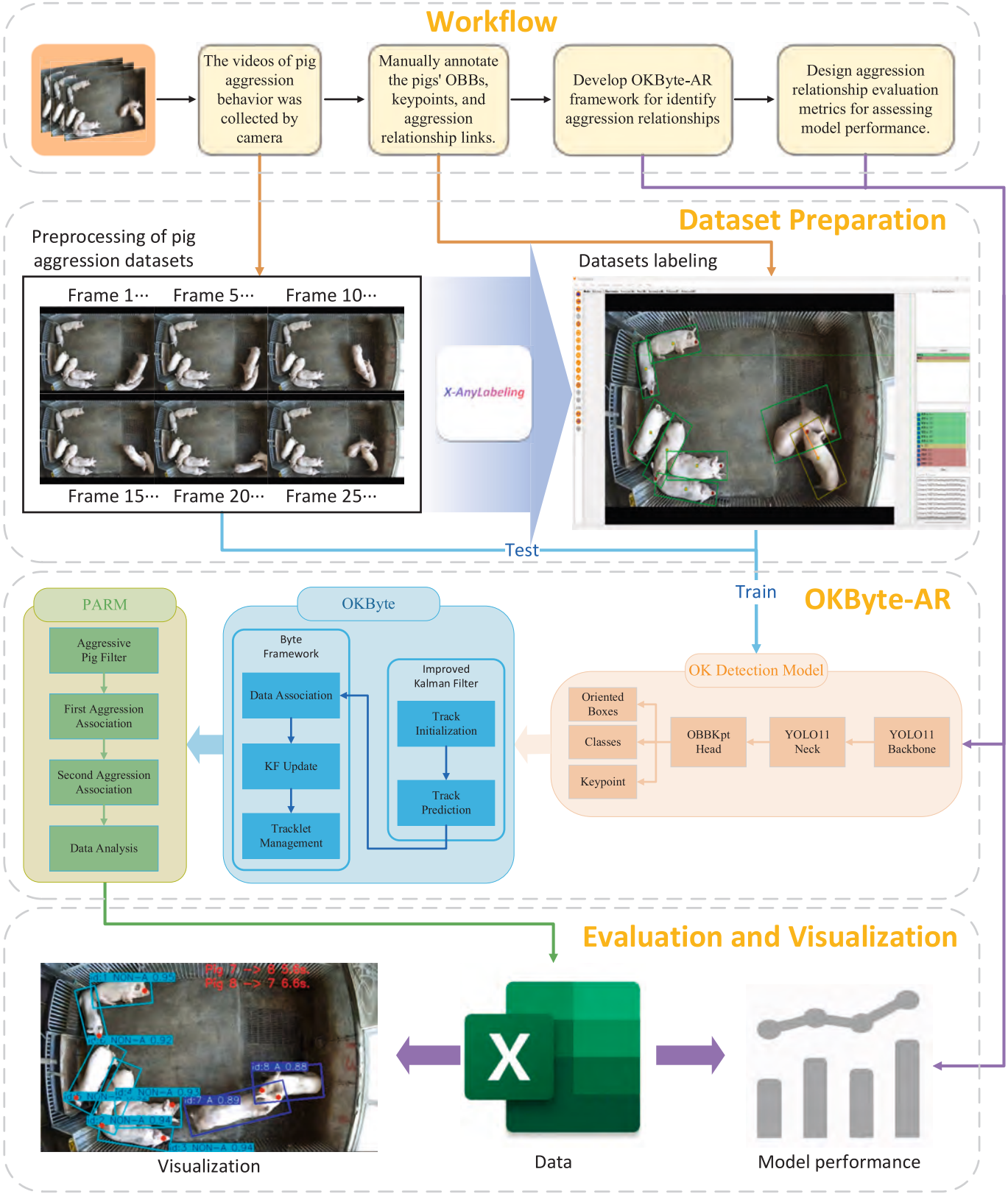


Fig. 1. Flowchart of the OKByte-AR for pig aggression relationship identification.

bounding box (OBB), which more accurately aligns with the object shape and minimizes redundant information. Inspired by these advances, this study adopts OBB for the localization of pigs.

Object detection techniques have achieved significant progress in the study of pig aggressive behavior. However, object detection can only extract spatial features of aggressive interactions and struggles to capture the temporal features, making it difficult to analyze the motion processes of pigs during aggressive interactions. Some studies have begun to explore the use of multi-target tracking (MOT) techniques to solve this problem. For instance, Wei et al. (2023) combined EMA-YOLOv8 with ByteTrack to track five types of aggressive behavior and calculate their durations for each pig. Liu et al. (2020) used a combination of CNNs and recurrent neural networks (RNNs) to extract spatiotemporal features and proposed a minimal unit tracking method to analyze aggression trajectories. These studies laid a foundation for dynamic behavior analysis; however, they did not identify the aggression relationships between pigs—i.e., the attacker and the attacked recognition in an aggression interaction.

Given the complexity of pig aggression, existing metrics are inadequate for evaluating the identification of aggression relationships. Therefore, we created a new metric, the Aggression Relationship Evaluation Score (ARES). ARES can capture two key characteristics of pig aggressive behavior: bilateral alternation and temporality. Bilateral alternation describes the phenomenon in which, during aggressive interactions, pigs may alternate roles between attacker and attacked or play both roles. Temporality refers to the duration of the pig attack interaction, which typically lasts several seconds to two minutes (McGlone 1985, D'Eath and Turner 2009). In complex scenarios, there are currently significant challenges to identifying identities in the interaction of aggressive behaviors.

Addressing the challenges mentioned above, we proposed OKByte-AR, a multi-stage MOT framework, which incorporated OBB and keypoint detection, MOT, and post-tracking processing for pig aggression relationship identification. Firstly, we used a combination of OBB and keypoint (OK) for pigs' detection, where OBB indicates the body and keypoint marks the mouth. Secondly, we developed the OKByte tracker to simultaneously track both the pigs' body OBBs and mouth keypoints. Thirdly, we designed a post-tracking processing module, the Pig Aggression Relationship Module (PARM), for pig aggression relationship identification and the duration of the aggressive interactions. Finally, we create the ARES metric to calculate the accuracy of aggressive relationships. Also, we used the OKByte-AR to record the duration of aggressive interactions, enabling the filtering of improbable interactions to accurately identify aggressive relationships.

The main contributions of this work are as follows: (1) To identify the attacker and the attacked in a pig aggression interaction while improving the detection accuracy, we designed the OK detection model by combining OBB and keypoint detection. (2) We proposed the OKByte-AR framework for pig detection, pig tracking, and aggression relationships identification. (3) To assess the accuracy of pig aggression relationship recognition, we created a new evaluation ARES metric. (4) We constructed a novel dataset for pig aggression identification, annotated with OBBs, keypoints, and aggression relationship links.

2. Materials and methods

The overall experimental process for identifying pig aggression relationships is shown in Fig. 1. First, we prepared the dataset by annotating all video recordings using the X-AnyLabeling (Wang 2023) tool. The annotation process incorporates three elements: OBBs, keypoints, and aggression relationship links. Specifically, OBBs are used to represent the body of each pig, keypoints mark the position of the pig's mouth, and aggression relationship links relationship between the attacker and the attacked. Second, we independently developed the OK detection model, OKByte tracker, and PARM, which were then integrated into a unified framework named OKByte-AR. Third, we evaluated the performance of each component in the OKByte-AR framework and visualized the corresponding output results.

2.1. Dataset preparation

In this study, pig aggression behavior videos were obtained from Ji et al. (2023). The dataset was collected from the Research Base for Pig Nutrition and Environmental Control in Rongchang District, Chongqing City. To create a standardized dataset, we preprocessed the raw video, retaining only segments that contained aggressive behaviors. Each segment includes a 3 to 5-second non-aggressive behavior window to enhance contextual information before and after the aggression event. We then annotated the dataset as illustrated in Fig. 2(a), each pig is labeled with an OBB and a mouth keypoint, both carrying an ID and category. If an aggressive interaction occurs, the attacker is linked to the attacked with an aggression relationship link as shown by the orange line in Fig. 2(b).

We labeled the behavioral categories based on the definitions of aggressive and non-aggressive behaviors outlined in Gao et al. (2023). The category of Aggression includes biting, knocking, treading, and chasing, and the category of NON-Aggression includes feeding, drinking, lying, etc. Specific behavioral descriptions and data examples are shown in Table 1 and Fig. 3.

Pigs' non-rigid bodies exhibit varying degrees of deformation during interactions (Chen et al., 2018), leading to multiple valid OBB annotation options. As shown in Fig. 4(a)–(c), when the pig is undeformed, it is easy to align the edges of the OBB with the pig's body. However, multiple annotation choices become possible under moderate or severe deformation, posing a challenge to constructing a standardized dataset.

To address this challenge, we propose a head–tail aligned annotation strategy. As shown in Fig. 5, the yellow dashed line represents the hypothetical connection between the pig's head and tail, and the red edge of the OBB indicates its long side. By aligning the red edge parallel to the yellow dashed line, the OBB can compactly and accurately represent the pig's body, effectively reducing the background, and also provides a standard way to construct the dataset.

Finally, we constructed a dataset consisting of 31 video clips with an average duration of 39 s, which was randomly split into training and validation sets at 7:3.

2.2. The OKByte-AR framework

Building upon the traditional tracking-by-detection (TBD) framework, we modified and extended it to develop OKByte-AR, as shown in the third row of Fig. 1. The traditional TBD framework consists of a detector for object detection and classification recognition tasks, and a tracker for ID identifying and behavior tracking task, while our framework adds a PARM post-tracking module to support pig aggression relationship recognition work after MOT task.

2.2.1. The OK detection model

Based on YOLO11 (Khanam and Hussain 2024), we developed the OK detection model, which integrated rotated bounding boxes and keypoints for joint detection. The overall architecture of the proposed model is illustrated in Fig. 6.

In the backbone and neck, our model follows the original YOLO11 architecture, which includes convolutional blocks, C3K2 modules, the SPPF module in the backbone, and a PAN structure with C2PSA modules in the neck. The C3K2 module enhances information flow by splitting feature maps and applying efficient small-kernel convolutions. The SPPF module captures multi-scale contextual information through sequential max pooling, improving small object detection. The C2PSA module in the neck introduces spatial attention to emphasize important regions, enhancing robustness under occlusion and dense scenarios.

In the head, we propose a novel OBBKpt Detect module, which includes four parallel branches for predicting object category, bounding

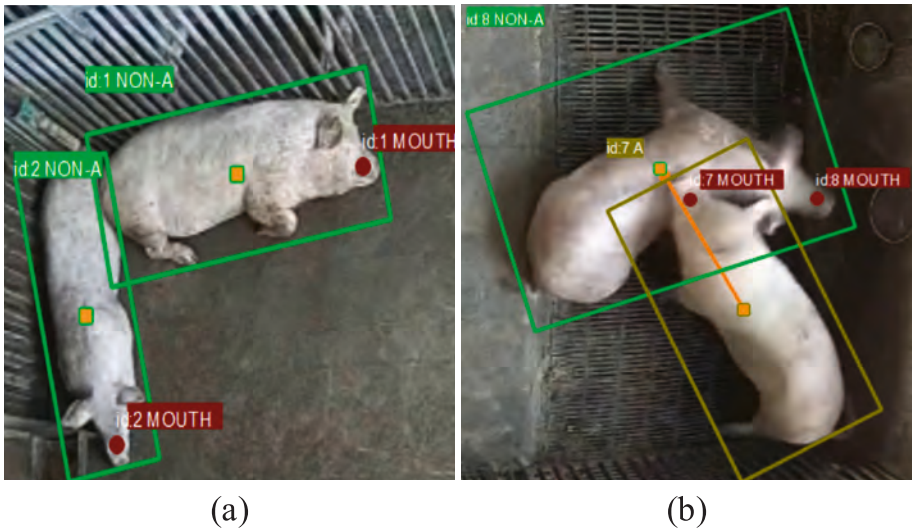


Fig. 2. Illustration of dataset annotation.

Table 1
Definition of aggressive behavior.

Classification	Type	Description
Aggression	Biting	The aggressive pig uses its mouth to bite the victims
	Knocking	The aggressive pig uses its head to hit the victims
	Treading	The aggressive pig uses its feet to trample the victims
	Chasing	The aggressive pig intentionally continues to pursue and attack other pigs.
Non-Aggression	others	Like feeding, drinking, lying, playing, mounting

box size and center, orientation angle, and keypoint location. This custom detection head enables the model to provide richer spatial cues for downstream relationship analysis.

2.2.2. *The OKByte tracker*

We designed the OKByte tracker based on the BYTE framework (Zhang et al., 2022), retaining its original efficiency while incorporating keypoint tracking and improving tracking accuracy. As illustrated in Fig. 7, the OKByte tracker consists of three main steps: input preprocessing, data association, and tracklet management.

(1) Input preprocessing.

The inputs consist of the detection results D of the OK detection model and the tracklets T_{t-1} of the previous frame $t-1$. The detection results D are separate into D_{high} and D_{low} based on the OBB confidence threshold τ. The tracklets T_{t-1} are processed by an improved Kalman Filter (KF) to predict tracks. The original BYTE framework inherited the 7-dimensional state vector from the SORT (Bewley et al., 2016) to predict each track, as in Eq. (1).

$$x = \left[x_c,\ y_c,\ s,\ a,\ \dot{x}_c,\ \dot{y}_c,\ s\right]^T \tag{1}$$

Where (x_c, y_c) are the 2D coordinates of the object center in the image plane, s is the bounding box scale (area) and a is the bounding aspect ratio. The remaining three variables represent their velocities.

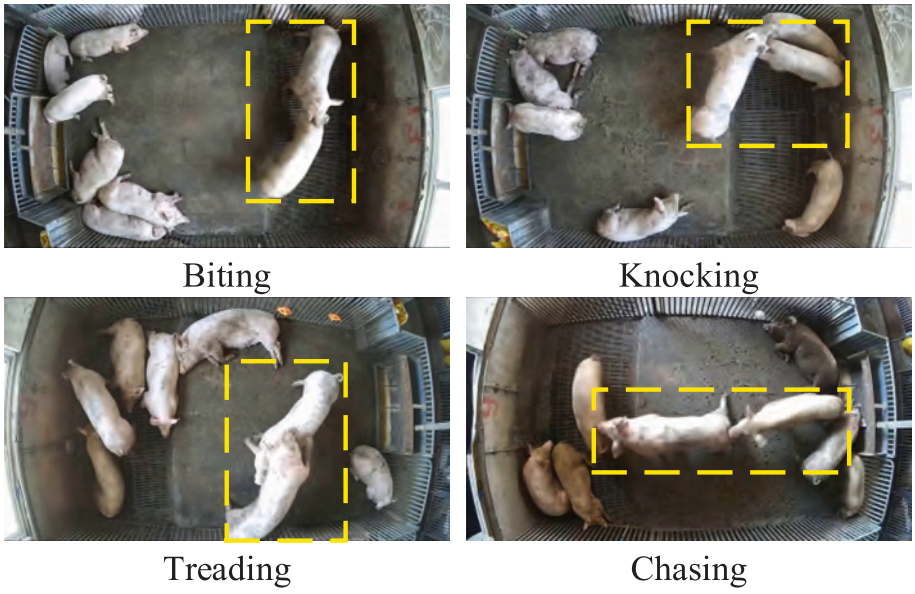
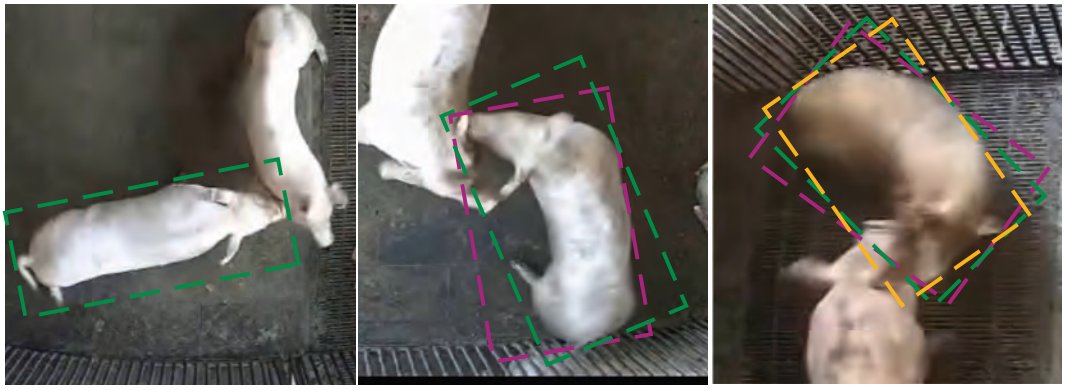
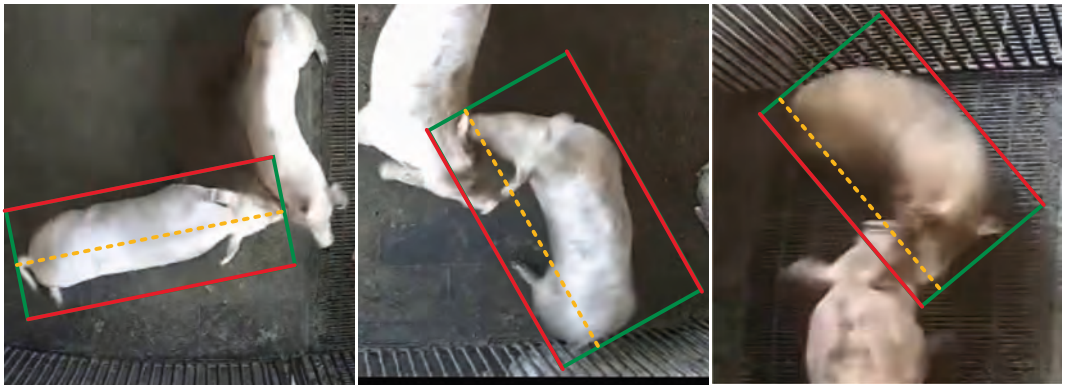


Fig. 3. Example of pig aggressive behavior.



(a) Undeformed (b) Moderate deformation (c) Severe deformation

Fig. 4. Labeling of OBB with different levels of deformation.



(a) Undeformed (b) Moderate deformation (c) Severe deformation

Fig. 5. Illustration of the head–tail alignment annotation strategy.

In contrast, our improved KF extends the state vector to 10-dimensions, as shown in Eq. (2).

$$x = \left[x_{obb},\ y_{obb},\ w,\ h,\ \dot{x}_{obb},\ \dot{y}_{obb},\ \dot{w},\ \dot{h},\ x_{kpt},\ y_{kpt}\right]^T \tag{2}$$

Where (x_{obb}, y_{obb}) are the 2D coordinates of the object center in the image plane, (w, h) are the width and height of the bounding box, and the next four variable $\left(\dot{x}_{obb}, \dot{y}_{obb}, \dot{w}, \dot{y}\right)$ represent their velocities. The remaining two variables $\left(x_{kpt}, y_{kpt}\right)$ represent the coordinates of the keypoint.

Replacing (s, a) with (w, h) enables more accurate track prediction, as illustrated in Fig. 8. The green boxes represent the tracking boxes using the improved KF, while the yellow dashed boxes represent the tracking boxes before the improvement. The tracking bounding box of the improved KF can fit the target more accurately.

The coordinates $\left(x_{kpt}, y_{kpt}\right)$ are used for keypoint tracking. Instead of using KF to predict the keypoint positions, we directly embed the detection results into the tracks, replacing the previous results whenever new ones are available.

(2) Data association

The first association is performed between the high-score detections D_{high} and the tracklets T_{t-1}, the similarity is computed by the OBB IoU distances. Then, the unmatched tracks T_{remain} from the first association and the detections D_{low} are matched by the second IoU-based association. Finally, we get the matched tracks T_t, unmatched tracks $T_{re-remain}$ and unmatched detection D_{remain} and $D_{re-remain}$, and enter them all into the Tracklet management module.

(3) Tracklet management.

The tracklet management module includes four situations: For the matched tracks T_t, update their trajectories using the improved KF. For the unmatched tracks $T_{re-remain}$, move them into the set T_{lost}. For the unmatched detections D_{remain}, create new tracks and assign them a new ID. For the unmatched detections $D_{re-remain}$, simply delete them. In the end, the current frame tracklets are output.

2.2.3. The PARM module

We created the PARM module to identify aggression relationships among pigs. The detailed workflow and corresponding pseudocode are presented in Fig. 9 and Algorithm 1. The PARM module consists of three main steps: input preprocessing, aggression association, and data analysis.

(1) Input preprocessing.

The input is the tracklets T_t of the current frame t from OKByte tracker. The aggression filter selects attacker tracks classified as Aggression from T_t and adds them to the queue T_a. Each T_{ai}, where i is the index of queue T_a, is then sequentially used for aggression association.

(2) Aggression association

Each attacker track T_{ai} undergoes two rounds of aggression associ-

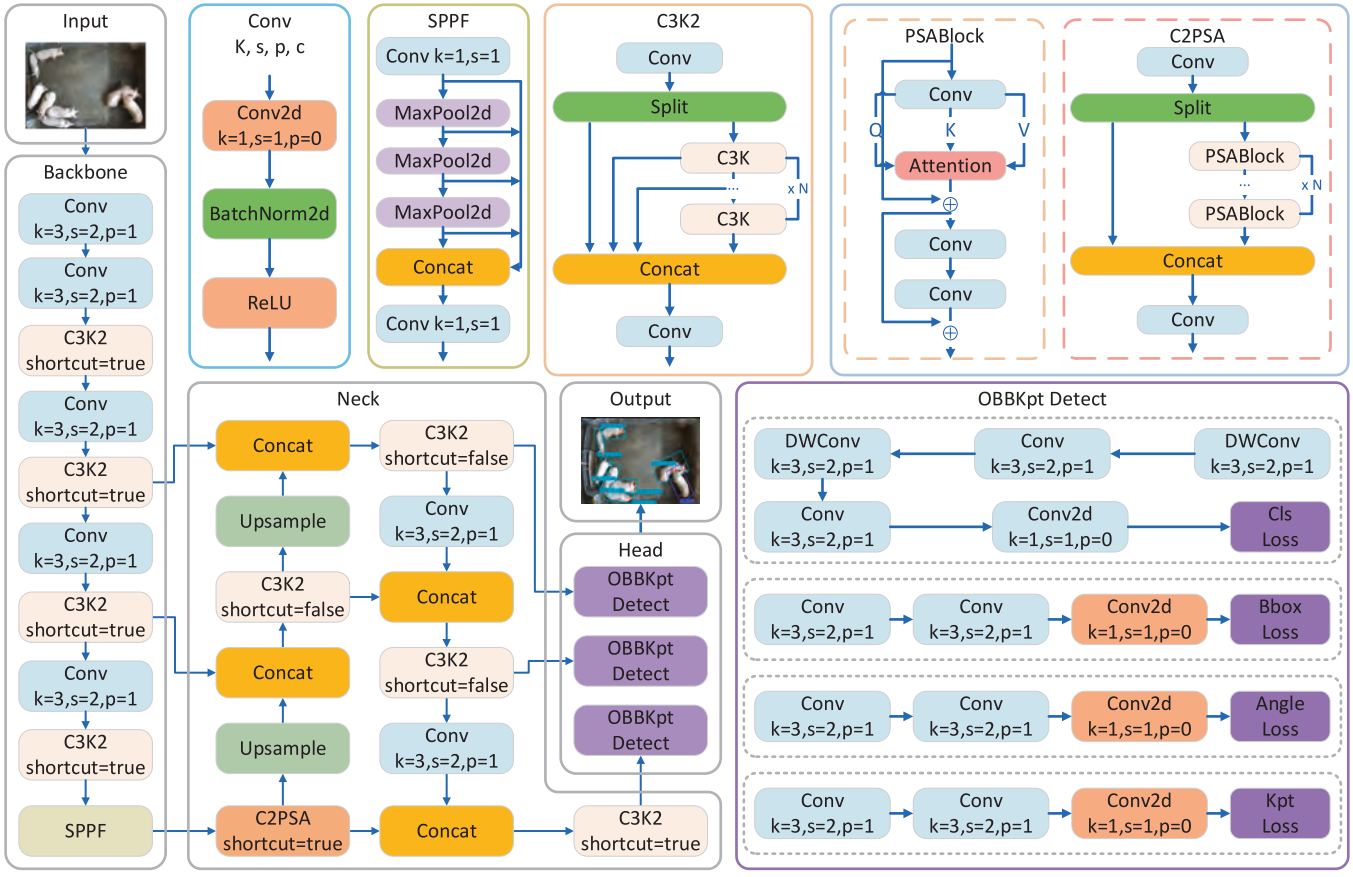


Fig. 6. The OK detection model.

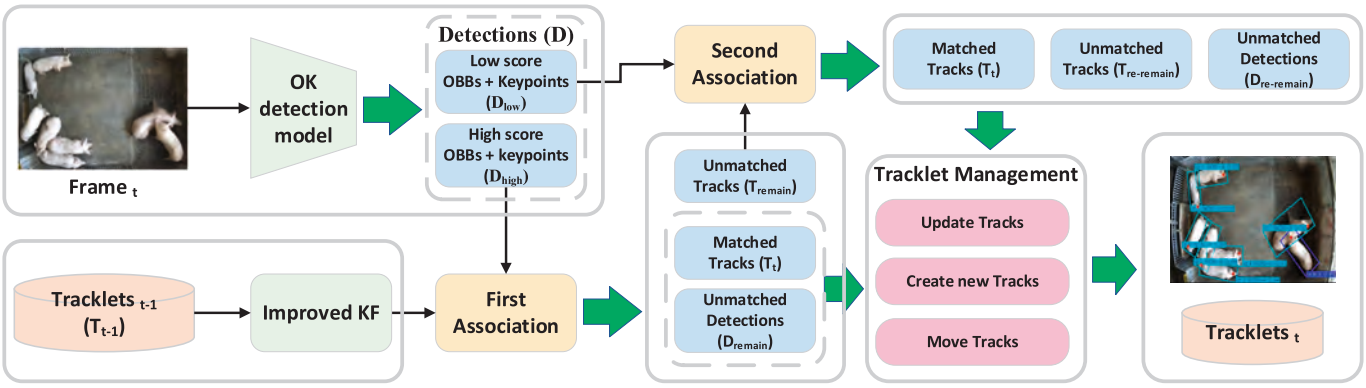


Fig. 7. Workflow of OKByte tracker.

ation with all other tracks $T_{t\backslash ai}$, where $T_{t\backslash ai}$ denotes all tracks in frame t except T_{ai}. The first aggression association is performed by matching the keypoint position of the attacker's track T_{ai} with the OBB positions of each track T_j in $T_{t\backslash ai}$. If the keypoint is inside an OBB or the distance between the keypoint and the OBB is less than the error margin ε, the matched relationship tuple (T_{ai}, T_j) is added to the list AR_1.We define this error margin ε because a pig's mouth is not a single point but a small area. Based on our camera height, the average mouth area of a pig is approximately 5 pixels, so ε is set to 5.

The keypoint of an attacker track T_{ai} may be inside the OBBs of multiple tracks T_j as in Fig. 10(a), or the distance between the keypoint and multiple OBBs is less than the error margin ε as in Fig. 10(b). Therefore, the second aggression association is to select the relationship tuples from AR_1, where the keypoint of each attacker track T_{ai} is most deeply inside the OBB of a track T_j, and records the corresponding time to be added to AR_2 as the final aggression relationship.

(3) Data analysis.

The PARM outputs identified aggression relationships AR_{final} along with their corresponding durations from AR_2 which durations lasting at least 2s. This threshold is introduced to filter out spurious aggression

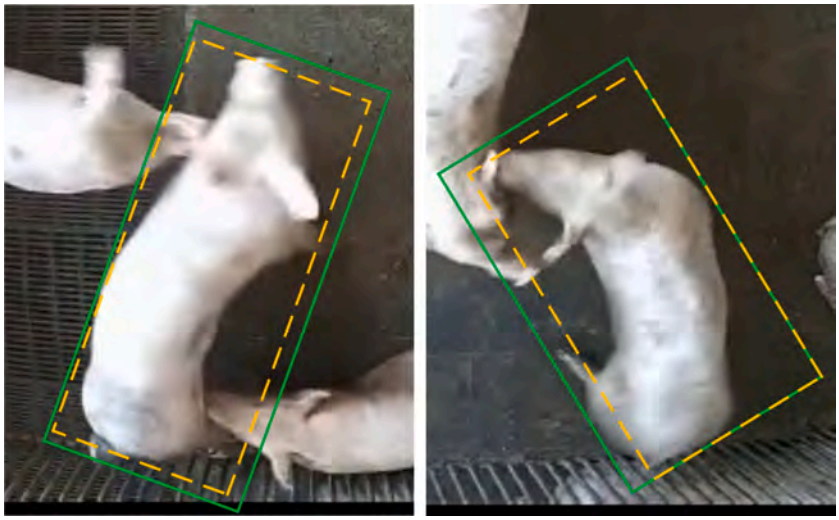


Fig. 8. Comparison of the effect before and after KF improvement.

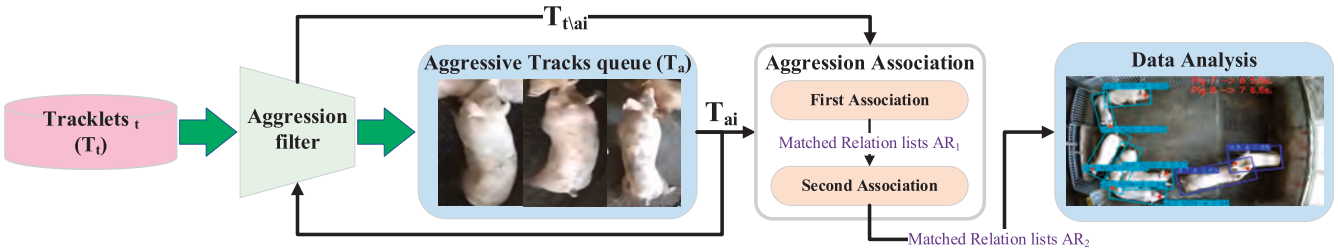


Fig. 9. The Workflow of PARM module.

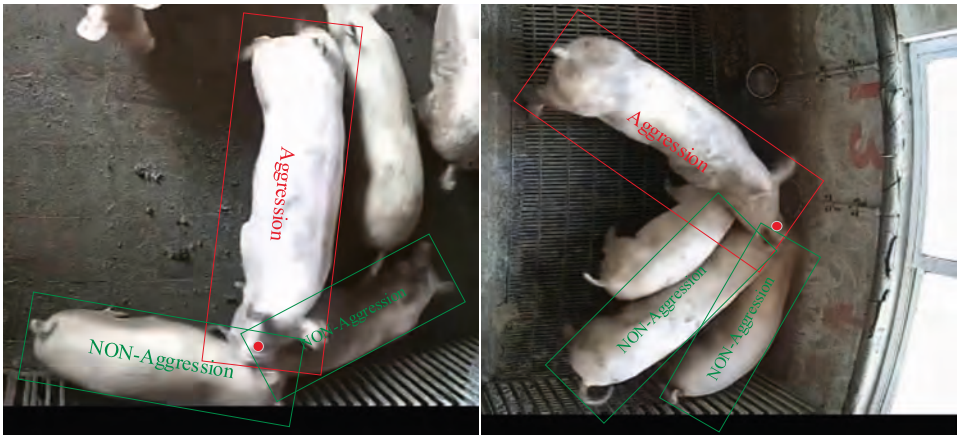


Fig. 10. Two scenarios that may be encountered in the second aggression association.

events.

Algorithm 1: Pseudo-code of PARM

Input: Current frame tracklets T_t

Output: Pig aggression relationships and their time duration AR_{final} *and time*

Initialization: $T_a = [T_{a1},\ T_{a2},\ T_{a3}, \cdots] \leftarrow \emptyset; AR_1 = [(T_{ai},\ T_j), \cdots] \leftarrow \emptyset; AR_2 = \{(T_{ai},\ T_j)\ :\ frame, \cdots\} \leftarrow \emptyset; AR_{final} = (T_{ai},\ T_j); time = frame/5$

for *frame in Video* **do**
  for T_i *in* T_t **do**
    /* **filter out tracks classified as Aggression** */
    $T_a \leftarrow filter(T_t,\ Aggression)$
  end
  /* **first aggression association** */
  for T_{ai} *in* T_a **do**
    /* **filter out tracks except** T_{ai} */
    $T_{t\backslash ai} \leftarrow filter(T_t,\ T_{ai})$
    for T_j *in* $T_{t\backslash ai}$ **do**
      if T_{ai}*.keypoint in* T_j*.obb* **or** T_{ai}*.keypoint to* T_j*. obb distance* $< \varepsilon$ **then**
        $AR_1 \leftarrow (T_{ai},\ T_j)$
      end
    end
  end
  /* **second aggression association** */
  for $(T_{ai},\ T_j)$ *in* AR_1 **do**
    /* **implemented using OpenCV's pointPolygonTest function** */
    $AR_2 \leftarrow (T_{ai},\ T_j) : frame + 1$ **if** T_{ai}*.keypoint* **is in the innermost of** T_j*.obb*
  end
  /* **output aggression relationships and time duration** */
  for AR_{final}, *frame in* AR_2 **do**
    if $frame \geqslant 10$ **then**
      $time = frame/5$
      Return AR_{final} *and time*
    end
  end
end

3. Results and analysis

3.1. Evaluation metrics

To objectively evaluate the performance of each task in the OKByte-AR, we comprehensively utilized detection metrics, MOT metrics, and our created aggression relationship accuracy metric.

The detection performance was evaluated using Precision (P), Recall (R), mAP50, and mAP50-95, where mAP50 represents the mean Average Precision at an IoU threshold of 0.5 and mAP50-95 across IoU thresholds from 0.5 to 0.95. In the tracking task, we selected four key evaluation metrics: Higher Order Tracking Accuracy (HOTA, Luiten et al., 2021), Multiple Object Tracking Accuracy (MOTA, Bernardin and Stiefelhagen 2008), Identification F1 Score (IDF1, Milan et al., 2016), and Identity Switch (IDSW, Milan et al., 2016) for performance assessment.

HOTA decomposes tracking into detection, association, and localization. MOTA offers an intuitive evaluation by considering false positives, false negatives, and identity switches, reflecting the tracker's effectiveness in object detection and trajectory maintenance. IDF1 focuses on the continuity and accuracy of the trajectory, responding to whether the objects remain consistently tracked across frames. IDSW measures the number of times a tracked object is incorrectly assigned a new identity, highlighting failures in identity association. A lower IDSW value indicates better tracking robustness, while higher HOTA, MOTA, and IDF1 scores suggest good overall tracking performance.

In the aggression relationship identification task, we created the Aggression Relationship Evaluation Score (ARES). Considering the temporality of pig aggressive behavior, ARES reflects this characteristic by summing the per-frame accuracy across time. It was as shown in Eq. (3).

Table 2
Performance comparison of the OK and the baseline YOLO11 detection model.

Detection model	Class	P ↑	R ↑	mAP50 ↑	mAP50-95 ↑
baseline YOLOv11 detection model	Aggression	68.80 %	72.20 %	74.90 %	59.30 %
	NON-Aggression	91.70 %	97.00 %	96.70 %	75.50 %
	ALL	80.20 %	84.60 %	85.80 %	67.40 %
OK detection model	Aggression	**86.80 %**	**79.30 %**	**83.10 %**	**75.00 %**
	NON-Aggression	**97.50 %**	**98.50 %**	**98.60 %**	**90.30 %**
	ALL	**92.10 %**	**88.90 %**	**90.90 %**	**82.70 %**

$$ARES = 1 - \frac{\sum_t (FP_t + FN_t)}{\sum_t GT_t} \tag{3}$$

In frame t, FP_t represents falsely identified aggression relationships, FN_t denotes aggression relationships not recognized by PARM, and GT_t is the total number of ground truth aggression relationships.

Moreover, considering the bilateral alternation characteristic of pig aggressive behavior, when two pigs are attacking each other simultaneously, we split it into two unidirectional aggression relationships for evaluation. Specifically, if Pig 1 and Pig 2 are both the attackers and the attacked at the same time, it is divided into "Pig 1 attacks Pig 2" and "Pig 2 attacks Pig 1". ARES provides a balanced evaluation of aggression relationship identification by penalizing the incorrectly identified and missed relationships, with higher values indicating better performance.

3.2. Experimental results

3.2.1. Performance of the OK detection model

We compared the detection performance of the OK detection model with the baseline YOLO11 detection model. The results are shown in Table 2. Across all categories, the OK detection model achieved 92.10 %, 88.90 %, 90.90 %, and 82.70 % for P, R, mAP, and mAP50-95, representing improvements of 11.90 %, 4.30 %, 5.10 %, and 15.30 %, respectively, over the baseline YOLO11 detection model. For categories Aggression and NON-Aggression, the OK detection model outperformed the baseline YOLO11 detection model by an average of 12.30 % and 6.00 %, respectively, across the four metrics.

We further observed some examples in Fig. 11 where the HBB representation contained a large amount of information from other pigs when the pig engaged in aggressive interactions, as shown by the yellow area. In contrast, the OBB representation had almost no redundant information. This is the fundamental reason why the OK detection model performs better than the baseline YOLO11 model, which also provides an excellent foundation for subsequent tracking and aggressive relationship identification.

In addition to object detection, we also evaluated the keypoint detection performance of the OK detection model. The results are summarized in Table 3. Across all categories, the OK detection model achieved 56.00 % for P, 53.50 % for R, 52.50 % for mAP50, and 29.20 % for mAP50-95. A comparison between categories Aggression and NON-Aggression revealed that the overall performance was adversely affected by the low scores in the NON-Aggression category.

Through video analysis, we observed that compared to aggressive pigs, non-aggressive pigs posed greater challenges for mouth keypoint localization, as illustrated in Fig. 12. The yellow markings indicate pigs whose keypoints are difficult to detect, while the green markings represent pigs whose keypoints are easily detectable. Due to camera angles and occlusions, the mouths of non-aggressive pigs are often partially or fully obscured. In contrast, aggressive pigs primarily use their heads or mouths to initiate attacks, making their mouth keypoints more prominent and easier to detect. However, since this study primarily focused on aggressive pigs, the scores in category Aggression were of greater significance.

Fig. 11. HBB and OBB representation.

Table 3
Performance of keypoint detection.

Class	P ↑	R ↑	mAP50 ↑	mAP50-95 ↑
A	**64.50 %**	**59.00 %**	**63.00 %**	**29.30 %**
NON-A	47.50 %	48.00 %	42.00 %	29.10 %
ALL	56.00 %	53.50 %	52.50 %	29.20 %

3.2.2. Performance of the OKByte

To evaluate the MOT performance, we compared the proposed OKByte tracker with ByteTrack (Zhang et al., 2022), OC-SORT (Cao et al., 2023), Deep OC-SORT (Maggiolino et al., 2023), and BoT-SORT (Aharon et al., 2022). The results are summarized in Table 4. OKByte achieved the highest HOTA of 85.78 %, MOTA of 98.50 %, and IDF1 of 99.25 %, significantly outperforming the other methods. Compared to the second-best BoT-SORT, OKByte improves HOTA by 1.75 %, MOTA by 1.10 %, and IDF1 by 3.69 %, while reducing IDSW from 37 to 2. These results validate the effectiveness of our OKByte in MOT, especially in scenarios where pigs frequently interact, overlap, or engage in aggressive behaviors. Furthermore, it lays a solid foundation for aggression relationship identification.

3.2.3. Performance of the PARM

To assess the effectiveness of our PARM, we evaluated the ARES value on the validation set, and the results are shown in Table 5 with an average score of 66.92 %. Sequences 01 and 10 achieved the best ARES values at 85.38 % and 84.62 %, respectively. In these sequences, the pigs involved in aggressive interactions exhibited prolonged, continuous attacking behavior, enabling PARM to associate their interactions effectively. As illustrated in Fig. 13(a), when pig2 attacked pig1, it continued without interruption for 24.4 s. This clear and sustained aggression led to a prediction nearly identical to the ground truth, with a total duration error of zero.

Sequences 02, 03, 04, 06, and 09 showed moderate performance,

Table 4
MOT Performance comparison of OKByte with other MOT algorithms.

Tracker	HOTA ↑	MOTA ↑	IDF1 ↑	IDSW ↓
ByteTrack	72.63 %	89.96 %	87.45 %	75
OC-SORT	79.05 %	93.89 %	89.43 %	84
Deep OC-SORT	77.54 %	93.10 %	88.22 %	78
BoT-SORT	84.03 %	97.39 %	95.55 %	37
OKByte	**85.78 %**	**98.50 %**	**99.25 %**	**2**

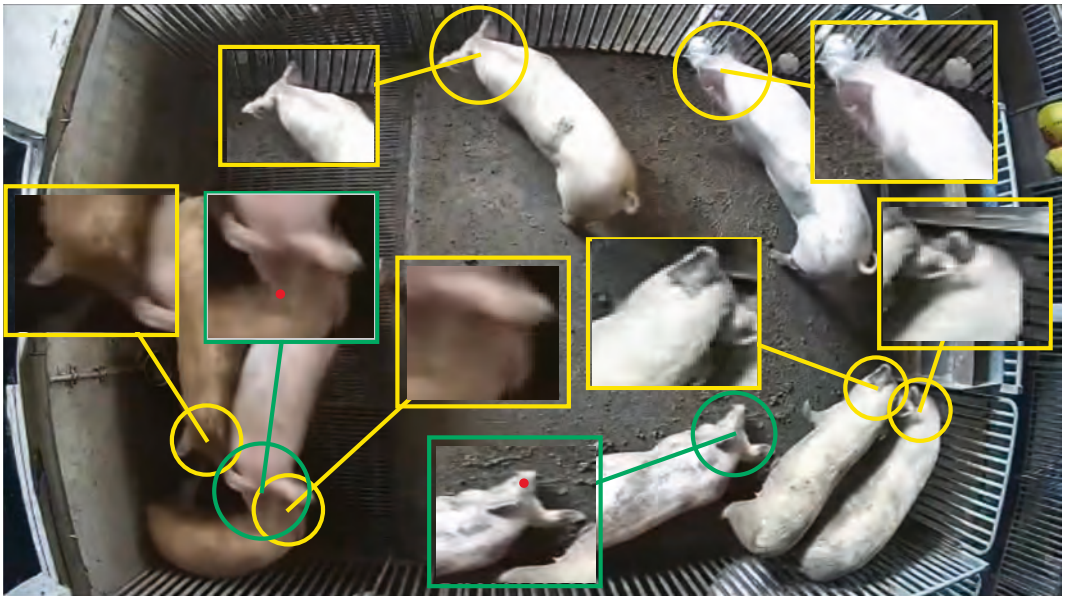


Fig. 12. Illustration of keypoint detection difficulty for A and NO categories.

Table 5
Performance of aggression relationship identification.

Sequence	01	02	03	04	05	06	07	08	09	10
ARES ↑	85.38 %	60.44 %	71.63 %	68.97 %	52.86 %	69.19 %	58.13 %	57.50 %	60.47 %	84.62 %
Average ↑	66.92 %									

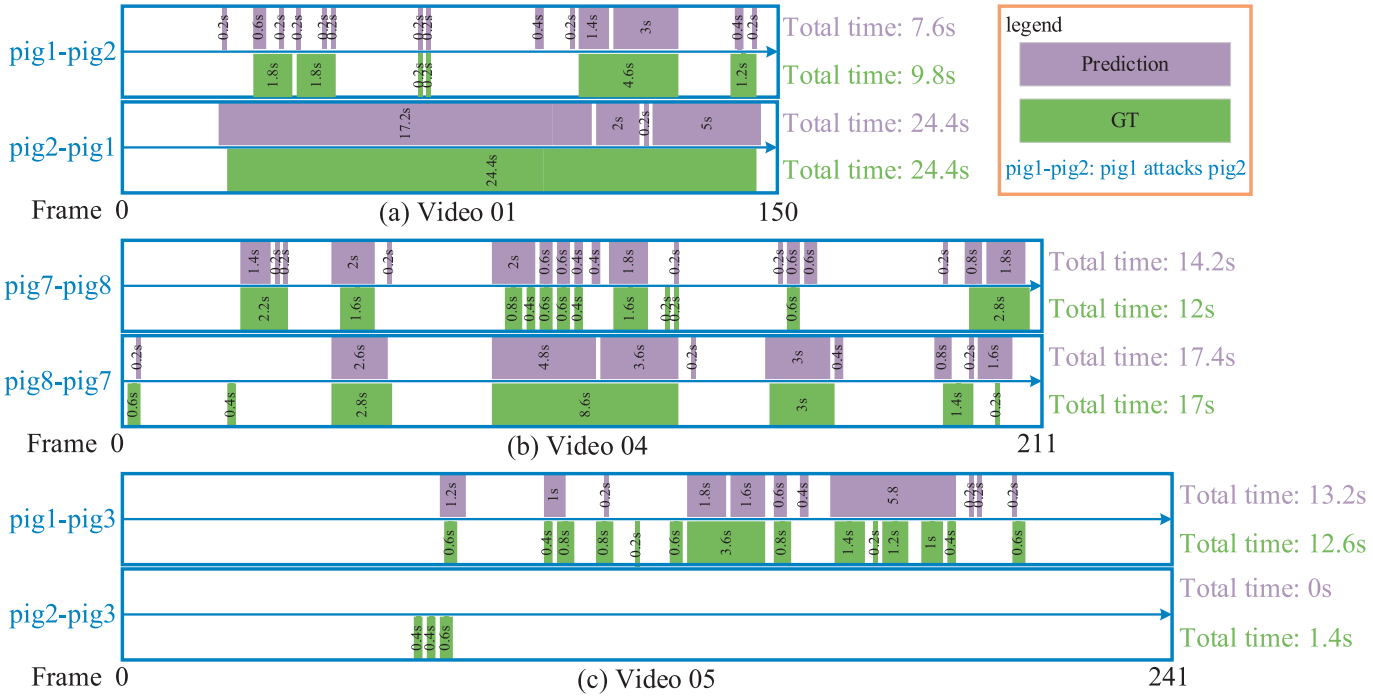


Fig. 13. Temporal diagrams of pig aggression relationships in video sequences with the highest, moderate, and lowest ARES value.

with an average ARES of 66.14 %. In these videos, pigs involved in aggressive interactions demonstrated a certain degree of intermittent attacks. Such discontinuity can make aggressive features less inconspicuous and affect the accuracy of the aggression association. As shown in Fig. 13(b), pig7 and pig8 exhibited different degrees of intermittent aggression. When pig7 attacked pig8, the attacks were highly fragmented, resulting in some deviation between prediction and ground truth, with a total duration error of 2.2 s. In contrast, pig8 attacked pig7 with a relatively low degree of interruption, and the total time error was only 0.4 s.

Sequences 05, 07, and 08 had lower performance, with ARES of 52.86 %, 58.13 %, and 57.50 %, respectively. In these sequences, the attack duration of the pigs was shorter and more fragmented compared with other sequences. Under such conditions, it is much more challenging to identify aggressive relationships clearly. As shown in Fig. 13 (c), pig2′s attack on pig3 lasted only 1.4 s, which was directly filtered out by PARM due to being shorter than the 2-s threshold. Additionally, pig1′s intermittent attacks on pig3 led to further prediction inaccuracies.

Despite the average ARES score being 66.92 % on the validation set, the qualitative results from Fig. 13(a)–(c) demonstrate a high level of alignment between the predicted and ground-truth aggression durations. The prediction errors for start and end frames were also minimal. Even in Sequence 05, which had the lowest ARES score, Fig. 13(c) shows that the predicted aggression duration still closely matches the ground truth, with only a 2-second error in total duration. This indicates that our method is sufficiently effective in identifying aggressive interactions among pigs.

4. Discussion

With the rapid advancements in computer vision, numerous deep learning-based methods have been developed for the analysis of pig aggressive behavior. However, most existing studies primarily focus on detecting and tracking aggressive actions, without addressing the identification of aggression relationships, a critical component for understanding interactive dynamics within group housing systems.

For instance, Xia et al. (2025), Yan et al. (2024), and Ji et al. (2023) developed detection models that identify general aggression events based on posture or motion cues. Wei et al. (2023) used MOT technology to record the duration of aggressive behavior in pigs. While these methods are effective for coarse-grained behavior classification, they can only identify the occurrence of aggressive events, thereby limiting their utility in detailed behavior analysis and welfare assessment.

Our previous work (Tu et al., 2024a, Tu et al., 2024b) employed an MOT framework to associate pig behaviors with individual identities, demonstrating the feasibility of applying MOT to pig tracking and behavior monitoring. However, directly applying MOT to the study of pig aggressive behavior has limitations, as it cannot capture the interactions between individual pigs' behaviors.

In this study, we developed the OKByte-AR framework, which integrates oriented bounding box (OBB) and keypoint detection, MOT, and post-tracking analysis into a unified pipeline for aggression relationship identification. The proposed OK detection model enhances robustness in crowded and occluded scenes by reducing background redundancy, thereby improving identity consistency in downstream multi-object tracking tasks and providing a crucial foundation for the task of identifying aggressive behavior interactions in pigs.

Despite the promising results, our method has several limitations

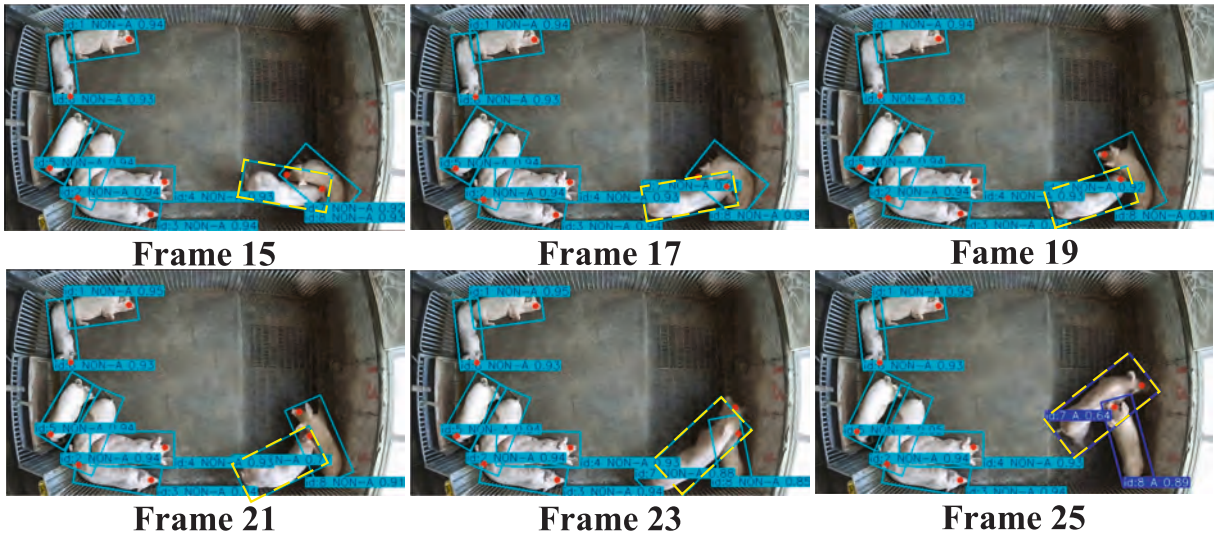


Fig. 14. Numerous misclassifications in Video 02.

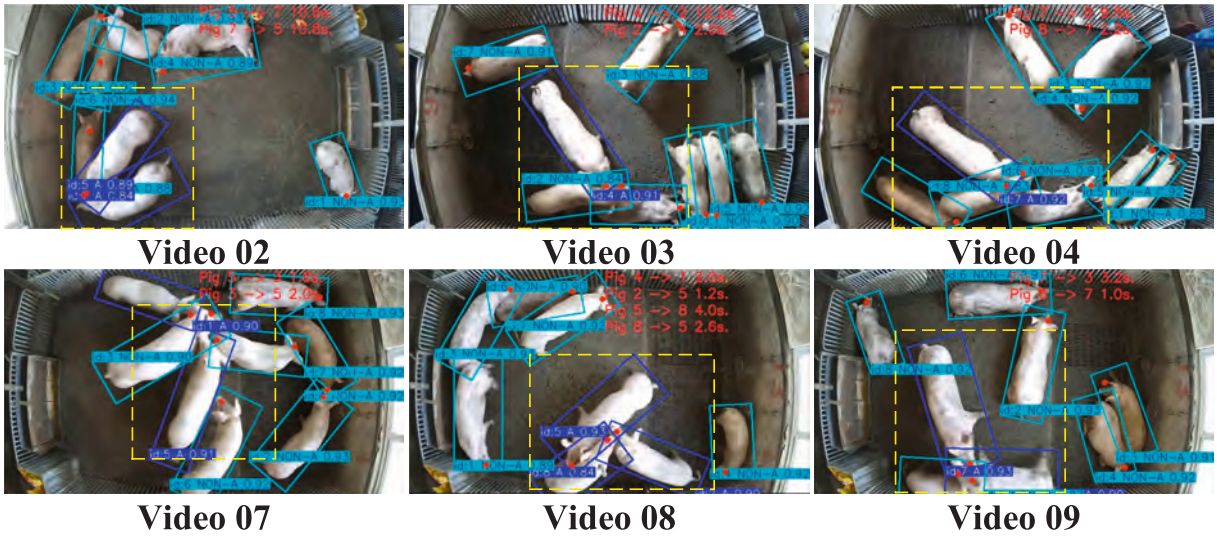


Fig. 15. Some examples of a keypoint within multiple OBBs.

that warrant further investigation.

First, as the proposed aggression relationship identification framework is built upon the outputs of an upstream MOT pipeline, its performance is inherently influenced by errors in detection and tracking. Specifically, incorrect classification of behavior categories during detection or identity switches during tracking can propagate downstream and negatively affect the accuracy of relationship identification. As shown in Fig. 14, it presents a segment from video 02, in which the yellow dashed box indicates that a pig in an aggressive state was misclassified as non-aggressive for an extended period.

Second, under crowded conditions, a single keypoint can fall within the boundaries of multiple oriented bounding boxes, especially when pigs are closely packed or overlapping, as shown in Fig. 15. This phenomenon is common in many videos, and such spatial positioning errors may lead to incorrect association, thereby reducing the precision of relationship recognition.

To address these limitations, future work will focus on developing an end-to-end aggression relationship identification framework, which directly integrates detection, tracking, and interaction recognition into a unified model, thereby reducing error propagation between stages. Furthermore, we plan to explore multimodal approaches to improve the accuracy of aggression relationship identification.

5. Conclusion

In this study, we proposed OKByte-AR, a multi-stage framework for aggression relationship identification, which integrated OBB and keypoint detection, MOT, and post-tracking processing. Experimental results demonstrated that our approach significantly improved detection and tracking performance compared to the baseline method while also proving effectiveness in aggression relationship identification. In terms of detection performance, our method achieved a 5.10 % increase in mAP50 and a 15.30 % increase in mAP50-95 compared to the baseline YOLO11 detection model, reaching 90.90 % and 82.70 %, respectively. For tracking performance, our approach outperformed four state-of-the-art MOT methods, achieving the highest HOTA, MOTA, IDF1, and lowest IDSW of 85.78 %, 98.50 %, 99.25 %, and 2, respectively. In aggression relationship identification, our method achieved an average ARES of 66.92 %. In conclusion, OKByte-AR has made significant progress in identifying pig aggression relationships to provide a smarter and more reliable monitoring system for modern pig farming.

7. Institutional review board statement

The animal study protocol was approved by the Animal Ethics Committee of South China Agricultural University (protocol code 2024F213 and date of approval: 14 March 2024).

CRediT authorship contribution statement

Shuqin Tu: Visualization, Supervision. **Yuefei Cao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Conceptualization. **Liang Mao:** Data curation. **Yun Liang:** Data curation. **Hairan Yang:** Data curation. **Baiyang Tang:** Data curation. **Fang Yuan:** Data curation.

Funding

The work was supported by Shenzhen Polytechnic University Smart Agriculture Innovation Application R&D Center (No. 602431001PQ), Huizhou Municipal Key Areas Research and Development Project (No. 2024BQ010007), National Natural Science Foundation of China (No. 62272320), Shenzhen Science and Technology Innovation Commission Foundation (No. 20220812222043002), and Shenzhen Polytechnic University Research Fund (No. 6025310045K).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Liang Mao reports financial support was provided by Shenzhen Polytechnic University Smart Agriculture Innovation Application R&D Center. Has patent pending to. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

i have shared the link to my code at the manuscript

References

Meese, G., Ewbank, R., 1973. The establishment and nature of the dominance hierarchy in the domesticated pig. Anim. Behav. 21 (2), 326–334.
Stookey, J.M., Gonyou, H.W., 1994. The effects of regrouping on behavioral and production parameters in finishing swine. J. Anim. Sci. 72 (11), 2804–2811.
Jensen, P., 1982. An analysis of agonistic interaction patterns in group-housed dry sows—aggression regulation through an “avoidance order”. Appl. Anim. Ethol. 9 (1), 47–61.
Fraser, J., Aricibasi, H., Tulpan, D., & Bergeron, R. (2023). A computer vision image differential approach for automatic detection of aggressive behavior in pigs using deep learning. *Journal of Animal Science, 101*, skad347.
Stukenborg, A., Traulsen, I., Puppe, B., Presuhn, U., Krieter, J., 2011. Agonistic behaviour after mixing in pigs under commercial farm conditions. Appl. Anim. Behav. Sci. 129 (1), 28–35.
Xia, X., Zhang, N., Guan, Z., Chai, X., Ma, S., Chai, X., Sun, T., 2025. PAB-Mamba-YOLO: VSSM assists in YOLO for aggressive behavior detection among weaned piglets. Artif. Intell. Agric. 15 (1), 52–66.
Yan, K., Dai, B., Liu, H., Yin, Y., Li, X., Wu, R., Shen, W., 2024. Deep neural network with adaptive dual-modality fusion for temporal aggressive behavior detection of group-housed pigs. Comput. Electron. Agric. 224, 109243.
Gao, Y., Yan, K., Dai, B., Sun, H., Yin, Y., Liu, R., Shen, W., 2023. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatio-temporal attention mechanism. Comput. Electron. Agric. 205, 107606.
Ji, H., Teng, G., Yu, J., Wen, Y., Deng, H., Zhuang, Y., 2023. Efficient aggressive behavior recognition of pigs based on temporal shift module. Animals 13 (13), 2078.
Chen, C., Zhu, W., Steibel, J., Siegford, J., Wurtz, K., Han, J., Norton, T., 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. Comput. Electron. Agric. 169, 105166.
Fu, Y., Wang, Z., Zheng, H., Yin, X., Fu, W., Gu, Y., 2025. Integrated detection of coconut clusters and oriented leaves using improved YOLOv8n-obb for robotic harvesting. Comput. Electron. Agric. 231, 109979.
Li, P., Chen, J., Chen, Q., Huang, L., Jiang, Z., Hua, W., Li, Y., 2025. Detection and picking point localization of grape bunches and stems based on oriented bounding box. Comput. Electron. Agric. 233, 110168.
Lu, J., Chen, Z., Li, X., Fu, Y., Xiong, X., Liu, X., Wang, H., 2024. ORP-Byte: a multi-object tracking method of pigs that combines Oriented RepPoints and improved Byte. Comput. Electron. Agric. 219, 108782.
Wei, J., Tang, X., Liu, J., Zhang, Z., 2023. Detection of pig movement and aggression using deep learning approaches. Animals 13 (19), 3074.
Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., Norton, T., 2020. A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. Biosyst. Eng. 195, 27–41.
McGlone, J.J., 1985. A quantitative ethogram of aggressive and submissive behaviors in recently regrouped pigs. J. Anim. Sci. 61 (3), 556–566.
D'Eath, R.B., Turner, S.P., 2009. The natural behaviour of the pig. In: *The Welfare of Pigs*. Springer, pp. 13–45.
Wang, W., 2023. Advanced Auto labeling solution with added Features. Retrieved from, Github repository https://github.com/CVHub520/X-AnyLabeling.
Chen, C., Zhu, W., Guo, Y., Ma, C., Huang, W., Ruan, C., 2018. A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs. Livest. Sci. 218, 70–78.
Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.
Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., . . . Wang, X. (2022). ByteTrack: Multi-object Tracking by Associating Every Detection Box. In *Computer Vision – ECCV 2022* (pp. 1-21).
Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. *Simple online and realtime tracking*. Paper Presented at the 2016 IEEE International Conference on Image Processing (ICIP).
Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B., 2021. Hota: a higher order metric for evaluating multi-object tracking. Int. J. Comput. Vis. 129 (2), 548–578.
Bernardin, K., Stiefelhagen, R., 2008. Evaluating multiple object tracking performance: the clear mot metrics. EURASIP J. Image Video Processing 2008 (1), 246309.
Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*.
Cao, J., Pang, J., Weng, X., Khirodkar, R., Kitani, K., 2023. *Observation-centric sort: Rethinking sort for robust multi-object tracking*. Paper Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Maggiolino, G., Ahmad, A., Cao, J., Kitani, K., 2023. *Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification*. Paper Presented at the 2023 IEEE International Conference on Image Processing (ICIP).
Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*.
Tu, S., Cai, Y., Liang, Y., Lei, H., Huang, Y., Liu, H., Xiao, D., 2024a. Tracking and monitoring of individual pig behavior based on YOLOv5-Byte. Comput. Electron. Agric. 221, 108997.
Tu, S., Cao, Y., Liang, Y., Zeng, Z., Ou, H., Du, J., Chen, W., 2024b. Tracking and automatic behavioral analysis of group-housed pigs based on YOLOX+ BoT-SORT-slim. Smart Agric. Technol. 9, 100566.

Computers and Electronics in Agriculture 229 (2025) 109727

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Estimation of passion fruit yield based on YOLOv8n + OC-SORT + CRCM algorithm

Shuqin Tu, Yufei Huang, Qiong Huang *, Hongxing Liu, Yifan Cai, Hua Lei

College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

ARTICLE INFO

Keywords:
Passion fruit yield
Multiple object tracking
YOLOv8n
OC-SORT

ABSTRACT

The precise estimation of passion fruit yield is crucial for efficient orchard management, but it poses challenges such as occlusion, light variations, and camera shake, which can lead to problems such as missed detection, error detection, and double counting of small fruits. In this study, we propose a robust computer vision algorithm named YOLOv8n + OC-SORT + CRCM (Central Region Counting Method) to accomplish three tasks: detection, tracking, and yield estimation of passion fruits. Firstly, we compare the passion fruit detection results using various YOLO series detection algorithms and choose YOLOv8n for detector. Then, OC-SORT algorithm is chosen as the tracker due to its effectiveness in addressing issues like occlusion, vertical shaking, and uneven speeds. Finally, we design CRCM counting algorithm for fruit counting for addressing challenges in estimating passion fruit yield. To validate the effectiveness of these methods, a real-world passion fruit video dataset including 24 videos for each with a 1-minute length was established. In the detection results on the test set, YOLOv8n detector achieved the best results with a mAP@0.5 (mean Average Precision) of 86.3 % and a model size of only 6.2 MB among YOLOv5n, YOLOv7 and YOLOv8n three detectors. Furthermore, the HOTA (higher order tracking accuracy) of OC-SORT tracker was 67.10 %, surpassing three mainstream trackers named BoT-SORT, Byte Track, and Strong SORT by 2.98 %, 4.71 %, and 8.82 %, respectively. In the fruit yield estimation, CRCM demonstrated an average counting accuracy of 87.0 %, surpassing ID number and Single Line Method (SLM) methods by 49.8 % and 10.5 %, respectively. In conclusion, the YOLOv8n + OC-SORT + CRCM algorithm effectively addresses issues of misidentification, missed detections of small fruits, and repeated counts, achieving stable, real-time, and accurate estimation of passion fruit yield.

1. Introduction

Passion fruit pulp has great potential for the development of different functional products and presents broad prospects in maintaining health (Pereira et al., 2023). The passion fruit is cultivated primarily in tropical and subtropical regions, which stands as a significant indigenous crop in South America, holding substantial economic value. Accurate estimation of passion fruit yield is crucial for farmers, aiding them in more effectively planning cultivation areas, selecting harvest times, and devising sales strategies. However, traditional manual estimation methods are tedious and time-consuming. In recent years, the widespread application of computer vision technology has offered effective support in fruit counting, presenting new possibilities for automating passion fruit yield estimation (Farjon et al., 2023; He, Fang, et al., 2022; Vasconez et al., 2020).

In fruit yield management, numerous researchers employed multiple object tracking (MOT) algorithms such as simple online and real-time tracking (SORT) or DeepSORT for fruit counting and achieved remarkable results (Gao et al., 2022; Tan et al., 2022; Villacrés et al., 2023). The MOT algorithms can implement the count of fruits by locating the fruit objects in the video and assigning a unique ID to each detected object (Bashar et al., 2022; Guo et al., 2022). For instance, Leiying He et al. (2022) utilized YOLO-v3 detection results, coupling detection boxes with trajectories using Cascade- MOT algorithm, enhancing the accuracy and robustness of fruit counting. Egi et al. (2022) employed YOLOv5 detector and DeepSORT tracker for counting tomato flowers and fruits. Yang et al. (2022) developed an advanced anchor-free object detection model using CenterNet to detect cotton seedlings and extract their identity embeddings, employing DeepSORT for positional and identity-based data association. The above proposed many fruit detection or tracking algorithms have enhanced the performance of automated fruit counting systems and improved the accuracy of fruit yield prediction in

* Corresponding author.
E-mail address: huang_qiong168@163.com (Q. Huang).

https://doi.org/10.1016/j.compag.2024.109727
Received 14 December 2023; Received in revised form 15 September 2024; Accepted 28 November 2024
Available online 4 December 2024

orchards.

However, tracking fruits in real-world scenarios often faces challenges such as occlusion and varying lighting conditions, leading to missed or false detection and frequent ID switches (Miranda et al., 2023). Improvements are generally pursued either through enhancing the detection efficacy of the detection model or the stability of tracking algorithms (Li et al., 2023; Xie et al., 2023). In terms of refining detection performance, Ge et al. (2022) introduced a visual object tracking network named YOLO-DeepSORT. Leveraging the YOLOv5s model with shufflenetv2 and CBAM attention mechanism, this approach aims to enhance detection accuracy. Furthermore, Wu et al. (2023) proposed an automated fruit counting method known as NDMFCS (Normal Detection Matched Fruit Counting System). This method defines fruits behind leaves and those on the ground as abnormal fruits and excludes them during the detection process to enhance counting accuracy. Concerning the stability of tracking algorithms, Farjon et al. (2023) proposed a motion displacement estimation algorithm OrangeSort based on the SORT algorithm to significantly reduce the error associated with the repetitive counting problem. Gao et al. (2022) chose tree trunks as single-target tracking targets for apple tracking to achieve higher accuracy and faster tracking speed than the commonly used fruit-based multi-target tracking methods. Our previous work focused on passion fruit yield counting using lightweight YOLOv5s and improved DeepSORT(Tu et al., 2024). In fruit yield management, with the emergence of more and more outstanding MOT algorithms, the fruit yield estimation is gaining increasingly in accuracy under natural scenarios.

At present, many studies estimate fruit quantity based on the number of IDs obtained from MOT algorithms. However, due to frequent ID switches, the resulting count often surpasses the actual number of fruits. Consequently, scholars have attempted to mitigate this issue by implementing various counting methods during the tracking process. Parico and Ahamed. (2021) compared ROI (Region of Interest) and unique ID methods. The ROI method counts fruits based on the number of tracked objects crossing the horizontal line, and the experimental results show that the fruit counts from the ROI method are closer to the artificial true counts. Rong et al. (2023) devised a specific tracking area counting method to overcome the problem of tracked tomato cluster ID switches. Zheng et al. (2023) proposed a nonuniform distributed counter to rectify citrus fruit counting during tracking, aiming to reduce counting errors resulting. In some fruit yield estimation, it may be inaccurate to adopt max IDs counted as the yield result for frequent ID error switches. Therefore, there is an urgent demand for researchers to study the different types of fruit yield counting algorithms according to the results of MOT.

Despite the widespread application of MOT methods in fruit counting and numerous improvements made specifically for fruit counting, there has yet to be comprehensive research that systematically compares and summarizes the mainstream detectors, tracking algorithms, and counting methods. This study aims to address common challenges in estimating passion fruit yield, including factors such as variations in lighting, background interference, camera angles, and camera shaking, to identify the optimal detector, tracker, and counting method, thereby determining the best strategy for this task. Specifically, the main objectives of this research include the following three points:

(1) A comprehensive comparison of the current mainstream YOLO series detection algorithms is conducted to obtain optimal detection performance. And we ultimately select YOLOv8n algorithm as the detector for this task.
(2) Mainstream fruit tracking algorithms, including OC-SORT, BoT-SORT, Byte Track and Strong SORT trackers, are compared and evaluated for passion fruit tracking. The best performing OC-SORT tracker is finally selected.
(3) A comprehensive comparison and evaluation of three commonly used counting methods, namely maximum ID number, single line method (SLM), and central region counting method (CRCM), are performed to determine the most suitable strategy for estimation of passion fruit yield. And CRCM is selected as the best appropriate passion fruit yield estimation method among three approaches.

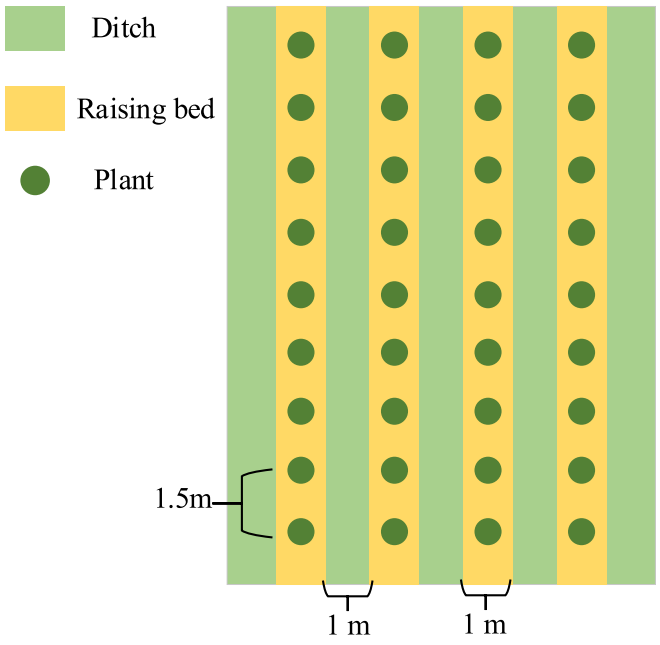


Fig. 1. Passion fruit cultivation.

2. Dataset description

2.1. Dataset collection

To evaluate the effectiveness of our approach, we collected the videos dataset in passion fruit orchards. These videos were captured on October 14, 2022, under clear weather conditions in a passion fruit plantation located in Zengcheng District, Guangzhou City, Guangdong Province, China. They employed the Ridge Cultivation Method to cultivate passion fruit, with a spacing of 1.5 m between plants, 1 m width for ditches, and a raising bed width of 1 m. Specific details regarding the layout of the passion fruit plantation are shown in Fig. 1. We used the mobile phones for the video shoot with walking along ditch, the angle of mobile phone was vertical with the passion fruit raising bed, and the distance of the phone from the fruit was 1 m and the height was 1.5 m.

The dataset comprises a total of 24 videos, all captured by individuals using smartphones, with each video lasting approximately 1 min. These videos have a resolution of 1080*1920 and a frame rate of 30 frames per second. They are divided into two groups: half were shot away from the sunlight, while the other half were shot toward the sunlight. We were able to capture clear footage of the fruits under favorable lighting conditions and appropriate angles, as depicted in Fig. 2(a). However, we found the following issues during the collection of the dataset: (1) the image would suddenly become dark when shooting toward the sunlight, as illustrated in Fig. 2(b). (2) the picture will be blurred when the camera moves quickly, as shown in Fig. 2(c). The ratio of images taken towards and away from the sun is 5:5. And the image taken by camera shake is about 200 images from the total volume of frames. We conducted experiments and analyzed results using these datasets to evaluation the effectiveness of our approach.

2.2. Data processing

The processing flow of the dataset is shown in Fig. 3. From all 24

Fig. 2. Part of the dataset.

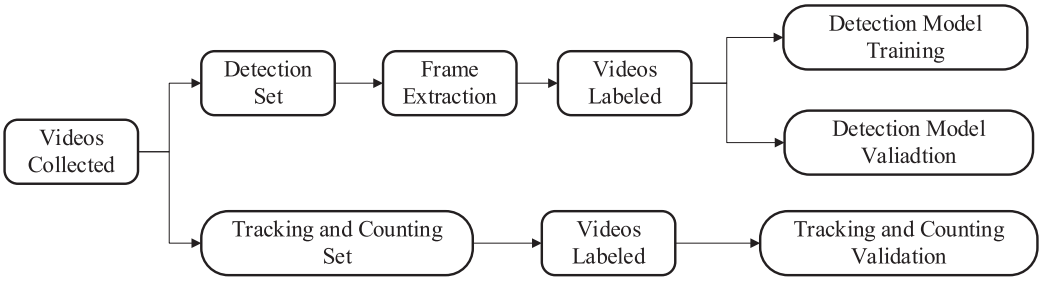


Fig. 3. Processing flow of the dataset.

videos, 5 videos were randomly selected for verifying the performance of the detection model and for the target tracking and counting tasks.

The processing of the detection set began with frame extraction, adjusting the video frame rate to 5 frames per second. And the videos were labeled using the DrakLabel software and cut into image sequences using the FFmpeg software. The labelled files of the videos were then converted into labelled file in YOLO and COCO format for the subsequent detection model training and validation process. Finally, all images and their corresponding labelled files were organized according to the division of the training and validation set.

For the five videos in tracking and counting set, they were directly labeled using DrakLabel software to be used as validation data for these tracking and counting tasks.

3. Methods

3.1. Method process and flow chart

We propose a passion fruit yield estimation algorithm based on YOLOv8n + OC-SORT + CRCM, the flow of which is shown in Fig. 4. Initially, YOLOv8n is used to detect each passion fruit and obtain its detection box and confidence in each frame of the video. Subsequently, the detection result is fed into OC-SORT tracker, where the tracking information of each passion fruit is achieved through three modules: observation-centered updating strategy, data association, and tracklet management. After the tracking process, ID number, SLM and CRCM are used to count passion fruits for yield estimation. Finally, the MOT visualization results are output, along with a text message of the passion fruit counting results in the upper left corner of each frame.

3.2. YOLOv8

The object detection algorithms of YOLO series employ a single-stage detection approach. This avoids the candidate box generation and filtering steps in traditional two-stage detection methods, thereby simplifying the algorithmic process (Redmon et al., 2018; Bochkovskiy et al., 2020). Over the years, the YOLO algorithm has undergone significant improvements, with multiple releases that have continually improved its efficiency.

YOLOv8 was released by Ultralytics in January 2023 (Talaat & ZainEldin, 2023). Compared to its predecessors, it brings a series of improvements: closing the Mosaic Augmentation, adoption of new convolutional modules, and anchor free detection. Firstly, YOLOv8 closes mosaic augmentation in the final ten training epochs. This

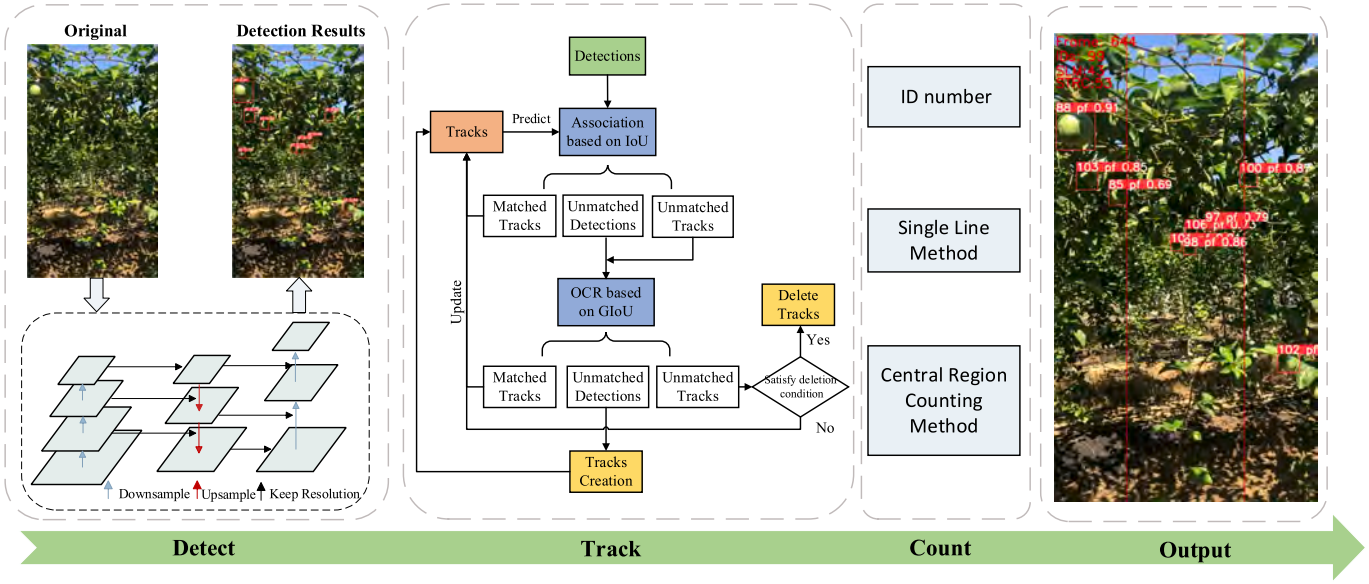


Fig. 4. Passion fruit yield estimation algorithm flow chart based on YOLOv8n + OC-SORT + CRCM.

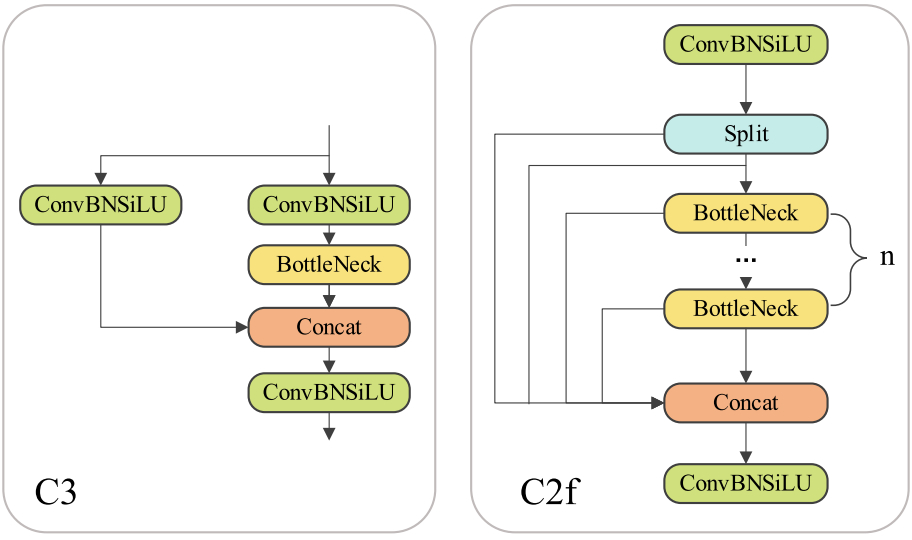


Fig. 5. C3 module and C2f module.

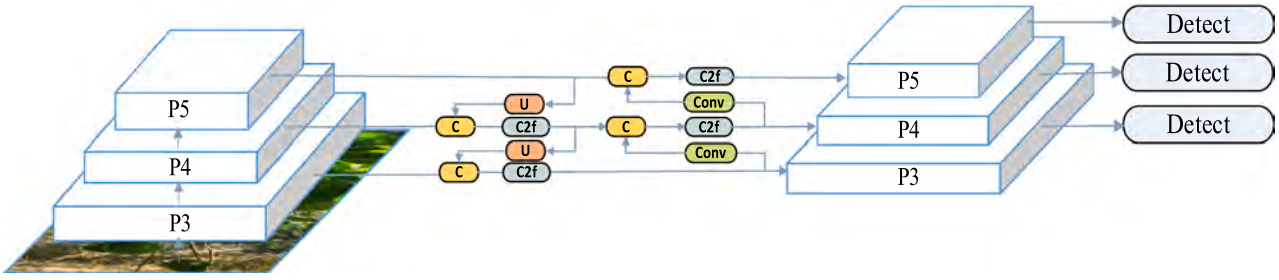


Fig. 6. YOLOv8 Architecture.

improvement is motivated by the observation that using this enhancement throughout the entire training process could decrease performance. Secondly, YOLOv8 replaces the C3 module with the C2f module. Fig. 5 illustrates the structures of the C3 and C2f modules. The C2f module is designed based on the C3 module and the efficient layer aggregation networks, allowing YOLOv8 to acquire richer gradient flow information while being lightweight. Lastly, anchor-free detection is employed. This approach directly predicts the object's center instead of relying on offsets based on known anchor boxes. This improvement effectively reduces the number of candidate detections, thus speeding up the post-processing step of non-maximum suppression.

Fig. 6 illustrates the structure of YOLOv8. The YOLOv8 model has

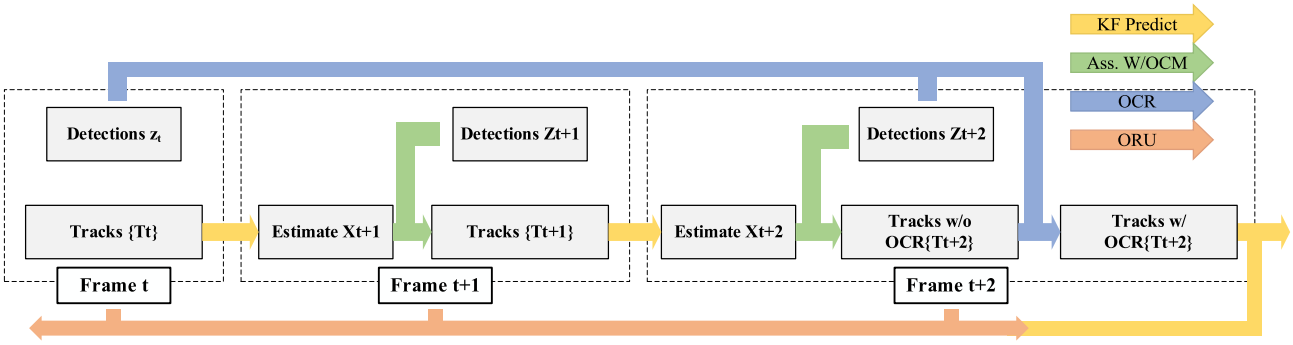


Fig. 7. Pipeline of OC-SORT.

completed structural adjustments, modifying the model's backbone, key building blocks, and fusion layers to render the model more compact and efficient. The backbone network employs the spatial pyramid pooling fast (SPPF) module, while the neck network retains the path aggregation network (PAN)-feature pyramid network (FPN) methods for feature fusion. In the detection head section, an anchor-free detection head is utilized. YOLOv8 was selected as the primary detector to meet the real-time and accurate detection requirements for passion fruit in practical scenarios.

3.3. OC-SORT

Currently, most target tracking methods rely heavily on the Kalman filter(KF), which employs a center-based estimation update strategy (Bewley et al., 2016; Wojke et al., 2017). These tracking algorithms update matched tracks without observation and put complete trust in the estimation for tracks. This may make it subject to severe noise when it suffers from occlusion or when the object motion is not completely linear. In fruit yield prediction work, the KF tracking algorithm achieved good performance under unconcluded, or slightly occluded and linear motion scenarios for not the long videos. However, in the passion fruit dense and target occluded scenario, the KF tracking algorithm may affect the tracking performance due to the existence of a certain amount of noise and the cumulative error in target tracking caused by long time target occlusion.

During the data collection process, we employed a handheld method for capturing images with a mobile device for convenience. This method easily introduces vertical shaking and uneven speeds, resulting in non-linear motions. Additionally, we observed instances where targets were occluded by leaves during the shooting process. Therefore, if we used the KF tracking algorithm to track passion fruit in heavily obscured scenes, the tracking results may be moderately affected.

To address the above issues, OC-SORT introduces three key improvements: Observation-Centric Re-Update (ORU), Observation-Centric Momentum (OCM), and Observation-Centric Recovery (OCR) (Cao et al., 2023). These enhancements aim to boost tracking performance in scenarios involving occlusion and non-linear object motion.

ORU: When a trajectory is re-associated with observations after a

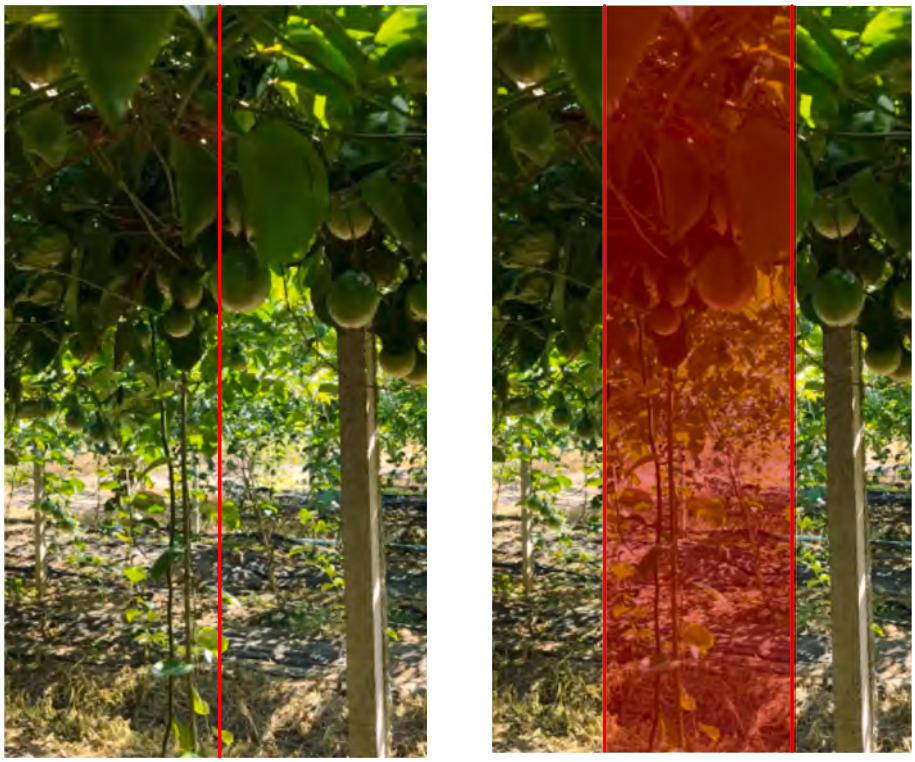


Fig. 8. Counting areas of SLM and CRCM.

period of being untracked, ORU strategy is employed to mitigate the accumulated errors during this untracked period. Initially, when generating a virtual trajectory, considerations are made regarding the observations at the beginning and end of the untracked period. The calculation for the virtual trajectory is represented by Equation (1).

$$\widetilde{Z}_t = traj_{virtual}(z_{t1}, z_{t2}, t), t_1 < t < t_2 \quad (1)$$

where z_{t1} corresponds to the last observation before the object was untracked, and z_{t2} represents the observation when the object is re-associated.

Subsequently, the parameters of the Kalman filter are readjusted based on the virtual trajectory, as outlined in Equation (2).

$$\text{re}-\text{update}\begin{cases} K_t = P_{t|t\text{-}1}H_t^T\left(H_tP_{t|t\text{-}1}H_t^T + R_t\right)^{-1} \\ \widehat{X}_{t|t} = \widehat{X}_{t|t-1} + K_t\left(\widetilde{z}_t - H_t\widehat{X}_{t|t-1}\right) \\ P_{t|t} = (I - K_tH_t)P_{t|t-1} \end{cases} \quad (2)$$

OCM: it uses observations to compute the direction of motion of a trajectory as part of the cost matrix. It aims to increase the directional consistency of trajectories and thus reduce erroneous associations

OCR: it adds a second round of GIoU-based correlation after the first round of correlation to improve the tracking accuracy.

Fig. 7 shows the flowchart of OC-SORT tracker. If any target is lost on frame t + 1 but on the next frame, it is recovered by referring to the OCR's observation of frame t. The OCR's observation of frame t is used as a reference for the recovery of the target. Retracing triggers an ORU of its KF parameter from frame t to frame t + 2.

3.4. Existing counting algorithms for yield estimation

The three methods, ID number, SLM, and CRCM were employed to count passion fruits for yield estimation, and we designed the CRCM approach for passion fruit planting scenes based on the specific tracking area counting (Rong et al., 2023).

(1) The ID number Counting method is relatively direct and simple, estimating the total number of fruits using the quantity of tracklets obtained from tracking algorithms.
(2) The SLM method counts passion fruits based on whether they cross predetermined vertical lines. Specifically, when a passion fruit is tracked crossing the central vertical line, the total fruit count increases by 1. In the experiment, the starting coordinates of the line are (540, 0), and the ending coordinates are (540, 1920), as illustrated in Fig. 8(a).
(3) The CRCM counts passion fruits based on whether the centroid of a passion fruit enters a predefined counting area. This method contains two counting rules: firstly, if a passion fruit enters the counting area from outside, the total count increases by 1; secondly, if a passion fruit that was previously in the counting area exits the area in the current frame, the total is added by 1. In this experiment, the coordinates of the upper left and lower right corners of the counting area are (300, 0) and (780, 1920), respectively as illustrated in Fig. 8(b).

4. Experiment and result analysis

4.1. Implementation details

The passion fruit yield estimation experiments in this study were divided into three parts. (1) Detection of passion fruit using YOLOv5s, YOLOv5m, YOLOv7-tiny(Wang et al., 2023), YOLOv8n and YOLOv8s models. (2) Passion fruit tracking using OC-SORT, BoT-SORT, ByteTrack, Deep OC-SORT, Strong SORT based on YOLOv8n model. (3) Passion fruit counting using the number of IDs, SLM and CRCM based on YOLOv8n and OC-SORT.

Table 1
The detection results based on different detectors.

Model	P/%↑	R/%↑	F1/%↑	mAP@.5/%↑	Size/MB↓
YOLOv5s	**81.3**	79.7	80.5	86.4	14.4
YOLOv5m	79.9	80.6	80.3	86.5	42.5
YOLOv7-tiny	81.0	80.9	**81.0**	**87.1**	12.0
YOLOv8n	80.4	80.4	80.4	86.3	**6.2**
YOLOv8s	80.4	**81.2**	80.8	86.9	22.5

In detection experiments, the parameters are set as follows: epoch is 200, batch size is 32, input image size is 640, learning rate is 0.01, optimizer is SGD, conf-thres is 0.25, and iou-thres is 0.45. We used migration to train the YOLOv8n model, merging the standard categories into a new category 'passion fruit'. In total, 6445 images were used for detection, of which 5051 were used as the training set and 1394 as the validation set. And we used the data enhancement operations including Mosaic, Mixup, random perspective and HSV augment. In the tracking experiments, the parameters used are the default parameters in each tracking algorithm. In the counting experiments, the start coordinates of the vertical line were set to (540, 0) and the end coordinates were set to (540, 1920) in SLC. CRCM sets the coordinates of the upper left corner of the region to (300, 0) and the lower right corner to (780, 1920).

4.2. Evaluation metrics

The evaluation of the target detection section involves assessing key metrics such as precision (P), recall (R), F-Measure (F1), mean average precision (mAP), and model size for evaluating detector performance. Precision (P), Recall (R), F1 and mAP collectively indicate the detector's overall performance(Padilla et al., 2020), with higher values signifying better performance. Model size serves as indicators of the detector's efficiency, with lower values denoting greater efficiency.

The MOT segment is evaluated using higher order tracking accuracy (HOTA), multiple object tracking accuracy (MOTA), identification F1 (IDF1), and ID switch (IDSW) (Luiten et al., 2021). HOTA, MOTA, IDF1, and IDSW are crucial metrics for assessing tracking algorithm performance, where higher values of HOTA, MOTA, and IDF1 indicate superior performance, while a smaller value of IDSW signifies better tracking accuracy.

For the counting component, the evaluation relied on counting accuracy and statistical average counting accuracy(Rong et al., 2023). Higher values of these metrics suggest that the counting method's predictions are closer to the true values.

4.3. The detection results based on different detectors

Table 1 presents a comprehensive evaluation of different models in detecting passion fruits in practical scenarios. These evaluations include metrics such as precision, recall, F1, and mAP@0.5. The performance of each model is quite similar, with an average precision of 80.6 %, an average recall of 80.56 %, an average F1 of 80.58 %, and an average mAP@0.5 of 86.64 %. However, when considering model size, significant differences can be observed among them. The YOLOv8n model is only half the size of the best-performing YOLOv7-tniy model and had similar detection performance. This means that the YOLOv8n model maintains high accuracy while having a smaller model size, making it more suitable for online, real-time applications, and deployment on portable devices. Therefore, considering the balance of accuracy and capacity, we chose to use the YOLOv8n model as the detector in the subsequent tracking experiments.

Fig. 9 showed the detection results of different detectors in a practical scenario. In the scenario, there were a total of 24 passion fruits, primarily concentrated in the upper part of the image, presenting a higher density. There were fruits overlapping, and some fruits were

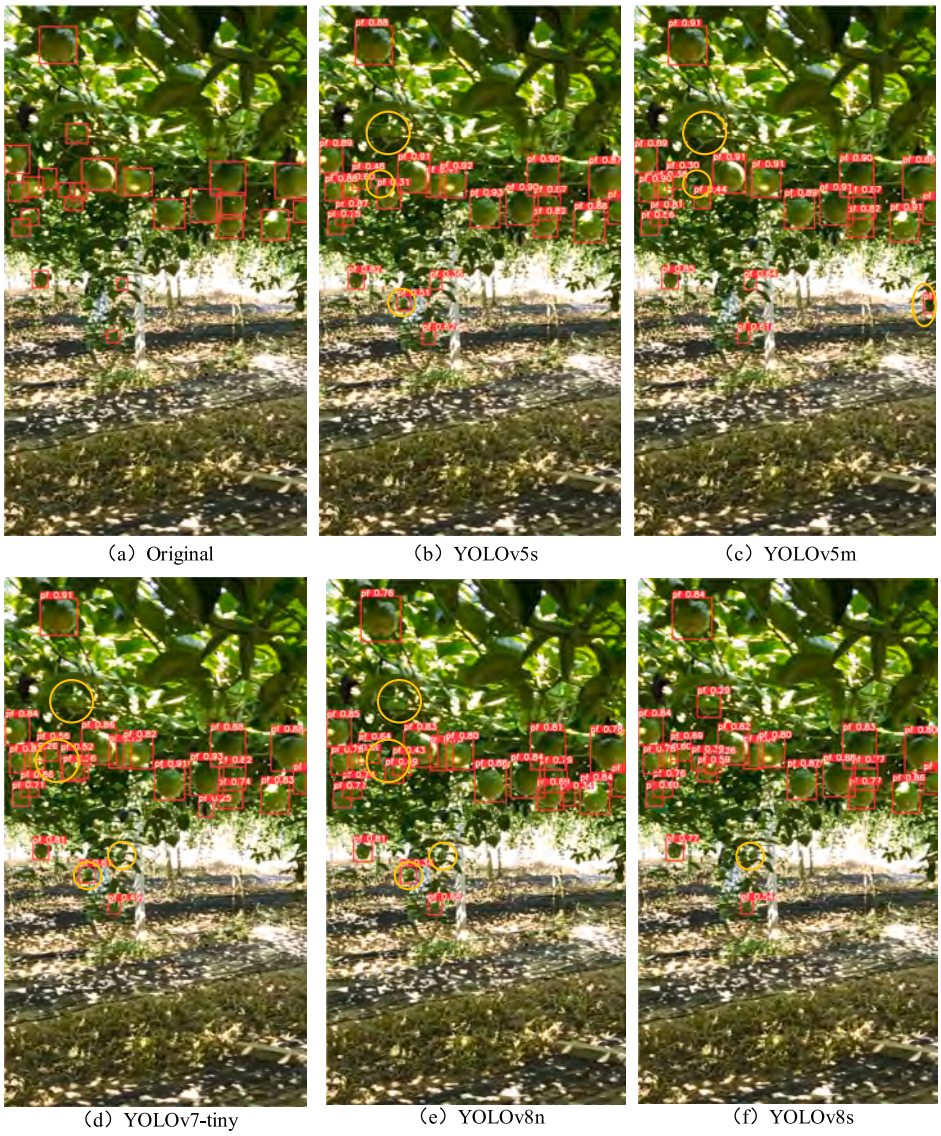


Fig. 9. The detection results of different detectors in practical scenario.

Table 2
The Tracking results based on different tracking algorithms.

Method	HOTA/%↑	MOTA/%↑	IDF1/%↑	IDSW↓	Time/ms↓
OC-SORT	**67.103**	**75.874**	**75.618**	354	19.0
BoT-SORT	64.126	74.782	71.889	716	106.6
Byte Track	62.396	72.429	74.644	**264**	**15.4**
Strong SORT	58.280	71.864	61.639	578	87.3

partially obscured by leaves. All detection models could detect most of the passion fruits in this scenario, but misdetections and missed detections were still present, which were marked with yellow boxes. Firstly, YOLOv5s, YOLOv5m, YOLOv7-tiny, and YOLOv8n did not detect the fruit in the upper left due to leaf occlusion. Secondly, on the left side of the central fruit-dense area, the fruits overlapped each other, resulting in under-detection by the models YOLOv5s, YOLOv5m, YOLOv7-tiny, and YOLOv8n. Finally, in the lower half of the image, all detection models missed or misidentified leaves as fruits due to the relatively small size of the fruits and their similar shape to leaves. When we evaluated the detection performance of passion fruits, we found that detection accuracy was influenced by various factors, including environmental lighting, the similarity between leaves and fruits, occlusion issues, and the small size of the fruits.

4.4. The tracking results based on different tracking algorithms

Table 2 presents a comparison of tracking results in test videos for OC-SORT, BoT-SORT(Aharon et al., 2022), Byte Track(Zhang et al., 2022), and Strong SORT(Du et al., 2023). From the perspectives of HOTA, MOTA, and IDF1, OC-SORT stands out with scores of 67.103 %, 75.874 %, and 75.618 %, respectively. Compared to BoT-SORT, Byte Track, and Strong SORT, OC-SORT exhibited improvements in HOTA by 2.977 %, 4.707 %, and 8.823 %, MOTA by 1.092 %, 3.445 %, and 4.01 %, and IDF1 by 3.729 %, 0.974 %, and 13.979 %, respectively. Furthermore, in terms of IDSW and runtime, Byte Track performed exceptionally well with an IDSW of 264 and an average runtime of 15.4 ms per frame. Taking all these metrics into consideration, OC-SORT shows significant advantages in comprehensive tracking metrics HOTA, MOTA, and IDF1, even though OC-SORT has a slight gap with Byte Track in IDSW and runtime. This demonstrates that OC-SORT possesses a unique competitive advantage in addressing the association and identification of passion fruits in practical scenarios.

Fig. 10 illustrated the visual results of target tracking in video 41 for OC-SORT, BoT-SORT, Byte Track, and Strong SORT. According to the

Fig. 10. Comparison of tracking results for OC-SORT, BoT-SORT, ByteTrack and Strong SORT in video 41.

Table 3
Tracking results based on YOLOv8n and OC-SORT on all test videos.

Test Videos	Fruit Number	HOTA/ %↑	MOTA/ %↑	IDF1/ %↑	IDSW↓	Time/ ms↓
10	136	65.776	78.811	81.848	12	17.6
21	329	67.412	77.592	75.360	65	21.4
41	203	70.081	72.808	77.969	97	17.4
61	348	63.365	72.659	71.201	128	18.5
90	188	67.475	78.845	74.612	52	20.2
All	1204	67.103	75.874	75.618	354	19.0

manually labeled ground truth file of video 41, the maximum IDs of the first frame, 301st frame, and 601st frame were 6, 42, and 120, respectively. At frame 3, the maximum IDs of all tracking algorithms were the same as the maximum ID in the ground truth. The differences in the visualization results of the algorithms were mainly found in frames 303 and 603. At frame 303, OC-SORT, BoT-SORT, Byte Track, and Strong SORT had maximum IDs differing from the maximum ID in the ground truth file by 4, 20, 5, and 24, respectively. These differences further increased in the 601st frame. OC-SORT, BoT-SORT, Byte Track, and Strong SORT had maximum IDs differing from the GT file by 20, 35, 39, and 44, respectively. From the above results, the maximum ID value obtained by OC-SORT was closer to the maximum ID value of the GT file than other tracking algorithms, indicating that it had stability in the continuous tracking of passion fruit. Based on the tracking evaluation index and visualization results, we decided to adopt the OC-SORT model as our main tracking algorithm in the subsequent tracking experiments.

4.5. Tracking results based on YOLOv8n and OC-SORT on all test videos

Table 3 presented the tracking results of all test videos based on YOLOv8n and OC-SORT. While the variation in the number of fruits did not seem to directly impact the tracking performance of the videos, there were discernible patterns when classifying videos based on shooting conditions. videos shot away from the sunlight (test videos 21, 41, 61) showed a decrease in tracking performance with an increase in the number of fruits. On the other hand, videos shot toward the sunlight (test videos 10, 90) exhibited a decrease in IDF1 and IDSW with an increase in fruit quantity. Moreover, despite test video 41 having a higher number of fruits compared to videos shot in sunlight, it displayed the best tracking performance. This suggested that the shooting conditions significantly influenced the tracking performance. In summary, these

a.530th frame in video 10 b.560th frame in video 10 c.590th frame in video 10

d.770th frame in video 21 e.800th frame in video 21 f.830th frame in video 21

Fig. 11. OC-SORT tracking results on test video 10, 21.

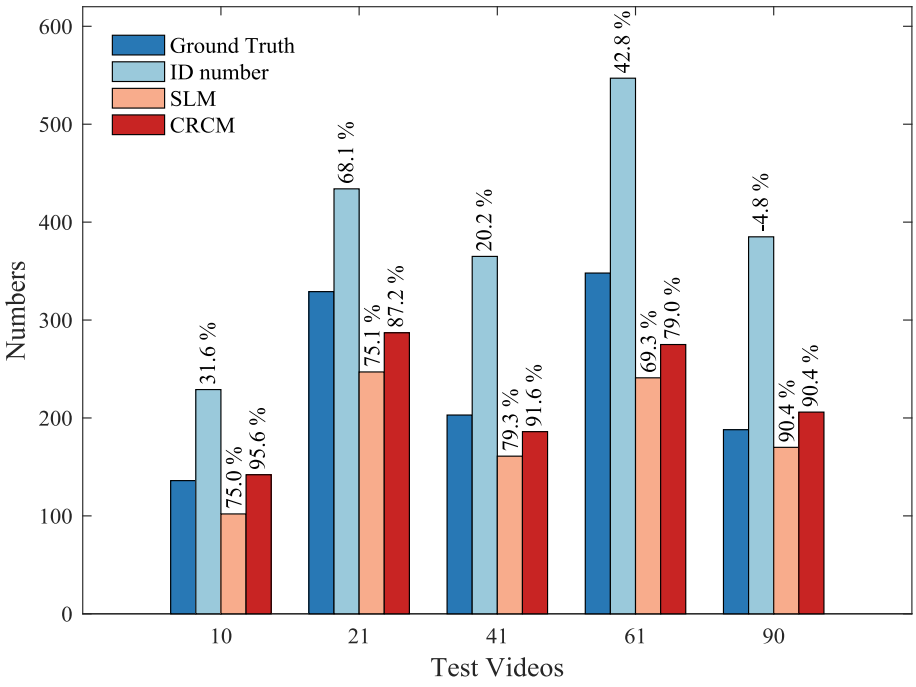


Fig. 12. The comparison results of the ground truth and the predicted fruit counting methods including ID number, SLM, and CRCM.

results implied that high fruit quantity and shooting toward the sunlight might have posed more challenges for object detection and tracking algorithms, such as increased interference between targets, mutual occlusion, and problems with frame exposure.

Fig. 11 presented the visual results based on OC-SORT in test videos 10 and 21. We studied the test videos for two different shooting conditions. Test video 10 was shot toward the sunlight, exhibiting occlusion of the leaves and darker scenes. Test video 21 was shot away from the sunlight, with certain areas showcasing densely packed and smaller-shaped fruits. OC-SORT tracked most of the fruits in both test videos 10 and 21. However, some fruits in the scene could not be tracked due to occlusion or inadequate lighting conditions. Fig. 11(a) to Fig. 11(c) exhibited the visual results for test video 10, there were untracked fruits in all three selected frames, due to leaf occlusion and low light issues. Conversely, Fig. 11(d) to Fig. 11(f) displayed the visual results for test video 21, with only one fruit not successfully tracked. The comprehensive analysis showed that the main reason for the poor tracking results was that the fruits in the scene were easily missed and the leaves were misidentified as fruits.

4.6. Evaluation of counting methods

Fig. 12 showed the comparison results of the ground truth and the predicted fruit counting methods including ID number, SLM, and CRCM. The x-axis represented the information for the test videos, while the y-axis represented the fruit counts. The text labels on the bar graphs were the counting accuracy of each prediction method for the test videos. Based on the results in Fig. 12, we found that CRCM method obtained highest counting accuracy than ID number and SLM in most videos. When compared to the SLM method, the CRCM improved the counting accuracy in each test video by 64 %, 19.1 %, 71.4 %, 36.2 %, and 95.2 %, respectively. When compared to the SLM method, the CRCM method exhibited counting accuracy improvements of 20.6 %, 12.1 %, 12.3 %, and 9.7 % in the first four test videos. Moreover, the average counting accuracy of ID number, SLM, and CRCM was 37.2 %, 76.5 %, and 87.0 %, respectively. In terms of counting accuracy and average counting accuracy, CRCM outperformed the ID number and SLM.

Fig. 13 presented the visualization results of the three different counting methods. Fig. 13(a)- Fig. 13(c) showed the visualization results of SLM, the counting results of ID number, and SLM, while Fig. 13(d)-(f) showed the visualization results of CRCM and the counting results of ID number, SLM, and CRCM. The fruit moved from right to left in the test video. In Fig. 13(a), it was clearly observed that the ID number method had a serious double counting problem. At frame 1774, the ID number count had reached 289, while only 157 fruits had been labeled according to the manually labeled ground truth file. In Fig. 13(b)- Fig. 13(c), the missed counting problem of the SLM method could be found. The SLM method was unable to count the newly detected fruits in the counted region, as shown by the yellow circles (fruits with ID 315 and 316), which led to missed counting. While the CRCM method effectively solved the problems of repeated counting and missed counting. In Fig. 13(e)- Fig. 13(f), the CRCM method accurately counted the fruits (with IDs 315 and 316) that were missed by the SLM method and was closest to the manually counted fruit count. Therefore, the CRCM method was more suitable for the passion fruit counting task compared to the ID count and the SLM method.

5. Discussion

In fruit yield prediction, target detection models play a key role in the performance of fruit yield estimation. There are significant differences in the fruit growing scenarios for various fruits, some of them simple and others are complicated. Therefore, in situations where it is not determined which target detector model is the most suitable, it is generally experimentally validated to choose the most appropriate target detector. To select the most suitable detector, this study also compared the results of the chosen YOLOv8n detector with those of the Faster R-CNN, YOLOv9s, YOLOv10n, and Deformable detection transformer(Zhu et al., 2021) detectors, and the results are shown in Table 4.

It is observed that YOLOv8n can be chosen the best detector for passion fruit detection according to accuracy and storage size. Specifically, the P, R, mAP50, mAP50-90, and size values of the YOLOv8n approach are 80.4 %, 80.4 %, 86.3 %,49.2 %, and 6.2 MB, respectively. Compared to the YOLOv10n method, the adopted YOLOv8n can improve by 2.0 % for mAP50 and 0.8 % for mAP50-95, while only adding storage space by 0.5 MB. Compared to Faster R-CNN, the

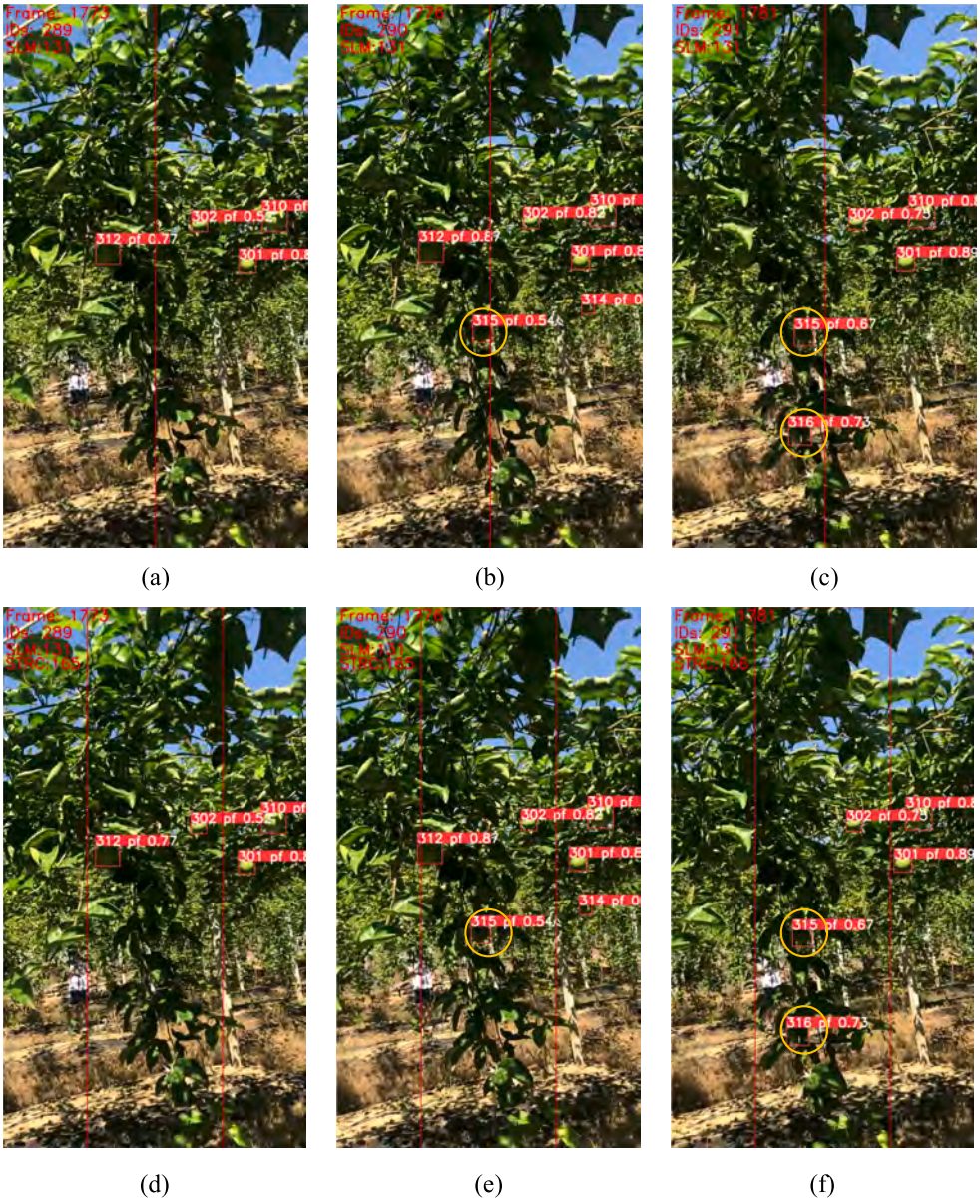


Fig. 13. The visualization results of the numbers of ID, SLM and CRCM.

Table 4
The comparison results of YOLOv8n with other detectors.

	P/%↑	R/%↑	mAP50/%↑	mAP50-95/↑	Size/MB↓
Faster R-CNN	–	–	**86.7**	41.1	335.3
YOLOv10n	79.8	78.5	84.3	48.4	**5.7**
YOLOv9s	78.1	80.4	86.0	48.5	15.2
Deformable DETR	–	–	85.0	44.2	524.2
YOLOv8n	80.4	80.4	86.3	**49.2**	6.2

YOLOv8n achieves decline of 0.4 % for mAP50. The size of storage space for the Faster R-CNN is 335.3 M, which is much larger than that of the YOLOv8n. Therefore, Faster R-CNN runs slower and does not satisfy real-time requirements. Additionally, Comparing YOLOv9s and DETR methods, the YOLOv8n achieves the best performance on all metrics. These comparative results further validate the effectiveness of the YOLOv8n in fruit detection and improving the accuracy of yield estimation.

The accuracy of maximum ID number approach for fruit yield estimation is 37.2 % in our study. There may be 2 main reasons about the inefficiency of maximum ID number method. One is that there are frequent incorrect exchanges of fruit IDs when passion fruit orchard passion fruit orchards provide a dense fruit production, resulting in low accuracy of ID counts. And our collected dataset is particularly dense and there is severe occlusion for the 1-minute video. Whereas in (Wu et al. 2023), it was shown that when calculating the number of apples in orchards, the use of removing duplicate fruit IDs and IDs of fruits lying on the ground makes it possible to achieve an average counting precision of more than 95 %. The reason may be that the passion fruit cultivation scenario differs significantly from that of the apple. Passion fruit is a vine grown graft (like the grape environment), and its fruits are particularly dense in the orchards with its very similarly colored leaves and fruits. Apples produce their fruits on a tree, and there is a significant difference in the color of the leaves and the fruit. Therefore, there is a relatively significant differences between passion fruit and apple scenarios, which can likely result in different results for the ID counting method for passion fruit versus apples.

Another reason is that the time duration of each video is varied, and there are more challenges with long video tracking. The average accuracy of ID counting method in our previous passion fruit yield estimation

(Tu S, et al.2024) is 86.19 % and it is not low. The reason may be that each video in our previous work is about 15 s in length and passion fruit orchard has sparse fruits. However, the length of each video in this study is about 60 s and dataset is particularly dense with fruits. When the illumination variations are sophisticated and mobile phone is employed for shooting, there may the camera moves quickly, resulting in unclear image condition, which caused the error ID counts.

The novelty of our work is not solely based on combining three known methods (YOLOv8n, OC-SORT, and CRCM) but on systematically evaluating and selecting the best combination of detection, tracking, and counting methods to address the specific challenges of passion fruit counting in real-world scenarios. The study identifies and implements the most effective solutions for overcoming issues like occlusion, lighting variations, and camera shaking, leading to more accurate and reliable yield estimation. This comprehensive approach, along with the creation of a real-world dataset and the performance validation against other methods, distinguishes this work from previous studies.

6. Conclusions

In this study, we actively explored the passion fruit yield estimation task in a real scenario with various challenging factors such as changes in lighting, background interference, camera shooting angles, as well as camera shake. To achieve optimal object detection performance, we conducted comparative analyses of detectors like YOLOv5, YOLOv7, and YOLOv8, ultimately selecting YOLOv8n as the most suitable detector, achieving precision of 80.4 %, recall of 80.5 %, F1 score of 80.4 %, mAP@0.5 of 86.3 %, with a model size of only 6.2 MB.

In fruit tracking, we compared tracking algorithms including OC-SORT, BoT-SORT, Byte Track, and Strong SORT. The results showed that OC-SORT performed the best, achieving 67.103 % of HOTA, 75.874 % of MOTA and 75.618 % of IDF1 respectively, ensuring more reliable fruit tracking.

Regarding passion fruit counting, we designed the CRCM algorithm and comprehensively compared with other two counting methods including ID count and SLM. In terms of average counting precision, the CRCM method performed remarkably well, achieving an average counting precision of 87.0 %, which was better than ID count (37.2 %) and SLM (76.5 %).

Considering that high fruit quantity and shooting toward sunlight may pose more challenges for passion fruit detection and tracking algorithms, such as increased mutual obstruction between fruits and frame exposure. Therefore, the utilization of the appropriate detectors YOLOv8n and tracker OC-SORT, coupled with the CRCM method, effectively resolved the issue of repetitive counting in the current passion fruit yield estimation methods. This approach enabled stable and real-time estimation of passion fruit yield. This research outcome can also be applied to similar fruit yield predictions, contributing positively to future fruit cultivation planning, farm management, and market supply chain dynamics.

CRediT authorship contribution statement

Shuqin Tu: Writing – review & editing, Resources, Funding acquisition, Formal analysis, Conceptualization. **Yufei Huang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Qiong Huang:** Supervision, Project administration, Funding acquisition. **Hongxing Liu:** Visualization, Resources, Investigation. **Yifan Cai:** Visualization, Resources, Investigation. **Hua Lei:** Resources, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This study was funded by Key Research and Development Program of Guangzhou (No. 2024B03J1358).

Data availability

Data will be made available on request.

References

Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust Associations Multi-Pedestrian Tracking (arXiv:2206.14651). arXiv. http://arxiv.org/abs/2206.14651.

Bashar, M., Islam, S., Hussain, K. K., Hasan, M. B., Rahman, A. B. M. A., & Kabir, M. H. (2022). Multiple Object Tracking in Recent Times: A Literature Review (arXiv: 2209.04796). arXiv. http://arxiv.org/abs/2209.04796.

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing (ICIP), 3464–3468.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv Preprint arXiv:2004.10934.

Cao, J., Pang, J., Weng, X., Khirodkar, R., & Kitani, K. (2023). Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking (arXiv:2203.14360). arXiv. http://arxiv.org/abs/2203.14360.

Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., Meng, H., 2023. StrongSORT: Make DeepSORT Great Again. IEEE Trans. Multimedia 1–14. https://doi.org/10.1109/TMM.2023.3240881.

Egi, Y., Hajyzadeh, M., Eyceyurt, E., 2022. Drone-Computer Communication Based Tomato Generative Organ Counting Model Using YOLO V5 and Deep-Sort. Agriculture 12 (9), 1290. https://doi.org/10.3390/agriculture12091290.

Farjon, G., Huijun, L., & Edan, Y. (2023). Deep-Learning-based Counting Methods, Datasets, and Applications in Agriculture—A Review (arXiv:2303.02632). arXiv. http://arxiv.org/abs/2303.02632.

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., Li, R., Fu, L., Zhang, Q., 2022. A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. Comput. Electron. Agric. 197, 107000. https://doi.org/10.1016/j.compag.2022.107000.

Ge, Y., Lin, S., Zhang, Y., Li, Z., Cheng, H., Dong, J., Shao, S., Zhang, J., Qi, X., Wu, Z., 2022. Tracking and Counting of Tomato at Different Growth Period Using an Improving YOLO-Deepsort Network for Inspection Robot. Machines 10 (6), 489. https://doi.org/10.3390/machines10060489.

Guo, S., Wang, S., Yang, Z., Wang, L., Zhang, H., Guo, P., Gao, Y., Guo, J., 2022. A Review of Deep Learning-Based Visual Multi-Object Tracking Algorithms for Autonomous Driving. Appl. Sci. 12 (21), 10741. https://doi.org/10.3390/app122110741.

He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., Majeed, Y., Dhupia, J., 2022a. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. Comput. Electron. Agric. 195, 106812. https://doi.org/10.1016/j.compag.2022.106812.

He, L., Wu, F., Du, X., Zhang, G., 2022b. Cascade-SORT: A robust fruit counting approach using multiple features cascade matching. Comput. Electron. Agric. 200, 107223. https://doi.org/10.1016/j.compag.2022.107223.

Li, Y., Ma, R., Zhang, R., Cheng, Y., & dong, chunwang. (2023). A tea buds counting method based on YOLOV5 and Kalman filter tracking algorithm. Plant Phenomics, plantphenomics.0030. doi: 10.34133/plantphenomics.0030.

Luiten, J., Ošep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., & Leibe, B. (2021). HOTA: A Higher Order Metric for Evaluating Multi-object Tracking. International Journal of Computer Vision, 129(2), 548–578. doi: 10.1007/s11263-020-01375-2.

Miranda, J.C., Gené-Mola, J., Zude-Sasse, M., Tsoulias, N., Escolà, A., Arnó, J., Rosell-Polo, J.R., Sanz-Cortiella, R., Martínez-Casasnovas, J.A., Gregorio, E., 2023. Fruit sizing using AI: A review of methods and challenges. Postharvest Biol. Technol. 206, 112587. https://doi.org/10.1016/j.postharvbio.2023.112587.

Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A Survey on Performance Metrics for Object-Detection Algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130.

Parico, A.I.B., Ahamed, T., 2021. Real Time Pear Fruit Detection and Counting Using YOLOv4 Models and Deep SORT. Sensors 21 (14), 4803. https://doi.org/10.3390/s21144803.

Pereira, Z.C., Cruz, J.M.D.A., Corrêa, R.F., Sanches, E.A., Campelo, P.H., Bezerra, J.D.A., 2023. Passion fruit (Passiflora spp.) pulp: A review on bioactive properties, health benefits and technological potential. Food Res. Int. 166, 112626. https://doi.org/10.1016/j.foodres.2023.112626.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv Preprint arXiv:1804.02767.

Rong, J., Zhou, H., Zhang, F., Yuan, T., Wang, P., 2023. Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion. Comput. Electron. Agric. 207, 107741. https://doi.org/10.1016/j.compag.2023.107741.

Talaat, F.M., ZainEldin, H., 2023. An improved fire detection approach based on YOLO-v8 for smart cities. Neural Comput. & Applic. 35 (28), 20939–20954.

Tan, C., Li, C., He, D., Song, H., 2022. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric. 193, 106683. https://doi.org/10.1016/j.compag.2021.106683.

Tu, S., Huang, Y., Liang, Y., Liu, H., Cai, Y., Lei, H., 2024. A passion fruit counting method based on the lightweight YOLOv5s and improved DeepSORT. Precis. Agric. https://doi.org/10.1007/s11119-024-10132-1.

Vasconez, J.P., Delpiano, J., Vougioukas, S., Auat Cheein, F., 2020. Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. Computers and Electronics in Agriculture 173, 105348. https://doi.org/10.1016/j.compag.2020.105348.

Villacrés, J., Viscaino, M., Delpiano, J., Vougioukas, S., Auat Cheein, F., 2023. Apple orchard production estimation using deep learning strategies: A comparison of tracking-by-detection algorithms. Comput. Electron. Agric. 204, 107513. https://doi.org/10.1016/j.compag.2022.107513.

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.-M., 2023. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023, 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple Online and Realtime Tracking with a Deep Association Metric (arXiv:1703.07402). arXiv. http://arxiv.org/abs/1703.07402.

Wu, Z., Sun, X., Jiang, H., Mao, W., Li, R., Andriyanov, N., Soloviev, V., Fu, L., 2023. NDMFCS: An automatic fruit counting system in modern apple orchard using abatement of abnormal fruit detection. Comput. Electron. Agric. 211, 108036. https://doi.org/10.1016/j.compag.2023.108036.

Xie, J., Hua, J., Chen, S., Liu, Z., Fu, X., He, P., Yin, H., Sun, D., Wang, W., Shen, J., Li, J., 2023. Litchitrack:A Fruit-Counting Method Suitable for High-Resolution Images of Dense Small Objects [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4457380.

Yang, H., Chang, F., Huang, Y., Xu, M., Zhao, Y., Ma, L., Su, H., 2022. Multi-object tracking using Deep SORT and modified CenterNet in cotton seedling counting. Comput. Electron. Agric. 202, 107339. https://doi.org/10.1016/j.compag.2022.107339.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box (arXiv:2110.06864). arXiv. http://arxiv.org/abs/2110.06864.

Zheng, Z., Xiong, J., Wang, X., Li, Z., Huang, Q., Chen, H., Han, Y., 2023. An efficient online citrus counting system for large-scale unstructured orchards based on the unmanned aerial vehicle. J. Field Rob. 40 (3), 552–573. https://doi.org/10.1002/rob.22147.

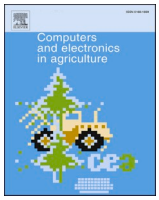
Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2021). Deformable DETR: Deformable Transformers for End-to-End Object Detection (arXiv:2010.04159). arXiv. http://arxiv.org/abs/2010.04159.

Computers and Electronics in Agriculture 221 (2024) 108997

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Tracking and monitoring of individual pig behavior based on YOLOv5-Byte

Shuqin Tu [a,b], Yifan Cai [a], Yun Liang [a,b,*], Hua Lei [a], Yufei Huang [a], Hongxing Liu [a], Deqin Xiao [a]

[a] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China
[b] Guangzhou Key Laboratory of Intelligent Agriculture, Guangzhou, China

ARTICLE INFO

Keywords:
Pig behavior tracking
YOLOv5
Behaviors analysis

ABSTRACT

In the precision-driven field of intelligent livestock breeding, the accurate monitoring and tracking of individual pig behaviors are crucial for assessing health and preventing disease. However, the complex environment of pig farms and the dynamic nature of pig behaviors can lead to issues such as false positives, false negatives, and identity switches, which pose significant challenges for the monitoring of individual pig behaviors. To overcome these challenges, this study proposes a multi-object tracking (MOT) of pigs' behavior based on YOLOv5 fusion Byte (YOLOv5-Byte). This method first uses YOLOv5 to detect individual pig behaviors (lie, stand, eat, other) and uses the Byte method to track the behaviors of individual pigs. According to the behavior tracking results, we statistically analyze the different behaviors' time of each pig for one minute. Moreover, we introduce an innovative metric by combining pig behavior information and ID value to make a more specific and precise behavior analysis of each pig. The experimental results on public and private datasets show that the best tracking results are achieved based on the YOLOv5-Byte method among the three detectors YOLOv5, YOLOv8 and YOLOX combing with the Byte tracker. On public datasets, the three MOT methods exhibit only minor differences in tracking performance between situations with and without behavior classification, with a slight decrease in metrics. On the private dataset, the YOLOv5-Byte method demonstrates superior performance in higher-order tracking accuracy (HOTA), multiple-object tracking accuracy (MOTA), and identity switches (IDs), with values of 76.5%, 94.4%, and 56. Compared to YOLOX-Byte and YOLOv8-Byte, it outperforms them with 4.2 and 9.2 percentage points higher in HOTA, 1.6 and 9.7 percentage points higher in MOTA, and 28 and 14 fewer IDs. Under varying lighting conditions, the experimental results show that YOLOv5-Byte has best tracking results, achieving the behaviors tracking accuracy of over 80%. The results demonstrate YOLOv5-Byte can accurately track pig behavior and conduct statistical behaviors analysis in intricate farm environments, which has the potential to provide a technical support for assessing the health of pigs.

1. Introduction

Pig farming has become a pillar industry in our country's livestock sector, and the health of pigs is a key concern in swine farming (Hu and Yu, 2022, Papakonstantinou et al., 2023). Numerous studies have indicated that pig behavior serves as a crucial indicator for the early detection of diseases and their diagnosis (Boyle et al., 2022, Fan et al., 2022, Mao et al., 2023). Negative interactions, including tail and ear biting, not only compromise animal welfare but also detrimentally affect the economic viability of the swine industry (Prunier et al., 2020, Grandin and Deesing, 2022). Therefore, effectively monitoring and analyzing pig behaviors have become a crucial research issue.

Currently, the traditional swine farming industry has long relied on manual monitoring, which are typically inefficient and costly. As computer technology continues to advance, monitoring systems have been effectively utilized. Monitoring systems can be divided into sensor-based contact methods, such as sensors installed on pig ears to collect behavioral data or microchips implanted in a pig's neck to obtain real-time body temperature (Larsen et al., 2021, Fuentes et al., 2022, Neethirajan, 2023). However, these methods can be stimulating to pigs and are not conducive to the management of large-scale pig farms. Video-based monitoring is another economical and productive method, providing non-contact monitoring. Compared to sensor-based contact methods, video-based non-contact monitoring has a lower impact on pigs and is more beneficial for animal welfare protection (Qiao et al., 2022, Yin et al., 2023).

In video surveillance, object detection is a crucial technology that facilitates the real-time identification and localization of multiple

* Corresponding author.
E-mail address: yliang@scau.edu.cn (Y. Liang).

https://doi.org/10.1016/j.compag.2024.108997
Received 18 December 2023; Received in revised form 24 March 2024; Accepted 28 April 2024
Available online 10 May 2024

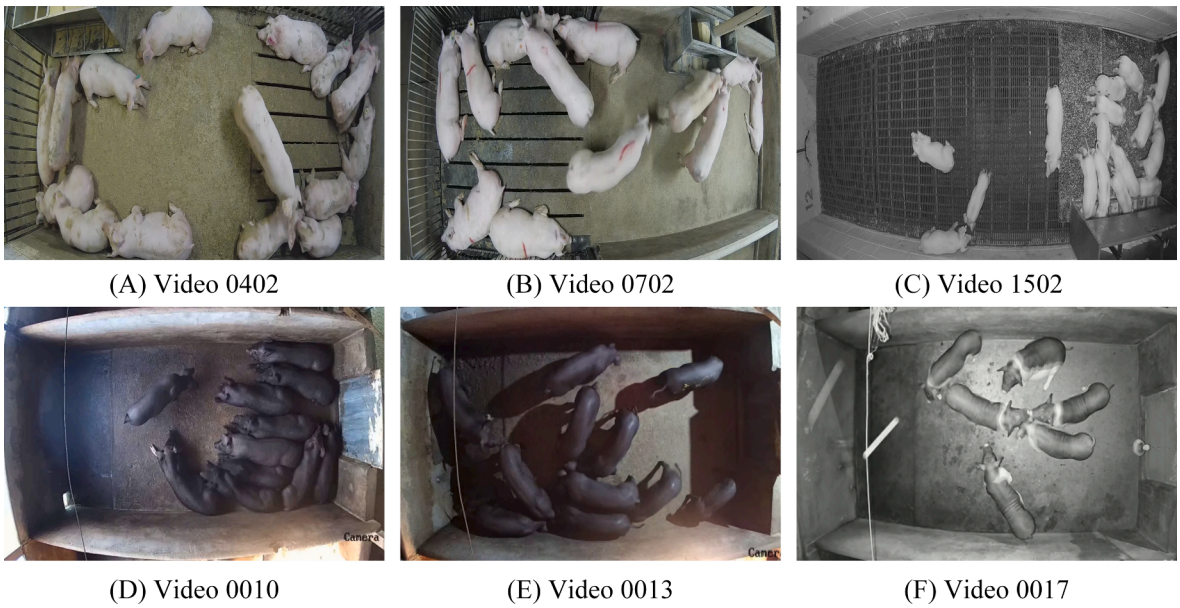


Fig. 1. The part examples of the dataset for pigs' behaviors tracking and analysis.

objects. Deep learning-based object detection algorithms are predominantly categorized into one-stage and two-stage methods. One-stage algorithms directly extract features within the network for classification and localization. The main mainstream detection algorithms include YOLOX (Ge et al., 2021) and YOLOv5.For example, Ji et al. (2022) utilized the improved YOLOX for pig recognition, which achieved a 90.9 % accuracy in seated pig recognition and an average precision of 95.7 % in overall posture recognition, effectively meeting the practical needs of pig farms. Kim et al. (2023) proposed the YOLOv5 model to estimate pig feeding behavior based on their postures. On the other hand, two-stage algorithms first use RPN (Region Proposal Network) to generate candidate boxes, and then classify and locate these candidate boxes. Notable methods include Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2017). For instance, Xiao et al. (2022) designed a Cascade Faster R-CNN pig detector to identify individual pigs and different parts of their bodies, achieving an accuracy of 98.4 % and providing technological references for group-raised pig behavioral recognition. The current one-stage target detection algorithm can not only detect object in real time, but also achieve good performance in terms of accuracy (Jiang et al., 2022, Diwan et al., 2023). The traditional two-stage target detection algorithms have good accuracy, while they are relatively slow and have a complicated training process; thus, the one-stage method is used as the detector for pig behavior tracking.

Multiple Object Tracking (MOT) is also a key technology in video surveillance. Most existing MOT algorithms can be broadly classified into three categories: Tracking By Detection (TBD), Joint Detection Embedding (JDE), and attention networks based on the Transformer series (Wojke et al., 2017, Sun et al., 2020, Wang et al., 2020). Researchers have begun to employ MOT technology to observe the behaviors of pigs. For example, Tran and Thanh (2023) utilized YOLOv7 and improved DeepSORT for object detection and tracking, analyzing the daily behavioral patterns of pigs, and constructing activity patterns for healthy pigs. Using an activity cycle-based approach, they achieved an accuracy of over 90 % in pig behavior pattern recognition within 30 min. Odo et al. (2023) introduced YOLOv4 and YOLOv7 to locate ear-biting regions, achieving detection accuracies of 98 % and 97.5 %, respectively. By combining these detections with DeepSORT and centroid tracking algorithms, they achieved 14 % and 34 % false-positive rates. Cowton et al. (2019) extracted pig behavior-related metrics from RGB cameras using Faster R-CNN and two different trackers, Sort and DeepSORT. Liu et al. (2020) simplified group-level behavior into pairwise interactions using tracking and detection algorithms. They combined convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to extract spatiotemporal features and classify behavior categories effectively, enabling the accurate localization and identification of tail-biting behavior.

Due to superior tracking capabilities of trackers using TBD paradigm, many researchers utilize them to monitor the status of pigs. Research results using the trackers of TBD paradigm for pig tracking are inevitably influenced by detectors. However, in real commercial pig farming environments, most detection algorithms are sensitive to changes in lighting conditions. The performance of pig tracking largely depends on the object detector used, and few researchers have evaluated which detectors are best suited for pig tracking. In addition, the above-mentioned studies not only concentrate on behaviors classification of pigs, but also monitor the status of individual pigs, thereby emphasizing the crucial role that detectors play in these processes.

Addressing the challenges mentioned above, this paper presents a tracking algorithm based on YOLOv5-Byte for each pig behavior analysis under real commercial pig farming environments. Firstly, we combined three different detection algorithms, YOLOv5, YOLOX, and YOLOv8, with the Byte tracker, and complete their comparison experiments in pig behavior classification and tracking (Zhang et al., 2022). Secondly, we assess the tracking performance of three detectors combing Byte tracker under varying lighting conditions. Finally, we utilized the optimal detector (YOLOv5) combining with Byte tracker to analyze the behavioral data of pigs, and evaluated the accuracy of behavioral tracking, providing valuable technical support for the early warning of pig health status.

The primary contributions of this paper can be summarized as follows:

(1) We had introduced a novel metric for calculating the behavioral tracking accuracy of each pig.

(2) We compared the pig tracking performance under situations with behavior and without behavior classification using YOLOv5, YOLOX, and YOLOv8 three detectors combining with the Byte tracker.

(3)We chose the YOLOv5-Byte tracking algorithm to evaluate pig behavior and conducted statistical analysis, ultimately calculating pig behavior tracking accuracy

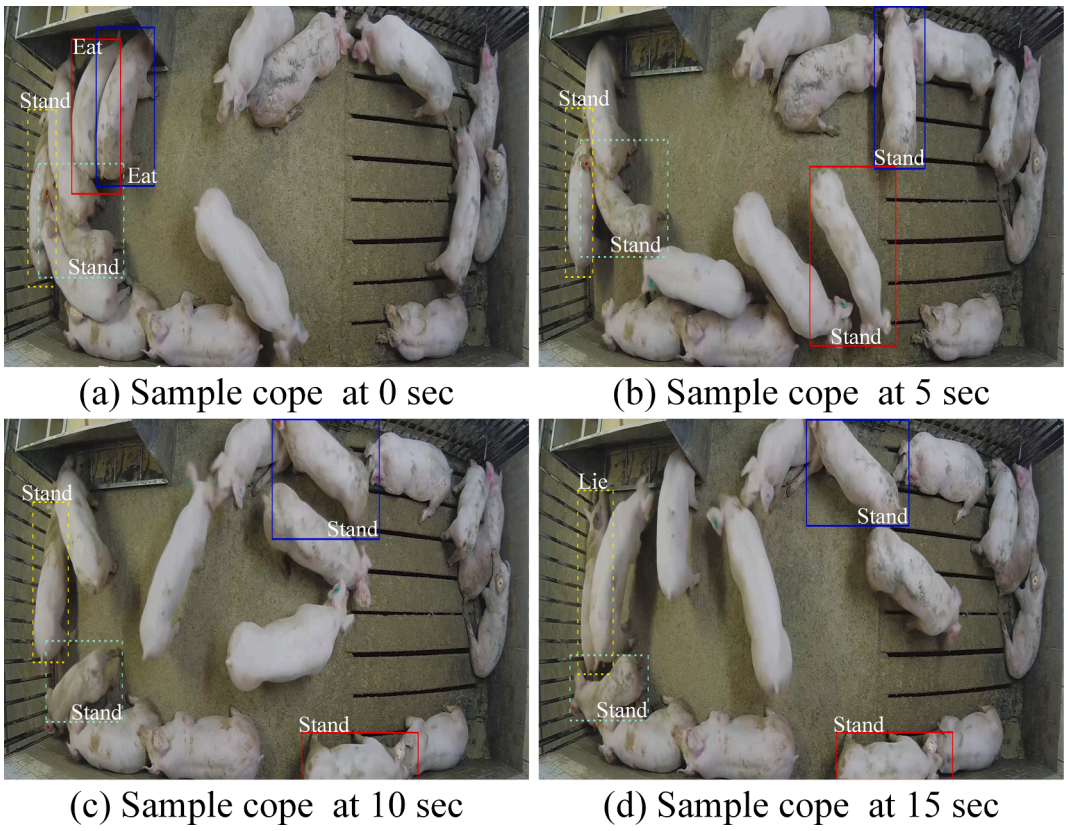


Fig. 2. Representation the challenges in pig behaviors within a 15-second period.

2. Material and methods

2.1. Materials

In this study, the dataset used includes two parts: one part is derived from the publicly available data set provided by (Psota et al., 2020),and the other part is obtained from Leshan City, Sanshui District, Foshan City, Guangdong Province, China. The main purpose of using these two datasets is to provide enough samples to ensure the validity and accuracy of our experiments. The public dataset contains 23 videos, and the private data set contains 18 videos. Each video is one minute long (5 frames/second), with a total of 300 frames. All videos have been cropped by FFmpeg software. We used the DarkLabel software to annotate the position information for each pig, as well as adding corresponding behavioral and identity information for each pig. It is worth noting that we annotated each pig's information in scenes with low light or even obstructed views. This method differs from that of MOT20, which discards objects under heavy obstructions or unclear visibility. Our processing method guarantees data accuracy by ensuring the consistent number of pigs in every frame.

we utilized the dataset for pig behaviors identification and for validating multi-object tracking algorithms. Fig. 1 showed some examples of the dataset. Fig. 1(A), (B), and (C) present the examples from the public datasets, and Fig. 1 (D), (E), and (F) display private dataset. We categorize the pigs' behaviors status into four types based on their representative features: 'stand', 'lie', 'eat', and 'other'. The detailed classifications aid us to understand pig behaviors patterns. However, there are significant challenges in performing behaviors recognition and analysis in the complex pig farm environment. Pigs exhibit unstable movement speeds, variable postures, and significant occlusions. These factors pose major challenges to detection and tracking. As illustrated in Fig. 2, four images are captured at 5-second intervals within a 15-second period. In Fig. 2 (a), numerous pigs densely clustered together, and pigs within the green and red bounding boxes are heavily occluded. Fig. 2 (a–d), pigs within blue and yellow bounding boxes cover larger activity areas, exhibit unpredictable directions, and display varying behaviors.

2.2. Method

We combine Byte tracker with the most advanced three detectors YOLOv5, YOLOX, and YOLOv8. The system process diagram was shown in Fig. 3. The framework of pigs' behaviors tracking consists of three components: (i) pig detection and behaviors classification, (ii) object tracking, (iii) behaviors analysis.

2.2.1. Detector and classification

YOLO (You Only Look Once) is a fast and accurate object detection algorithm. The research adopts the TBD paradigm. The TBD paradigm relies on detection performance, we use three different detectors from the YOLO series, YOLOv5, YOLOX, and YOLOv8, as detectors to compare the behaviors classification and the tracking effects. YOLOv5, YOLOX, and YOLOv8 have all achieved excellent results on the COCO datasets. Their three different detectors can be divided into different structures according to the depth of the network or the size of the parameters. To balance speed and accuracy, we selected the medium model M of detectors for model training. The YOLOv5, YOLOX and YOLOv8 network structures include four stages. The first stage of the model is input procedure, which uses a variety of data enhancement methods to increase the diversity of samples to the model to improve robustness. The second stage is the backbone network, which is based on Darknet-53 and enhances the ability to extract backbone features to varying degrees. The third part is the detection neck (PAFPN), which adopts the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures to enhance the model's feature representation capability and receptive field. The fourth stage is the decoupled head (Prediction), used for object detection and classification, aiming to

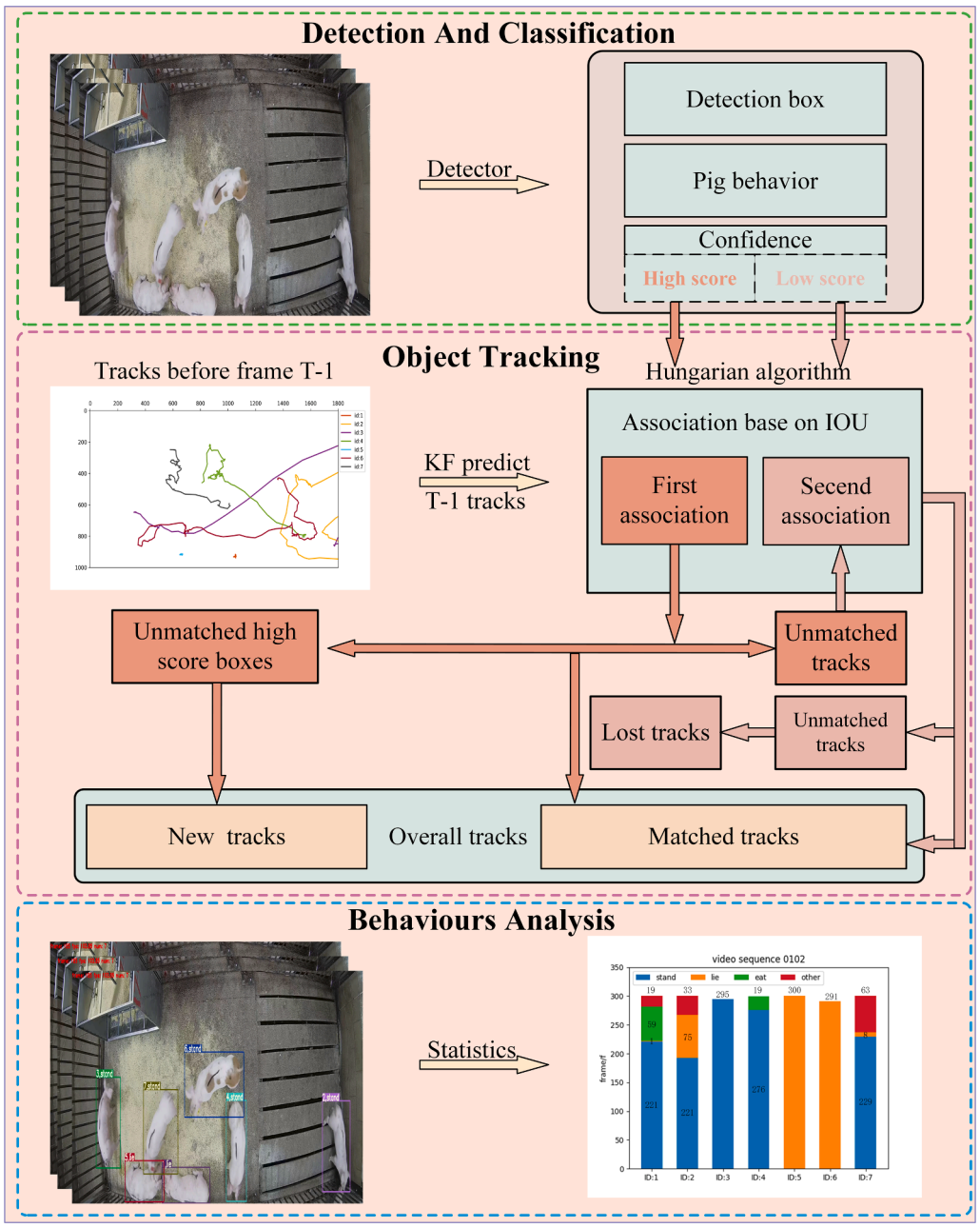


Fig. 3. System process diagram.

improve convergence speed and accuracy.

2.2.2. *Byte tracker for MOT*

Compared with the current mainstream tracking algorithms, Byte will retain all detection results and track objects according to detection scores from high to low, reducing losses during the tracking process. Its core is composed of the Kalman Filtering (KF) and the Hungarian algorithm. KF is widely utilized in fields such as radar, autonomous driving, and MOT tasks. Its function is to update the predicted value through its measured value to achieve accurate estimation. Byte tracker uses KF to predict the tracks information of the previous frame, matches the tracks with the detection box based on the IOU (intersection over union ratio) value, and then obtains the final matching result through the Hungarian algorithm. In pig behaviors tracking, it is necessary to estimate two states of the tracks: the mean and the covariance. The mean represents the position information of the pig, which is composed of the center coordinates of the bounding box (cx, cy), aspect ratio r, height h, and their respective velocity change values. It is represented as an 8-dimensional vector x = [cx, cy, r, h, vx, vy, vr, vh], with all velocity values initialized to 0. The covariance represents the uncertainty in the pig's position information and is represented by an 8x8 diagonal matrix.

KF consists of two stages: first, predicting the trajectory of the next frame; second, revising the predicted position according to the detection value. The prediction phase is defined by Eq (1) and (2). In Eq (1), x represents the mean of the track at time t-1, and F is the state transition matrix, which predicts x' at time t. In Eq (2), P represents the covariance of the track at time t-1, and Q is the system noise matrix, which predicts P' at time.

$$x' = Fx$$

$$P' = FPF^T + Q$$

The update phase is based on the detected pig positions at time t, correcting their associated tracks states, and is composed of Eq (3), (4),

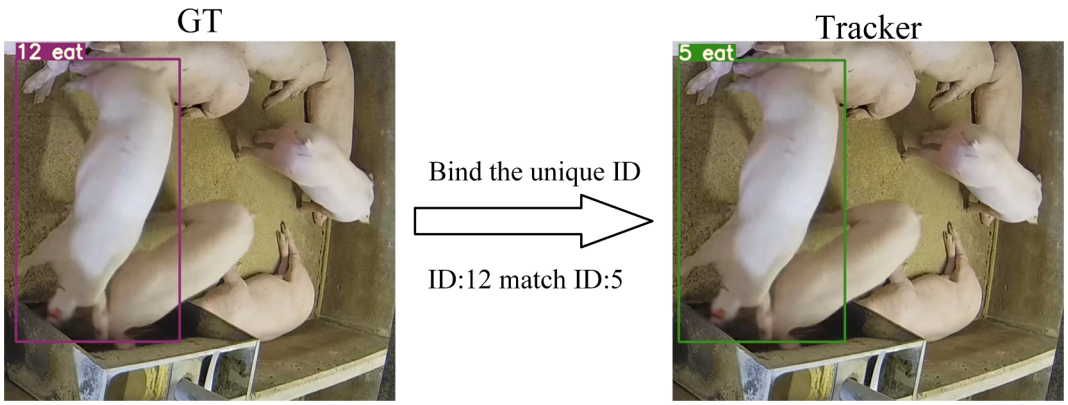


Fig. 4. Examples of statistical methods for pig behaviors.

(5), (6), and (7). In Eq (3), z represents the mean of the detection, z = [cx, cy, r, h], where H is the measurement matrix. Its function is to map the mean vector of the tracks to the detection space and calculate the mean error between the detection and the tracks values.

$$y = z - Hx'$$

In Eq (4), R is the noise matrix of the detection. This equation first maps the covariance matrix to the detection space and then adds the noise matrix R to enhance its reliability. Eq (5) calculates the Kalman gain, K, which is used to estimate the importance of measurement errors.

$$S = HP'H^T + R$$

$$K = P'H^T S^{-1}$$

The mean x, and the covariance matrix P are updated by Eq (6) and (7), where I is the identity matrix.

$$x = x' + Ky$$

$$P = (I - KH)P'$$

When the KF has completed the predicting the tracks from time t-1 to the current time, the IOU distance matrix between the highly scored box and the predicted trajectory is matched using the Hungarian algorithm. If the high-scoring detection boxes fail to match and exceed the threshold for initializing the tracks, new tracks are created. For unmatched trajectories, matching will continue in the second round. In the second round, low score detection boxes are matched with the unmatched trajectories using IOU. If there is no successful match for over 60 frames, the tracks are deleted. Finally, the trajectory T is updated jointly by the newly initialized trajectory and the matching trajectory.

2.2.3. Behaviors analysis

After the video tracking was completed, we generated a statistical chart that included the behaviors information and ID of each pig. To evaluate the performance of the tracker and detector, we introduced the Behavior-Tracking Accuracy (BTA) metric, which measures the degree of agreement between the ID assigned by the tracker and the predicted behaviors with the ground truth (GT). The BTA metric had certain differences from the existing MOTA and IDs metrics. The MOTA metric mainly focuses on the accuracy of object detection, and the IDs metric focuses on the frequency changes of individual pig ID during the tracking process.

In the construction of our evaluation metric, we incorporated the advantages of MOTA and IDs. We combined pig behaviors information and ID information into one standard, which allows us to make a more specific and precise assessment of each pig. In an ideal tracking scenario, the ID of each pig should remain constant, and every pig should be tracked. When the ID of a pig changes, it suggests a possible temporary loss of tracking. In this scenario, we were only judging its behaviors, and couldn't accurately reflect the unique behaviors pattern of this pig, as its individual uniqueness has been lost in the process of ID change. Therefore, we only recorded the number of matches between its behaviors and GT when the pig's ID remains stable.

BTA is defined in the equation (8), the variable BT represents the total number of pig matches, TC represents the total number of video frames. In calculating the BTA value, we needed to note that the ID randomly assigned by the tracker may differ from our annotated ID. To solve this problem, we used the IOU matching method to confirm the unique corresponding ID. For example, in Fig. 4, the pig in the annotated purple box was given the ID 12, while the green box assigns the ID 5 by the tracker. Through IOU matching, we could conclude that the pig with ID 12 and the pig with ID 5 were the same pig, and in the entire calculation process, the ID 12 always corresponded to the ID 5. After this unique corresponding ID was determined, the behaviors of this pig could be evaluated. In all matching processes, if the ID of the pig does not change, we will compare its behaviors information with the GT. As shown in Fig. 4, if the annotated behaviors of the pig in the purple box is 'eat', and the predicted behaviors in the green box was also judged as 'eat', then we consider this as a successful BT match. By accumulating all successful behaviors matches times, we could calculate the value of BTA.

$$BTA = \frac{\sum_t BT}{TC}$$

2.3. Evaluation metrics

This study evaluates the performance of the pig behaviors recognition model using four evaluation metrics: Precision (P), Recall (R), F1 Score and mean Average Precision (mAP). TP represents True Positives; FN stands for False Negatives; FP represents False Positives. The standard for correct detection is set as the ratio of IOU not lower than 0.5. The definitions of these metrics are given by Eq (9)-(12).

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 = \frac{2PR}{P + R}$$

$$\mathrm{m}AP = \frac{\sum_{c=1}^{C} AP(c)}{C}$$

In pig behaviors tracking tasks, we selected three main evaluation metrics: High Order Tracking Accuracy (HOTA)(Luiten et al., 2021), Multiple Object Tracking Accuracy (MOTA), and Identification Average Rate (IDF1) for performance comparison. HOTA decomposes the

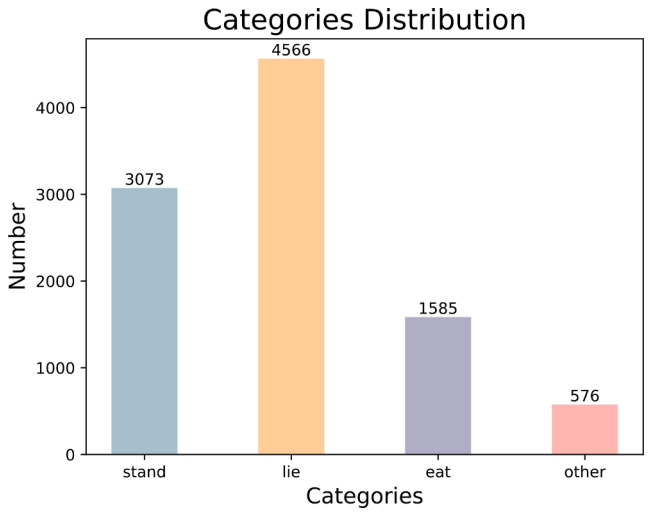


Fig. 5. Categories distribution of the training Set.

tracking task into three sub-tasks, detection, association, and localization, to obtain a more balanced metric that is more consistent with human visual perception. MOTA provides an intuitive evaluation method that reflects the tracker's performance in object detection and tracks maintenance. IDF1 focuses on the continuity and accuracy of tracking object.

The HOTA calculation formula is shown in Eq (13), DetA is used to evaluate the inspection accuracy, and AssA is used to evaluate the correlation accuracy. C is a point belonging to TP, A(c) represents the association accuracy defined by Eq (14), where TPA(c) represents the accuracy of correct association, FPA(c) represents the predicted tracks accuracy of incorrect association, and FNA(c) represents none Predicted tracks accuracy of correlation prediction.

$$HOTA = \sqrt{DetA \cdot AssA} = \sqrt{\frac{\sum_{c \in TP} A(c)}{TP + FN + FP}}$$

$$A(c) = \frac{TPA(c)}{TPA(c) + FPA(c) + FNA(c)}$$

The MOTA calculation is shown in Eq (15), FP_t represents the number of False Positives at time t,FN_t represents the number of False Negatives at time t,$IDSW_t$ represents the number of ID change at time t, and GT_t represents the total number of all objects.

$$MOTA = 1 - \frac{\sum_t (FP_t + FN_t + IDSW_t)}{\sum_t GT_t}$$

The calculation formula of IDF1 is shown in Eq (16), IDTP represents the true positive ID number, IDFP represents the false positive ID number, and IDFN represents the false negative ID number.

$$IDF1 = \frac{2IDTP}{2IDTP + IIDFP + IDFN}$$

Additionally, this study evaluates the methods performance through other four metrics: Identity switches (IDs), False Positives (FP), False Negatives (FN) and Frames Per Second (FPS).

3. Results and analysis

3.1. Experimental platform and parameter settings

To investigate the performance of different detectors in tracking pigs in natural environments, five experiments were conducted: (1) pigs behaviors recognition (2) the impact of pig behaviors classification on MOT, (3) pig behaviors tracking based on YOLOv5-Byte, (4) behaviors tracking comparison under different lighting conditions. (5) statistical analysis and evaluate pig behavior information of each pig. For this experiment, we standardized the training image size to 640 × 640 pixels for three different detectors. The learning rate, batch size, and epoch were set to 0.01, 8, and 200. The experiments were conducted on a Linux platform with Ubuntu 18.04 operating system. The hardware configuration included a 12th Gen Intel(R) i9-12900KF CPU, NVIDIA GeForce RTX 3090 GPU, and 32 GB of RAM. PyTorch version 1.11.1, Python version 3.7, and CUDA version 11.3 were utilized for the experiments.

In this study, we adopted a multi-object tracking algorithm based on the TBD paradigm. The performance of target detection will directly impact the subsequent tracking effect. To effectively train the detector, we segmented the video into images, extracting one every ten frames, to ensure the diversity of sample features. The obtained images were allocated into a training set (889 images), a validation set (254 images), and a testing set (128 images), following a ratio of 7:2:1. As depicted in Fig. 5, the distribution of samples in the training set is as follows: the numbers of samples for the behaviors—'stand', 'lie', 'eat', and 'other'—are 3073, 4566, 1585, and 576 respectively. The number of 'other' samples numerous is nearly 8 times less than the most numerous 'lie', which poses a certain difficulty for training the detector.

Table 1
Detection results of YOLOv5, YOLOX, and YOLOv8.

Model	P /%↑	R /%↑	F1/%↑	mAP/%↑
YOLOv5	92.0	87.8	89.8	91.8
YOLOX	91.0	88.2	89.5	90.5
YOLOv8	89.8	86.1	87.1	89.9

Table 2
Detection results for each category by YOLOv5.

Class	P /%↑	R /%↑	F1/%↑	mAP/%↑
stand	94.2	90.4	92.2	93.6
lie	97.8	86.6	91.8	92.9
eat	91.8	89.9	90.8	92.5
other	84.2	84.2	84.2	88.4

3.2. Comparison of pig behaviors recognition results of different detector

In Table 1, we presented the performances of three detectors. The three detectors achieved good results with precision, recall, F1, and mAP. YOLOv5 obtained 92.0 %,87.8 %,89.8, and 91.8 % in precision, recall, F1, and mAP, respectively. YOLOv5 surpasses YOLOX by 1 % in precision and 0.3 % in F1 scores, also outperforms YOLOv8. It is found that YOLOv5 achieved the best detection results among three detectors. Then, we chose YOLOv5 detector to analyze the results of pig behaviors categories. In Table 2, we provided detailed analysis for these results of YOLOv5 detector. The P and R metrics of the 'other' category both stand at 84.2 %, markedly differing from the scores of the other categories. The reason is that the "other" category has fewest samples compared with three categories. Despite the challenges posed by sample imbalance, the YOLOv5 model still showed excellent performance in key metrics such as precision and mAP. These results indicate that the YOLOv5 model provides a good foundation for improving behaviors tracking accuracy.

3.3. The impact of pig behaviors classification on MOT

To assess the impact of pig behaviors on MOT, we combined the Byte method with different detectors and compared their performance on public datasets, as shown in Table 3. Without considering behaviors classification, YOLOv5-Byte attained 81.1 % and 96.3 % in HOTA and MOTA, respectively. Compared to the other two methods, YOLOv5-Byte

Table 3
Tracking results for pigs and its behaviors tracking.

Method	Behaviors	HOTA/%↑	MOTA/%↑	IDF1/%↑	IDs↓	FN↓	FP↓	FPS↑
YOLOv5-Byte		81.1	96.3	94.5	48	1297	252	100
	√	80.1	96.0	93.8	54	1578	228	100
ByteTrack		80.6	95.6	95.9	37	956	899	91
	√	78.9	95.0	92.9	68	980	1130	91
YOLOv8-Byte		77.6	92.7	93.0	52	2572	591	68
	√	77.3	92.3	92.7	52	2582	755	68

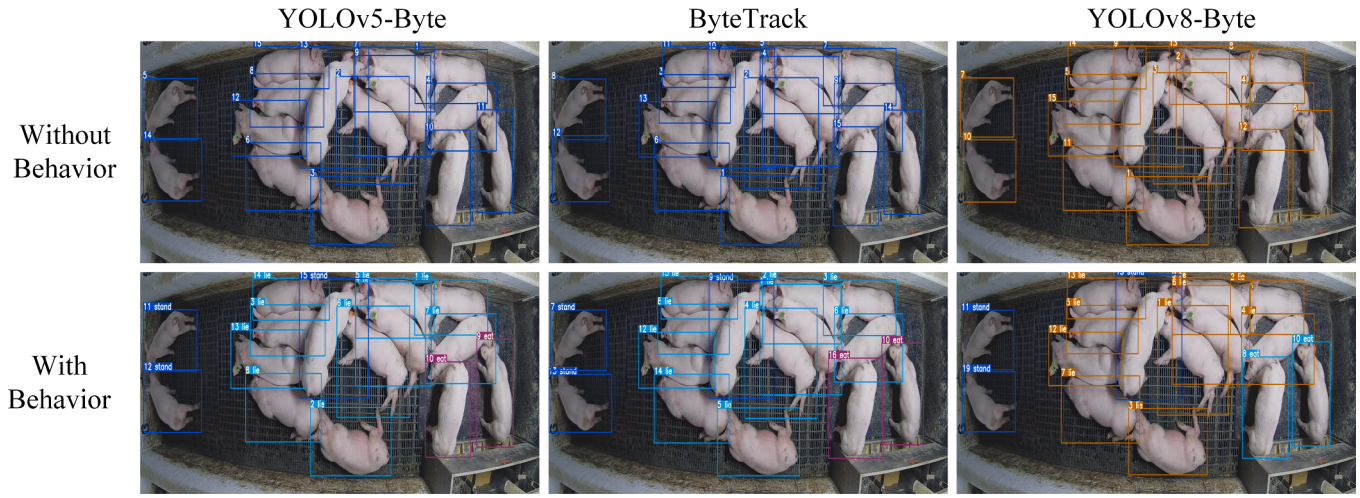


Fig. 6. Visual comparison results of pig tracking and its behaviors tracking.

Table 4
Pig behaviors tracking results on the private datasets.

Method	HOTA/%↑	MOTA/%↑	IDs↓	IDF1/%↑	FN↓	FP↓	FPS↑
YOLOv5-Byte	76.5	94.4	56	87.9	1113	119	107
ByteTrack	72.3	92.8	84	83.0	1055	467	93
YOLOv8-Byte	67.3	85.7	70	81.6	2861	241	86

improved by 0.5 % and 3.5 % in HOTA and 0.7 % and 3.6 % in MOTA. Furthermore, YOLOv5-Byte achieved the best results in pig tracking with behaviors classification, with HOTA and MOTA values of 80.1 % and 96.0 %, respectively. When considering pig behaviors classification, its HOTA values decreased by 1.0 %, 1.7 %, and 0.3 %, and MOTA values decreased by 0.3 %, 0.6 %, and 0.4 %, respectively, compared to results without behaviors tracking. Overall, YOLOv5-Byte demonstrated good performance in pig behaviors tracking compared to the other two methods.

The study results indicated that YOLOV5-Byte performs well in pig behaviors tracking, and although the tracking performance slightly decreased compared to tracking pigs without behaviors, the impact was not significantly detrimental. Behaviors tracking provided essential technical support for assessing the health of pigs in production and farming and was more suitable for practical livestock needs.

We presented some visual results of the YOLOv5-Byte, ByteTrack, and YOLOv8-Byte algorithms in Fig. 6. For pig tracking without behavior classification, all three algorithms correctly identified the number of pigs (as shown in the upper line of Fig. 5). For pig tracking with behaviors classification, the maximum ID numbers obtained by the YOLOv5-Byte, ByteTrack, and YOLOv8-Byte three algorithms were 15 (equal to GT), 16 and 19 (as shown in the bottom row of Fig. 5), respectively. And only YOLOv5-Byte demonstrated consistent tracking without any IDs error, whereas both ByteTrack and YOLOv8-Byte

Table 5
Pig behaviors tracking results with YOLOv5-Byte.

No.	HOTA/%↑	MOTA/%↑	IDs↓	IDF1/%↑	FN↓	FP↓
0010	79.4	97.9	6	94.7	58	4
0011	73.2	94.6	5	85.4	164	8
0012	62.7	86.3	21	75.1	392	39
0013	74.8	95.4	8	81.6	133	11
0014	84.5	100	0	100	0	0
0015	83.0	94.2	4	89.9	100	0
0016	87.6	100	0	100	0	0
0017	87.1	99.9	0	99.9	0	1
0018	56.5	81.4	12	64.6	266	56
Overall	76.5	94.4	56	87.9	1113	119

occurred error IDs. Therefore, YOLOv5-Byte achieved the best accuracy in pig tracking with behaviors classification, and the behaviors tracking had a minimal impact on MOT results.

To evaluate the generalization ability and practical effectiveness of YOLOv5-Byte, we conducted a comparative analysis on private datasets under different scenarios, as presented in Table 4. YOLOv5-Byte achieved 76.5 %, 94.4 % and 56 in HOTA, MOTA, and IDs, respectively, outperforming ByteTrack and YOLOv8-Byte by 4.2 % and 9.2 % in HOTA, 1.6 % and 8.7 % in MOTA and decline 28 and 14 in IDs, respectively. Additionally, YOLOv5-Byte outperformed ByteTrack and YOLOv8-Byte in IDF1, FP, and FPS metrics. These results highlighted the exceptional generalization abilities and precise tracking performance of YOLOv5-Byte across various datasets.

The MOT results based on YOLOv5-Byte for each test video were presented in Table 5. For video 0014, HOTA, MOTA, IDF1, and IDs were 84.5 %, 100 %, 100 %, and 0, respectively. In contrast, for video 0018, the corresponding metrics were 56.5 %, 81.4 %, 64.6 %, and 12, revealing a notable difference compared to the other video results. This reason could be attributed to environmental factors. The video 0014 was recorded on daylight with enough sunlight and slow movement

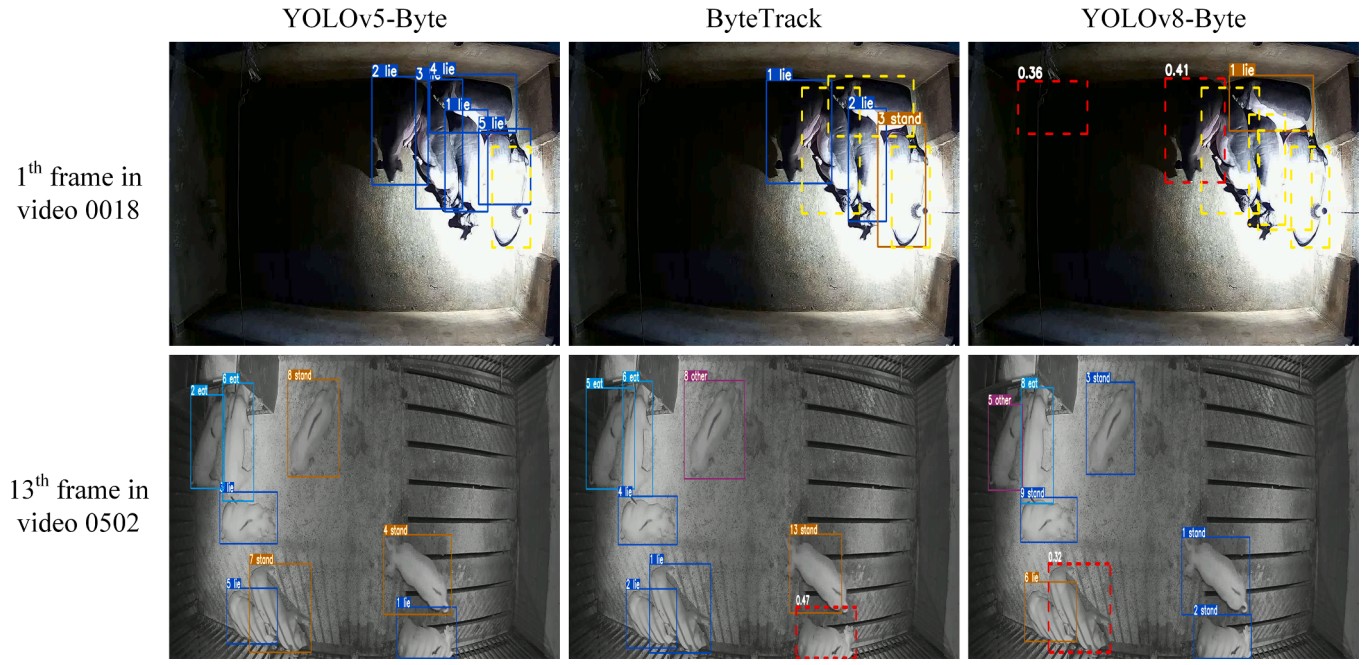


Fig. 7. Tracking comparison of three algorithms in night scenarios.

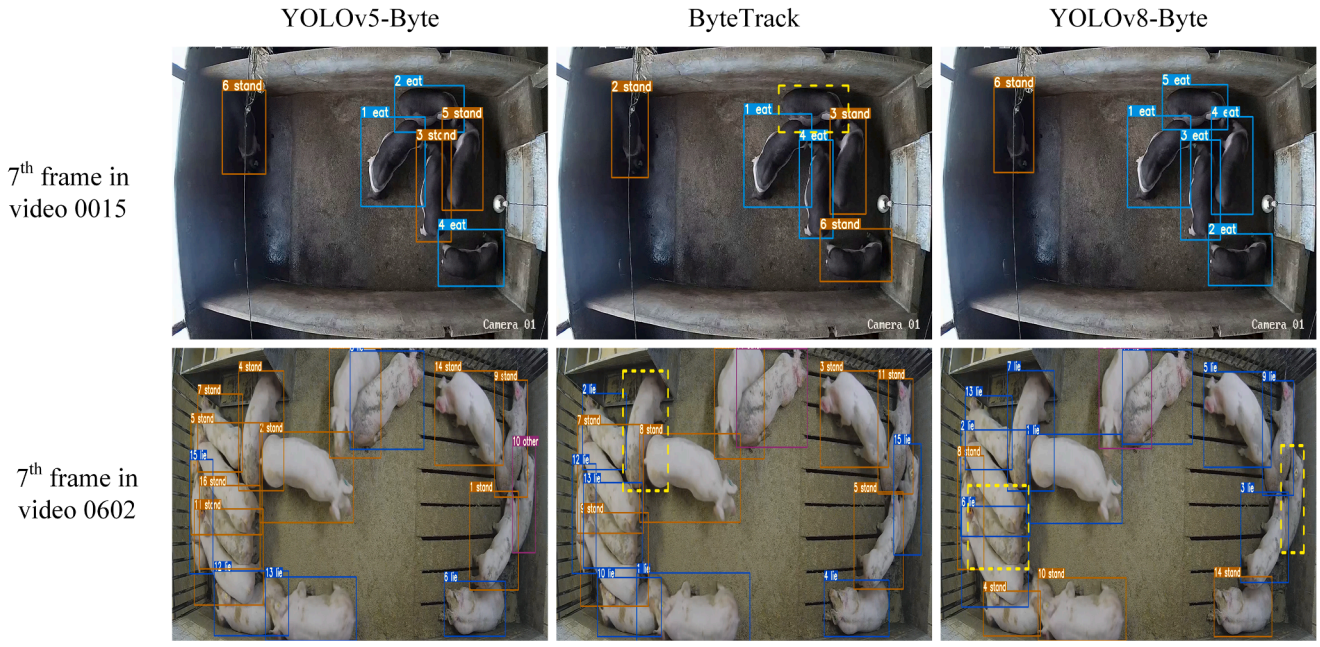


Fig. 8. Tracking comparison of three methods in daytime scenarios.

condition, while video 0018 was recorded at night as show the upper line of Fig. 6, leading to strong exposure occurred in the illuminated area and inadequate lighting in other areas. In conclusion, these observations demonstrated that lighting conditions pose the significant impact for the efficacy of MOT results.

3.4. Behaviors tracking comparison under different lighting conditions

To compare the performance of three methods for MOT under night conditions, two nighttime videos were selected for MOT tests. Video 0018 was recorded at 10:00 PM, where many pigs gathered to sleep, with partial exposure. Video 0502 was recorded at 4:00 AM, with multiple pigs eating frequently. The visual results of the three methods were shown in Fig. 7.

In the 1-th frame of Video 0018 as shown the upper row of Fig. 7, the YOLOv5-Byte method exhibited a false negative for one pig, as indicated by the yellow bounding box, the ByteTrack method displayed false negatives for three pigs, and the YOLOv8-Byte method produced false positives for two pigs and false negatives for four pigs, successfully tracking only one pig. When tracking pigs in the 13-th frame of Video 0502 (as shown the bottom row of Fig. 6), the YOLOv5-Byte method successfully matched the maximum object ID to the real numbers with 8 without missed target. The ByteTrack failed to accurately track all objects with IDs, resulting in a maximum pig ID with 13 that differed by 5

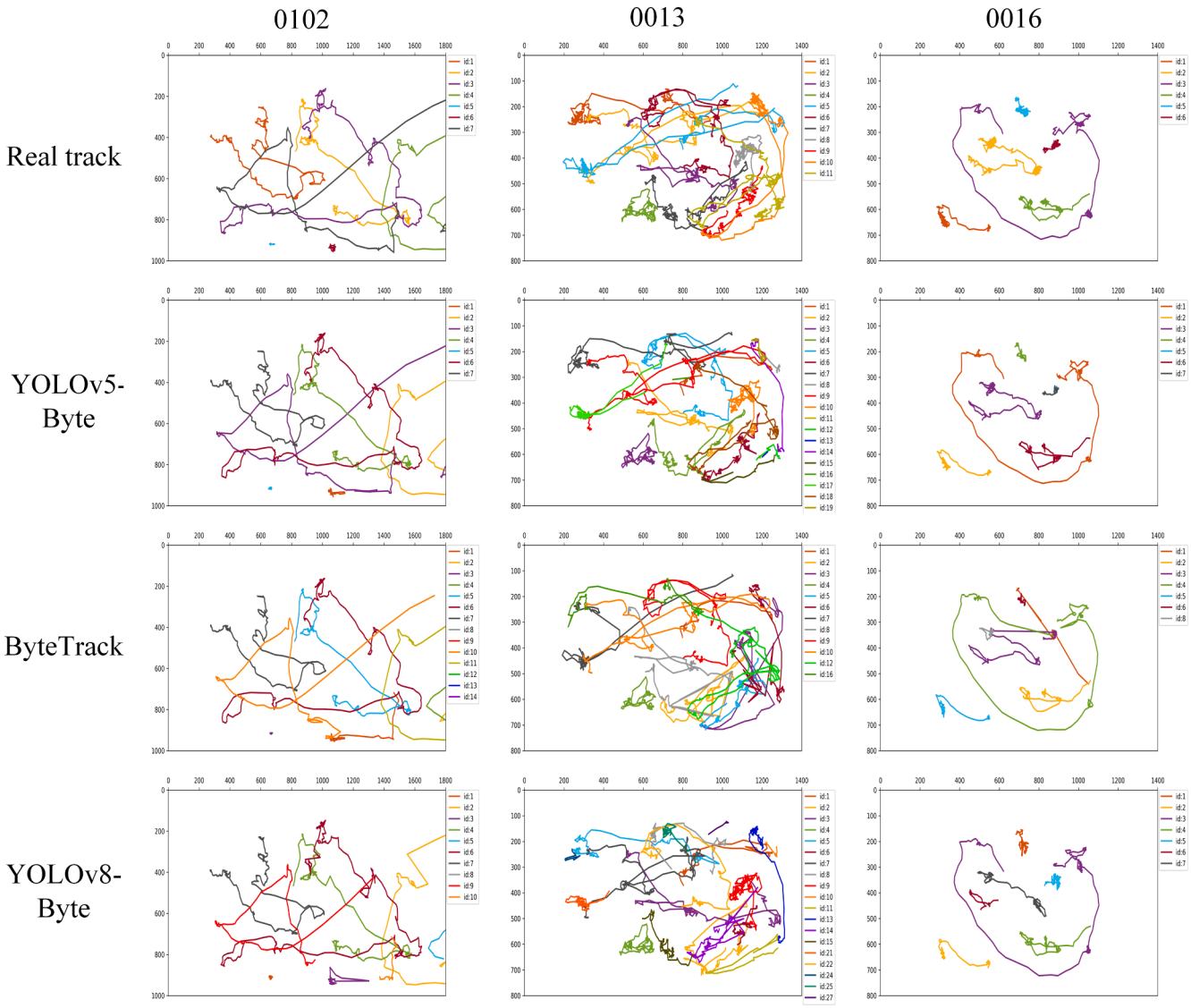


Fig. 9. Representation the visualization of pig activity tracks of three methods.

from real numbers. The YOLOv8-Byte method failed to track some pigs due to the low confidence of the detector, as shown in the red box. The analysis results suggested that the YOLOv5-Byte method achieved superior performance in low-light conditions, potentially due to its better detection capabilities in night environments compared to the other two methods. This enabled it to achieve the best MOT performance.

The visual results for pig behaviors tracking using the three methods under daytime conditions were shown in Fig. 7. The results indicated that the YOLOv5-Byte method achieved the best tracking performance under enough lighting conditions. The YOLOv5-Byte method successfully tracked all pigs in both the 7-th frame of Videos 0015 and 0602 (shown in the first row of Fig. 7), without incorrect IDs appear. However, the ByteTrack method failed to track one pig in both the 7-th frame of Videos 0015 and 0602, as indicated by the yellow bounding boxes in the middle row of Fig. 8. The YOLOv8-Byte method tracked all pigs in the 7-th frame of Video 0015 but missed two pigs in the 7-th frame of Video 0602 as shown in the last row of Fig. 7.

Based on the above analysis of the results under varying lighting conditions, it was found that the YOLOv5-Byte method exhibits the best MOT performance, regardless of the lighting conditions. Therefore, it was more suitable for intelligent analysis of group pigs under video surveillance.

Table 6
0102 video behaviors tracking accuracy.

Individual	1	2	3	4	5	6	7
Match ID	7	4	6	2	5	1	3
BT	243	194	295	276	300	291	228
BTA(%)	81	64.7	98.3	92.0	100.0	97.0	76.0

3.5. Statistical analysis

In addition to quantitative analysis, we presented estimated tracks of pig activities for 0102,0013, and 0016 videos. The pig activity tracks of three methods and their real tracks were shown in Fig. 9. Among the three methods, the pig activity tracks of YOLOV5-Byte were the closest to the real pig tracks. In the ByteTrack method, there was a difference between the maximum pigs ID and the actual number of pigs in videos 0102, 0013 and 0016. In the YOLOV8-Byte method, there was a significant issue of incorrect pig ID change in videos 0102, 0013 and 0016. In video 0016, the movement tracks of pigs with ID 6 and 7 are discontinuous.

We calculated the pig BTA values using YOLOv5-Byte according to Eq (8). The obtained BTA results of YOLOv5-Byte for videos 0102, 0013,

Table 7
0013 video behaviors tracking accuracy.

Individual	1	2	3	4	5	6	7	8	9	10	11
Match ID	7	9	2	3	1	5	4	10	6	8	11
BT	297	300	189	300	107	291	270	244	118	69	0
BTA(%)	99.0	100.0	63.0	100.0	35.7	97.0	90.0	81.3	39.3	23	0

Table 8
0016 video behaviors tracking accuracy.

Individual	1	2	3	4	5	6
Match ID	2	3	1	6	4	5
BT	295	277	283	128	273	76
BTA(%)	98.3	92.3	94.3	42.7	91	25.3

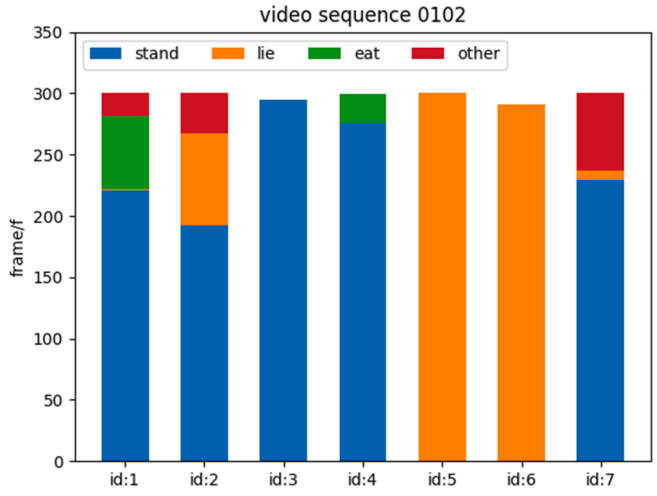


Fig. 10. Time statistics of four different behaviors for seven pigs in the video 0102.

and 0016 were presented in Tables 5, 6, and 7, respectively. In the three videos, over half of the object pigs achieved BTA values with over 80 %, with a few pigs having BTA equal to 100 %. In Table 6, pigs with ID 3 to 6 had BTA values of over 92 %. In Table 7, pig with ID 2 and pig with ID 4 had BTA values of over 92 %. In Table 8, pigs with ID 1, 2, 3, and 5 had BTA values of over 91 %. To analyze the different behaviors tracking information for each individual pig, we selected video 0102 for four behaviors time statistics, as shown in Fig. 10. Pig with ID 5 exhibited a single behavior with an accuracy of 100 %, while pig with ID 1 displayed varied behaviors with an accuracy exceeding 80 %. These results indicate that YOLOv5-Byte achieves a high accuracy in pig behaviors tracking.

4. Discussion

With the continuous development of computer vision, MOT technology is widely used, especially in people and vehicles tracking. Some researchers established public datasets such as MOT20 and KITTI (Voigtlaender et al., 2019) to evaluate tracking performance. Animal tracking was more difficult than human or vehicle tracking due to animals have very similar appearance features and unpredictable behaviors. Zhang et al. (2023) introduced a dataset specifically designed for multi-object animal tracking and evaluated the performance of different methods. This research provided a reference for a detailed understanding of animal tracking and offered valuable insights for further research on animal behaviors tracking. Compared to animal tracking, there were fewer studies on pig tracking. In large-scale breeding farms, the number of pigs was large, and it was easy to cause occlusion. Farms were usually indoors, and lighting conditions may be unstable. These factors could easily cause challenges to pig tracking. Guo et al. (2023) studied three different tracking methods, JDE, FairMOT, and YOLOv5s with DeepSORT, and optimized pig re-identification (re-ID) which improved tracking performance and reduced the IDs. The studies above focused on pig tracking and did not consider pig behaviors tracking. In this study, we introduced behaviors tracking of pigs and conducted statistics and analysis on pig four behaviors. Our previous work explored an improved DeepSORT algorithm for pig behaviors tracking. Compared to the original DeepSORT method, the improved approach achieved increases of 1.8 % in MOTA and 6.8 % in IDF1 while reducing the number of IDs by 80 % (Tu et al., 2022). However, our previous work did not conduct corresponding statistics and analysis on pig behaviors tracking.

Therefore, we proposed an effective YOLOv5-Byte method for monitoring pig behaviors, it can achieve real-time tracking performance and provide good tracking accuracy. Our study had some limitations, which are that the test videos were short and did not cover all time periods. We addressed common issues in practical production, including changes in night lighting and real-time monitoring of pig behaviors. Our study aimed to provide the necessary technical support for continuous tracking of pig behaviors, to better assess the health condition of pigs. Our next step research work is to study long-term pig behaviors tracking and analysis.

5. Conclusion

In this study, we firstly investigated the task of tracking pig behaviors in a commercial environment. In our study, we employed three diverse detectors, YOLOv5, YOLOX, and YOLOv8, to identify pigs' behaviors and locate pigs, then combined with the Byte tracker to track the pig behaviors, to evaluate the performance of each detector in pig behaviors tracking. we compared the performance of the three methods without and with behaviors classification under different lighting conditions. The results show that YOLOv5-Byte performs best compared to the other two methods. Finally, we selected YOLOv5-Byte as the behavioral statistics method. In the experiments of pig behaviors classification, we found that the tracking performance slightly decreased compared to tracking pigs without behaviors classification, but it did not have a significant impact. In experiments with different lighting conditions, YOLOv5-Byte achieved the best performance compared to the other two methods. In the statistical analysis of pig behaviors, we proposed a BTA metrics to evaluate pig behaviors tracking performance and BTA of YOLOv5-Byte exceeded 80 %. The proposed model can provide real-time analysis of pig behaviors and effectively support monitoring in large-scale pig farms.

CRediT authorship contribution statement

Shuqin Tu: Funding acquisition. **Yifan Cai:** Writing – original draft. **Yun Liang:** Funding acquisition. **Hua Lei:** Investigation. **Yufei Huang:** Resources. **Hongxing Liu:** Investigation. **Deqin Xiao:** Software.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shuqin TU reports financial support was provided by Guangzhou Institute of Science and Technology. This study was funded partly by key

R&D project of Guangzhou (202206010091,2024B03J1358, 2023B03J1363), Meizhou Tobacco Technology Project of Guangdong Province (202304)。If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Boyle, L.A., Edwards, S.A., Bolhuis, J.E., Pol, F., Semrov, M.Z., Schuetze, S., Nordgreen, J., Bozakova, N., Sossidou, E.N., Valros, A., 2022. The evidence for a causal link between disease and damaging behavior in pigs. Front. Veterin. Sci. 8, 771682 https://doi.org/10.3389/fvets.2021.771682.

Cowton, J., Kyriazakis, I., Bacardit, J., 2019. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. IEEE Access 7, 108049–108060. https://doi.org/10.1109/access.2019.2933060.

Diwan, T., Anirudh, G., Tembhurne, J.V., 2023. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimed. Tools Appl. 82, 9243–9275. https://doi.org/10.1007/s11042-022-13644-y.

Fan, B., Bryant, R., Greer, A., 2022. Behavioral Fingerprinting: Acceleration Sensors for Identifying Changes in Livestock Health. 5, 435-454. doi: 10.3390/j5040030.

Fuentes, S., Viejo, C.G., Tongson, E., Dunshea, F.R., 2022. The livestock farming digital transformation: implementation of new and emerging technologies using artificial intelligence. Animal Health Res. Rev. 23, 59-71, Pii s1466252321000177. doi: 10.1017/s1466252321000177.

Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J.J.A., 2021. YOLOX: Exceeding YOLO Series in 2021. abs/2107.08430. doi: 10.48550/arXiv.2107.08430.

Girshick, R., 2015. Fast r-cnn, Proceedings of the IEEE international conference on computer vision, pp. 1440-1448.doi: 10.48550/arXiv.1504.08083.

Grandin, T., Deesing, M.J., 2022. Chapter 13 - Genetics and animal welfare. In: Grandin, T. (Ed.), Genetics and the Behavior of Domestic Animals, (Third Edition). Academic Press, pp. 507–548. https://doi.org/10.1016/B978-0-323-85752-9.00013-5.

Guo, Q., Sun, Y., Orsini, C., Bolhuis, J.E., de Vlieg, J., Bijma, P., de With, P.H.N., 2023. Enhanced camera-based individual pig detection and tracking for smart pig farms. Comput. Electron. Agric. 211, 108009 https://doi.org/10.1016/j.compag.2023.108009.

Hu, Y., Yu, Y., 2022. Scale difference from the impact of disease control on pig production efficiency. Animals 12, 2647. https://doi.org/10.3390/ani12192647.

Ji, H., Yu, J., Lao, F., Zhuang, Y., Wen, Y., Teng, G., 2022. Automatic position detection and posture recognition of grouped pigs based on deep learning. Agriculture-Basel 12, 1314. https://doi.org/10.3390/agriculture12091314.

Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B., 2022. A review of Yolo algorithm developments. Procedia Comput. Sci. 199, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135.

Kim, T., Kim, Y., Kim, S., Ko, J., 2023. Estimation of number of pigs taking in feed using posture filtration. Sensors 23, 238. https://doi.org/10.3390/s23010238.

Larsen, M.L.V., Wang, M., Norton, T., 2021. Information technologies for welfare monitoring in pigs and their relation to welfare quality(R). Sustainability 13, 692. https://doi.org/10.3390/su13020692.

Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., Norton, T., 2020. A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. Biosyst. Eng. 195, 27–41. https://doi.org/10.1016/j.biosystemseng.2020.04.007.

Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixe, L., Leibe, B., 2021. HOTA: A higher order metric for evaluating multi-object tracking. Int. J. Comput. Vis. 129, 548–578. https://doi.org/10.1007/s11263-020-01375-2.

Mao, A., Huang, E., Wang, X., Liu, K., 2023. Deep learning-based animal activity recognition with wearable sensors: Overview, challenges, and future directions. Comput. Electron. Agric. 211, 108043 https://doi.org/10.1016/j.compag.2023.108043.

Neethirajan, S., 2023. Artificial intelligence and sensor innovations: enhancing livestock welfare with a human-centric approach. Human-Centric Intelligent Syst. https://doi.org/10.1007/s44230-023-00050-2.

Odo, A., Muns, R., Boyle, L., Kyriazakis, I., 2023. Video analysis using deep learning for automated quantification of ear biting in pigs. IEEE Access 11, 59744–59757. https://doi.org/10.1109/access.2023.3285144.

Papakonstantinou, G.I., Arsenakis, I., Pourlis, A., Papatsiros, V.G., 2023. Animal health and productivity of organic greek pig farms: the current situation and prospects for sustainability. Animals 13, 2834. https://doi.org/10.3390/ani13182834.

Prunier, A., Averos, X., Dimitrov, I., Edwards, S.A., Hillmann, E., Holinger, M., Ilieski, V., Leming, R., Tallet, C., Turner, S.P., Zupan, M., Camerlink, I., 2020. Review: Early life predisposing factors for biting in pigs. Animal 14, 570-587, Pii s1751731119001940. doi: 10.1017/s1751731119001940.

T. Psota, E., Schmidt, T., Mote, B., C. Pérez, L., 2020. Long-Term Tracking of Group-Housed Livestock Using Keypoint Detection and MAP Estimation for Individual Animal Identificat. 20, 3670. doi: 10.3390/s20133670.

Qiao, Y., Guo, Y., Yu, K., He, D., 2022. C3D-ConvLSTM based cow behaviour classification using video data for precision livestock farming. Comput. Electron. Agric. 193, 106650 https://doi.org/10.1016/j.compag.2021.106650.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/tpami.2016.2577031.

Sun, P., Jiang, Y., Zhang, R., Xie, E., Cao, J., Hu, X., Kong, T., Yuan, Z., Wang, C., Luo, P. J.A., 2020. TransTrack: Multiple-Object Tracking with Transformer. abs/2012.15460. doi: 10.48550/arXiv.2012.15460.

Tran, D.D., Thanh, N.D., 2023. Pig health abnormality detection based on behavior patterns in activity periods using deep learning. Int. J. Adv. Comput. Sci. Appl. 14 https://doi.org/10.14569/IJACSA.2023.0140564.

Tu, S., Zeng, Q., Liang, Y., Liu, X., Huang, L., Weng, S., Huang, Q., 2022. Automated behavior recognition and tracking of group-housed pigs with an improved DeepSORT method. Agriculture-Basel 12, 1907. https://doi.org/10.3390/agriculture12111907.

Voigtlaender, P., Krause, M., Osep, A., Luiten, J., Sekar, B.B.G., Geiger, A., Leibe, B., 2019. Mots: Multi-object tracking and segmentation. In: Proceedings of the ieee/cvf conference on computer vision and pattern recognition, pp. 7942–7951. https://doi.org/10.48550/arXiv.1902.03604.

Wang, Z., Zheng, L., Liu, Y., Li, Y., Wang, S., 2020. Towards real-time multi-object tracking. In: European Conference on Computer Vision. Springer, pp. 107–122. https://doi.org/10.48550/arXiv.1909.12605.

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric, 2017 IEEE international conference on image processing (ICIP). IEEE, pp. 3645-3649.doi: 10.48550/arXiv.1703.07402.

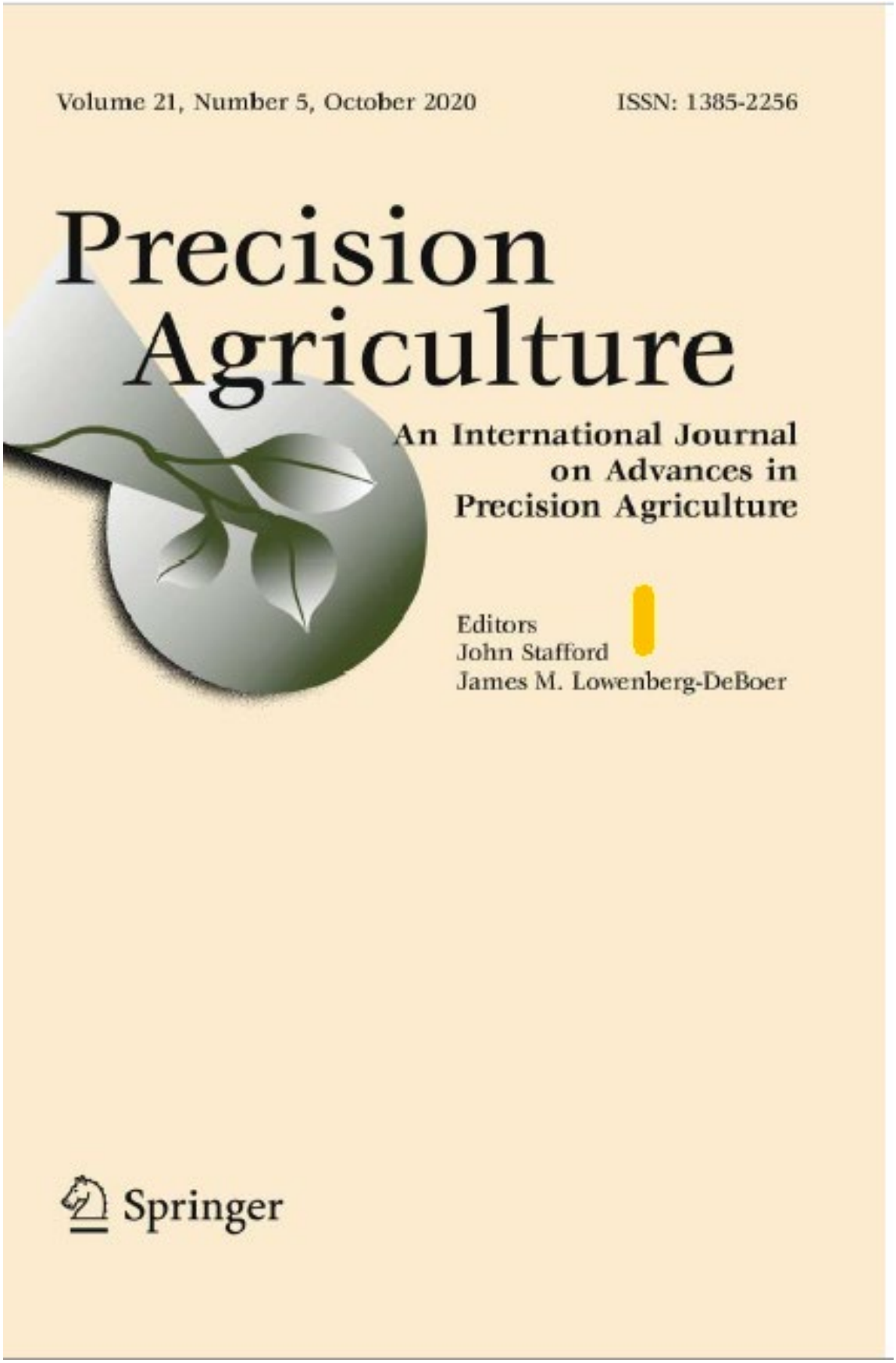
Xiao, D., Lin, S., Liu, Y., Yang, Q., Wu, H., 2022. Group-housed pigs and their body parts detection with Cascade Faster R-CNN. Int. J. Agric. Biol. Eng. 15, 203–209. https://doi.org/10.25165/j.ijabe.20221503.6286.

Yin, M., Ma, R., Luo, H., Li, J., Zhao, Q., Zhang, M., 2023. Non-contact sensing technology enables precision livestock farming in smart farms. Comput. Electron. Agric. 212, 108171 https://doi.org/10.1016/j.compag.2023.108171.

Zhang, L., Gao, J., Xiao, Z., Fan, H., 2023. AnimalTrack: A benchmark for multi-animal tracking in the wild. Int. J. Comput. Vis. 131, 496–513. https://doi.org/10.1007/s11263-022-01711-8.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X., 2022. ByteTrack: Multi-object Tracking by Associating Every Detection Box. In: Avidan, S. (Ed.), Computer Vision – ECCV 2022. Springer Nature, Switzerland, Cham, pp. 1–21. https://doi.org/10.48550/arXiv.2110.06864.

2.2.Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images



目录

目录

Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images

出版物信息: Tu Shuqin; Pang, Jing; Liu, Haofeng; Zhuang Nan; Chen, Yong; 等.

摘要 (ENGLISH)

The accurate and reliable fruit detection in orchards is one of the most crucial tasks for supporting higher level agriculture tasks such as yield mapping and robotic harvesting. However, detecting and counting small fruit is a very challenging task under variable lighting conditions, low-resolutions and heavy occlusion by neighboring fruits or foliage. To robustly detect small fruits, an improved method is proposed based on multiple scale faster region-based convolutional neural networks (MS-FRCNN) approach using the color and depth images acquired with an RGB-D camera. The architecture of MS-FRCNN is improved to detect lower-level features by incorporating feature maps from shallower convolution feature maps for regions of interest (ROI) pooling. The detection framework consists of three phases. Firstly, multiple scale feature extractors are used to extract low and high features from RGB and depth images respectively. Then, RGB-detector and depth-detector are trained separately using MS-FRCNN. Finally, late fusion methods are explored for combining the RGB and depth detector. The detection framework was demonstrated and evaluated on two datasets that include passion fruit images under variable illumination conditions and occlusion. Compared with the faster R-CNN detector of RGB-D images, the recall, the precision and F1-score of MS-FRCNN method increased from 0.922 to 0.962, 0.850 to 0.931 and 0.885 to 0.946, respectively. Furthermore, the MS-FRCNN method effectively improves small passion fruit detection by achieving 0.909 of the F1 score. It is concluded that the detector based on MS-FRCNN can be applied practically in the actual orchard environment.

索引

主题: Feature extraction; Foliage; Orchards; Fruits; Artificial neural networks; Convolution; Harvesting; Occlusion; Image processing; Illumination; Passion fruit; Sensors; Feature maps; Image acquisition; Neural networks

Precision Agriculture (2020) 21:1072–1091
https://doi.org/10.1007/s11119-020-09709-3

Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images

Shuqin Tu[1] · Jing Pang[2] · Haofeng Liu[1] · Nan Zhuang[1] · Yong Chen[3] · Chan Zheng[1] · Hua Wan[1] · Yueju Xue[2]

Published online: 27 January 2020

Abstract
The accurate and reliable fruit detection in orchards is one of the most crucial tasks for supporting higher level agriculture tasks such as yield mapping and robotic harvesting. However, detecting and counting small fruit is a very challenging task under variable lighting conditions, low-resolutions and heavy occlusion by neighboring fruits or foliage. To robustly detect small fruits, an improved method is proposed based on multiple scale faster region-based convolutional neural networks (MS-FRCNN) approach using the color and depth images acquired with an RGB-D camera. The architecture of MS-FRCNN is improved to detect lower-level features by incorporating feature maps from shallower convolution feature maps for regions of interest (ROI) pooling. The detection framework consists of three phases. Firstly, multiple scale feature extractors are used to extract low and high features from RGB and depth images respectively. Then, RGB-detector and depth-detector are trained separately using MS-FRCNN. Finally, late fusion methods are explored for combining the RGB and depth detector. The detection framework was demonstrated and evaluated on two datasets that include passion fruit images under variable illumination conditions and occlusion. Compared with the faster R-CNN detector of RGB-D images, the recall, the precision and F1-score of MS-FRCNN method increased from 0.922 to 0.962, 0.850 to 0.931 and 0.885 to 0.946, respectively. Furthermore, the MS-FRCNN method effectively improves small passion fruit detection by achieving 0.909 of the F1 score. It is concluded that the detector based on MS-FRCNN can be applied practically in the actual orchard environment.

Keywords Multiple scale feature extractor (MSFE) · RGB-D detector · Region proposal network (RPN) · Passion fruit

✉ Yueju Xue
tsq5_6@scau.edu.cn

[1] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

[2] College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China

[3] Hainan Harvest Land Agriculture Technology Company Limited, Chengmai 571900, China

Introduction

Passion fruit is popularly grown in the tropics and sub-tropics, and valued for its captivating flavour, nutritional benefits and medicinal properties (Pongener et al. 2014). There has been continuous increase in its area and production in China. Early yield estimation of immature green fruit is critical for supporting higher level agriculture tasks for robotic harvesting. It may also help growers to identify growth conditions of immature passion fruit and thus implement precise site-specific fertilization based on their conditions and allocate labor based on yield maps (Han et al. 2016; Häni et al. 2019; Chen et al. 2017). Detection and counting of immature green passion fruit are essential steps in creating yield maps. So, there is an increasing demand for automatic detection of fruit and creating accurate early yield maps at the immature green stage from tree canopy images of passion fruit.

Computer vision may provide a promising tool for agricultural machinery-based fruit detection and counting of various types of fruit. For fruit detecting, most studies have been conducted to increase fruit detection accuracies and improve robustness to variable lighting conditions and occlusions. The fusion of blob analysis and the Circular Hough Transform (CHT) method was developed by Gongal et al. (2016) for apple detection under the occlusion of fruit in canopy images by leaves, branches, and other fruits. The method achieved an accuracy of 0.82 on estimating crop load on trees with dual side imaging. The Faster R-CNN was developed and used by Bargoti and Underwood (2017) for fruit detection, including mangoes, almonds, and apples on outdoor orchard images. A framework was presented by Stein et al. (2016) to identify, track, and localize mango. Fruits were detected on the test set using a state-of-the-art faster R-CNN detector with a F1 score of 0.89. A method was designed by Lu et al. (2018) to detect the immature green citrus fruit images under occlusions with F1 = 0.813 based on local binary pattern (LBP) features and adaptive boosting (AdaBoost) technology. To reduce the influence of insufficient light on the segmentation result, a global monomorphic filter was used to conduct image enhancement, 93% of apples were accurately identified in 50 images with uneven illumination distribution (Xu and Lv 2017). A deep semantic segmentation architecture (MangoNet) which was a deep convolutional neural network used semantic segmentation was proposed for detection and counting of mangoes in an open mango field. The MangoNet was capable of segmenting images of different input sizes, which displayed robustness to various conditions of illumination, scale mango density and occlusion (Kestur et al. 2019). For fruit counting, an end-to-end computer vision system was developed by incorporating a fully convolutional network (FCN) to detect a variety of apples under different lighting conditions and estimate fruit counts from apple clusters having arbitrarily complex geometry. The yield estimation system achieved an overall accuracy of 0.912–0.948 across different datasets. Häni et al. (2019) proposed three methods for apple detection and counting based on recent deep learning approaches and classical methods. These methods achieved remarkable yield accuracies ranging from 0.956 to 0.978. A fruit detection and counting method was proposed by Chen et al. (2017) based on a deep learning strategy for variable unstructured environments Song et al. (2014). proposed a two-step method for recognizing and counting pepper fruits in cluttered greenhouses, which obtained a strong correlation of 0.946 with manual counting data. MangoYOLO was developed by A. Koirala et al. (2019a, b) based on features of YOLOv3 and YOLOv2 for mango fruit, which achieved a F1 score of 0.968 and average precision of 0.983 on a test set independent of the training set with a detection speed of 8 ms per 512×512 pixel image tile. However, due to the occlusion of the target fruits by foliage, branches or other fruits and the non-uniform and unstructured

environment (Anand Koirala et al. 2019a, b; Gongal et al. 2015), recognizing and counting fruits on tree canopy is a key challenge in developing yield estimation systems.

Detecting and counting green fruit is a very difficult task due to the following issues: (1) the colours of green fruits and leaves in one image are often very similar; (2) most of detection algorithms are sensitive to variable lighting conditions and occlusions of fruits in orchard environment; (3) immature green fruit is smaller size than mature fruits. For the two problems above, many solutions have been studied by using RGB-D sensors and information fusion technology (Gongal et al. 2015). To avoid the color similarities between fruit and leaves, Gan et al. (2018) built a multi-modal system using a thermal camera and a color camera to detect immature green citrus fruit. They demonstrated that the image fusion of the two camera types improved fruit detection compared to using a colour image alone. The study improved the detection precision from 0.866 to 0.955 and recall from 0.781 to 0.904. To overcome the influence of highly variable illumination conditions, an adaptive thresholding algorithm combined with RGB-D sensor was proposed by Vitzrabin and Edan (2016) to detect sweet-peppers with detection rates of 0.909. The thresholding algorithm uses combinations of the nine thresholds where one threshold for each dimension of RGB images (three dimension) was calculated using three lighting level (low, medium, and high). A study of fusing RGB-D and radiometric information obtained with the Microsoft Kinect for Windows v2 (Kinect v2) (Yang et al. 2015) was proposed for Fuji apple detection. This study achieved an F1 score of 0.898 and an AP of 0.948 when all channels were used (Gené-Mola et al. 2019). These studies showed that combining depth images with color images could potentially improve fruit detection.

Moreover, the previous published works have illustrated the use of Faster R-CNN with VGG-16 model for fruit detection by fusing RGB and depth images (Tu et al. 2018). The emphasis of this paper is on maturity discrimination, which is used for fruit picking by robots. So, the collected data were close to the acquisition equipment and small objects need no attention in the images. However, the basic Faster R-CNN methods in this paper have serious omissions when used in yield estimation system and were not suitable for automatic counting. The reason is possibly that the collected data are far from the acquisition equipment and fruit objects for yield estimation are smaller in the image than ones acquired for a fruit-picking robot. Further, the small objects have fewer pixels and the finite pixels contain few object features, it is difficult to detect the small fruit by the basic Faster R-CNN model when the ROI-pooling layer of Faster R-CNN model builds features only from one single high-level feature map. So, small object detection is crucial in yield estimation tasks.

In recent years, there is considerable research interest in the application of small object detection. Due to the inability of the regional proposal network to accurately detect objects for the low resolution from small objects, Eggert et al. (2017) proposed a modification to Faster R-CNN model that leverages higher-resolution feature maps and generates anchor proposals for small objects. The scheme can improve the performance of small objects detection on the Flicker dataset including 32 logo-classes(Romberg et al. 2011). To trade off the quality of the detector on large objects with that on small objects, Kisantal et al. (2019) proposed to oversample those images with small objects and augment each of those images by copy-pasting small objects many times. The approach achieved 9.7% relative improvement on the instance segmentation and 7.1% on the object detection of small objects. To improve the detection performance for small objects, Cai et al. (2018) developed a multi-stage object detection architecture (the Cascade R-CNN) by creating a series of detectors trained to be sequentially more selective against close false positives. The Cascade R-CNN surpassed all single-model object detectors on the challenging COCO dataset

which contains photos of 80 objects types and all objects are labeled using per-instance segmentations to aid in precise object localization(Lin et al. 2014). Deep learning methods are representation-learning methods with multiple levels of representation, where low-level features represent the presence or absence of edges at particular orientations and locations in the image and high-level features describe abstract semantic information (LeCun et al. 2015). To construct the inherent multi-scale features of deep convolutional networks, the feature pyramid network (FPN) architecture that has rich semantics at all levels scale was exploited by (Lin et al. 2017), which showed significant improvement as a generic feature extractor in small objects detection applications (Le et al. 2016).

Detection of small objects in yield estimation system is challenging since their size in the input is small and they are more difficult to recognize due to lower resolution. In addition, there are few studies, references, and also no standard fruit dataset on automatic detection of small objects for yield estimation. Therefore, a multiple scale Faster R-CNN (MS-FRCNN) approach was proposed based on color (RGB) and depth images for small passion fruit detection and counting the numbers of passion fruits in an image under varying illumination and random occlusion conditions. To test the performance of the MS-FRCNN according to engineering method (Koen 1985), RGB-D dataset was used to conduct extensive experiments and extra dataset focusing on small objects was built for yield estimation system. Compared with the faster R-CNN detector of RGB-D images (Tu et al. 2018), the MS-FRCNN approach combines the features of the 3th, 4th, and 5th convolution layers to form a multiscale feature vector which contributes to the development of a solution for RGB-D and small fruit objects detection used in yield estimation. Detection precision, recall, F1 score and the run time of detection were employed to evaluate the results of the MS-FRCNN approach.

Materials and methods

RGB-D image acquisition

RGB-D images were obtained from passion fruit farm in HeYuan from July 29th to August 12th, 2016 (Tu et al. 2018) and Guangzhou Conghua from Aug. 12th to Sep. 12th, 2017, Guangdong Province, CHINA. To evaluate the proposed detection method, 2275 RGB-D images were captured using a Kinect v2 device in two natural passion fruit orchards. The images were saved in JPG format with 1 920 × 1 080 pixels in RGB and 512 × 424 pixels in depth. Fruit size ranged from 10 × 10 to 100 × 100 pixels. The average fruit size per image was 50 × 50 pixels. Image acquisition was performed between 8:00 am to 10:00 am and 6:00 pm local time. The illumination was affected by the sunlight change, wind, and cloud in the sky when capturing the images. For the reason that the underlying time-of-flight (ToF) technology is unsuitable in strong sunlight conditions, the Kinect V2 sensor was used to avoid strong sunlight and work in shady areas (Yang et al. 2015). Trellis-grown vines of the passion fruit can protect the Kinect sensor from a direct source of light. The Kinect was moved along Trellis-grown vines such that the camera was positioned approximately 2 m from the vine.

Building data library

Among the 2275 RGB-D images of passion fruit acquired, 1900 RGB-D images were randomly selected to generate the training dataset and the validation dataset used during creation of model, and the other 100 RGB-D images formed the extra test dataset for different model comparisons. The rest of the 275 RGB-D images formed the dataset including small objects fruit for testing yield estimation system. An example image of these classes is shown in Fig. 1, where (a) shows an RGB image with 1 920 × 1 080 pixels and (b) shows the corresponding depth image with 512 × 424 pixels, under normal illumination. The depth image in Fig. 1c was pre-processed by applying image enhancement and the median filter, to smooth the objects' contours in the target image and to remove the noise interference.

Locations of fruit in the 1900 RGB-D images (5621 RGB fruits and 3352 depth fruits image with 50 × 50 pixels) were manually labeled with Labelme software for the class name of fruits and stored. The training of the model was run from a workstation (CPU, Intel® Xeon(R) CPU E3-1245 v3 @ 3.40 GHz × 8; Memory, 32 GB) equipped with a GPU (GeForce GTX TITAN X/PCIe/SSE2). Parameters in the MS-FRCNN model were fine-tuned and the final model was used for detecting fruit in images acquired by the Kinect2.0 device.

Overview of the proposed algorithm

A flowchart of the detection of passion fruit is shown in Fig. 2. Multiple scale faster R-CNN detectors for RGB and depth data respectively were introduced and fused both detections of RGB-D image linearly. The linear operation was described in Eq. (1).The implementation of the detection algorithm was based on MS-FRCNN detection.

As can be seen in the Fig. 2, the detection algorithm consists of four phases: RGB-D data augmentation, building the RGB-detector and depth-detector based on the MS-FRCNN algorithm with FPN, fusing the RGB detector and depth detector linearly and counting passion fruits only using RGB test images. In the following paragraphs, each phase of the image analysis algorithm is discussed in more detail.

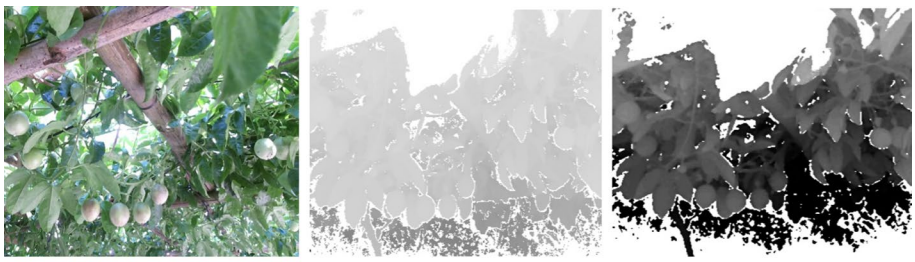


Fig. 1 An RGB image and the corresponding depth image. **a** The RGB image, **b** the depth image and **c** the pre-processed depth image

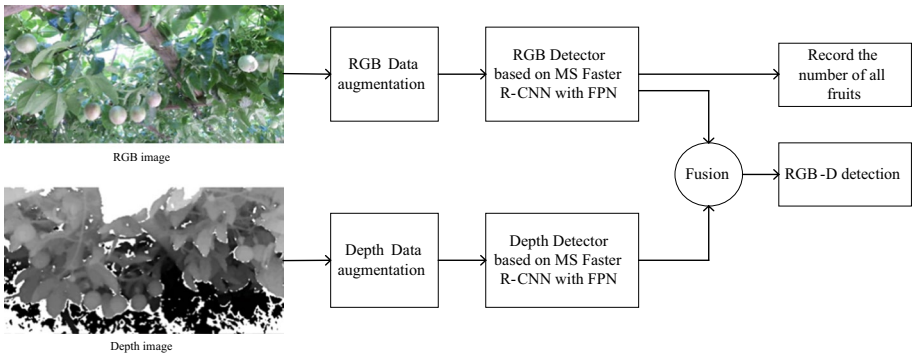


Fig. 2 Flow chart of the detection algorithm based on RGB-D images

Data augmentation

The limited training dataset was not sufficient to train a deep and robust MS-FRCNN. A reasonable way to enlarge the training dataset size was to use data augmentation, which is a common way to expand the variability of the training data by artificially enlarging the dataset using label-preserving transformations with a strong generalization ability (Ren et al. 2015). Different combinations of transformations are commonly used as data augmentation to teach the network the desired invariance and robustness properties. Typical augmentation techniques used in the computer vision community include left–right flipping, image re-scaling, and changes to image colour. This paper applied the left–right flipping, rotation, and elastic deformations (Simard et al. 2003) to the fruit images during training. Based on those 1900 RGB-D images during training of model, the dataset after data augmentation is listed in Table 1.

The basic Faster R-CNN algorithm

The base model of Faster R-CNN framework is a residual network (ResNet) proposed by He et al. (2016),which is easier to optimize, and can gain accuracy from considerably increased depth. The architecture of Faster R-CNN model consists of two modules. The first, called the regional proposal network (RPN), is a fully convolutional network for generating object proposals that will be fed into the second module. The second module is the Fast R-CNN detector which classifies individual proposals and regresses a bounding box around the fruits. The key idea is to share the same convolutional layers for the RPN and Fast R-CNN detector up to their own fully connected layers. The output

Table 1 Number of images used for training and testing after data augmentation

Type of images	Original data	Augmented data	Number of images in training set	Number of images in test set
RGB	1 900	8 651	6 055	2 596
Depth	1 900	3 352	2 346	1 006

from the final convolutional layer is a high- dimensional feature map, which is then connected to two sibling layers including a box-regression layer and a box-classification layer. These layers define the RPN, and they detect and classify Regions of Interest (RoIs) in the image. The RoIs are subsequently propagated through the fully-connected layers which refine their object probability and associated bounding box label. The output from the network is a set of bounding boxes withprobability of fruit. To these, a probability threshold is applied, followed by non-maximum suppression (NMS) where a bounding box with the maximum classification score is selected and its neighboring boxes are suppressed to handle overlapping detections.

Figure 3 illustrates the ResNet-based Faster R-CNN detection flowchart based on RGB-D images. There are five parts including feature extraction of ResNet, RPN network, two fully-connected networks (Fc6 and Fc7), one softmax classifier layer and NMS. ResNet consists of several basic residual blocks which provide a shortcut connection between layers. This shortcut connection makes it possible to train hundreds or more layers while achieving enhanced performance. ResNet was developed with many different numbers of layers including 50 and 101 which is used in this paper (He et al. 2016). The number of output proposals is set at 300. NMS with a threshold of 0.3 removes duplicate predictions.

In the RPN, the convolution layers of a pre-trained network are followed by a 3×3 convolutional layer. This corresponds to mapping a large spatial window or receptive field in the input image to a low-dimensional feature vector at a sliding-window location. Two 1×1 convolutional layers are then added for classification and regression branches of all spatial windows.

To deal with different scales and aspect ratios of objects, anchors are introduced in the RPN. For k region proposals, these region proposals are parameterize relative to k reference boxes, called anchors. An anchor is at each sliding location of the convolutional maps and thus at the center of each spatial window. Each anchor is associated with a scale and an aspect ratio. The scales (32^2, 64^2,128^2, 256^2, and 512^2 pixels) and three aspect ratios (1:1, 1:2, and 2:1), leading to use of $k = 15$ anchors at each location in the study.

For training RPNs, a binary class label (of being an object or not) is assigned to each anchor. A positive label was assigned to two kinds of anchors: (i) the anchor/anchors with the highest Intersection over-Union (IoU) overlap with a ground-truth box, or (ii) an anchor that has an IoU overlap higher than 0.7 with any ground-truth box. A negative label to a non-positive anchor is assigned if its IoU ratio is lower than 0.3 for all ground-truth boxes. Anchors that are neither positive nor negative do not contribute to the training objective.

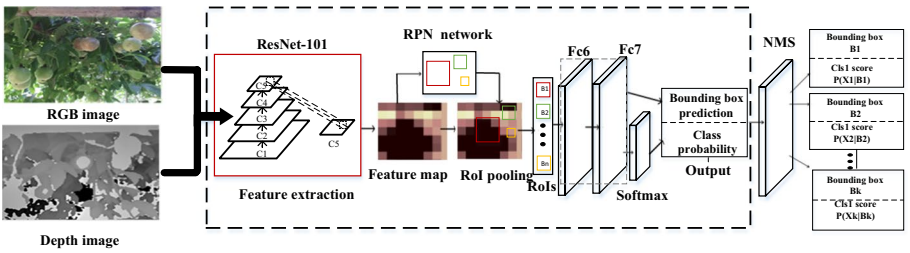


Fig. 3 The ResNet-based Faster R-CNN detection flowchart based on RGB-D images

The multiple scale Faster R-CNN detector

The Faster R-CNN achieves state-of-the-art performance in detecting persons, animals, or vehicles. However, the detection network in Faster R-CNN has trouble detecting such small objects as fruits in the trees. The main reason is that the ROI-pooling layer and RPN build features only from the feature map of the last convolution layer which omitted low-resolution features maps useful for small object detection. Therefore, a combination of both global and local features, i.e., multi scaling, to enhance the global context and local information in the ResNet-based Faster R-CNN network can help robustly detect the objects of interest. In order to enhance the capability of the network, feature maps from shallower convolution feature maps, i.e., conv3, conv4, and conv5 are incorporated for RPN and ROI pooling. Then the network was able to detect lower level features containing a higher proportion of information in ROI regions.

Figure 4 shows multiple scale feature extractor which replaces the original feature extraction in the basic Faster R-CNN (Fig. 4). The feature extraction involves a bottom-up pathway, a top-down pathway, and lateral connections. The bottom-up pathway is the feedforward computation of the backbone convolutional neural networks (ConvNet), which computes a feature hierarchy consisting of feature maps at several scales with a scaling step of 2. The top-down pathway generates higher resolution features by upsampling spatially coarser feature maps from higher pyramid levels. These features are then enhanced with features from the bottom-up pathway via lateral connections. Each lateral connection merges feature maps of the same spatial size from the bottom-up pathway and the top-down pathway. The detailed process was described in the following.

The output of these last residual blocks from conv1,conv2, conv3, conv4, and conv5 outputs was defined as $\{C_1,C_2,C_3,C_4,C_5,\}$. First, a coarsest-resolution feature map (P_5) is generated by attaching a 1*1 convolutional layer on C_5, and a feature map is obtained by a factor of 2 on P_5 up sample operation. The upsampled feature map is then merged with the corresponding bottom-up C_4, by element-wise addition to obtain the feature map P4 (Fig. 4 lateral connections). This process is iterated until the finest resolution feature map is generated. This final set of feature maps is called $\{P_3,P_4,P_5\}$, corresponding to $\{C_3,C_4,C_5\}$ that are respectively of the same spatial sizes.

The MS-FRCNN detector adapted RPN and Fast R-CNN by replacing the single-scale feature map with multiple scale feature maps. A publically available pre-trained ResNet-101

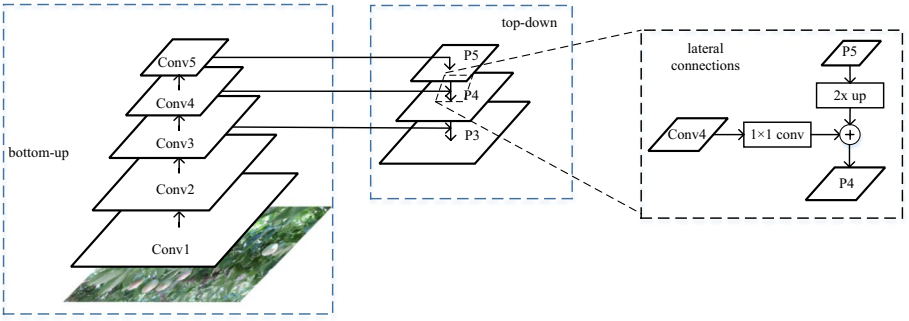


Fig. 4 Multiple scale feature extractor including bottom-up pathway, top-down pathway and lateral connections

model was used (He et al. 2016). The MS-FRCNN detector used the same training process and loss function in RPN for bounding box proposal generation and in the Fast R-CNN for object detection with the basic Faster R-CNN detector (Lin et al. 2017).

Fusion of RGB and depth detector

The combination of RGB data and depth data appears promising: depth data are robust to illumination changes but sensitive to low-signal strength returns and suffer from limited depth resolution. Image data are rich in colour and texture, have a high angular resolution but break down quickly under non-ideal illumination.

To take advantage of the richness of RGB-D data, late fusion was proposed by Sa et al. (2016), which combined the classification information from the two modalities, colour and NIR imagery. The RGB-D detector was trained separately by an RGB Faster R-CNN detector trained on colour images and a depth Faster R-CNN detector trained on depth images. If no depth data are available, the detector degrades to the regular RGB detector. The linear strategy described in Eq. (1) was used to combine the RGB-based Faster R-CNN detector with the depth-based Faster R-CNN detector by late fusion.

$$p = \begin{cases} p_{RGB}\, p_D = 0 \\ \lambda p_D + \ (1-\lambda) p_{RGB}\, p_D \neq 0 \wedge p_{RGB} \neq 0 \\ p_D\, p_{RGB} = 0 \end{cases} \tag{1}$$

where p is resulting probability of detecting passion fruit of RGB-D images, λ is adjusted by cross validation method in the experiment (used 0.4 in the paper), p_D is resulting probability of detecting depth images and p_{RGB} is resulting probability of detecting RGB images.

Implementation details

The proposed multi-modal object detection networks can be trained end-to-end with back-propagation and stochastic gradient descent (SGD). For RPN networks, each mini-batch arises from a single image that contains many positive and negative example anchors. The input image is resized such that its shorter side has 800 pixels. Synchronized SGD is used to train the model on 8 GPUs. Each mini-batch involves 2 images per GPU and 512 anchors per image with a weight decay of 0.001 and a momentum of 0.9 which is used for the gradient updating of weights and bias in the model. These weights are used in convolutional neural network and fully connected layers network of model. The learning rate which is a parameter for the gradient descent is 0.01 for the first 6 k mini-batches and 0.001 for the next 1 k, and 0.0001 for the next 1 k. To reduce redundancy, non-maximum suppression (NMS) is performed over the proposals according to their ensemble objectness scores with an IoU threshold of 0.5, which leaves about 300 proposal regions per image. The pre-trained models include ResNet-50 and ResNet-101. The input size (network resolution) for MS-FRCNN framework is 1333*800.

Results and discussion

In this section, the MS-FRCNN method was qualitatively and quantitatively evaluated on four experimental settings: (1) the performance of RGB and depth detector under different feature extractors was assessed; (2) the performance of RGB-D fusion detector was evaluated; (3) the results of RGB detection in occlusion and overlap situations were analyzed; (4) the results of RGB detection and counting in small fruit objects for yield estimation system (5) the performance between MS-FRCNN with HOG + SVM, HOG + Adaboost and basic Faster R-CNN methods was compared.

Given this threshold, the precision (P), recall (R) and $F1$ score are computed as:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F1 = \frac{2 * Precision * Recall}{Precison + Recall} \quad (2)$$

where TP is the number of true positives (correct detections), FP is the number of false positives (false detection), and FN is the number of false negatives (miss).

In this paper, the precision-recall curve with the corresponding $F1$ score was used as the evaluation metric for fruit detection. Fruit is considered as detected if the intersection over union (IoU) between the prediction and ground truth bounding boxes is greater than 0.5. It is worth noting that this threshold was chosen to be smaller than the ImageNet (Deng et al. 2009) challenge due to the relatively small fruit size with respect to the image resolution. ImageNet is a large-scale hierarchical image database which include 14 million images annotated. Although the threshold affects the performance evaluations (the smaller the threshold is, the higher the $F1$ score produced), the identical threshold is consistently used for all experiments and comparisons presented in this paper.

The results and analyzes for RGB and depth detectors based on MS-FRCNN

The public ResNet-50 and ResNet-101 including 50 and 101 of layers with multiple scale feature extractor named as ResNet-50-MSFE and ResNet-101-MSFE were used (Alom et al. 2019). The detection performance of RGB and depth detectors based on MS-FRCNN was separately demonstrated in Table 2. The extra validation data include 100 RGB images with 1008 fruits and depth images with 527 fruit which did not involve in training and testing of the model.

Figure 5 shows the precision-recall curves of the MS-FRCNN systems based on ResNet-50-MSFE and ResNet-101-MSFE for RGB and depth detection. The markers denote the points where precision and recall are identical. RGB detector can achieve better detection performance than depth detector which suffered from the depth images containing noise from sunlight. The ResNet-101-MSFE model outperforms the ResNet-50-MSFE. It is noted that the ResNet-101-MSFE approach contains twice as many conv layers than the ResNet-50-MSFE and requires more resources, such as computation time and GPU memory space. Comparative experiments were conducted to analyse the contribution of MSFE between ResNet-101 and ResNet-50.

The results of RGB and depth detectors generated using different kinds of MSFE are shown in Table 2. The precision, recall, F1 and detection speed values were calculated using ResNet-101 and ResNet-50 for RGB and depth detectors. On RGB and depth fruit

Table 2 The results of RGB and depth detectors based on ResNet-50-MSFE and ResNet-101-MSFE

Type of detector	Feature extractor	Correctly identified (TP)	Missed fruit (FN)	Incorrectly identified (FP)	Precision	[a]Recall	F1	[b]Detection speed(s)
RGB	ResNet-50-MSFE	930	78	98	0.904	0.923	0.913	0.122
	ResNet-101-MSFE	932	76	96	0.907	0.925	0.916	0.175
Depth	ResNet-50-MSFE	434	93	98	0.816	0.824	0.819	0.101
	ResNet-101-MSFE	445	82	87	0.837	0.845	0.841	0.145

[a]Recall is the ratio of the number of correct detections to that of annotated passion fruits. The total number of annotated fruits in the RGB and Depth images are 1008 and 527

[b]Detection speed is the average processing speed (frames per second) for the extra validation dataset containing 100 images

detectors, using the ResNet-101-MSFE obtained better detection performance compared with the ResNet-50-MSFE. The highest F1 of 0.916 was obtained using ResNet-101-MSFE for RGB detector; while the lowest F1 of 0.819 was obtained using ResNet-50-MSFE for depth detector. Compared with the depth detector, the RGB detector achieves approximately 10% and 7% improvement of F1 using ResNet-50-MSFE and ResNet-101-MSFE.

The detection time of RGB and depth detectors is 0.175 s and 0.145 s respectively using ResNet-101-MSFE. Compared with ResNet-101-MSFE, the detection time of RGB and depth detectors using ResNet-50-MSFE is 0.122 s and 0.101 s respectively, showing the faster detection speed. The detection times of ResNet-50-MSFE and ResNet-101-MSFE models are fast enough for real-time detection application in the field (Detection time < 0.2 s). Comprehensive consideration of detection accuracy and time performance, the Faster R-CNN method based on ResNet-101-MSFE would be conducted and analysed on RGB-D images.

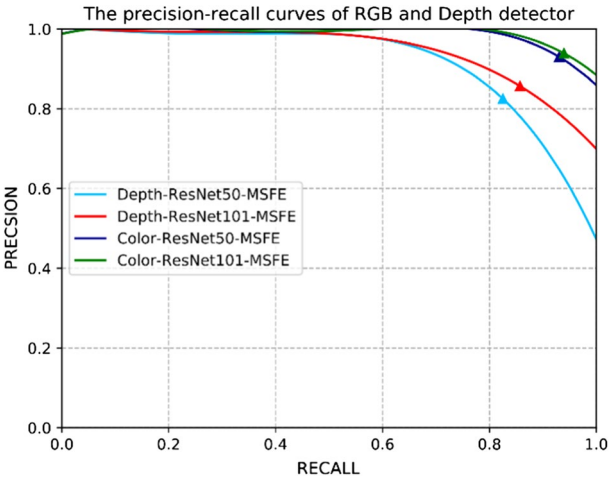


Fig. 5 Precision-recall curves RGB and Depth detector using ResNet-50-MSFE and ResNet-101-MSFE

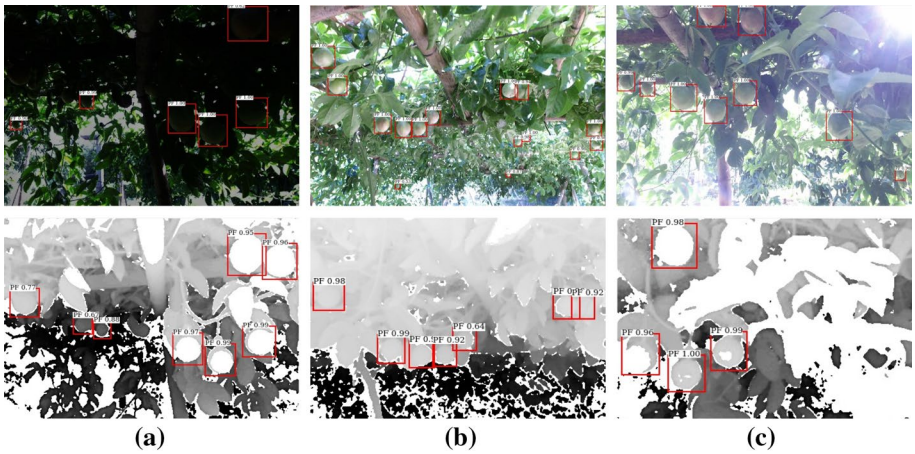


Fig. 6 Detection results of passion fruit in RGB-D data under different light. **a** Weak illumination **b** normal illumination **c** bright light

Some examples detected via the MS-FRCNN approach are shown in Fig. 6 where text on red rectangle box includes the abbreviation of passion fruits (PF) and their corresponding confidence values. The figure illustrates several passion fruits detected using ResNet-101-MSFE feature extractor under three different lighting conditions. Both colour and depth detectors achieved good detection accuracy. However, for small passion fruit detection, the colour detector can achieve good better detection accuracy rate than the depth detector. This occurred because ResNet-101-MSFE feature extractor can extract low-level features for detecting the smaller object; consequently, many small passion fruits were correctly detected by the RGB detector under normal illumination and bright light conditions (e.g., Fig. 6b, c top row). Under Weak illumination, some fruits were not correctly detected by the RGB detector (e.g., Fig. 6a top row); while using the Depth detector, fruits (e.g., Fig. 6a bottom row) could be detected quite reliably. This suggests that a detection system that combines depth and colour detection can be used under different illumination conditions; this further suggests that multimodality can help detect fruits in situations that cannot be handled by single-cue detectors.

Detection results from fusion of RGB and depth detectors based on ResNet-101-MSFE

The results of the detection from information fusion were evaluated and compared with previous results. Table 3 shows the numbers of the correctly identified fruit, missed fruit, incorrectly identified fruit, the computed precision, recall rates and F1 value. Comparing these results with Table 2, precision increased from 0.907 using only RGB detector based on ResNet-101-MSFE to 0.962; a 0.68% and 3.07% increase in the recall and F1 when using RGB-D detector comparing with only RGB detector, respectively.

Examples are shown in Fig. 7, in which confirmed fruit regions are marked with red rectangles by RGB & depth images and confirmed fruit regions are marked using blue rectangles by fusing RGB and depth images. RGB detector can usually achieve better detection

Table 3 Precision and recall from the detections after information fusion of color and depth images

Manual fruit count	Correctly identified (TP)	Missed fruit (FN)	Incorrectly identified (FP)	Precision	Recall	F1
1020	950	70	38	0.962	0.931	0.964

performance than depth detector (e.g., Fig. 7a middle and bottom row). However, under weak illumination, depth detector can achieve better results than RGB one (e.g., Fig. 7b top row). According to Eq. (1), some missed fruit regions in the color images were confirmed as fruit after adding depth information (e.g., Fig. 7c top and bottom row). So, RGB-D detector improved passion fruit detection compared with RGB and depth detector (e.g., Fig. 7c).

Some of the passion fruit samples being heavily occluded by the trunks, branches and leaves could yield higher similarities between some positive and negative training samples, forcing such test samples to fall on the decision boundaries of the softmax classifiers which is a generalization of logistic regression used for multi-class classification (https://cs231n.github.io/linear-classify/#softmax). So, it is difficult to classify positive and negative samples with high colour similarities correctly, and the detector obtains a lower classification accuracy. Therefore, the detection of highly overlapping passion fruits and occluded regions requires additional research.

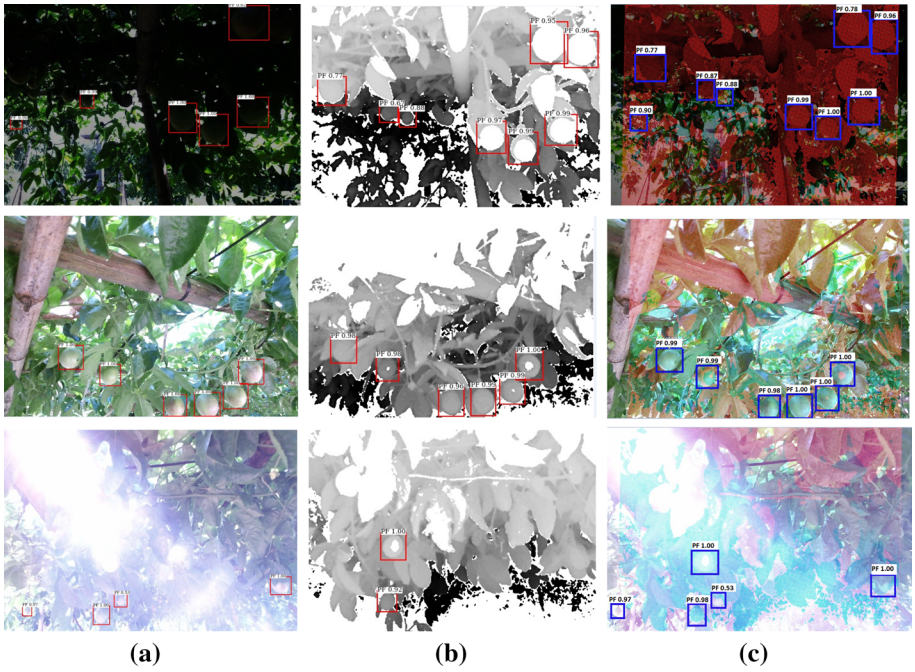


Fig. 7 Three detection results examples after information fusion. The images **a** show the results from Faster R-CNN detection in the RGB images. The images **b** are the corresponding depth images. The images **c** are the detection results after information fusion. In **c**, blue rectangles are confirmed fruit locations from RGB and depth images

Performance evaluation of passion fruit detection under occlusion and overlap conditions

In this study, color images were used as the primary information resource for detecting overlapping passion fruits and occluded regions due to their higher reliability compared to depth images under normal illumination. The fruits in a color image using MS-FRCNN were detected. The numbers of output label are four set as non-overlapped fruit, overlapped fruit, occlusion fruit, and background. The confidence value of classifier was set greater than 0.5.

Table 4 shows a summary of the detection results. In the table, the precision and recall values were calculated when the output classes were non-overlap, occlusion, and overlap. The precision in non-overlapped fruit detection achieved the best precision (0.962) and recall (0.972). Compared with the performance of the occlusion and overlapped fruit detection, the detection performance in non-overlapped status was approximately 10% better in terms of the average precision and recall.

Some examples are shown in Fig. 8. The figure illustrates several passion fruits detected with different probabilities ($p \geq 0.5$), under non-overlapped (No-ov), occlusion (Occ) and overlap (Ov) conditions. The detected object is represented using a bounding box, which is determined by a minimum enclosing rectangle containing all the foreground pixels after Faster R-CNN detection operation. The numeral marked in white under each bounding box shown in Fig. 8 refers to the calculated relative probabilities, and higher values of p (ranging from 0 to 1) indicate higher localization precision values of the detected citrus fruit. The bounding boxes marked in red in Fig. 8 refer to the detection results with higher matching results ($p \geq 0.5$). False and missed detections still occurred (Table 4) in non-overlapped, occlusion and overlap conditions. Most of the false detections and some missed detections occurred because it is difficult for the training model to learn all the complex occlusion and overlap conditions.

The results of RGB detection and counting in small fruit objects for yield estimation system

In this study, the rest of 275 color images including many small objects were used as the primary information resource for testing yield estimation system. The total number of annotated fruits in the 275 images was 7032 fruits. The number of fruits per image ranged from 10 to 100.

Table 5 shows the numbers of the correctly identified fruit, missed fruit, incorrectly identified fruit, the computed precision, recall rates and F1 value between the MS-FRCNN approach and Faster R-CNN(ResNet101). The correctly identified fruit numbers are 6205 and the F1 score is 0.909 using MS-FRCNN. Compared with the Faster R-CNN(ResNet101), the MS-FRCNN is approximately 3% better in terms of the precision.

Table 4 The results of passion fruit detection under non-overlapped, occlusion and overlap conditions

Output label	Manual fruit count	Correctly identified (TP)	Missed fruit (FN)	Incorrectly identified (FP)	Precision	Recall	F1
Non-overlapped	643	625	18	25	0.962	0.972	0.967
Occlusion	311	285	26	45	0.864	0.916	0.889
Overlapped	54	40	14	12	0.769	0.741	0.755

Fig. 8 Examples of passion fruit detection by RGB for non-overlapped, occlusion and overlapped conditions. **a** No overlap, **b** a relatively strong occlusion, **c** a relatively strong overlap. The red rectangles with solid line, blue rectangles with dashed line and yellow rectangles with dotted line indicate the three classes (No overlap, occlusion and overlap), respectively

Table 5 Precision and recall from the detections for MS-FRCNN and Faster R-CNN(ResNet101)

Method	Correctly identified (TP)	Missed fruit (FN)	Incorrectly identified (FP)	Precision	Recall	F1
MS-FRCNN	6 205	827	409	0.938	0.882	0.909
Faster R-CNN(ResNet101)	6 053	979	639	0.905	0.861	0.882

Some examples are shown in Fig. 9. These test data were collected far from the acquisition equipment, so fruit objects are small in the image for yield estimation. The bounding boxes marked in red in Fig. 9 refer to the detection results. The dense small fruit objects (e.g., Fig. 9a, b) can be well detected within 5 meters from the camera under normal illumination. Furthermore, many small passion fruits in the range of pixels [20×20, 50×50] can be correctly detected under normal and weakly illumination. For fruits with resolution less than 10×10 pixels, there were some omissions and errors in the MS-FRCNN method.

Result comparison with other three methods

Histogram of oriented gradients (HOG) proposed by Dalal et al. (2006) is a powerful descriptor for object detection and recognition. In fruit detection, HOG has been widely used and frequently achieved excellent performance. The MS-FRCNN approach was compared with HOG + SVM, HOG + Adaboost and Faster R-CNN based on VGG-16 methods. For a fair comparison, the same training and test datasets acquired from a passion fruit farm in HeYuan from July 29 to August 12, 2016 were select to be used for evaluation. Then, a comparison was made by analysing the F1 and the time costs of the different methods about RGB-D images in Table 6.

Fig. 9 Examples of passion fruit detection and counting with dense small fruit objects, using MS-FRCNN with ResNet-101-MSFE feature extractor. **a** The high density small fruit objects with the number of fruits over 80. **b** The moderate density small fruit objects with the number of fruits between [50 80]

Table 6 Comparison of results for the MS-FRCNN and other three methods on RGB-D images

Approach	Precision	Recall	F1	Detection speed (s)
HOG + Adaboost	0.927	0.605	0.733	0.215
HOG + SVM	0.851	0.619	0.717	0.504
Faster R-CNN(VGG-16)	0.922	0.850	0.885	0.072
The MS-FRCNN	0.962	0.931	0.946	0.175

Table 6 shows that HOG + Adaboost, HOG + SVM and Faster R-CNN(VGG-16) achieve a F1 value of 0.733,0.717 and 0.885, respectively. The highest F1 value of 0.946 was obtained with the MS-FRCNN approach, which has 0.21 higher than HOG + Adaboost and 0.23 points higher than HOG + SVM. HOG + Adaboost and HOG + SVM methods obtained low detection accuracy because the HOG method extracted feature struggled to deal with occlusion and was sensitive to noise points. Therefore, the HOG methods failed to detect most passion fruits in the test dataset where more than half of the images were captured under uniform and sufficient illumination conditions. The Faster R-CNN(VGG-16) detector was better at detecting passion fruits, possibly because RPN can extract not only coarse-to-fine feature information but also their spatial positions across deep CNN models. Compared with the performance of Faster R-CNN(VGG-16) detector, the MS-FRCNN approach brought approximatel y a 6% enhancement in the F1

value, which can be significant in image detection tasks. The reason is likely that Faster R-CNN(VGG-16) detector had poor performance on small objects.

The average processing time for four methods was also evaluated. The Faster R-CNN(VGG-16) takes 0.072 s in total, which is 3×, 7×and 2×faster than HOG + Adaboost, HOG + SVM, and the MS-FRCNN approach, as shown in Table 4 (right). The MS-FRCNN approach tends to lead to slower but more accurate models, requiring 0.175 s per image compared with Faster R-CNN(VGG-16) requiring 0.072 s. Although the testing time of the MS-FRCNN is acceptable and fits the targets, the number of regions of interest in the MS-FRCNN approach requires additional and further research to improve the detection efficiency.

Discussion

A new MS-FRCNN method was developed for passion fruit detection by combining color and depth information. Although previous studies have demonstrated that the basic faster R-CNN method can achieve good detection accuracies with an F1 of 0.885, the method does not work well to detect small fruit objects in natural outdoor scenes. To solve the problems, a multiple scale feature extractor was developed in this study by combining conv3, conv4, and conv5 as feature maps. The method is shown to be successful by significantly improving the precision and recall of small fruit detection. Using the MS-FRCNN method, the precision, recall and F1 values increase from 0.922 to 0.962, 0.850 to 0.931 and 0.885 to 0.946, respectively.

RGB detector can achieve better detection performance than the depth algorithm under normal illumination. The depth images display more distinguishable features between fruit and leaves on some occasions. The depth detector can detect the fruits hidden in dimly lighted areas, while in the color images many of them are less visible. Though fruits in depth images appeared to be more distinct in the example images, color images are still more robust in different situations (such as overlap and occlusion conditions) and provide more consistent detection results.

Conclusions and future perspectives

In this paper, a new approach for the detection of passion fruits using RGB-D images was developed, and some pilot tests were performed. The major contributions of this study can be summarized as follows:

(1) The MS-FRCNN is a novel method for fruit detection using colour and depth images. The MS-FRCNN method can be used for detection of passion fruits in natural canopies. The deep learning methods seem to generalize well in new conditions and can be trained on smaller set of images through transfer learning. Therefore, it should be easy to adopt a detection method based on the MS-FRCNN for detection of new fruit species.
(2) An RGB-D image-based machine vision system was developed and evaluated. In the detection step, the fruit detector performed better when RGB information was combined with depth information. The RGB-D detector demonstrated the F1 of 0.946, outperforming the RGB and depth detectors by more than 3% and 10%, respectively.

The Kinect device might be subsequently used for fruit detection in natural outdoor conditions if the device is protected from direct exposure to sunlight.

(3) The algorithm developed in this study was able to identify small passion fruits for testing yield estimation system in the canopy with a complex background, under uneven non-overlap, overlap and occlusion conditions. In these experiments, the proposed MS-FRCNN provides a significant boost in detection performance and average processing time compared with HOG + Adaboost, HOG + SVM, and the basic Faster R-CNN methods.

Further improvement will be considered by exploring more effective detection model such as YOLO-V3 in videos by incorporating adversarial learning strategies (Goodfellow et al. 2014) for recognizing passion fruits with less time to locate fruits with higher precision for yield estimation, especially for those fruits that grow in dense clusters. The adversary nets framework creates examples on the outdoor scene with different occlusions and deformations, such that these occlusions/deformations make it difficult for original object detector to classify. In addition, the detection performance of passion fruit in different growth stages will be involved and analyzed.

Acknowledgements This work was supported by the Science and Technology Planning Project of Guangdong Province (2015A020224038 and 2015A020209148), and the National Natural Science Foundation of China (31600591 and 61772209).

References

Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., & Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *Journal of medical imaging (Bellingham, Wash.), 6*(1), 014006. https://doi.org/10.1117/1.Jmi.6.1.014006.

Bargoti, S., & Underwood, J. (2017). *Deep fruit detection in orchards. In Proceedings of the IEEE International Conference on Robotics & Automation* (pp. 3626–3633, https://doi.org/10.1109/ICRA.2017.7989417.

Cai, Z. W., Vasconcelos, N., & IEEE. (2018). *Cascade R-CNN: Delving into High Quality Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6154–6162,). New York: IEEE.

Chen, S. W., Skandan, S. S., Dcunha, S., Das, J., Kumar, V. J. I. R., & Letters, A. (2017). Counting apples and oranges with deep learning: A data driven approach. *IEEE Robotics and Automation Letters, 2*(2), 781–788.

Dalal, N., Triggs, B., & Schmid, C. (2006). *Human detection using oriented histograms of flow and appearance. In Proceedings of the European Conference on Computer Vision* (pp. 886–893, https://doi.org/10.1007/11744047_33. Graz, Austria.

Deng, J., Dong, W., Socher, R., Li, L. J., & Li, F. F. (2009). *ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.

Eggert, C., Brehm, S., Winschel, A., Dan, Z., & Lienhart, R .(2017). *A closer look: Small object detection in faster R-CNN. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 421–426, https://doi.org/10.1109/ICME.2017.8019550.

Gan, H., Lee, W. S., Alchanatis, V., Ehsani, R., & Schueller, J. K. (2018). Immature green citrus fruit detection using color and thermal images. *Computers and Electronics in Agriculture, 152*, 117–125. https://doi.org/10.1016/j.compag.2018.07.011.

Gené-Mola, J., Vilaplana, V., Rosell-Polo, J. R., Morros, J.-R., Ruiz-Hidalgo, J., & Gregorio, E. (2019). Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities. *Computers and Electronics in Agriculture, 162*, 689–698. https://doi.org/10.1016/j.compag.2019.05.016.

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. *Computers and Electronics in Agriculture, 116*, 8–19.

Gongal, A., Silwal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2016). Apple crop-load estimation with over-the-row machine vision system. *Computers and Electronics in Agriculture, 120*, 26–35. https://doi.org/10.1016/j.compag.2015.10.022.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Bing, X., Warde-Farley, D., Ozair, S., et al. (2014). *Generative adversarial nets. In Proceedings* of the *International Conference on Neural Information Processing Systems* (pp. 2672–2680).

Han, L., Lee, W. S., & Wang, K. J. P. A. (2016). Immature green citrus fruit detection and counting based on fast normalized cross correlation (FNCC) using natural outdoor colour images. *Precision Agriculture, 17*(6), 1–20.

Häni, N., Roy, P., & Isler, V. (2019). A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics,*. https://doi.org/10.1002/rob.21902.

He, K., Zhang, X., Ren, S., & Jian, S. (2016). Identity *mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision* (pp. 630–645, https://doi.org/10.1007/978-3-319-46493-0_38.

Kestur, R., Meduri, A., & Narasipura, O. (2019). MangoNet: A deep semantic segmentation architecture for a method to detect and count mangoes in an open orchard. *Engineering Applications of Artificial Intelligence, 77*, 59–69. https://doi.org/10.1016/j.engappai.2018.09.011.

Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., & Cho, K. (2019). Augmentation for small object detection. *Computer Vision and Pattern Recognitionar*. https://arxiv.org/abs/1902.07296.

Koen, B. V. (1985). *Definition of the Engineering Method*. ASEE Publications, Suite 200, 11 Dupont Circle,Washington, DC 20036: ASEE Publications.

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019a). Deep learning—Method overview and review of use for fruit detection and yield estimation. *Computers and Electronics in Agriculture, 162*, 219–234. https://doi.org/10.1016/j.compag.2019.04.017.

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019b). Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precision Agriculture, 20*(6), 1107–1135. https://doi.org/10.1007/s11119-019-09642-0.

Le, T. H. N., Zheng, Y., Zhu, C., Luu, K., & Savvides, M. (2016). Multiple Scale Faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection. *Proceedings of the Computer Vision and Pattern Recognition Workshops*. https://doi.org/10.1109/CVPRW.2016.13.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539.

Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., et al. (2014). *Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision* (pp. 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.

Lin, T. Y., Dollár, P., Girshick, R., He, K., & Belongie, S. (2017). *Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2117–2125, https://doi.org/10.1109/CVPR.2017.106.

Lu, J., Lee, W. S., Gan, H., & Hu, X. (2018). Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis. *Biosystems Engineering, 171*, 78–90. https://doi.org/10.1016/j.biosystemseng.2018.04.009.

Pongener, A., Sagar, V., Pal, R. K., Asrey, R., Sharma, R. R., & Singh, S. K. (2014). Physiological and quality changes during postharvest ripening of purple passion fruit (Passiflora edulis Sims). *Fruits, 69*(1), 19–30. https://doi.org/10.1051/fruits/2013097.

Ren, W., Yan, S., Yi, S., Dang, Q., & Gang, S. (2015). Deep image: Scaling up image recognition. *Computer Science*. https://doi.org/10.1038/nature0693.

Romberg, S., Pueyo, L. G., Lienhart, R., & Zwol, R. V. (2011). *Scalable logo recognition in real-world images. In Proceedings of the 1st International Conference on Multimedia Retrieval*. https://doi.org/10.1145/1991996.1992021.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). DeepFruits: A fruit detection system using deep neural networks. *Sensors (Basel), 16*(8), 1222. https://doi.org/10.3390/s16081222.

Simard, P., Steinkraus, D., & Platt, J. C. (2003). *Best practices for convolutional neural networks applied to visual document analysis. In 7th International Conference on Document Analysis and Recognition (ICDAR 2003)*, (Vol. 2, pp. 958–962). Edinburgh, Scotland, UK.

Song, Y., Glasbey, C. A., Horgan, G. W., Polder, G., Dieleman, J. A., & van der Heijden, G. W. A. M. (2014). Automatic fruit recognition and counting from multiple images. *Biosystems Engineering, 118*, 203–215. https://doi.org/10.1016/j.biosystemseng.2013.12.008.

Stein, M., Bargoti, S., & Underwood, J. (2016). Image based mango fruit detection, localisation and yield estimation using multiple view geometry. *Sensors (Basel), 16*(11), 1915. https://doi.org/10.3390/s16111915.

Tu, S., Xue, Y., Zheng, C., Qi, Y., Wan, H., & Mao, L. (2018). Detection of passion fruits and maturity classification using Red-Green-Blue Depth images. *Biosystems Engineering, 175*, 156–167. https://doi.org/10.1016/j.biosystemseng.2018.09.004.

Vitzrabin, E., & Edan, Y. (2016). Adaptive thresholding with fusion using a RGBD sensor for red sweet-pepper detection. *Biosystems Engineering, 146*, 45–56. https://doi.org/10.1016/j.biosystemseng.2015.12.002.

Xu, L., & Lv, J. (2017). Recognition method for apple fruit based on SUSAN and PCNN. *Multimedia Tools and Applications, 77*(6), 7205–7219. https://doi.org/10.1007/s11042-017-4629-6.

Yang, L., Zhang, L. Y., Dong, H. W., Alelaiwi, A., & El Saddik, A. (2015). Evaluating and improving the depth accuracy of kinect for windows v2. *IEEE Sensors Journal, 15*(8), 4275–4285. https://doi.org/10.1109/jsen.2015.2416651.

2.3 Behavior Tracking and Analyses of Group-Housed Pigs Based on Improved ByteTrack

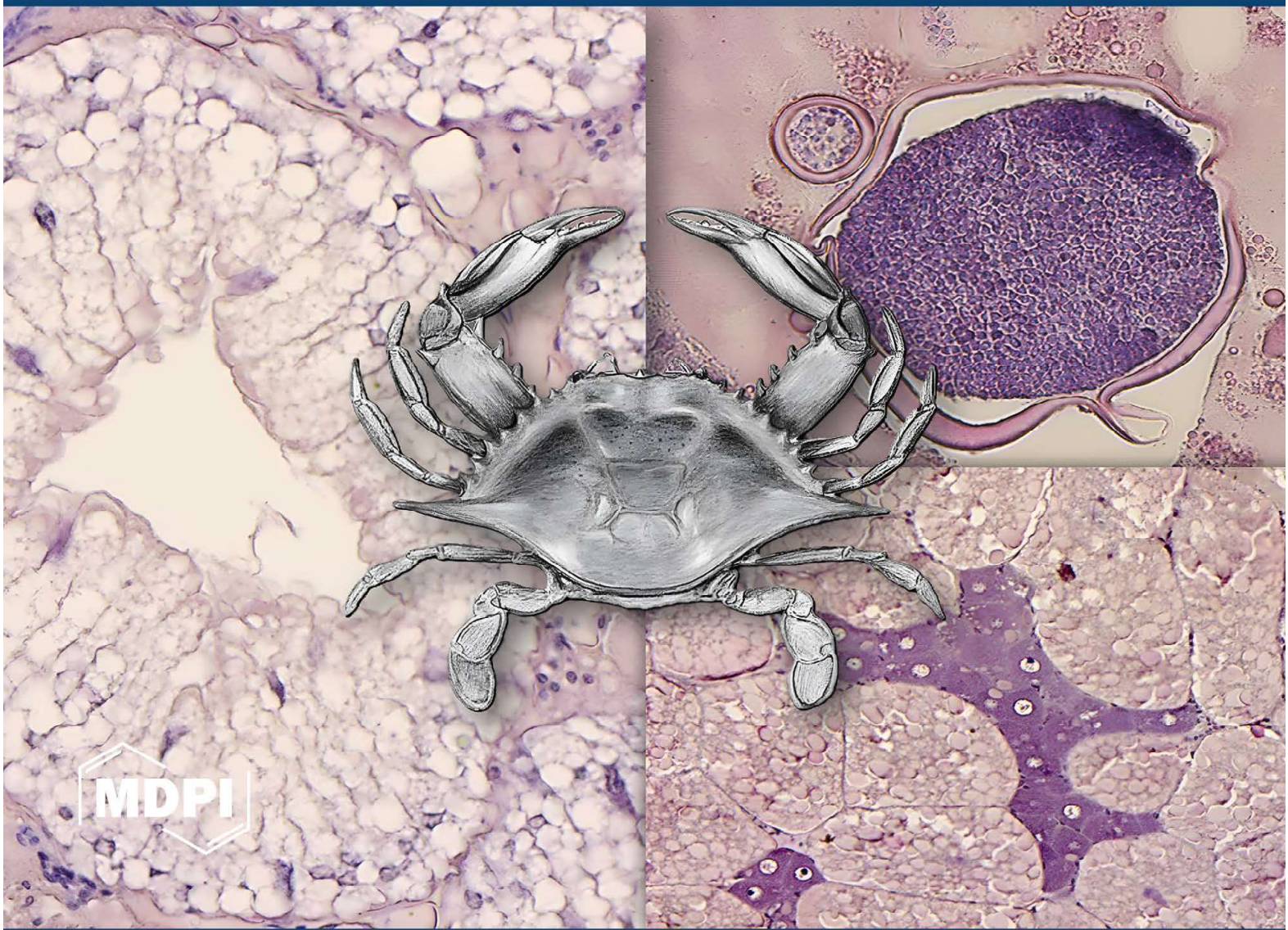
animals

Impact Factor 2.7
CiteScore 4.9
Indexed in PubMed

ISSN 2076-2615

The Influence of Diet on Captive-Reared Blue Crabs: A Focus on Reproduction

Volume 14 · Issue 22 November-2 2024



MDPI

Report Information from ProQuest

July 14 2025 08:45

目录

目录

目录

目录

目录

目录

目录

目录

animals

MDPI

Article

Behavior Tracking and Analyses of Group-Housed Pigs Based on Improved ByteTrack

Shuqin Tu [1], Haoxuan Ou [1], Liang Mao [2,*], Jiaying Du [1], Yuefei Cao [1] and Weidian Chen [1]

[1] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China; tsq5_6@scau.edu.cn (S.T.); ouhaoxuan123@163.com (H.O.); dujiaying2001@163.com (J.D.); cyf_1208@163.com (Y.C.); chenwd2543@163.com (W.C.)

[2] Institute of Applied Artificial Intelligence of the Guangdong-Hong Kong-Macao Greater Bay Area, Shenzhen Polytechnic University, Shenzhen 518055, China

* Correspondence: maoliangscau@szpu.edu.cn

Simple Summary: This study develops a new method to automatically track and analyze pig behavior in complex environments. We use Pig-ByteTrack algorithm for real-time tracking and perform trajectory interpolation post-processing on the original algorithm to solve tracking problems caused by light changes, occlusion, and collisions between pigs. Finally, a set of time statistical algorithms for behavior categories are designed. The method helps farm managers detect abnormalities and health problems of pigs in a timely manner. Experimental results show that the method performs well in behavior recognition and tracking, accurately records pig behavior, and provides technical support for monitoring the health and welfare of pig herds.

Abstract: Daily behavioral analysis of group-housed pigs provides critical insights into early warning systems for pig health issues and animal welfare in smart pig farming. In this study, our main objective was to develop an automated method for monitoring and analyzing the behavior of group-reared pigs to detect health problems and improve animal welfare promptly. We have developed the method named Pig-ByteTrack. Our approach addresses target detection, Multi-Object Tracking (MOT), and behavioral time computation for each pig. The YOLOX-X detection model is employed for pig detection and behavior recognition, followed by Pig-ByteTrack for tracking behavioral information. In 1 min videos, the Pig-ByteTrack algorithm achieved Higher Order Tracking Accuracy (HOTA) of 72.9%, Multi-Object Tracking Accuracy (MOTA) of 91.7%, identification F1 Score (IDF1) of 89.0%, and ID switches (IDs) of 41. Compared with ByteTrack and TransTrack, the Pig-ByteTrack achieved significant improvements in HOTA, IDF1, MOTA, and IDs. In 10 min videos, the Pig-ByteTrack achieved the results with 59.3% of HOTA, 89.6% of MOTA, 53.0% of IDF1, and 198 of IDs, respectively. Experiments on video datasets demonstrate the method's efficacy in behavior recognition and tracking, offering technical support for health and welfare monitoring of pig herds.

Keywords: behavioral analysis algorithm; multi-object tracking (MOT); long-term video tracking; Pig-ByteTrack

Citation: Tu, S.; Ou, H.; Mao, L.; Du, J.; Cao, Y.; Chen, W. Behavior Tracking and Analyses of Group-Housed Pigs Based on Improved ByteTrack. *Animals* **2024**, *14*, 3299. https://doi.org/10.3390/ani14223299

Academic Editor: Giuseppe De Rosa

Received: 12 October 2024
Revised: 2 November 2024
Accepted: 4 November 2024
Published: 16 November 2024

1. Introduction

With the increasing demand for animal products and the growing social concern for animal welfare, effective monitoring and analysis of animal welfare is increasingly becoming a hot research priority. The health status of pigs will determine the development and economic efficiency of pig farming. Meanwhile, clinical or subclinical signs of most swine diseases are usually accompanied by behavioral abnormalities before the appearance of symptoms [1]. Currently, with the development of image processing technology, the integration of manual observation and computer vision monitoring is the main management method in large-scale pig farms, this still requires a large amount of labor. And

with the rapid development of Multi-Object Tracking (MOT) technology in the field of video surveillance, the need for manual labor can be significantly reduced through the application of MOT for the pig industry, improving the efficiency and cost-effectiveness of pig farm management.

In recent years, several outstanding MOT algorithms have been proposed. For example, the prevailing trackers including simple online and real-time tracking (SORT [2]) and deep simple online and real-time tracking (DeepSORT [3]), have been widely used for the MOT of pedestrians and vehicles. SORT is a data association method based on a Kalman filter (KF) and a Hungarian algorithm to associate the detected bounding-box results between adjacent frames. DeepSORT introduces the comparison of appearance features, which is added to the motion model in SORT. This enhances the performance for a longer duration of occlusion. Zhang et al. proposed Fair Multi-Object Tracking (FairMOT) based on the anchor-independent target detection architecture CenterNet, which obtains high accuracy in detection and tracking for MOT datasets [4]. Sun et al. proposed an end-to-end Transtrack MOT method based on TransFormer, which can perform target detection and tracking tasks simultaneously [5]. The framework performed well in complex scenarios and achieved excellent performance on MOT16, MOT17, and MOT20 datasets. Zhang et al. proposed the ByteTrack algorithm, which adds a correlation phase to low-scoring detection frames to improve tracking performance on many pedestrian tracking datasets [6]. Aharon et al. proposed the Bayesian online tracking with sorting (BOT-SORT) algorithm, which achieves accurate tracking of multiple targets by using appearance and motion information for re-identification [7]. It could effectively deal with challenges such as target occlusion and appearance changes. Due to the excellent performance of these MOT algorithms, they are widely applied in various fields, including autonomous driving, traffic management, drones, aviation, medical image analysis, agriculture [8–17], and so on.

Nowadays, more and more studies of MOT technology have been used in livestock detection and tracking. For example, Guo et al. improved the joint detection and embedding (JDE), FairMOT, and you only look once version 5 small (YOLOv5s) + DeepSORT algorithms, respectively, to improve the performance of individual animal tracking, especially to reduce the number of identity switches, thus ensuring timely animal welfare [18]. Zheng et al. proposed a MOT method (YOLO-BYTE), which aims to address the problem of missed and false detections caused by the complex environment in the detection and tracking of individual cows [19]. Lu et al. proposed a MOT method based on a rotating bounding box detector to recognize the number of pigs and monitor their health status, which improved the adaptability of the tracking technique in complex scenarios and reduced the switching of pig identities [20]. Zou et al. proposed an improved YOLOv3+DeepSORT algorithm to achieve accurate MOT of yellow-feathered broilers in large-scale broiler farms, which provides a technical reference for analyzing the behavioral perception and health status of broilers [21]. However, all these methods focused on target detection and tracking tasks, little research was conducted for further behavioral automated analysis based on the MOT results.

In addition, most work regarding pig detection and tracking involves studies on the pig tracking of short-duration videos or identifying postural behaviors. For example, Alameer et al. utilized a MOT technique for automated diagnosis and intelligent monitoring of health status in 1 min videos of pigs on a pig farm [22]. Huang et al. designed the HE-Yolo (High Effect Yolo) model to identify the postural behaviors of fenced pigs in real-time from 10 to 100 s of video [23]. Huang et al. proposed an improved pig counting algorithm (MPC-YD) based on theYOLOv5 + DeepSORT model, which aims to solve the problems of partial feature detection, tracking loss due to fast movement, and video counting errors of pigs [24]. Our previously published work proposed an improved DeepSORT algorithm for multi-target behavioral tracking in 1 min pig videos. The approach could improve tracking accuracy under complex scenarios and reduce error IDs due to overlapping and occlusion between pigs [25]. All the above methods demonstrated superior performance in pigs' MOT tasks for short-time videos. However, there is little research on pigs' behavioral

tracking for long-time videos. This is because in the long-term MOT of pigs, factors such as the occlusion of objects and influence of light appear with greater probability, which leads to a decrease in the accuracy and stability of detection and tracking of pigs; it is very challenging to recognize the behavior of pigs on this basis. Therefore, it is extremely challenging to record the behaviors of individual pigs for the long term completely by MOT methods.

To address the above challenges, we proposed an approach to complete target detection, tracking, and behavioral automated analysis of three tasks for group-housed pigs. First, we used the YOLOX-X detector in the detection task to detect the target and output for the four behavioral categories (lie, stand, eat, and other), locations, and confidence values of the pigs. Then, we employed a trajectory interpolation post-processing strategy in the data association part to minimize ID errors caused by occlusion to improve the tracking accuracy. Finally, we designed an automated behavioral analysis algorithm to calculate four behavioral times of each pig in each pen based on its ID and behavioral information.

In this study, our main objective was to develop an automated method for monitoring and analyzing the behavior of group-reared pigs to detect health problems and improve animal welfare promptly. Specific objectives include the following:

(1) We proposed the Pig-ByteTrack algorithm, integrating trajectory interpolation to reduce false alarms and enhance tracking stability, to improve the accuracy of behavior monitoring in pig farming.
(2) We designed a behavioral analysis algorithm to calculate the temporal distribution of behaviors for each pig, leveraging tracking IDs and categories, to enable detailed behavioral analysis within individual enclosures.
(3) We constructed a 10 min long-term pig dataset with real pig house videos and validated our methodology's effectiveness through comparative tracking experiments, to ensure the practical applicability and reliability of our approach under real-world conditions.

2. Method

To address the challenges of accurate precision in pig tracking and behavioral analyses under complex environments, we proposed the process of tracking and behavioral time statistics for group-housed pigs, as depicted in Figure 1. Firstly, we introduced a novel target tracking algorithm called Pig-ByteTrack, which is used to detect the pigs, classify the behavior categories, and assign the target pig ID for input video sequences. Then, the MOT results were generated, which included the following three parts: pig ID, the pigs' location, and the pigs' behavioral categories. Additionally, a behavioral analysis algorithm was designed to calculate the frequency of each behavior for each pig based on the pig's ID and behavior category information. Finally, based on a behavioral analysis algorithm, we designed the program to visualize the frame number, categories, and frequency for each pigs' behaviors, thereby obtaining statistical results of pig behaviors in the videos.

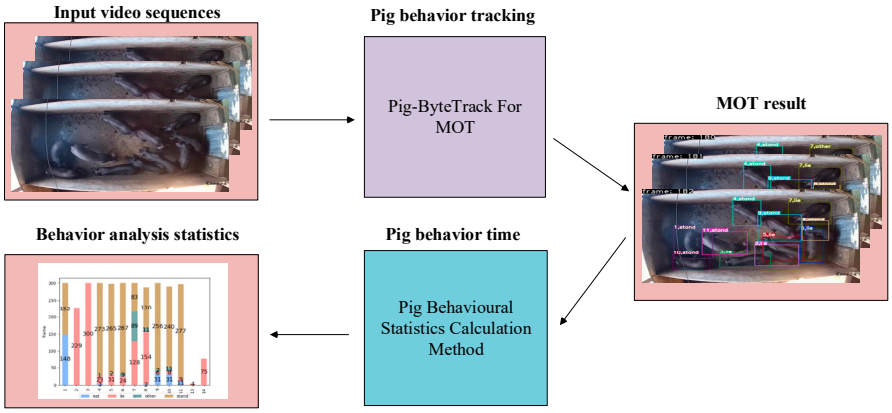


Figure 1. Process diagram of tracking and behavioral time statistics for group-housed pigs.

Figure 1. Process diagram of tracking and behavioral time statistics for group-housed pigs.

2.1. Pig-ByteTrack Algorithm of Group-Housed Pigs

Pig-ByteTrack is divided into two main phases: object detection and multi-target tracking. The workflow is shown in Figure 2. Firstly, the YOLOX-X detector is used to detect all pig targets for each frame of input video sequences, and the detection results contained pigs' confidence value, bounding box (BB), and behavioral category. Then, the

2.1. Pig-ByteTrack Algorithm of Group-Housed Pigs

Pig-ByteTrack is divided into two main phases: object detection and multi-target tracking. The workflow is shown in Figure 2. Firstly, the YOLOX-X detector is used to detect all pig targets for each frame of input video sequences, and the detection results contained pigs' confidence value, bounding box (BB), and behavioral category. Then, the MOT tracker uses the improved BYTE data association algorithm to match the high-scoring detection boxes to the trajectory, and the unmatched trajectory is linked to the low-scoring boxes. The improved Byte includes two operations: the Hungarian Matching Algorithm and the Kalman Filter (KF) prediction, respectively. Finally, the target trajectory output for the images of consecutive video frames is obtained by the MOT tracker.

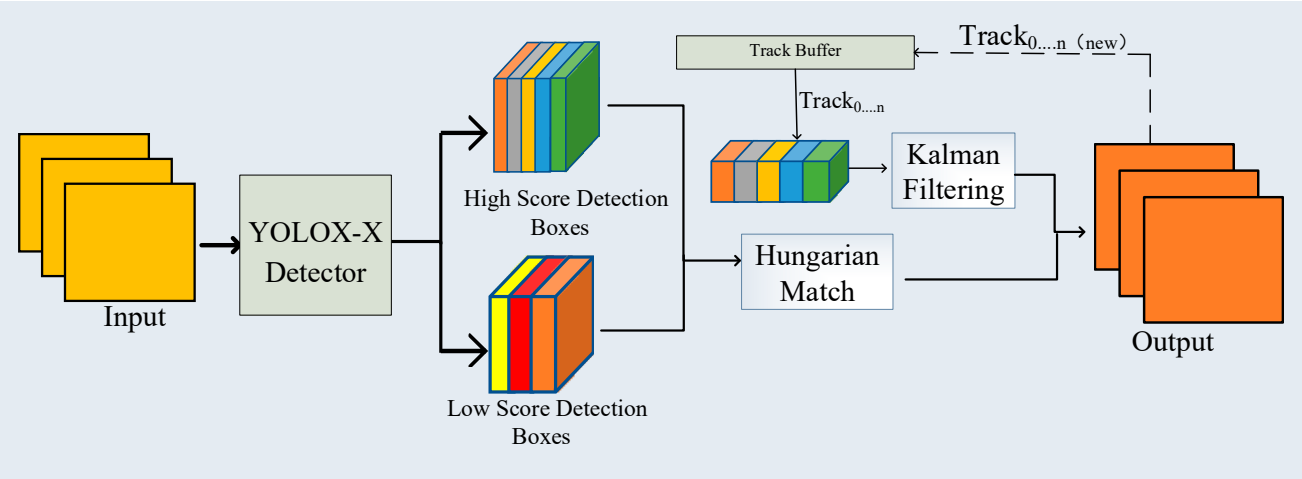


Figure 2. Flow chart of Pig-ByteTrack algorithm.

2.1.1. The Original ByteTrack Algorithm

The ByteTrack algorithm utilized the detector of YOLOX to complete the BB regression and behavior recognition. The detector of YOLOX is an improved version of YOLO [26,27] with a simple scheme and better performance without the anchor mechanism. YOLOX decoupled the YOLO detection head into distinct feature channels for box coordinate regression and target classification. Then, the ByteTrack algorithm utilized the BYTE data association strategy for MOT task. Its processing flow chart is shown in Figure 3.

The detailed process is as follows:

(1) The object detection results are divided into high-score and low-score detection boxes.

In the detection results, if the confidence value of the detection box is greater than the high score box threshold, the detection box is placed in the high score detection box set (D_{high}). If the confidence value of the detection box is less than the high score box threshold and greater than the low score box threshold, the detection box will be placed in the low score detection box set (D_{low}).

(2) The high-scoring detection boxes D_{high} are matched with the existing tracks for the first IoU data association.

The IoU distance matrixes between high-scoring detection boxes D_{high} and the set of trajectories are calculated and then used to match using the Hungarian algorithm, which produces three kinds of outputs including matched track set, unmatched high-scoring detection boxes, and unmatched tracks. The matched track set contains successfully matched detection boxes updated with their Kalman filter. The unmatched high-scoring detection boxes and unmatched tracks are placed in D_{remain} and T_{remain} sets, respectively.

(3) The low score detection boxes D_{low} are associated with the unmatched trace-in T_{low} for the second IoU data.

We calculate the IoU distance matrixes between D_{low} and the unmatched tracks T_{remain}, which output three kinds of sets, including unmatched tracks, unmatched low-scoring detection boxes, and matched track sets. The unmatched trajectories are placed in T_{lost} set. The unmatched low-score detection boxes will be processed background and boxes in Figure 3 will be deleted directly. For those successfully matched trajectories, its Kalman filter is updated and placed in the current frame trajectory set.

(4) Trajectories creation, deletion, and merging.

For the detection boxes D_{remain}, if the confidence values are greater than the tracking threshold, new tracks are created for them and are merged into the current set of trajectories for the current frame, if not, they are neglected. The stored trajectories that exceed 30 frames in T_{lost} are deleted. All trajectories for the current frame are outputted and fed to the next image frame as the existing trajectories.

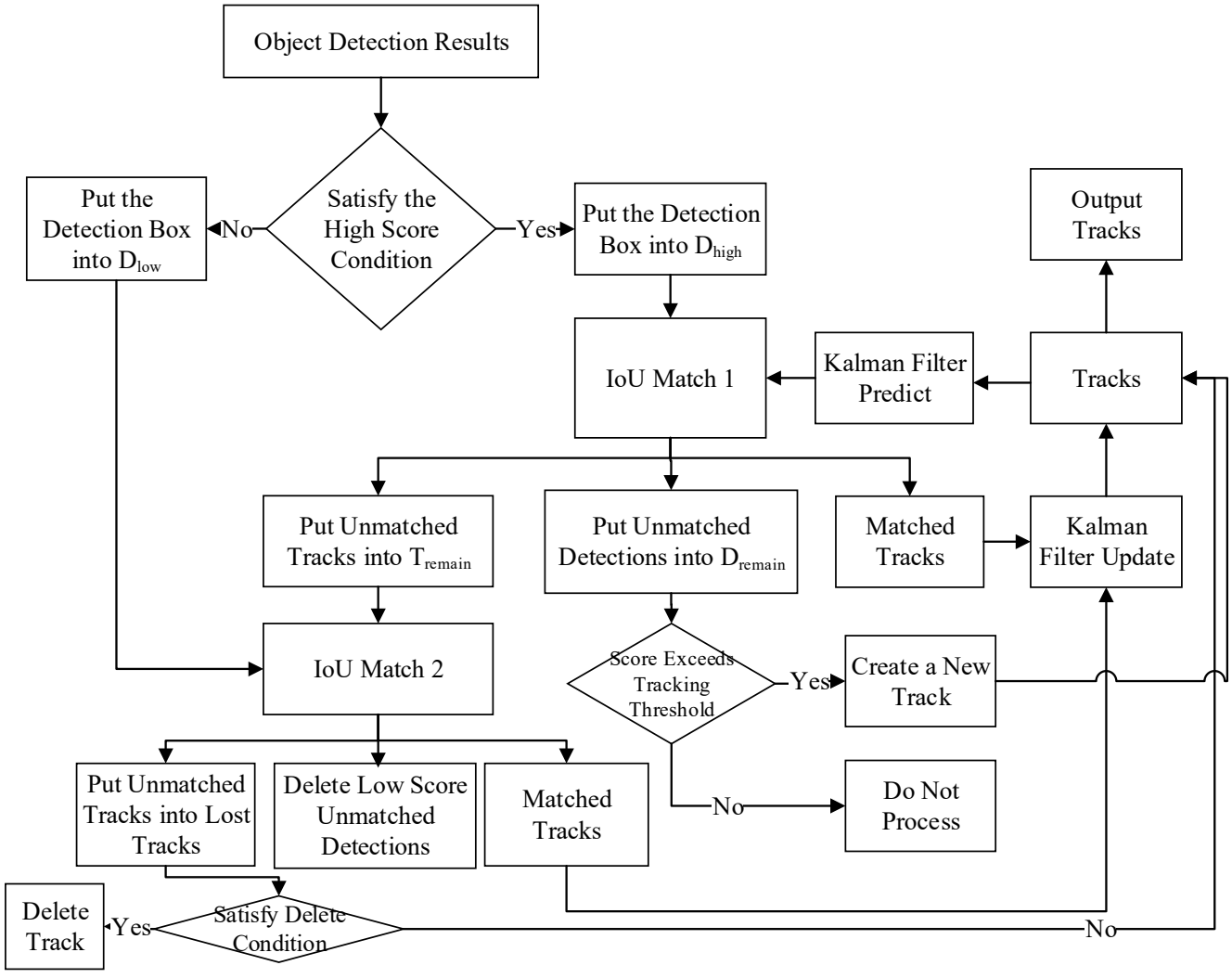


Figure 3. The flow chart of the Byte data association algorithm.

2.1.2. The Pig-ByteTrack Algorithm

To implement stable behavior tracking of group-housed pigs, based on ByteTrack, the improvement of the Pig-ByteTrack algorithm consists of two steps as follows:

(1) The design of suitable detection anchor boxes and the improvement of tracking boxes, which reduces the kinds of outputs, including unmatched tracks, unmatched high-scoring detection boxes, and unmatched tracks. The invalid tracks will be removed successfully matched detection boxes updated with the Kalman filter. The unmatched high-score detection boxes and unmatched tracks are placed in D_{remain} and T_{remain} sets, respectively.

(2) The high-scoring detection boxes D_{high} are matched with the existing tracks for the first IoU data association.

The IoU distance matrixes between high-scoring detection boxes D_{high} and the set of trajectories are calculated and then used to match using the Hungarian algorithm, which outputs three kinds of outputs, including unmatched tracks, unmatched high-scoring detection boxes, and unmatched tracks. The invalid tracks will be removed successfully matched detection boxes updated with the Kalman filter. The unmatched high-score detection boxes and unmatched tracks are placed in D_{remain} and T_{remain} sets, respectively.

appropriate ratio of anchor boxes for pigs. At the same time, four behavioral classes of pigs (lie, stand, eat, and other) are added to the BYTE tracker for tracking.

The visualization of the Pig-ByteTrack and ByteTrack is shown in Figure 4. The tracking box of the original ByteTrack in Figure 4a only shows the ID number of each pig, whereas the tracking box of the Pig-ByteTrack in Figure 4b can reflect the behavioral categories and ID value of each pig at the same time.

(a) Tracking boxes of original ByteTrack (b) Tracking box of Pig-ByteTrack

Figure 4. Comparison of tracking box between Pig-ByteTrack and ByteTrack.

(2) The implementation of the trajectory interpolation post-processing strategy for BYTE tracker.

To avoid error IDs due to severe occlusion between group-housed pigs, we propose the trajectory interpolation post-processing strategy to significantly improve the stable tracking performance of occluded targets. The process is as follows:

Suppose a trajectory T is lost due to occlusion between t_1 frame and t_2 frame, if the current trajectory T is at exactly t frame ($t_1 < t < t_2$), the interpolation box B_t of the trajectory T can be calculated by Equation (1) as follows:

$$B_t = B_{t1} + (B_{t2} - B_{t1})\frac{t - t_1}{t_2 - t_1} \quad (1)$$

where B_t denotes the track box coordinates of the t frame (containing four values, one for the top left and one for the bottom right coordinates). Moreover, B_{t1} denotes the track box coordinates of the t_1 frame, B_{t2} denotes the track box coordinates of the t_2 frame.

2.2. The Behavioral Analysis Algorithm

Based on the video sequence tracking results, we design and implement the pig's four behavioral time calculation algorithm as shown in Algorithm 1. The algorithm flow is as follows:

(1) An array of a behavioral category named $[A_1, A_2, A_3, A_4]$ is designed for each track, which creates statistics for the number of all tracks for the four categories of the stand, lie, eat, and other behaviors. The statistic is added as an attribute in each of the tracks.
(2) For each frame of video, the BYTE data association algorithm first obtains information about the YOLOX-X detection results of each BB named D, including the category information mentioned above. Then, the behaviors analysis algorithm creates an array of frames named $[a_1, a_2, a_3, a_4]$ based on the categories (stand, lie, eat, and other behaviors of each BB) for each pig ID. And if the behavior category belongs to stand, a_1 is set to 1 and the other parameters are set to 0, and so on.

(3) After Associating T with D using the BYTE operation, we can revise the values $[A_1, A_2, A_3, A_4]$ if the tracklet and detection BB match successfully. The formula of revised A is as follows:

$$\begin{bmatrix} A_{1new} \\ A_{2new} \\ A_{3new} \\ A_{4new} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \quad (2)$$

If there is no match between the detection BB and the track, or if the confidence value of the detection BB is greater than the high score threshold, we can set the value of $[A_1, A_2, A_3, A_4]$ for this tracklet to 0. Finally, after summing the number of frames obtained for the four behaviors with different pig IDs, we divide it by the frame rate to obtain the time for the different behaviors of each pig ID.

Algorithm 1. Pseudo-code of Behavior Category Time Statistics of Pigs

Input: A video sequence V; object detector Det; the k frame f_k; the detection BB D; $category$ includes lie, eat, stand, other; variable a and A: one-dimensional array including four elements for time statistics; T: tracklet information and four behavior category time statistics; tracking score threshold η is set 0.75; Frames per second Fps;
Output: Tracks T of the video

Initialization: $T \leftarrow \varnothing$
for frame f_k in V **do**
  $D \leftarrow Det(f_k)$
  Initialize time-count array including four elements for time statistics $a \leftarrow [0, 0, 0, 0]$
  Set variable category $\_index \leftarrow D\{category\}$
  $[category\_index] \leftarrow 1$
  Associate T with D using BYTE:
    if succeed to match **then**
      Call the Update or Re-activate function to update the status of tracks
      Set variable $A \leftarrow T\{category\_\_time\_array\} + a$
      $T\{category\_\_time\_array\} \leftarrow A$
    end
    if D failed to match and $D >$ **then**
      Call the function to create a new track
      Initialize time-count array A$\leftarrow [0, 0, 0, 0]$
      $T\{category\_\_time\_array\} \leftarrow A$
    end
end
$T\{category\_\_time\_array\} = T\{category\_\_time\_array\}/Fps$

Return T

3. Experiment

All the experiments in this study were conducted on the same computer using Linux as the experimental platform with Ubuntu 20.04 operating system, using Python 3.7 as the programming language, Pytorch 1.9.1 as the deep learning framework, and CUDA version 11.1. The GPU server is RTX3090, and the memory is 64 GB. We select HOTA, MOTA, IDF1, and IDs as the pig MOT evaluation metrics.

$HOTA$ calculation is shown in Equation (3) as follows:

$$HOTA = \sqrt{\frac{\sum_{c \in TP} A(c)}{TP + FN + FP}} \quad (3)$$

where c is a point belonging to TP, according to which we can always determine a unique Ground Truth trajectory, and $A(c)$ represents the association accuracy. TP refers to the number of positive samples; FN refers to the positive samples predicted by the

model to be negative; and FP represents the negative samples predicted by the model as positive samples.

$MOTA$ calculation is shown in Equation (4) as follows:

$$MOTA = 1 - \frac{\sum_t (FP + FN + IDS)}{\sum_t g_t} \tag{4}$$

where FP represents the total number of false detections in frame t; FN represents the total number of missed detections in frame t; IDS represents the number of times the target label ID switched during tracking in frame t; and g_t represents the number of targets observed at frame t.

$IDF1$ calculation is shown in Equation (5) as follows:

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{5}$$

where $IDTP$ represents the total number of targets correctly tracked with unchanged ID; $IDFP$ represents the total number of targets incorrectly tracked with unchanged ID; and $IDFN$ represents the total number of targets lost in tracking with unchanged ID.

Additionally, the model performance is evaluated with the number of ID Switches (IDs) in this study. Higher values of $HOTA$, $MOTA$, and $IDF1$, and a lower value of IDs indicate better model performance.

3.1. Dataset

For a comprehensive analysis of group-housed pig behavior, this study collected two sets of pig behavior video datasets from different scenarios, categorized as public and private datasets. The public dataset [28] comprised video clips of pigs of different breeds, recorded in both daytime and nighttime environments, with each pigpen house including 7 to 20 pigs, with a total of 4 annotated 10 min videos. The private dataset was collected in September 2022 from a commercial pig farm in Foshan, where each pigpen houses 6 to 11 black and spotted pigs. Pigs in the study were categorized into three age groups: nursery (3–10 weeks), early fattening (11–18 weeks), and late fattening (19–26 weeks). Weight distributions across these stages align with standard growth curves. In the nursery phase, pigs weighed 7–10 kg at 3 weeks, 10–15 kg at 4–5 weeks, and 15–25 kg at 6–10 weeks. Early fattening pigs weighed 25–35 kg at 11 weeks, 35–50 kg at 12–15 weeks, and 50–70 kg at 16–18 weeks. In the late fattening phase, pigs weighed 70–90 kg at 19 weeks, 90–110 kg at 20–23 weeks, and 110–130 kg at 24–26 weeks. These estimates may vary due to breed, husbandry practices, and diet. The study included two breeds: black and spotted pigs. Welfare principles were adhered to, with each pig allocated 1.2 square meters of pen space. The housing environment was controlled for temperature and humidity to ensure suitability for the animals, and the facility was equipped with adequate ventilation for herd density. Bedding conditions comprised both solid and metal grid flooring. Empirical evidence suggests that pigs display comparable behaviors on solid and grid floors, indicating that flooring preferences do not significantly affect basic behavioral patterns and thus do not impact the study's findings.

This dataset annotated 18 1 min video segments, with 9 video segments for training and 9 for testing. Each video segment had a frame rate of 5 frames per second, resulting in 300 images in a 1 min video and 3000 images in a 10 min video segment. All datasets were annotated using DarkLabel1.3 software. Our classifications of pig activity levels were determined by direct observation in cooperation with farm staff. Specifically, the classifications of 'low', 'medium', and 'high' activity levels were based on daily observations and empirical judgments of staff on pig behavior. These categories, although subjective, reflect the routine practice of actual pig farming and provide a practical benchmark for our study. In future studies, we plan to introduce more objective quantitative methods to enhance the accuracy and consistency of the categories. We divided the activity level

of pigs into three categories according to the time period: high activity during the day, medium activity during the day (or night), and low activity during the day (or night). The activity level of pigs was defined as follows: during the day (07:00–17:00), pigs have more frequent behaviors such as eating and walking, this time was defined as a high activity level. During the day (07:00–17:00) or night (18:00–06:00), pigs have no high activities such as eating and walking. This time was defined as the medium activity level during the day or night. During the day (07:00–17:00) or night (18:00–06:00), pigs had few eating and walking behaviors, mostly lying behavior. This time period was defined as low activity during the day or night. The detailed test dataset is shown in Table 1.

Table 1. Test dataset.

Dataset	Sequence	Day	Night	Activity Level	Number of Pigs
1 min videos	10	√	–	Medium	11
	11	√	–	High	10
	12	–	√	Low	11
	13	√	–	High	11
	14	√	–	High	6
	15	√	–	Medium	6
	16	√	–	Medium	6
	17	–	√	Low	6
	18	–	√	Low	6
10 min videos	01	√	–	Medium	7
	02	–	√	Low	8
	03	√	–	High	14
	04	–	√	Low	15

The "√" indicates that the value is true, and the "–" indicates that the value is false.

The experiment consisted of the following four parts: (1) the comparison of tracking results of Pig-ByteTrack, ByteTrack, and TransTrack in the private dataset; (2) the analysis of MOT results using Pig-ByteTrack for 1 min videos in the private dataset; (3) the tracking results analysis of Pig-ByteTrack for 10 min videos; and (4) the behavioral analysis for four video segments in the private dataset.

3.2. Experimental Results and Performance Comparison

In this section, we evaluated the achievement of experiment goals in detail. To achieve these goals, the following research tools and methods were used:

(1) The YOLOX-X detection model was used for target detection and behavioral recognition in pigs.
(2) The Pig-ByteTrack algorithm was designed to track behavioral information for each pig.
(3) The Automated Behavioral Analysis algorithm was developed to calculate the temporal distribution of behaviors.

3.2.1. Results Comparison of Pig-ByteTrack, ByteTrack, and TransTrack

The results comparison of the Pig-ByteTrack with the ByteTrack and TransTrack in the private dataset is shown in Table 2. The Pig-ByteTrack achieved the best performance with HOTA, MOTA, IDF1, and IDs of 72.9%, 91.7%, 89% and 41, respectively. Compared with the ByteTrack, the results of the Pig-ByteTrack improved by 1.5%, 1.1%, 1.1% and 14 in HOTA, MOTA, IDF1, and IDs, respectively. Compared with TransTrack, its HOTA, MOTA, and IDF1 were improved by 23.4%, 4.4%, and 21%, and its IDs decreased 212, respectively, with significant performance improvement. Combining the above results, we found that Pig-ByteTrack can achieve stable behavioral tracking of group-reared pigs.

Table 2. The performance comparison between Pig-ByteTrack and the other 2 MOT methods.

Algorithms	HOTA(%) ↑	MOTA(%) ↑	IDF1(%) ↑	IDs ↓
TransTrack	49.5	87.3	68	255
ByteTrack	71.4	90.6	87.9	55
Pig-ByteTrack	**72.9**	**91.7**	**89.0**	**41**

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable. and bolded results represent the method used in this study.

The tracking results for Pig-ByteTrack, ByteTrack, and TransTrack in Videos 10, 11, 12, and 13 were shown in Figure 5. The Pig-ByteTrack method not only provided the ID number but also the behavioral class of each tracked pig, whereas ByteTrack and TransTrack only displayed the pig's ID without behavioral categorization. Pig-ByteTrack consistently achieved accurate pig behavior tracking across all four videos, with minimal IDs. In Video 10, ByteTrack failed to accurately identify a pig within the red dashed box at frame 17. In Video 11, ByteTrack incorrectly changed an ID indicated by the arrow in frame 26, where the pig was changed from ID11 to ID12. TransTrack missed detecting two black pigs within the red dashed box. Video 12 was captured in a challenging night scene, where the color of the pig is very similar to the shadow, making it difficult to identify the pig. Pig-ByteTrack had no missed detection, and ByteTrack and TransTrack both exhibited missed detection within the red dashed box. In Video 13, the same problem occurred with ByteTrack and TransTrack.

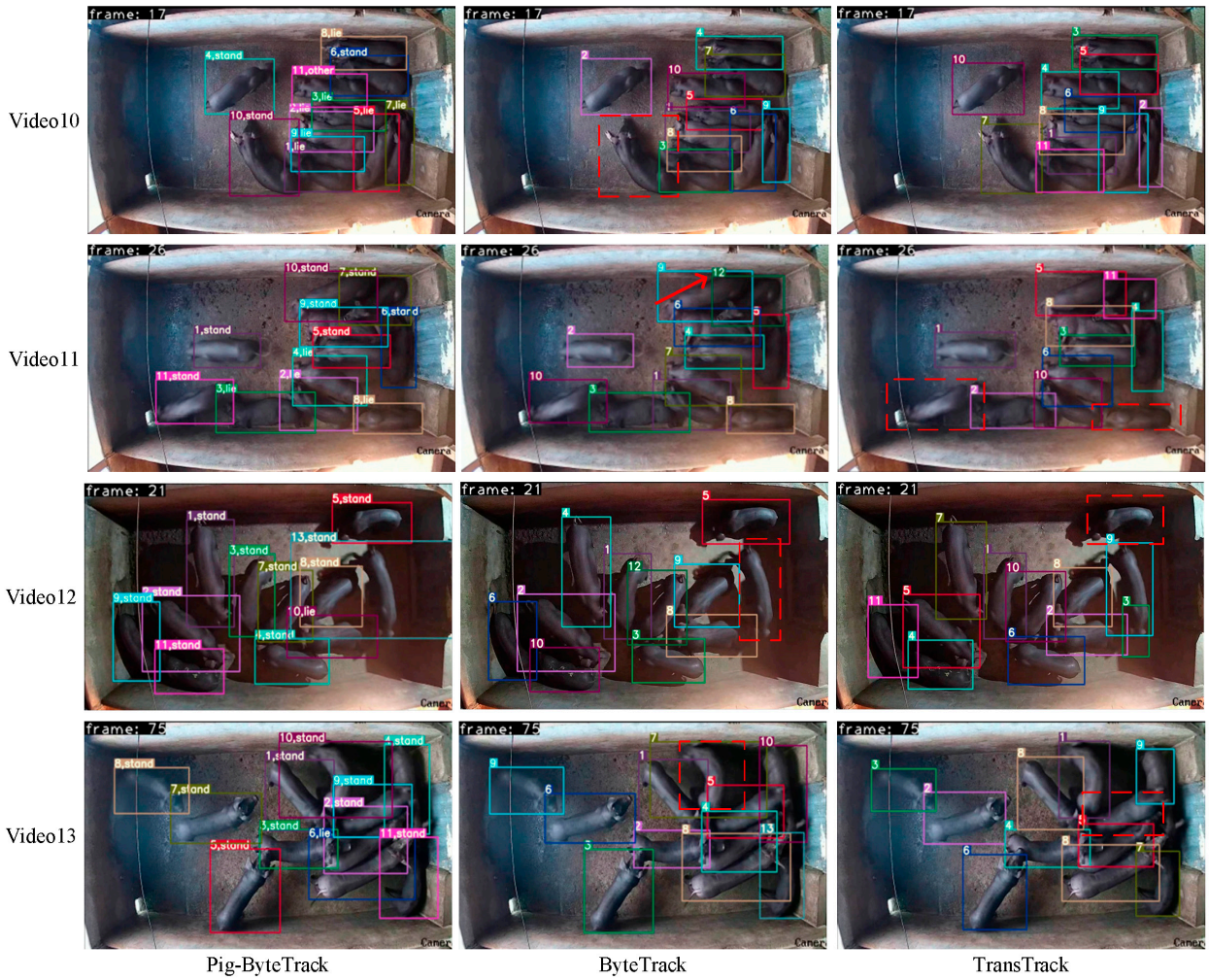


Figure 5. Comparison of Pig-BytetTrack, ByteTrack and TransTrack results on private datasets.

3.2.2. Results of Pig-ByteTrack for 1 min Video in Private Dataset

Pig-ByteTrack performed well in the various video tests with high averages. Table 3 shows its performance in each test video. Pig-ByteTrack performed well in Videos 14, 15, 16, and 17 with HOTA scores of 77.1%, 81.7%, 82.5%, and 79.6%, respectively, and MOTA scores consistently above 97%. In Video 17, both HOTA and IDF1 obtained max values of 97.9% and 99.0%, respectively, with zero IDs. These results showed Pig-ByteTrack could achieve good accuracy in this scenario. However, in Video 18, Pig-ByteTrack performed relatively poorly with 56.0%, 67.7%, and 64.2% for HOTA, MOTA, and IDF1, respectively, and the number of IDs was as high as 17. These results suggested that Pig-ByteTrack is affected in scenes with low lighting at night. Overall, Pig-ByteTrack showed excellent accuracy and stability in the pig MOT task. Further optimization and improvement is needed in complex situations such as severe occlusion and insufficient light.

Table 3. The results of each 1 min video tracking for Pig-ByteTrack.

Video	HOTA(%) ↑	MOTA(%) ↑	IDF1(%) ↑	IDs ↓
10	80.3	94.5	95.0	4
11	72.5	94.9	87.8	3
12	63.0	88.7	84.9	10
13	66.2	97.9	84.9	6
14	77.1	97.9	95.0	0
15	81.7	97.1	90.9	1
16	82.5	97.6	98.8	0
17	79.6	97.9	99.0	0
18	56.0	67.7	64.2	17
Average	72.9	91.7	89.0	41

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable.

Figure 6 shows the tracking results for Pig-ByteTrack, ByteTrack, and TransTrack from frames 6 to 116 of Video 11, where the real number of pigs was 11. The Pig-ByteTrack method achieved good tracking throughout and additionally provided behavioral categories for each pig tracked. ByteTrack and TransTrack both experienced tracking loss (pigs in the red dashed box in Figure 6). The tracking performance of Pig-ByteTrack was more stable and accurate than that of the other two methods.

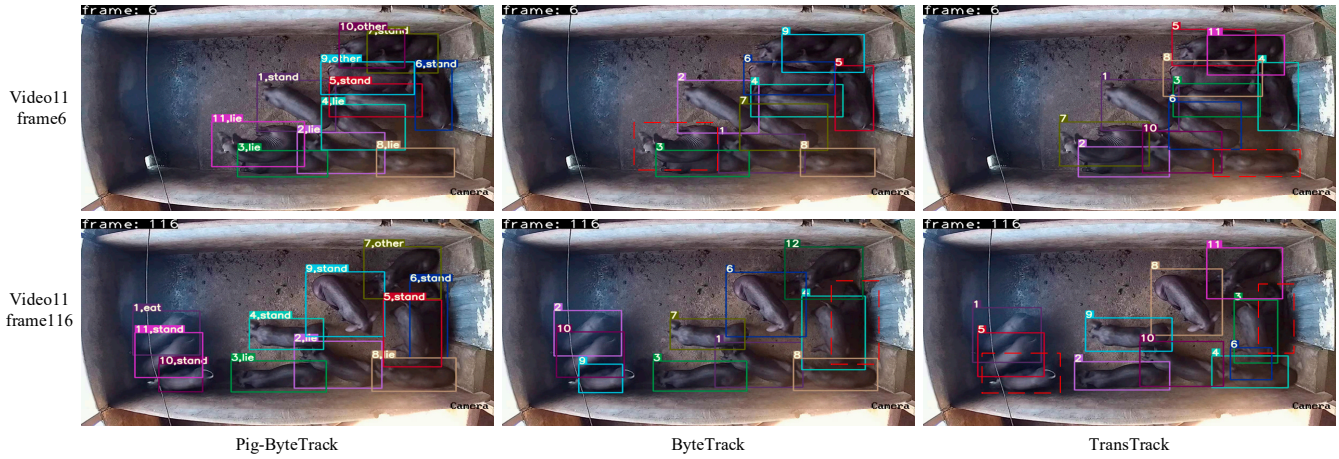


Figure 6. The visualized tracking results comparison of Pig-BytetTrack, ByteTrack, and TransTrack.

3.2.3. Results of Pig-ByteTrack for 10 min Video Dataset

The results of Pig-ByteTrack in the 10 min videos are shown in Table 4. We found that the tracking results of the 10 min videos were much less than those of the 1 min videos by comparing the results with Table 3. Video 02 had the highest HOTA rate, reaching 69.1% precision, while Video 03 had an HOTA rate of only 50.8%. The reason was that

the pig's active level in Video 03 was the highest among the four videos, as the pig moves frequently, resulting in frequent problems such as target occlusion, environmental light changes, etc. The Pig-ByteTrack achieved the best IDF1 value with 67.0% in Video 04 and the worst IDF1 value with 47.8% in Video 01. The reason for this phenomenon is that Video 04 is in the nighttime period, and most of the pigs are resting.

Table 4. The 10 min video tracking results for Pig-ByteTrack.

Video	HOTA(%) ↑	MOTA(%) ↑	IDF1(%) ↑	IDs ↓
01	66.1	90.8	47.8	14
02	69.1	95.1	53.4	29
03	50.8	88.4	50.8	72
04	69.1	95.1	67.0	83
Average	59.3	89.6	53.0	198

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable.

Figure 7 shows the visualized tracking results of Pig-ByteTrack for a 10 min video. The 10 min video had many issues with lost tracking and IDs (as indicated by the red arrow in Figure 7). We thought that this was because the 10 min video produces the situations such as many pig occlusions, light changes, and so on. The probability of its problems gap in tracking results. Therefore, the current method is still difficult to meet the requirements for a long time, which is a direction that needs to be improved in the future.

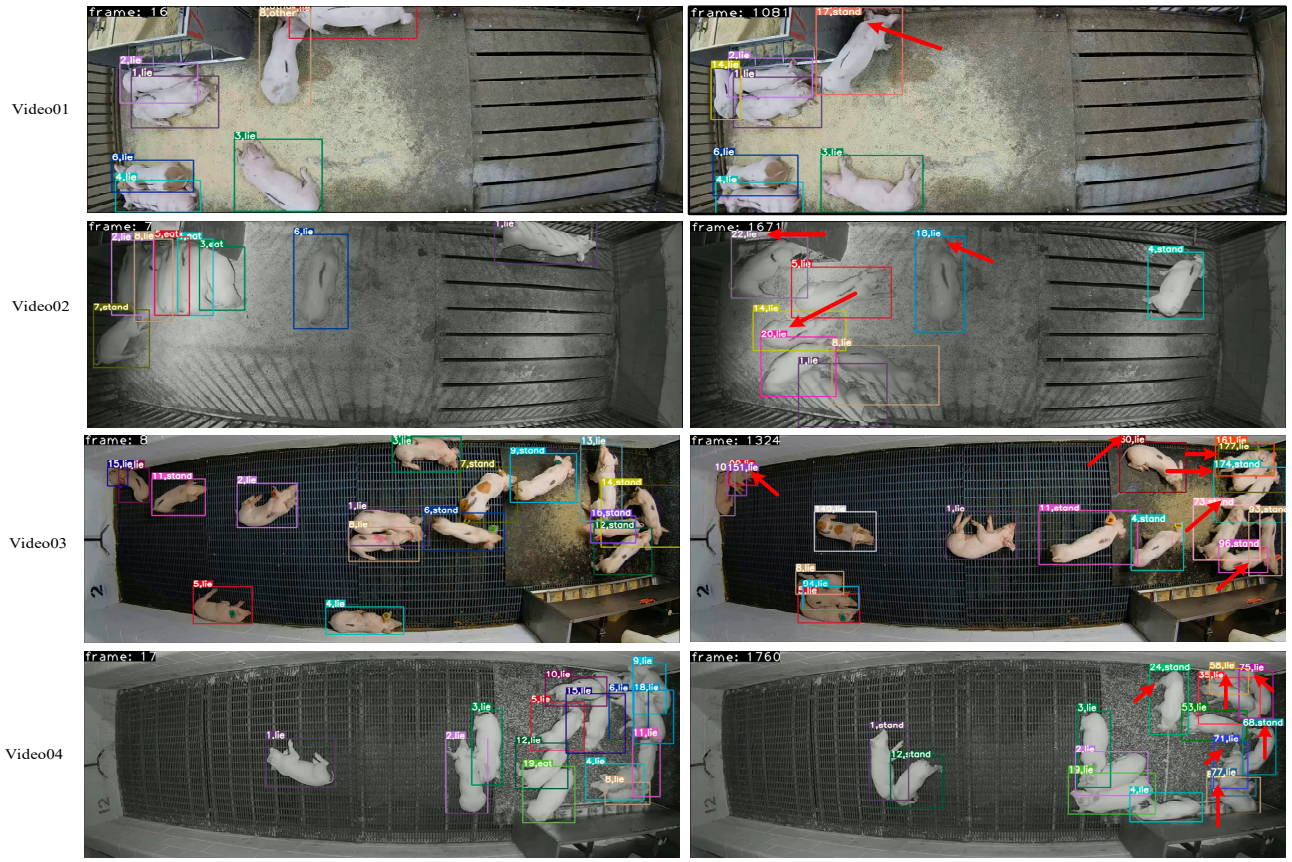


Figure 7. The visualized tracking results of Pig-BytetTrack in the 10 min videos. (The red arrows in the figure indicate pigs with id transformations).

Figure 7. The visualized tracking results of Pig-BytetTrack in the 10 min videos. (The red arrows in the figure indicate pigs with id transformations).

3.2.4. Behavioral Analysis for Four Video Segments in Private Dataset Algorithm

The histogram results using our designed behavioral analysis algorithm are shown in Figure 7. It presented the number of frames and distribution of different behavior classes indicated by each ID number results for the 1 min time videos, including Videos 14, 15, 16, and 17. In each sub-graph of Figure 7, the horizontal coordinates represented the ID number of the pigs, and the vertical coordinates represented the number of frames for each behavior. Four different colors represented the four behavioral classes of pigs: stand, lie, eat, and other, which were represented by light blue, dark blue, red, and yellow, respectively.

When we used the Pig-ByteTrack method to count the frame numbers of different behaviors of group-housed pigs, the loss of IDs was encountered, but this did not affect the consistency of the statistical results with the actual situation. In some time periods, pigs were masked, resulting in changes in identity, such as id3 and id7 in video15, and id6 in video16. In video14 of Figure 8, id1–4 mainly performed eat behaviors, while id5 and id6 performed lie and other behaviors, respectively, which proved that the pigs were highly active in this video. Moreover, in video15 of Figure 8, the pigs mainly engaged in eat and lie behaviors, with a similar proportion of the two, which indicated that the pigs were generally active. For video16 in Figure 8, most of the pigs exhibited stand behaviors, and only id3 and id5 engaged in other behaviors, which inferred that the pigs were medium active. Lastly, in video17 in Figure 8, the pigs, except id5, mainly exhibited stand behavior, proving the low activity level of the pigs. According to Table 1, we found that the results of videos 14–17 are completely in line with the actual situation.

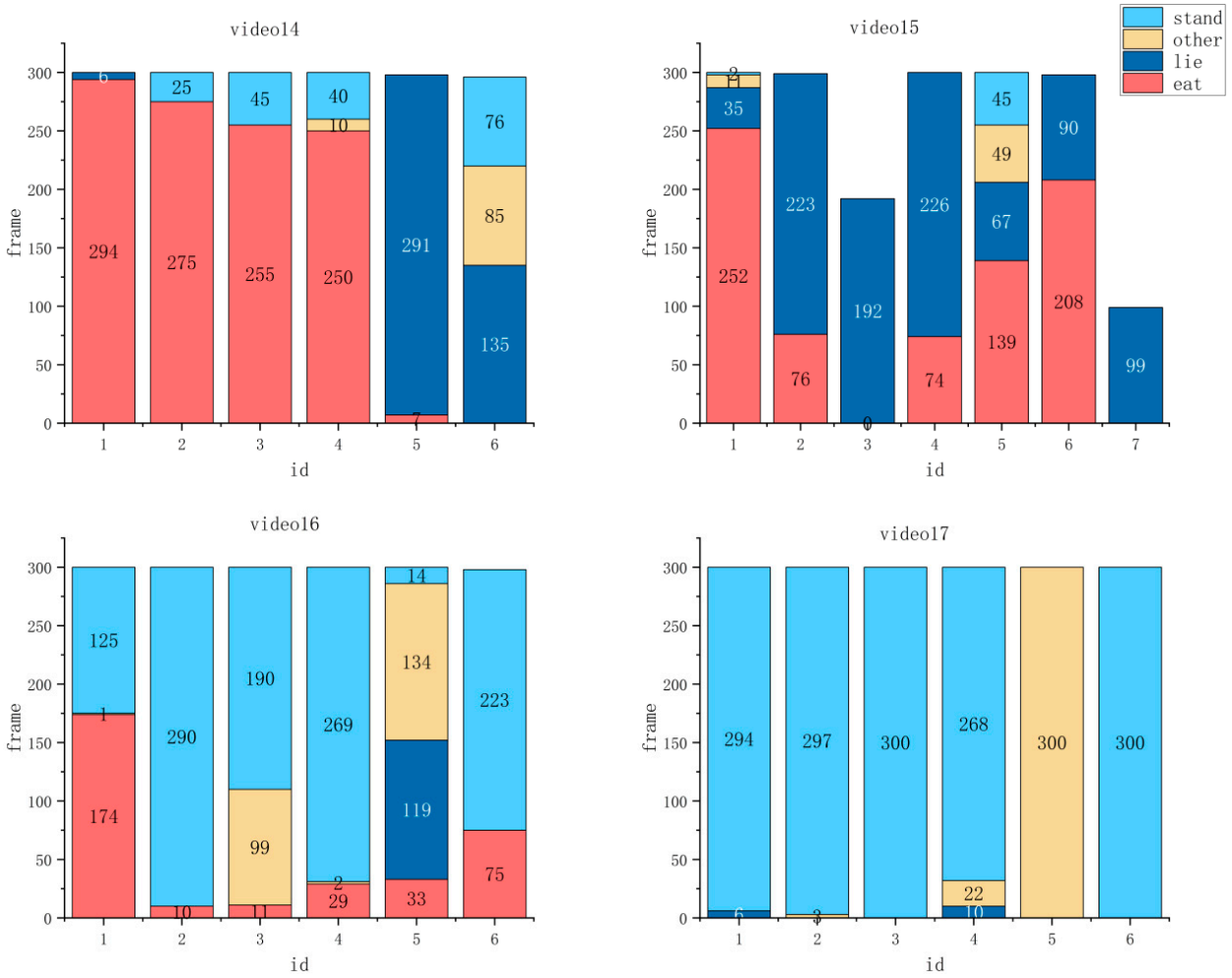


Figure 8. Pig behavior statistics for videos 14–17.

By the pig behavior statistics algorithm, our research provides professionals with a tool that enables them to identify and initially determine the cause of abnormal behavior. For example, if our system monitors a pig not eating for an extended period of time, this may indicate that the pig is experiencing a loss of appetite. This change in behavior could be due to a health issue, environmental discomfort, or a feed problem. If a pig is monitored lying down for an extended period of time without normal activity, this may indicate a possible injury or discomfort. With this behavioral data, our system helps farm managers and veterinary professionals quickly identify pigs of concern and provide an initial indication of the cause of abnormal behavior. This provides valuable time for further diagnosis and timely intervention.

4. Conclusions

In this paper, a Pig-ByteTrack method was proposed for target detection, behavioral classification, and multi-target tracking of group-reared pigs. The Pig-ByteTrack method achieves an HOTA of 72.9%, an MOTA of 91.7%, an IDF1 of 89%, and IDs of 41 for the behavioral tracking of pigs. Compared with ByteTrack, the HOTA, MOTA, IDF1, and IDs of the Pig-ByteTrack method were improved by 1.5%, 1.1%, 1.1%, and 14, respectively. The advantage of the Pig-ByteTrack method was tracking all detection frames and then dividing the detection frames into high-scoring and low-scoring detection frames. Meanwhile, the post-processing strategy of trajectory interpolation is used in data correlation to maximize the accuracy of tracking under occlusion. Based on Pig-ByteTrack, we designed an algorithm for statistical analysis of pig behavior. The algorithm could calculate the number of times each pig behaves in each pen according to the pig's ID and category information and could visualize the behavioral statistics time of pigs. Finally, we conducted a multi-target tracking study on the 10 min-long video of a pig, which provided a technical reference for subsequent long-time video tracking studies in this field.

Through the analysis of the method in this paper, we could consider that future research could focus on enhancing the model's long-term tracking performance. This could be achieved by delving deeper into the optimization of deep learning models, which entails refined network structures and fine-tuning hyperparameters. Furthermore, the integration of advanced data analysis and machine learning techniques presented an opportunity to develop predictive models for assessing the health status of pigs based on their behaviors. By continuously monitoring various parameters such as activity levels, feeding patterns, movement trajectories, and physiological indicators like body temperature and heart rate, it became feasible to implement real-time tracking systems for monitoring the health conditions of the pigs.

Author Contributions: Conceptualization, S.T. and L.M.; methodology, H.O.; software, H.O.; validation, H.O., W.C. and Y.C.; formal analysis, J.D.; investigation, H.O.; resources, H.O.; data curation, H.O.; writing—original draft preparation, H.O.; writing—review and editing, H.O.; visualization, H.O.; supervision, S.T.; project administration, H.O.; funding acquisition, H.O. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the Guangdong Province Rural Science and Technology Commissioner Project, zen tea reliable traceability and intelligent planting key technology research and development, promotion and application (KTP20210199), Special Project of Guangdong Provincial Education Department, research on abnormal behavior recognition technology of pregnant sows based on graph convolution (2021ZDZX1091), and Shenzhen Polytechnic University Smart Agriculture Innovation Application R&D Center (602431001PQ).

Institutional Review Board Statement: The animal study protocol was approved by the Animal Ethics Committee of South China Agricultural University (protocol code 2024F213 and date of approval: 14 March 2024).

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are included in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, S.; Jiang, H.; Qiao, Y.; Jiang, S.; Lin, H.; Sun, Q. The Research Progress of Vision-Based Artificial Intelligence in Smart Pig Farming. *Sensors* **2022**, *22*, 6541. [CrossRef] [PubMed]
2. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 26 September 2016; pp. 3464–3468.
3. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3645–3649.
4. Zhang, Y.; Wang, C.; Wang, X.; Zeng, W.; Liu, W. FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking. *Int. J. Comput. Vis.* **2021**, *129*, 3069–3087. [CrossRef]
5. Sun, P.; Jiang, Y.; Zhang, R.; Xie, E.; Cao, J.; Hu, X.; Kong, T.; Yuan, Z.; Wang, C.; Luo, P. TransTrack Multiple Object Tracking with Transformer. *arXiv* **2020**, arXiv:2012.15460.
6. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022.
7. Aharon, N.; Orfaig, R.; Bobrovsky, B.-Z.J.A. BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *arXiv* **2022**, arXiv:2206.14651.
8. Dorum, E.S.; Kaufmann, T.; Alnaes, D.; Richard, G.; Kolskar, K.K.; Engvig, A.; Sanders, A.M.; Ulrichsen, K.; Ihle-Hansen, H.; Nordvik, J.E.; et al. Functional brain network modeling in sub-acute stroke patients and healthy controls during rest and continuous attentive tracking. *Heliyon* **2020**, *6*, e04854. [CrossRef]
9. Arulmozhi, E.; Bhujel, A.; Moon, B.E.; Kim, H.T. The Application of Cameras in Precision Pig Farming: An Overview for Swine-Keeping Professionals. *Animals* **2021**, *11*, 2343. [CrossRef]
10. Bhujel, A.; Arulmozhi, E.; Moon, B.E.; Kim, H.T. Deep-Learning-Based Automatic Monitoring of Pigs' Physico-Temporal Activities at Different Greenhouse Gas Concentrations. *Animals* **2021**, *11*, 3089. [CrossRef]
11. Pandey, S.; Kalwa, U.; Kong, T.; Guo, B.; Gauger, P.C.; Peters, D.J.; Yoon, K.J. Behavioral Monitoring Tool for Pig Farmers: Ear Tag Sensors, Machine Intelligence, and Technology Adoption Roadmap. *Animals* **2021**, *11*, 2665. [CrossRef]
12. Wang, H.; Zhang, S.; Zhao, S.; Wang, Q.; Li, D.; Zhao, R. Real-time detection and tracking of fish abnormal behavior based on improved YOLOV5 and SiamRPN++. *Comput. Electron. Agric.* **2022**, *192*, 106512. [CrossRef]
13. Ma, J.; Liu, D.; Qin, S.; Jia, G.; Zhang, J.; Xu, Z. An Asymmetric Feature Enhancement Network for Multiple Object Tracking of Unmanned Aerial Vehicle. *Remote Sens.* **2023**, *16*, 70. [CrossRef]
14. Mar, C.C.; Zin, T.T.; Tin, P.; Honkawa, K.; Kobayashi, I.; Horii, Y. Cow detection and tracking system utilizing multi-feature tracking algorithm. *Sci. Rep.* **2023**, *13*, 17423. [CrossRef]
15. Myat Noe, S.; Zin, T.T.; Tin, P.; Kobayashi, I. Comparing State-of-the-Art Deep Learning Algorithms for the Automated Detection and Tracking of Black Cattle. *Sensors* **2023**, *23*, 532. [CrossRef] [PubMed]
16. Xu, Q.; Lin, X.; Cai, M.; Guo, Y.-a.; Zhang, C.; Li, K.; Li, K.; Wang, J.; Cao, D. End-to-End Joint Multi-Object Detection and Tracking for Intelligent Transportation Systems. *Chin. J. Mech. Eng.* **2023**, *36*, 138. [CrossRef]
17. Zhang, K.; Liu, Y.; Mei, F.; Jin, J.; Wang, Y. Boost Correlation Features with 3D-MiIoU-Based Camera-LiDAR Fusion for MODT in Autonomous Driving. *Remote Sens.* **2023**, *15*, 874. [CrossRef]
18. Guo, Q.; Sun, Y.; Orsini, C.; Bolhuis, J.E.; de Vlieg, J.; Bijma, P.; de With, P.H.N. Enhanced camera-based individual pig detection and tracking for smart pig farms. *Comput. Electron. Agric.* **2023**, *211*, 108009. [CrossRef]
19. Zheng, Z.; Qin, L. PrunedYOLO-Tracker: An efficient multi-cows basic behavior recognition and tracking technique. *Comput. Electron. Agric.* **2023**, *213*, 108172. [CrossRef]
20. Lu, J.; Chen, Z.; Li, X.; Fu, Y.; Xiong, X.; Liu, X.; Wang, H. ORP-Byte: A multi-object tracking method of pigs that combines Oriented RepPoints and improved Byte. *Comput. Electron. Agric.* **2024**, *219*, 108782. [CrossRef]
21. Zou, X.; Yin, Z.; Li, Y.; Gong, F.; Bai, Y.; Zhao, Z.; Zhang, W.; Qian, Y.; Xiao, M. Novel multiple object tracking method for yellow feather broilers in a flat breeding chamber based on improved YOLOv3 and deep SORT. *Int. J. Agric. Biol. Eng.* **2023**, *16*, 44–55. [CrossRef]
22. Alameer, A.; Buijs, S.; O'Connell, N.; Dalton, L.; Larsen, M.; Pedersen, L.; Kyriazakis, I. Automated detection and quantification of contact behaviour in pigs using deep learning. *Biosyst. Eng.* **2022**, *224*, 118–130. [CrossRef]
23. Huang, L.; Xu, L.; Wang, Y.; Peng, Y.; Zou, Z.; Huang, P. Efficient Detection Method of Pig-Posture Behavior Based on Multiple Attention Mechanism. *Comput. Intell. Neurosci.* **2022**, *2022*, 1759542. [CrossRef]
24. Huang, Y.; Xiao, D.; Liu, J.; Tan, Z.; Liu, K.; Chen, M. An Improved Pig Counting Algorithm Based on YOLOv5 and DeepSORT Model. *Sensors* **2023**, *23*, 6309. [CrossRef] [PubMed]

25. Tu, S.; Zeng, Q.; Liang, Y.; Liu, X.; Huang, L.; Weng, S.; Huang, Q. Automated Behavior Recognition and Tracking of Group-Housed Pigs with an Improved DeepSORT Method. *Agriculture* **2022**, *12*, 1907. [CrossRef]
26. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
27. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
28. Psota, E.T.; Schmidt, T.; Mote, B.; Pérez, L.C. Long-Term Tracking of Group-Housed Livestock Using Keypoint Detection and MAP Estimation for Individual Animal Identification. *Sensors* **2020**, *20*, 3670. [CrossRef] [PubMed]


scientific reports
nature portfolio

scientific reports

Check for updates

OPEN

The urine formed element instance segmentation based on YOLOv5n

Shuqin Tu[1], Hongxing Liu[1], Liang Mao[2]✉, Chang Tu[3]✉, Wenwei Ye[3], Huiming Yu[4] & Weidian Chen[1]

Accurate and efficient detection and segmentation of the urine formed element plays a vital role in the clinical diagnosis and treatment of many diseases, such as urinary system diseases, kidney diseases, and other diseases. However, artificial microscopy is subjective, and time- consuming. The mainstream detection and instance segmentation algorithms lack adequate accuracy and speed for the urine formed element due to small and dense targets. Therefore, this study proposes a quick one-stage urine formed element instance segmentation model based on YOLOv5n. The approach first employs a backbone architecture to extract features named shallow graphical features and semantic features from urine cells. Next, the neck network combines shallow graphical features with different deep semantic features, obtaining multi-scale, and multi-level features. Finally, according to these multi-level features, the head network of YOLOv5n integrates a small FCN network into the YOLOv5 detector. It obtains the location, classification, and segmentation results of the targets. To validate the superiority of this approach in terms of speed and accuracy, a special urine formed element dataset including 500 images was created. Experimental results show that the YOLOv5n method achieves a Mean Average Precision (mAP) at intersection over the union threshold of 0.5 (mAP50) with 91.8%, and Frames Per Second (FPS) of 63.3. Compared to Mask R-CNN and YOLOv8, its FPS increased by 62.6 and 60.9, respectively, resulting in nearly a hundred-fold speedup, and its mAP50 also increased by 3.6 and 1.4% points in accuracy, respectively. Additionally, the YOLOv5n obtains a superior balance of accuracy and speed in comparisons with SOLOv2, BoxInst, and ConvNeXt V2. This study developed a new automated analysis of urinary particles based on deep learning, and this method is expected to be used for the automated analysis and detection of the urine formed element. The experimental results also demonstrate that YOLOv5n can achieve more accurate and faster instance segmentation of urine formed element, providing technical support for clinical disease diagnosis.

Keywords YOLOv5n, Urine formed element, Fast, Instance segmentation

Urinalysis reveals many issues and diseases in the human body, and urinary sediment examination, as an important part of urine analysis, plays a significant role in the diagnosis, treatment, and cure of kidney diseases[1,2]. Urinary sediment mainly consists of substances such as red blood cells, white blood cells, and crystals[3]. Among these, red blood cell counting helps diagnose hematuria, bladder inflammation, kidney tuberculosis, and other conditions[4,5]. White blood cell counting assists in diagnosing bladder infections, kidney infections, urinary system obstructions, urinary retention, and more[6]. The formation of crystals reflects the oversaturation of substances in urine caused by metabolic and hereditary diseases, as well as drug exposure. Accurate analysis of crystals in urine helps identify the risk of stones and prevent their recurrence[7,8]. Manual counting of red and white blood cells and crystals in urinary sediment examination images is time-consuming, and complex images can lead to subjective errors in judgment. Using computer technology to assist in urine cell instance segmentation is crucial for accurate and fast urinalysis.

Detection and segmentation of cells in microscopic images have received extensive research attention. However, most existing methods focus only on cell detection or segmentation. For example, Vu et al.[9] used a multiscale deep residual aggregation network to fuse the watershed algorithm for cell nucleus segmentation. Kutlu et al.[10] employed the regional convolutional neural networks for white blood cell detection and classification. Wang et al.[11] used an automatic hierarchical patch-based deep learning framework to detect and classify bone marrow cells. Zheng et al.[12] utilized saliency detection and CenterNet for white blood cell detection. While these

[1]College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China. [2]Institute of Applied Artificial Intelligence of the Guangdong-Hong Kong-Macao Greater Bay Area, Shenzhen Polytechnic University, Shenzhen 518055, China. [3]Department of Cardiovascular Medicine, Dongguan Songshan Lake Central Hospital, Dongguan 523321, China. [4]Tianjin Hualing Medical Technology Company Limited, Tianjin 301700, China. ✉email: maoliangscau@szpu.edu.cn; 330012338@qq.com

approaches have shown promise in individual detection and segmentation tasks, they face challenges in precisely distinguishing different cells. Single detection lacks cell contour information, and single segmentation lacks information about different objects of the same class. However, instance segmentation provides individual-level segmentation and recognition, allowing quantitative analysis of cell features, making it increasingly important in cell segmentation.

Instance segmentation is a task that combines object detection and segmentation. Instance segmentation methods can be divided into two categories: two-stage and one-stage instance segmentation methods. Two-stage instance segmentation methods have dominated cell instance segmentation[13]. For example, Yi et al.[14] introduced an instance segmentation method that combined SSD and U-Net with attention mechanisms for neurons, achieving an AP@0.5 of 89.06. Lin, Norouzi[15] developed a method for cell segmentation using Mask R-CNN and shape-aware loss, achieving IoU scores of 91.90% and 94.90% on DIC-C2DH-HeLa and PhC-C2DH-U373 datasets, respectively. Atıcı, Koçer[16]employed Mask R-CNN for instance segmentation of blood cells, achieving a mAP50 of 91.39%. While these methods achieve high accuracy, they require extensive post-processing and are not suitable for tasks with high time requirements. To overcome these challenges, some researchers have recently explored the use of one-stage instance segmentation methods for cell instance segmentation. For example, Luo et al.[17] integrated the Ghost module and enhanced the FPN structure based on YOLACT, introducing a model called YOLACT-CIS, which achieved an AP of 87.41% on the blood cell dataset. Wang et al.[18] proposed an improved BlendMask method for cell nuclei, attaining an AP50 of 78.3 and an FPS of 7.50 on the DSB2018 dataset.

Building on this foundation, this paper presents a one-stage quick urine cell instance segmentation method based on the YOLOv5 detector, which integrates a small FCN network into the YOLOv5 detector and achieves fast and accurate results on urine cell datasets. Compared to traditional methods, this approach achieves remarkable speed while maintaining accuracy. The contributions of this paper are as follows:

(1) We propose the YOLOv5 instance segmentation and create a urine cell dataset containing 500 images, which was labeled to make it publicly available.

(2) We compare different variants of the YOLOv5 instance segmentation method and select the most suitable instance segmentation algorithm for the urine cell dataset.

(3) We conduct comparison experiments of YOLOv5n with YOLOv8, MS R-CNN, SOLOv2, BoxInst, and ConvNeXt V2 methods.

(4) We explore the impact of incorporating lightweight methods and attention mechanisms for small objects on the proposed method.

Related works

Two-stage instance segmentation method

Two-stage instance segmentation methods have become mainstream due to their high accuracy. Since the advent of Mask R-CNN, many researchers have conducted experiments based on it. In the field of cell instance segmentation, Ren et al.[19] addressed the issues of significant variations in image signal intensity and the blurriness of adjacent cell boundary information by modifying the anchor sizes in Mask R-CNN and introducing a multi-task U-Net network. They also took the union of masks with maximum overlap and achieved an IoU of 45.02% on a tumor cell dataset. Loh et al.[20] tackled the limitations of traditional microscopes in malaria infection screening and rapid diagnostic tests by using Mask R-CNN for automatic red blood cell segmentation, achieving a speedup of 15 times compared to manual counting, with an accuracy of 82.00%. Qiu et al.[21] addressed the challenges posed by the variable shapes and sizes of multiple myeloma (MM) cells, the proximity of the cytoplasm to the background, and the overlap between the cells of interest and normal cells. They fused a feature selection pyramid structure based on Cascade Mask R-CNN, obtaining a dice similarity coefficient of 85.62 ± 13.95 on bone marrow tumor cell data. Mitate et al.[22] tackled the challenge of accurately detecting nuclei in oral cytology images by using SWM and Mask R-CNN for instance segmentation of oral cell nuclei, achieving the highest detection rate of 93.14% and an error rate of 2.70%. Additionally, Bai et al.[23] addressed the challenges of the significant time and resource demands of mainstream instance segmentation techniques by combining YOLO and U-Net for cell instance segmentation, achieving an MIoU of 88.70%. Despite the good results of two-stage instance segmentation methods on many datasets, they are challenged by complex post-processing requirements and significant time overhead, making them less suitable for tasks with strict real-time requirements.

One-stage instance segmentation method

One-stage instance segmentation methods are gaining momentum due to their efficiency. In the field of cell instance segmentation, Prangemeier et al.[24] addressed the limitations in speed and efficiency of existing methods by combining Transformer and CNN to propose a method dedicated to fast cell instance segmentation called Cell-DETR. Cell-DETR has two variants, Cell-DETR A and Cell-DETR B, achieved Jaccard coefficients of 0.82 and 0.84, respectively. Priego-Torres et al.[25] tackled the issue of heterogeneity in breast tumors leading to uneven distribution and diverse characteristics of tumor cells by using SOLOv2 for tumor cell instance segmentation, achieving a mAP50 of 68.00%. Zhao et al.[26] addressed the challenges posed by variations in shape, size, and color due to different subtypes of white blood cells and staining techniques by proposing a multiscale and multi-staining WBC instance segmentation network (MISS-WISN) and achieved a mAP50 of 79.96 on datasets with five white blood cell subtypes. Liu et al.[27] tackled the difficulty of efficient cell segmentation in complex cell image analysis systems by proposing an efficient end-to-end cell segmentation algorithm, ECS-Net, which achieved an AP50 of 82.3 on the LIVECell dataset. Although one-stage instance segmentation methods enable real-time segmentation, they may have slightly lower accuracy compared to two-stage methods.

Method

The overall structure of YOLOv5n

This study proposes an instance segmentation approach based on the YOLOv5n detector and FCN segmentation network, and its overall structure is illustrated in Fig. 1, consisting of four components: the input images, the backbone network, the neck network, and the head network. Firstly, the backbone network obtains different types of feature information including shallow graphical features and semantic features from the input images. Then, the neck network combines shallow graphical features with deeper semantic features to produce multi-scale and multi-level features. Finally, the head network includes detection and FCN segmentation modules, generating results for target localization, classification, and segmentation masks.

Detailed Architecture of YOLOv5n instance segmentation approach

Backbone network

The primary function of the backbone network is to extract features from input images while gradually reducing the feature maps. The backbone network consists of 10 layers, 5 standard Convolution Module (Conv) layers, 4 cross-stage partial bottlenecks with 3 convolutions (C3) layers, and a terminal Spatial Pyramid Pooling with Features (SPPF) layer. The Conv layers consist of a convolutional layer, a normalization layer, and a SiLU activation function, which together extract and organize features. The C3 layers include 3 Conv modules and 1 bottleneck module, serving as crucial feature extraction modules. The Conv and C3 layers reduce feature dimensionality to help the convolutional kernels better understand feature information and then expand dimensionality to extract more detailed features. SPPF layer includes 2 Conv, 3 MaxPool2d, and 1 Concat layers, and its primary purpose is to integrate features at multiple scales.

Neck network

The neck network structure is illustrated in layers 10–23 of Fig. 1. Layers 10–17 of the neck network use up-sampling operation to increase the scale of feature maps, enabling feature fusion from bottom to top. Meanwhile, layers 18–23 of the neck network use the down-sampling operation to obtain feature maps at different scales, facilitating a better fusion of shallow graphical features with deep semantic features. Eventually, the neck network forms multi-scale and multi-level features, by acquiring relatively shallow features from the backbone network and fusing them with deep semantic features.

Head network

The head network consists of the object detection head and the FCN segmentation head. The object detection head is responsible for target localization and classification tasks. It comprises three object detection heads with

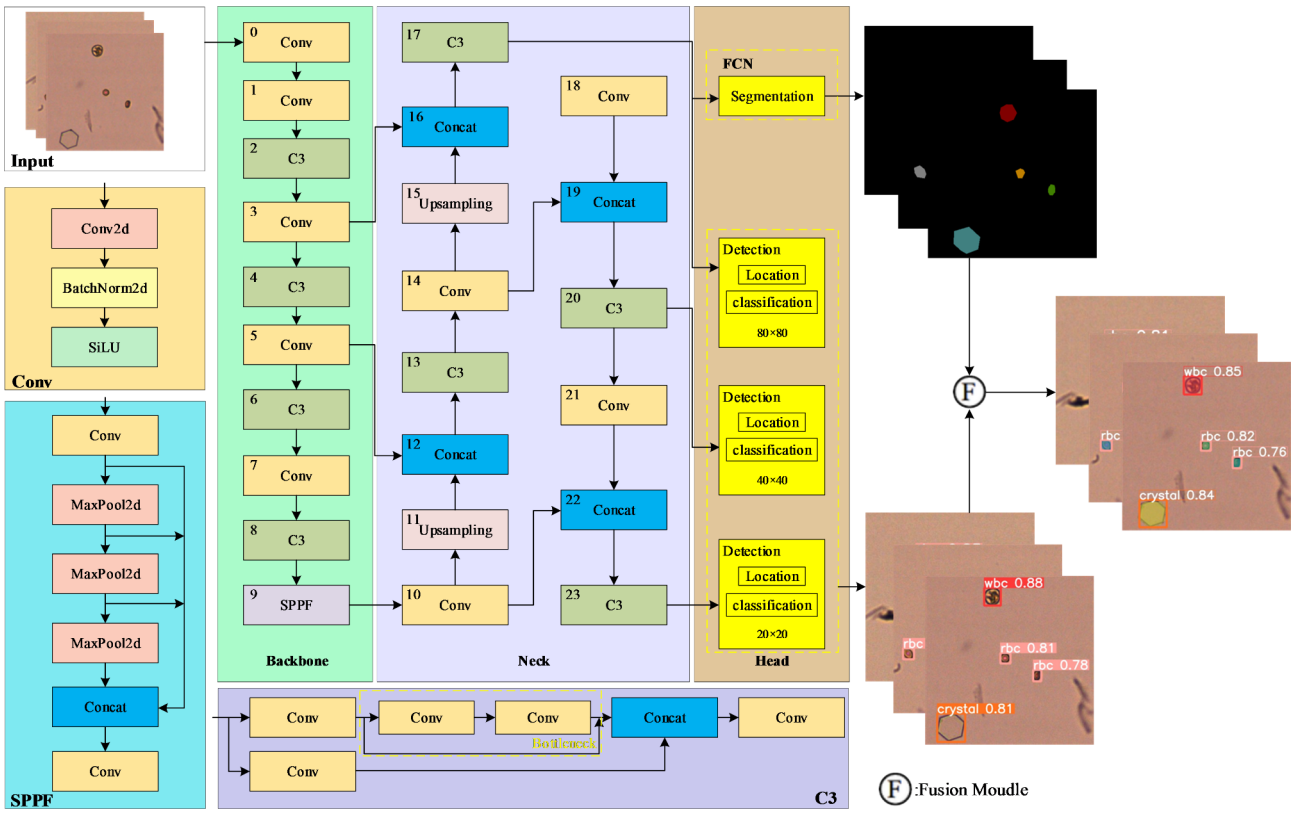


Fig. 1. The overall architecture details of YOLOv5n instance segmentation.

three different scales: 20×20, 40×40, and 80×80 (as shown in Fig. 1). By combining these three heads, more accurate object detection information can be obtained.

The structure of the FCN, as shown in Fig. 2, includes a small fully convolutional neural network composed of three Conv modules and an up-sampling module. The output (C3) from the neck network is first taken as input, and deep feature information is extracted using the first Conv modules with 3×3 kernel size. Then, the up-sampling module with 2× size enlarges the feature map's size for the extracted depth feature information according to the input image's resolution. Finally, two Conv modules with 3×3 and 1×1 kernel sizes are used to extract features from the upsampled feature map to achieve accurate instance segmentation results.

Loss function

The overall loss function of YOLOv5n consists of four components and is defined as in Eq. (1).

$$L_{\text{total}} = L_{box} + L_{cls} + L_{obj} + L_{seg} \tag{1}$$

Where L_{box} is the bounding box regression loss, L_{cis} is the classification loss, L_{obj} is the object detection loss, and L_{seg} is the segmentation loss. L_{cis} and L_{obj} both utilize BCEWithLogitsLoss, shown in Eq. (2).

$$BCEWithLogitsLoss\left(z, y\right) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \cdot \log\left(\sigma\left(z_i\right)\right) + \left(1 - y_i\right) \cdot \log\left(1 - \sigma\left(z_i\right)\right)\right) \tag{2}$$

Where N is the sample size, Z_i represents the raw score, y_i denotes the true label, and σ(.) represents the sigmoid function. The objective of this loss function is to minimize the difference between the model's predicted probabilities and the true labels for training the model to better classify the data.

The bounding box regression loss is defined by Eq. (3).

$$L_{box} = 1 - IoU\left(b, b^{gt}\right) + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha\nu \tag{3}$$

Where $IoU\left(b, b^{gt}\right)$ variables are consistent with the description in[28]. b represents the predicted box coordinates, b^{gt} represents the true box coordinates. $\rho^2\left(b, b^{gt}\right)$ represents the Euclidean distance between the predicted and true box centers, and c represents the diagonal distance of the smallest enclosing box that can contain both the predicted and true boxes. Additionally, α serves as a weight coefficient, and ν is a parameter used to measure aspect ratio consistency. The α and ν are calculated as in Eqs. (4) and (5).

$$\alpha = \frac{\nu}{\left(1 - IoU\left(b, b^{gt}\right)\right) + \nu} \tag{4}$$

$$\nu = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{\boldsymbol{w}_p}{h_p}\right) \tag{5}$$

Where w_p and h_p respectively represent the width and height of the predicted box, while w^{gt} and h^{gt} represent the width and height of the true box.

The segmentation loss (L_{seg}) is defined as in Eqs. (6) and (7).

$$L_{seg} = -\frac{1}{N}\sum_{i=1}^{N}\frac{1}{A_i}\left(m_i^t \log\left(m_i^p\right) + \left(1 - m_i^t\right)\log\left(1 - m_i^p\right)\right)\sum_{j=1}^{C_i} Crop\left(b_j\right) \tag{6}$$

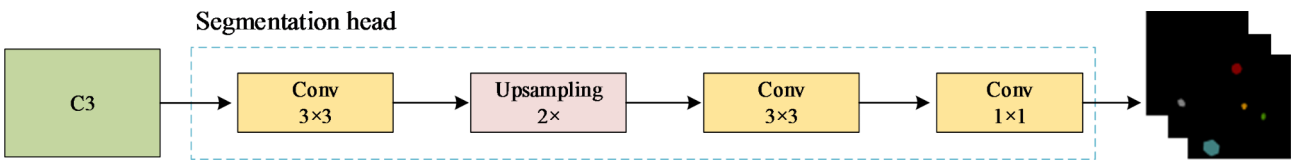


Fig. 2. FCN Head Structure of YOLOv5n.

$$Crop\left(b_j\right) = \begin{cases} 1if b_i \in R_{\mathrm{i}} \\ 0if b_i \notin R_i \end{cases} \tag{7}$$

Where is the number of masks, A_i is the area of the predicted bounding box, m_i^t represents the truth mask, m_i^p is the predicted mask, C_i is the number of pixels contained within the mask, b_j is the pixel coordinates, and R_i represents the coordinate region corresponding to the predicted bounding box.

Statement

All methods were carried out in accordance with relevant guidelines and regulations.

The experimental protocols involving human participants were approved by Tianjin Hualing Medical Technology Company Limited .

Code Availability

The code used in this study is publicly available and can be accessed via GitHub at the following link: https://github.com/superman-liu/YOLOv5-cell. This repository includes all the necessary scripts, training configurations, and setup instructions for replicating the experiments and utilizing the model for similar tasks.

Experiment

Dataset

The dataset was provided by Tianjin Hualing Medical Technology Company Limited, and all data usage has been granted patient consent. The dataset is available for access at https://terabox.com/s/1afaHP7g_Yd1Dn0qLNM8AhQ. The dataset consists of 500 images with a resolution of 2048*1536. The images are obtained using the RZ1100 urine analyzer as shown in Fig. 3 (a) and the originate image is shown in Fig. 3 (b). We randomly select 400 images for training and 100 images for testing. LABELME software is selected to annotate the instances of the images, the version of LABELME is v5.3.1. The software can be downloaded from https://github.com/labelmeai/labelme. The labeling results from LabelMe are shown in Fig. 3(c) and Fig. 3(d). The accuracy of the annotation is validated by three experts.

There are three types of instances in the dataset, including 7252 red blood cells (RBCs), 1441 white blood cells (WBCs), and 202 crystals. As shown in Fig. 3 (b), the object surrounded by the brown box is a crystal, the object surrounded by the red box is a WBC, and the object surrounded by the green box is an RBC. The training set includes 6952 instances, including 5725 RBCs, 1102 WBCs, and 165 crystals. The test set includes 1903 instances, including 1527 RBCs, 339 WBCs, and 37 crystals. We transfer the weights of a backbone network pre-trained on the ImageNet dataset to our model to improve task performance on a smaller dataset.

Evaluation metrics and implementation details

For the evaluation of instance segmentation algorithms, performance metrics are required to assess the algorithm models. In this study, precision (P), recall (R), F1 score, mAP50, mAP50-95, and FPS are used to measure the performance of instance segmentation algorithms. Precision and recall are divided into precision (P(box)) and recall (R(box)) for object detection, and precision (P(seg)) and recall (R(seg)) for segmentation. For different algorithms, in the comparative experiments of the Mask R-CNN series, P(box), R(box), and F1(box) score are adopted to evaluate the quality of detection, while FPS is used to assess the inference speed of the algorithm. Segmentation quality is evaluated using mAP50 and mAP50-95. In the ablation experiments of the YOLOv5n series, detection quality is measured by F1(box) score, while the algorithm's inference speed is assessed using the FPS metric. Segmentation quality, on the other hand, is evaluated using P(seg), R(seg), F1(seg), mAP50, and mAP50-95.

The experimental configuration environment is shown in Table 1. The experiments utilize stochastic gradient descent (SGD) as the optimization algorithm. The initial learning rate is set to 0.01, SGD momentum is set to 0.937, and the weight decay is set to 0.0005. Considering the GPU memory limitations during training, the input image size is set to 1280, the batch size is set to 4, and a total of 200 epochs are executed.Additionally, the IoU training threshold was set to 0.20, the box loss gain was set to 0.05, the class loss gain was set to 0.5, the object loss gain was set to 1.0, and the anchor-multiple threshold was set to 4.0.

The results of YOLOv5n instance segmentation

In this section, we will present the experimental results for three types of urine cell instances on YOLOv5n. Table 2 presents the results of instance segmentation for the three types of urine formed elements. Overall, the segmentation performance is slightly better than the detection performance. Specifically, the segmentation accuracy for RBC is 93.1%, with a segmentation recall of 98.0%, and an mAP50 of 98.3%. Although RBCs are the smallest among the three instances, their morphology and size are relatively stable, contributing to high segmentation performance.

WBC also achieves an overall accuracy and recall above 90.0%, with mAP50-95 reaching the highest at 65.6%. While WBCs are stable in shape and size, their internal composition varies, which leads to relatively lower accuracy compared to RBCs. In contrast, among the three types of urine formed elements, the accuracy for crystal achieves the lowest value, with a segmentation accuracy of 74.7% and a mAP50 of 79.5%. This is because crystals are very scarce in the entire dataset and often overlap with other objects.

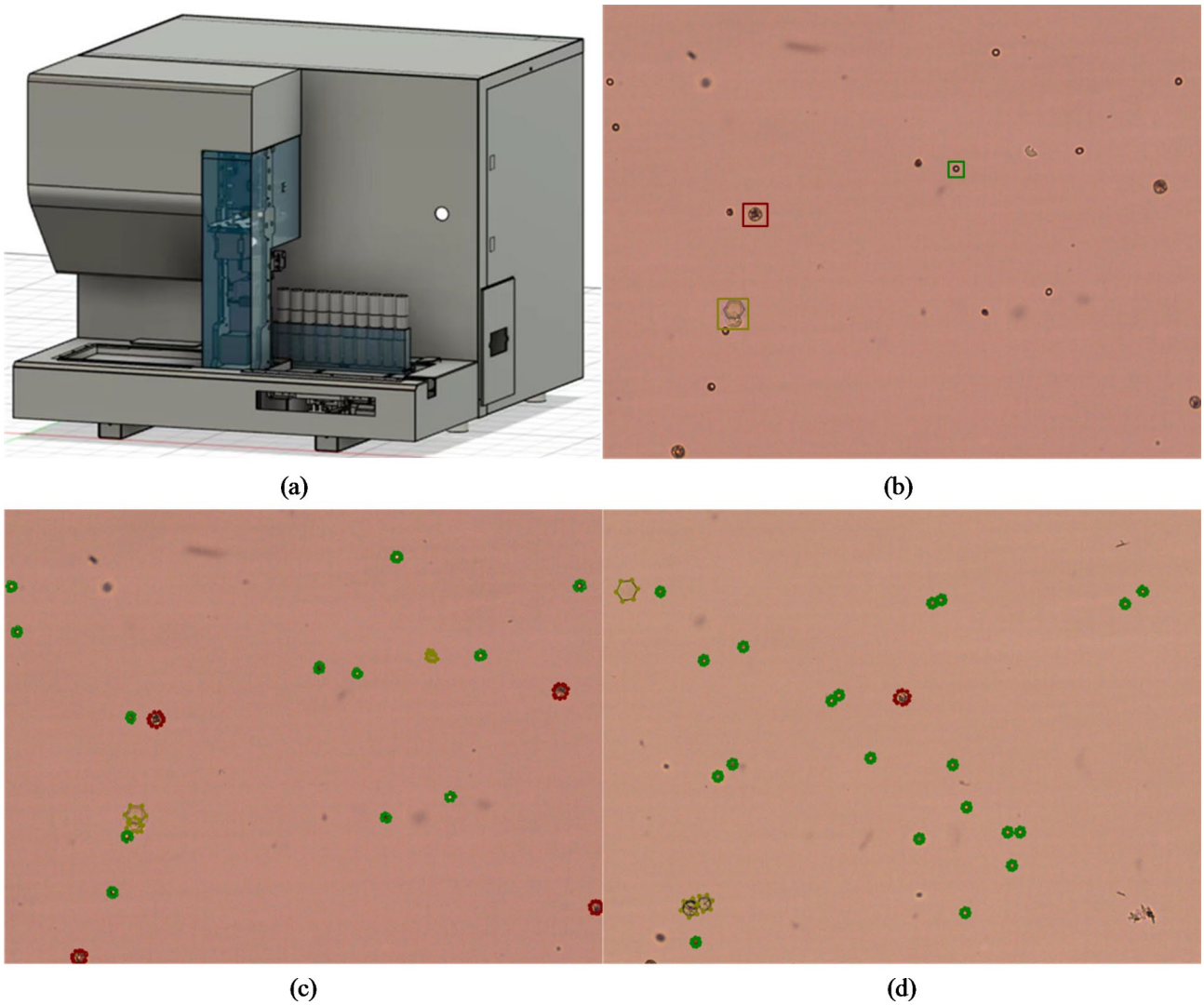


Fig. 3. Urine analyzer and the obtained urine images. (**a**) The RZ1100 urine analyzer. (**b**) The originate image. (**c**) and (**d**) The labeling results from LabelMe.

Configuration	Parameter
CPU	Intel Core i5-11400 F
GPU	Nvidia GeForce RTX 3060
Operating system	Windows10
Accelerated environment	CUDA 11.3、CUDNN 8.9.0
Development environment	Pycharm2022

Table 1. Experimental configuration.

class	*P*(box)	*R*(box)	*P*(seg)	*R*(seg)	mAP50	mAP50-95
RBC	0.929	0.980	0.931	0.980	0.983	0.539
WBC	0.912	0.950	0.918	0.953	0.974	0.656
Crystal	0.695	0.757	0.747	0.811	0.795	0.473

Table 2. The experimental results for three types of urine cell instances on YOLOv5n.

Figure 4 displays the confusion matrix results of YOLOv5n instance segmentation on the test set. From Fig. 4, it can be observed that the classification performance for RBC is the best, achieving the highest value of 0.980, meanwhile, it tends to be confused with the background, due to the presence of impurities in the background that resemble the morphology of RBCs. The classification performance for WBC is the second, with a value of 0.950, and it is less likely to be confused with other categories. However, the classification performance for crystal is the worst, with a value of 0.730, but it is not easily confused with the background.

Figure 5 illustrates the segmentation performance of the YOLOv5n instance segmentation approach, including the F1-Confidence curve, Precision-Confidence curve, Precision-Recall(P-R) curve, and Recall-Confidence curve. From the F1-Confidence curve (as shown in Fig. 5(a)), it can be observed that the F1 score reaches its peak at 0.890 when the confidence is 0.275, remains relatively stable between 0.275 and 0.800, and gradually decreases after surpassing 0.800, with the crystal curve showing relatively poor results. From the Precision-Confidence curve (as shown in Fig. 5(b)), we find that the precision values increase with the rise of confidence value, reaching its highest value of 1.00 at the confidence with 0.887. Moreover, the P-R curve (as shown in Fig. 5(c)) exhibits a high level of performance, with an overall mAP50 value of 0.918. From the Recall-Confidence curve (as shown in Fig. 5(d)), Recall decreases as the confidence increases, achieving the highest value of 0.970 at confidence 0.00 and a result of 0.915 at confidence 0.50. From these images, we observe that YOLOv5n has achieved satisfactory results in the instance segmentation of RBCs and WBCs.

Figure 6 displays the visualization results of the YOLOv5n instance segmentation approach on the test set. Among Fig. 6(a) shows the input image, Fig. 6(a) presents the Ground Truth (GT) of the input image, and Fig. 6(c) shows the segmentation results using the YOLOv5n instance segmentation approach. From Fig. 6(c), it can be observed that YOLOv5n can successfully segment almost all objects, with confidence scores generally above 0.80. Whether the cells are individual or adjacent to each other, YOLOv5n instance segmentation approach can accurately segment them.

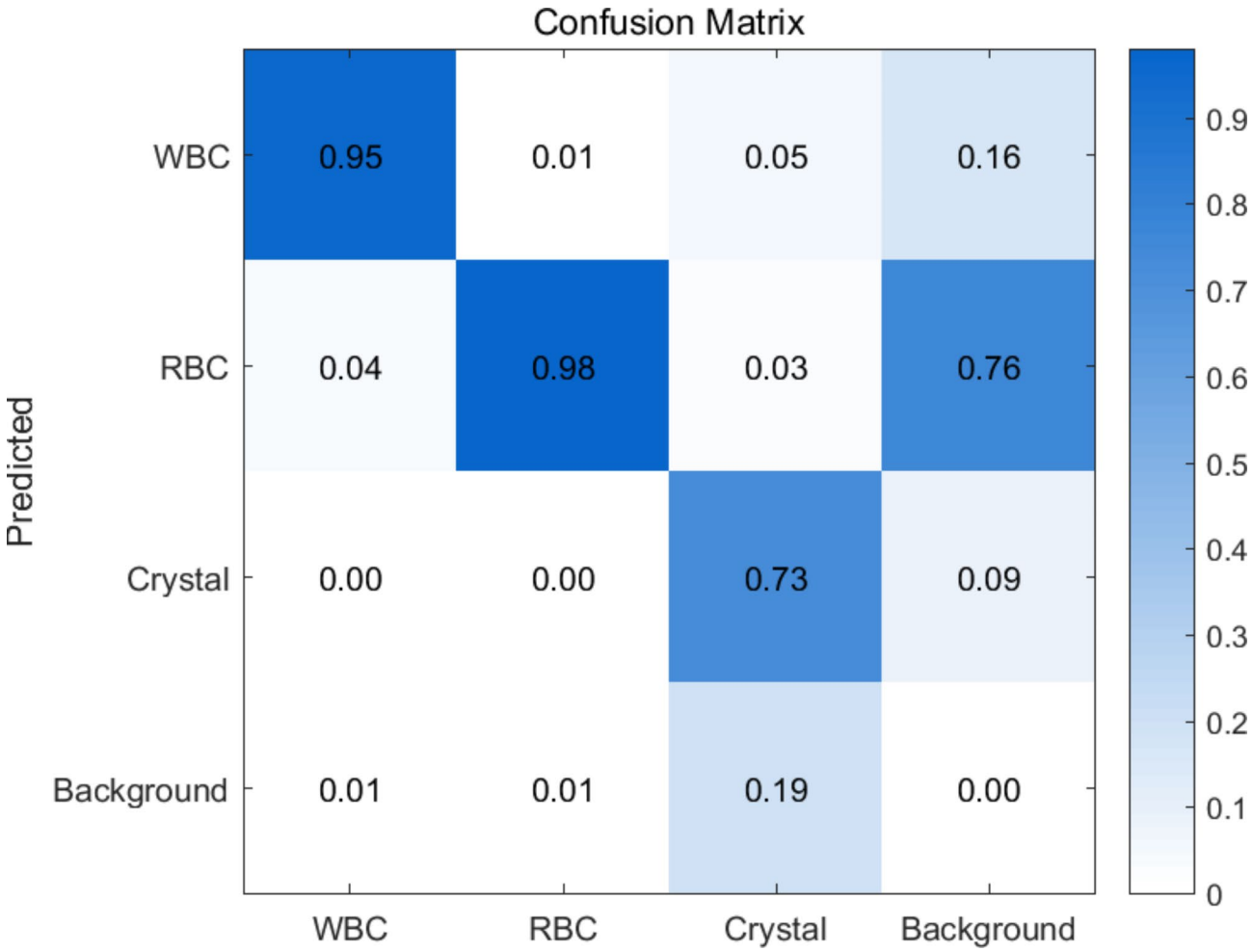


Fig. 4. Confusion Matrix.

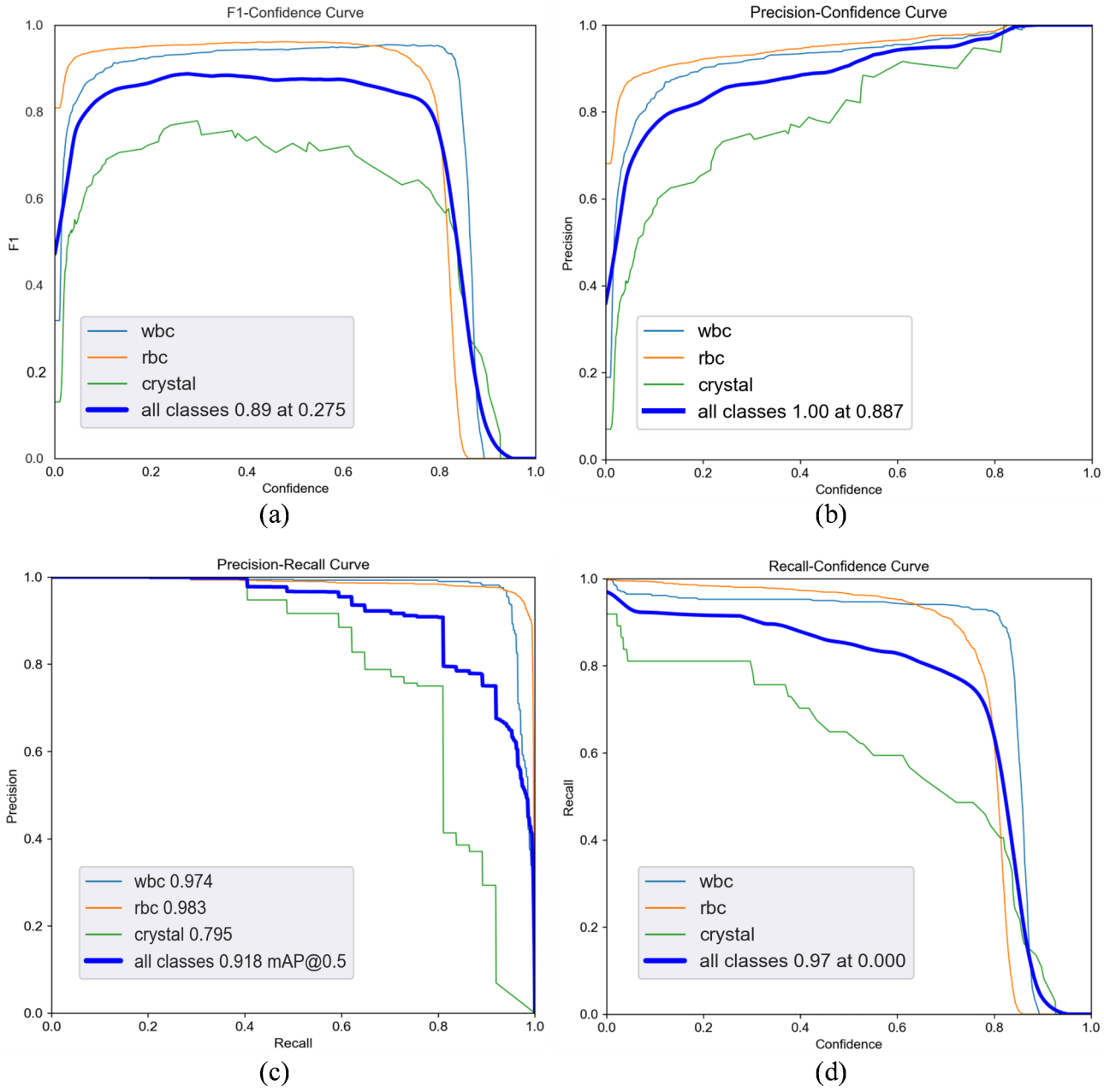


Fig. 5. The Mask curves of YOLOv5n. (**a**) F1-Confidence Curve. (**b**) Precision-Confidence Curve. (**c**) Precision-Recall Curve. (**d**) Recall-Confidence Curve.

Comparative analysis

Comparison of results between different variants of YOLOv5

To investigate the impact of different depths and parameter sizes of the network on experimental results, a comparative experiment was conducted on the five variants of YOLOv5, including YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The comparative experimental results are shown in Table 3. From Table 3, it can be observed that YOLOv5n achieved results of 84.5% for P(box), 89.5% for R(box), 86.9% for F1(box), 86.5% for P(seg), 91.5% for R(seg), 88.8% for F1(seg), 91.8% for mAP50, 55.6% for m50-95, and 63.3 for FPS. YOLOv5n demonstrated good performance in terms of R(box), R(seg), F1(seg), mAP50, mAP50-95, and FPS. Compared to other variants, YOLOv5n showed slightly inferior detection performance, and it excelled in segmentation performance among all variants. The F1(box) score of YOLOv5n decreased by 1.3%, 0.3%, 2.6%, and 2.8% compared to YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, respectively. However, in terms of F1(seg) score, compared to YOLOv5s, YOLOv5m, and YOLOv5l, YOLOv5n improved by 1.1%, 3.3%, and 1.3%, respectively, and only 0.3% lower than with YOLOv5x model. YOLOv5n outperformed the other four variants in mAP50 and mAP50-95 by 2.1%, 4.4%, 2.9%, 1.5%, 5.5%, 4.0%, 5.0%, and 3.8%, respectively. In terms of speed, YOLOv5n significantly outperformed the other four variants, achieving an FPS of 63.3, which is 27.1,

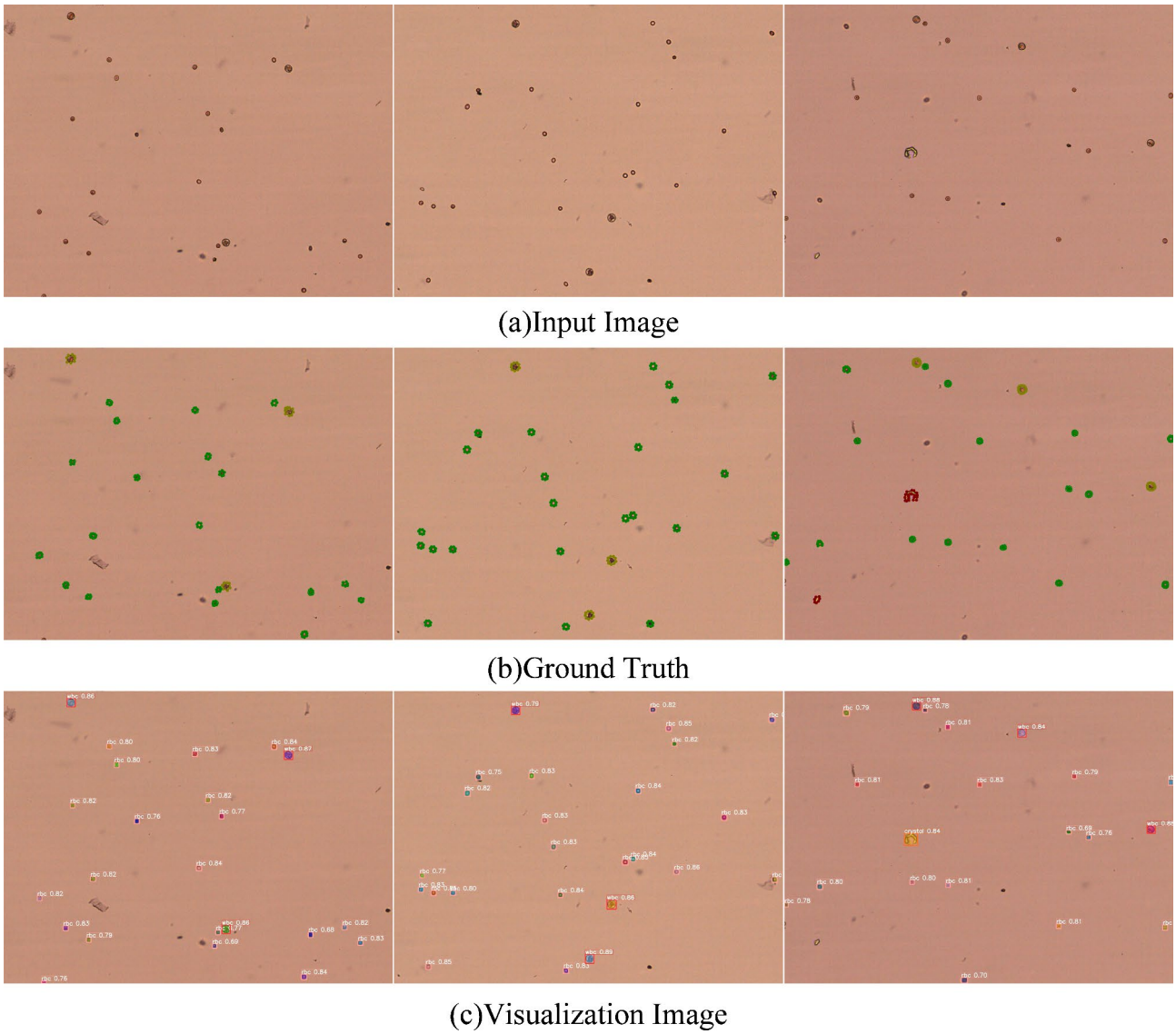


Fig. 6. Visualization results of Yolov5n.

Method	*P*(box)	*R*(box)	F1(box)	*P*(seg)	*R*(seg)	F1(seg)	mAP50	mAP50-95	FPS
YOLOv5n	84.5	89.5	86.9	86.5	**91.5**	88.8	**91.8**	**55.6**	**63.3**
YOLOv5s	**91.5**	85.2	88.2	**91.0**	84.6	87.7	89.7	50.1	36.2
YOLOv5m	86.7	87.7	87.2	84.8	85.9	85.5	87.4	51.6	19.0
YOLOv5l	91.0	88.0	89.5	89.0	86.0	87.5	88.9	50.6	12.7
YOLOv5x	88.9	**90.6**	**89.7**	88.1	89.9	**89.1**	90.3	51.8	6.9

Table 3. The experimental results of different variants of YOLOv5.

44.3, 50.6, and 56.4 higher than the other four variants, respectively. The YOLOv5n algorithm is superior to other algorithms in the comprehensive performance of accuracy and speed.

The instance segmentation P-R curves for the five variants of YOLOv5 are shown in Fig. 7. YOLOv5n achieves the best results, with the P-R curves corresponding to RBCs, WBCs, and crystals obtaining the highest performance among all variants. Specifically, the P-R curve area for RBCs and WBCs ranks first among all variants with close-to-1 results of 0.974 and 0.983, respectively. The P-R curve area for crystals, with a result of 0.795, surpasses YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x by margins of 0.074, 0.030, 0.032, and 0.022, respectively.

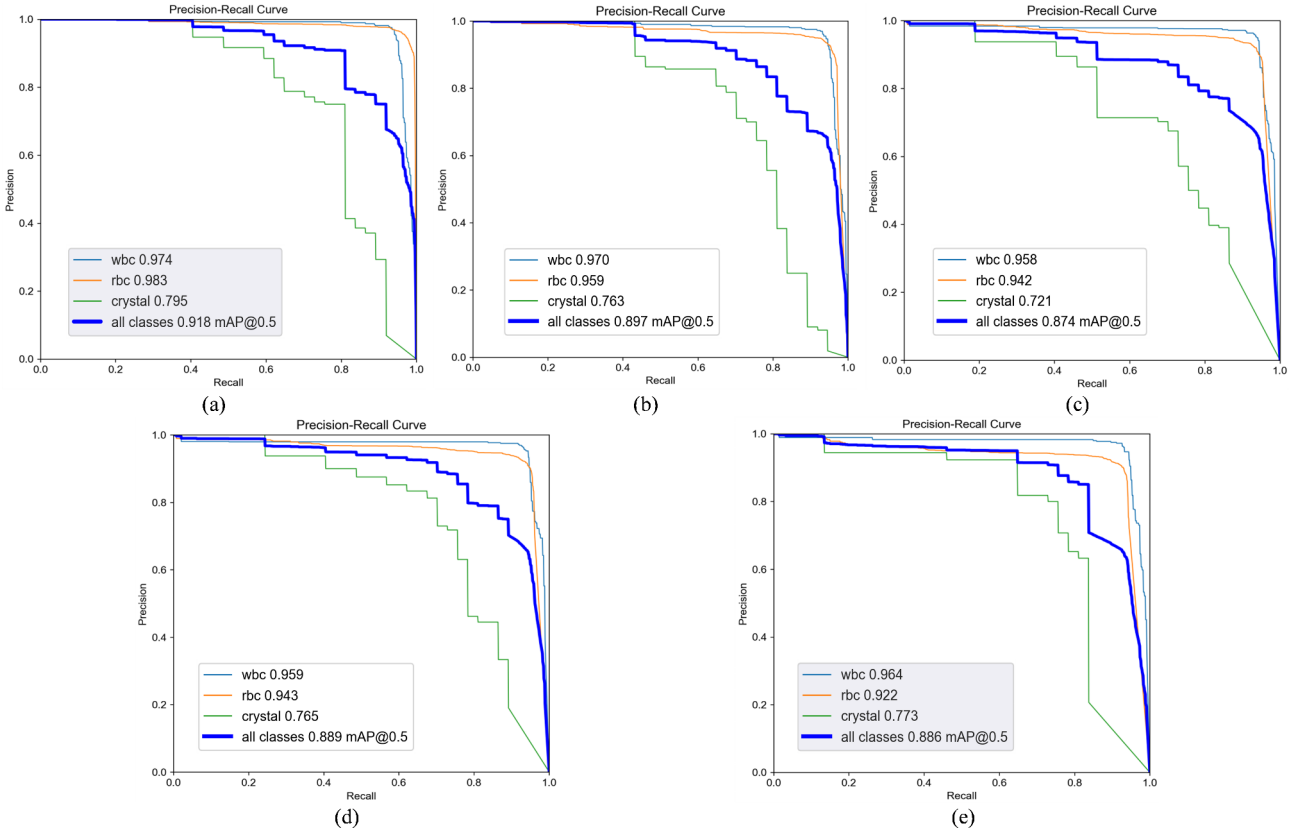


Fig. 7. P-R Curves for Different Variants of YOLOv5. (**a**) Results for YOLOv5n. (**b**) Results for YOLOv5s. (**c**) Results for YOLOv5m. (**d**) Results for YOLOv5l. (**e**) Results for YOLOv5x.

Taking all factors into consideration, based on the experimental results, YOLOv5n is a preferable choice that strikes a balance between accuracy and performance, while YOLOv5s is suitable for applications with lower efficiency requirements. In the field of medical applications, efficiency is of paramount importance. Considering the trade-off between speed and segmentation quality, we ultimately choose the YOLOv5n model for urine sediment segmentation application.

The comparative results of the five variants of YOLOv5 for crystal detection are shown in Fig. 8. The target in the red boxes represents the objects of interest (as shown in Fig. 8(a)-(b) and Fig. 8(h)-(i)), and the white color indicates confidence scores of the classifications. In Fig. 8(b), the object within the red boxe corresponds to the Ground Truth of a crystal. By comparing Fig. 8(c)-(g) with Fig. 8(b), it is observed that YOLOv5n and YOLOv5x exhibit excellent performance in crystal sementation. Notably, YOLOv5n is particulary adept at handling edge details, and although its classification confidence is slightly lower than that of YOLOv5x, it remains above 0.7, indicating a high level of accuracy. Furthermore, the comparison between Fig. 8(j)-(n) and Fig. 8(i) demonstrates that YOLOv5n and YOLOv5x continue to maintain high-quality performance in crystal segmentation, with their classification confidence scores remaining at an acceptable high level. Figure 9 illustrates the capability of YOLOv5n in detecting and segmenting red blood cells (RBCs) and white blood cells (WBCs). In terms of both classification confidence and segmentation quality, YOLOv5n shows outstanding performance.

In conclusion, YOLOv5n not only excels in the detection and segmentation of crystals but also demonstrates high accuracy and reliability in the detection of RBCs and WBCs. Therefore, YOLOv5n is highly suitable for automated detection and segmentation tasks of urinary formed elements.

Comparison of results between YOLOv5n and other methods

The comparison results of YOLOv5n with Mask R-CNN[29], Mask Scoring R-CNN (MS R-CNN)[30], MaskDis R-CNN method[31], and the latest YOLO[32] series method named YOLOv8 are presented in Table 4. The Mask R-CNN methods include two backbone networks: Resnet-50-FPN and Resnet-101-FPN. The YOLOv5n backbone network adopts CSPDarknet53, and the YOLOv8 backbone network adopts Darknet53. From Table 4, it can be observed that YOLOv5n achieves the highest performance in terms of segmentation quality (represented by mAP50 and mAP50-95) and speed (measured by FPS) with scores of 91.8%, 55.6%, and 63.3% respectively. While F1(box) value of YOLOv5n is 86.8%, which indicates that yolov5n has achieved moderate performance in detection. In terms of mAP50 and mAP50-95, YOLOv5n outperforms Mask R-CNN, MS R-CNN, and MaskDis R-CNN by 3.6%, 2.9%, 0.8% and 3.6%, 2.3%, 3.3% respectively, and outperforms YOLOv8 by 1.4% and 4.5%. In terms of speed, the FPS of YOLOv5n is 63.3, which shows that YOLOv5n significantly outperforms the other 7 methods. YOLOv8 achieves the best detection performance, with F1 of 90.4, while its segmentation quality is

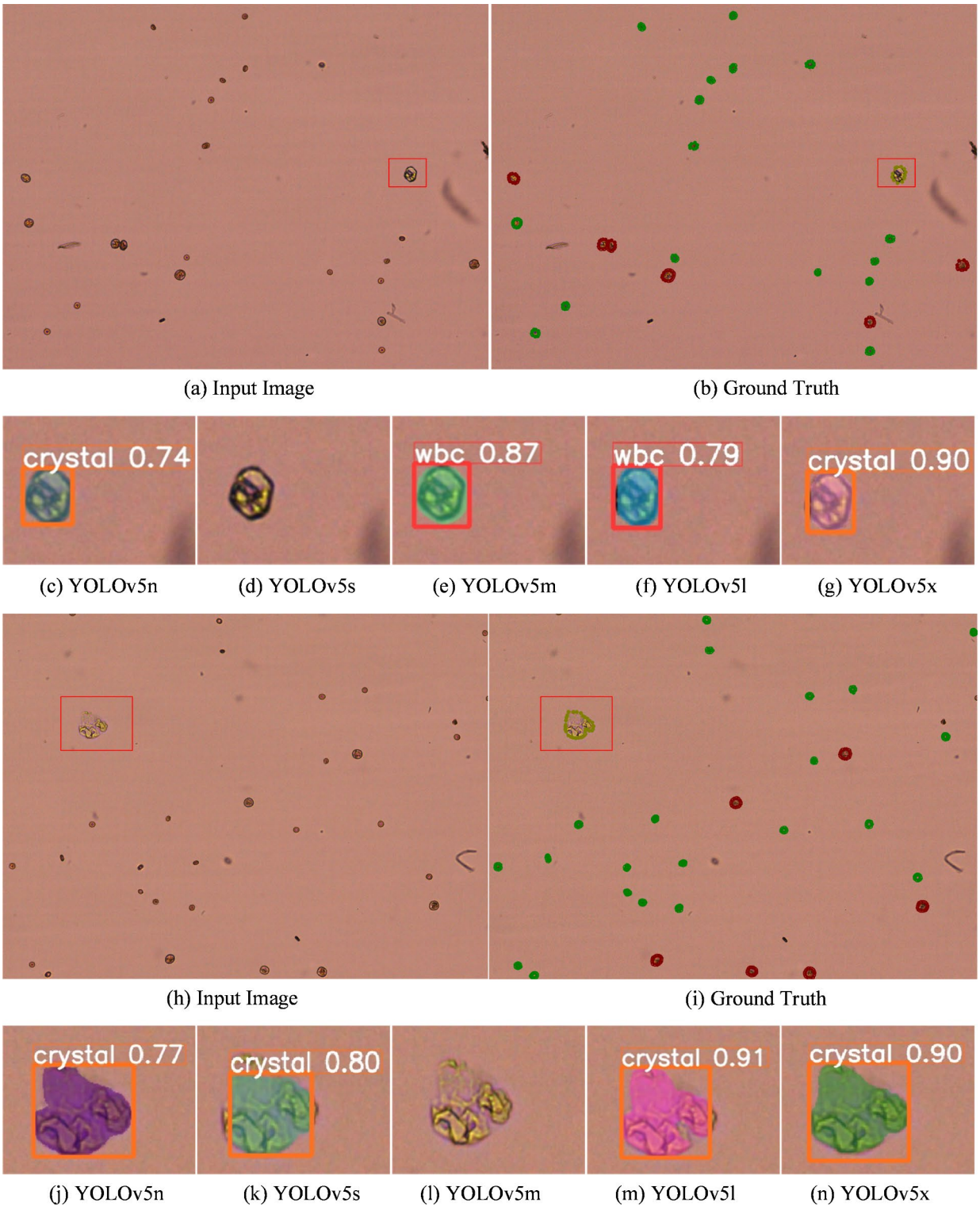


Fig. 8. Visual results of crystal detection for the five variants of YOLOv5.

slightly lower than that of YOLOv5n, and its speed is much slower than YOLOv5n. Overall, YOLOv5n algorithm is superior to other algorithms in the comprehensive performance of speed and segmentation accuracy.

Furthermore, the urine formed element instance segmentation performance of YOLOv5n is compared with Mask R-CNN(Resnet-101-FPN), MS R-CNN (Resnet-101-FPN), and YOLOv8n method. Exemplar segmentations of these methods are shown in Fig. 10. It can be observed that Mask R-CNN, MS R-CNN, YOLOv5n, and YOLOv8 have achieved quite good segmentation results for WBC and RBC. But Mask R-CNN

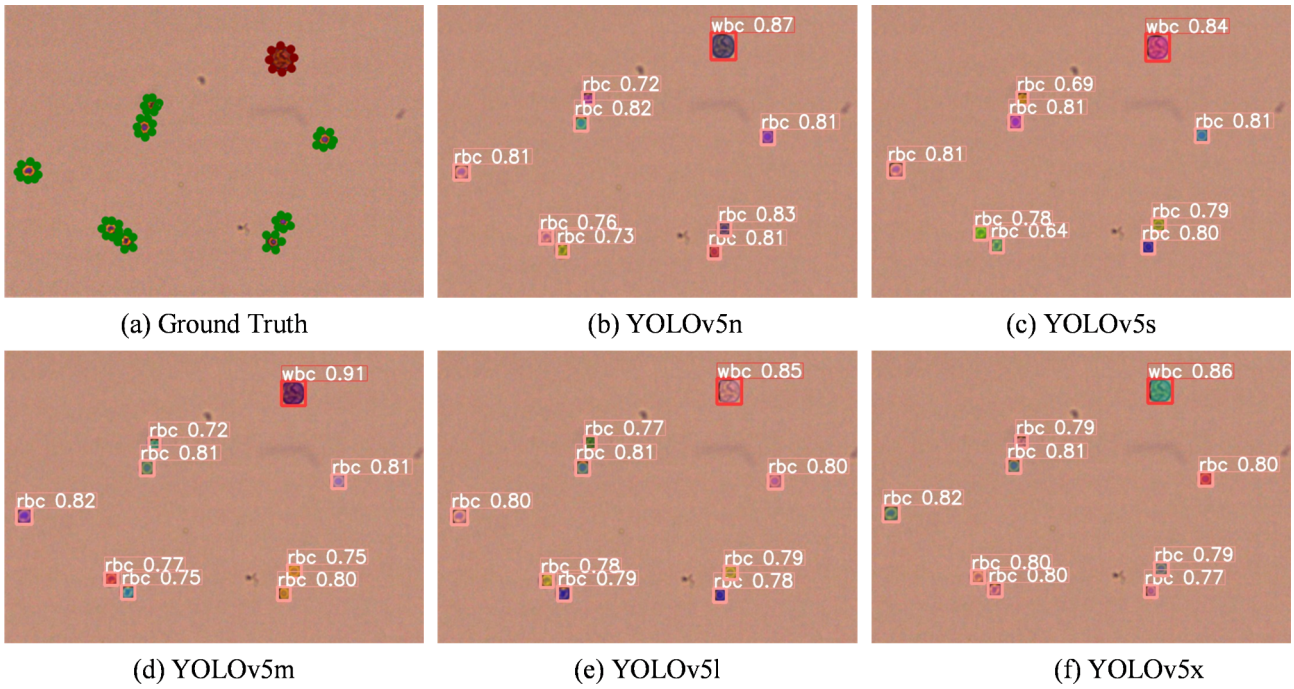


Fig. 9. Visual results of RBC and WBC detection for the five variants of YOLOv5.

Method	Backbone	FPS	mAP50	mAP50-95	*P*(box)	*R*(box)	F1(box)
Mask R-CNN	Resnet-50-FPN	0.7	87.4	52.0	88.2	84.8	84.6
Mask R-CNN	Resnet-101-FPN	0.7	88.2	52.0	89.9	85.8	87.7
MS R-CNN	Resnet-50-FPN	0.6	88.8	52.2	88.4	84.1	86.1
MS R-CNN	Resnet-101-FPN	0.6	88.9	53.3	90.7	80.1	84.0
MaskDis R-CNN	Resnet-50-FPN	0.6	88.8	52.0	91.5	85.9	88.4
MaskDis R-CNN	Resnet-101-FPN	0.6	91.0	52.3	91.1	83.5	86.7
YOLOv8	Darknet53	2.4	90.4	51.1	**92.8**	88.2	**90.4**
YOLOv5n	CSPDarknet53	**63.3**	**91.8**	**55.6**	84.5	**89.5**	86.8

Table 4. Results of the comparison between YOLOv5n and other methods.

missed the detection of the crystal (as shown on the left of Fig. 10(b) of the black arrow), resulting in not segmenting the object.

YOLOv8 has missed the left RBC target (as shown in the left of Fig. 10(d)), the segmentation boundaries in Mask R-CNN and MS R-CNN appear to be relatively rough, whereas YOLOv5n and YOLOv8 have made improvements in this aspect. The superior performance of YOLOv5n in segmentation tasks can be attributed to the incorporation of advanced feature extraction and fusion technologies, namely Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). In contrast, the feature fusion and extraction mechanisms in the Mask R-CNN series are relatively traditional. This enables YOLOv5n to achieve better results when handling complex scenes and small objects.Moreover, YOLOv5n integrates the mask confidence scores into the segmentation evaluation metrics, ensuring not only accurate segmentation but also improving the quality of the segmentation boundaries. Consequently, among all methods, YOLOv5n exhibits the best performance, guaranteeing accurate segmentation and enhancing the boundary quality of the segmented regions.

Furthermore, we compare the performance of YOLOv5n with SOLOv2[33], BoxInst[34], and ConvNeXt V2[35] using the metrics mAP50-95, mAP50, and FPS. As show in Table 5, SOLOv2 achieves an mAP50-95 of 39.5, an mAP50 of 74.1, and an FPS of 14.1. BoxInst performs better with an mAP50-95 of 44.6, an mAP50 of 88.0, and an FPS of 12.8. ConvNeXt V2 further improves the mAP scores with an mAP50-95 of 55.6 and an mAP50 of 91.0, though it operates at a lower FPS of 5.2. In contrast, YOLOv5n not only matches the highest mAP50-95 score of 55.6 but also surpasses all others with an mAP50 of 91.8. Remarkably, YOLOv5n achieves these results with an outstanding FPS of 63.3, significantly higher than the other models. This demonstrates that YOLOv5n not only provides superior accuracy but also offers exceptional speed, making it highly advantageous for real-time applications in instance segmentation of formed elements in urine.

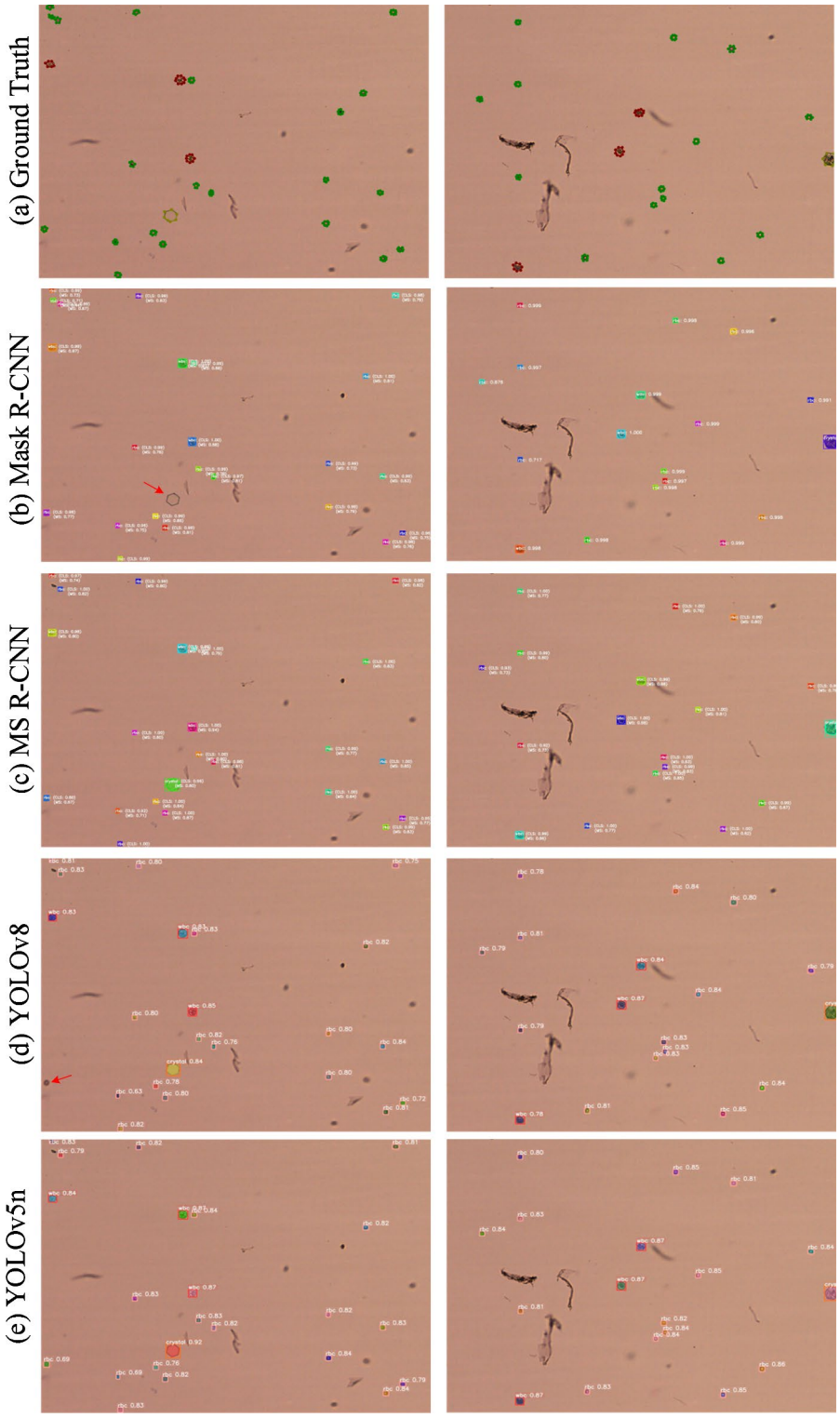


Fig. 10. Visualization results of YOLOv5n compared with other methods.

Ablation experiment

In computer vision applications, object detection and segmentation are often interrelated tasks, and the quality of detection significantly affects the performance of segmentation. To improve the performance of detection and maintain real-time processing speed, this study improved the basic YOLOv5n model by replacing the backbone network with the EfficientFormerV2[36] named YOLOv5n-EfficientFormerV2-s0 and EfficientViT[37] (YOLOv5n-EfficientViT-B0 and YOLOv5n-EfficientViT-B1) architectures. Experimental results show that YOLOv5n-EfficientFormerV2-s0(shown in the fifth line of Table 6) decreases the quality of both detection

Method	Backbone	mAP50-95	AP50	FPS
SOLOv2	Resnet-50-FPN	39.5	74.1	14.1
BoxInst	Resnet-50-FPN	44.6	88.0	12.8
ConvNeXt V2	ConvNeXt V2-B-FCMAE	**55.6**	91.0	5.2
YOLOv5n	CSPDarknet53	**55.6**	**91.8**	**63.3**

Table 5. Comparison of YOLOv5n with state-of-the-art Instance Segmentation models.

Method	*P*(seg)	*R*(seg)	F1(seg)	mAP50	mAP50-95	FPS	F1(box)
YOLOv5n-Omni-ResNet101-1x	90.0	87.7	**88.8**	91.0	50.5	7.8	88.8
YOLOv5n-Omni-ResNet101-2x	86.8	88.2	87.5	91.6	50.7	7.5	87.3
YOLOv5n-Omni-MobileNet-1x	90.3	86.6	88.4	89.4	50.0	24.0	89.8
YOLOv5n-Omni-MobileNet-4x	**91.6**	84.7	88.0	90.9	50.3	23.9	**90.0**
YOLOv5n-EfficientFormerV2-s0	84.0	84.8	84.4	88.9	50.2	14.9	84.5
YOLOv5n-EfficientViT-B0	85.3	84.7	85.0	88.1	48.6	41.3	86.2
YOLOv5n-EfficientViT-B1	89.2	86.4	87.8	89.7	49.1	29.9	87.9
YOLOv5n-Context Aggregation	83.3	85.2	84.2	88.2	48.5	47.8	86.4
YOLOv5n	86.5	**91.5**	**88.8**	**91.8**	**55.6**	**63.3**	86.9

Table 6. Experimental results of incorporating other methods into YOLOv5n.

and segmentation, with a significant reduction in processing speed. Compared to YOLOv5n, using YOLOv5n-EfficientViT-B0 model, its F1(box) decreased by 2.4%, its F1(seg) decreased by 4.4%, its mAP50 decreased by 2.9%, and its FPS decreased by 48.4. We did not attempt larger EfficientFormerV2 weights on YOLOv5n, which might further decrease processing speed.

Compared to YOLOv5n, YOLOv5n-EfficientViT-B0 (as shown in the sixth line of Table 6) achieves a decrease in detection and segmentation quality. Its F1(seg) decreased by 3.3%, mAP50 decreased by 3.7%, and FPS decreased by 22.0. However, the YOLOv5n-EfficientViT-B1 model obtains an improvement in detection quality of F1(box) and segmentation precision of P(seg), while achieving a decrease in segmentation recall with R(seg) and processing speed with FPS. Compared to YOLOv5n, its F1(box) increased by 1.0%, its P(seg) increased by 2.7%, its R(seg) decreased by 5.1%, its mAP50 decreased by 2.1%, and its FPS decreased by 33.4. In summary, replacing the backbone network with EfficientFormerV2 and EfficientViT can improve slightly the quality of detection and segmentation when using larger network weights, but they also increase computation time.

Additionally, we studied the impact of the small object module on segmentation quality in the urine sediment dataset. Specifically, we integrated the Context Aggregation[38] and Omni-Dimensional Dynamic Convolution[39] (Omni) modules into the YOLOv5n backbone architecture to evaluate their effects. The YOLOv5n-Context Aggregation model is formed by adding the Context Aggregation module after the 23rd layer of the neck network on YOLOv5n. The YOLOv5n-Omni-ResNet101-1x, YOLOv5n-Omni-ResNet101-2x, YOLOv5n-Omni-MobileNet-1x, and YOLOv5n-Omni-MobileNet-4x models were built by incorporating Omni, ResNet101 and MobileNet networks as YOLOv5n backbone networks, with training strategies set to 1x, 2x, and 4x as described in[36].

The experimental results show the YOLOv5n-Context Aggregation model hardly improves detection quality but results in a decrease in segmentation quality and processing speed as shown in the eighth line of Table 6. Compared to YOLOv5n, using the YOLOv5n-Context Aggregation model, its F1(box) decreased by 0.5%, its F1(seg) decreased by 4.6%, its mAP50 decreased by 3.6%, and its FPS decreased by 15.5. In contrast, the Omni module shows more promising effects according to the results in the first to fourth line of Table 6. Whether using ResNet101 or MobileNet on YOLOv5n, we observe a significant improvement in detection quality and segmentation precision, although they increase computation time due to additional parameters. Among the four models, YOLOv5n-Omni-MobileNet-4x achieves the best results. Compared to YOLOv5n, using YOLOv5n-Omni-MobileNet-4x approach, we observed an increase of 3.1% in F1(box), 5.1% in P(seg), a decrease of 6.8% in R(seg), a decrease of 0.8% in F1(seg), a decrease of 0.9% in mAP50, and a decrease of 39.4 in FPS. In conclusion, the small object module shows potential for improving small object segmentation precision, but this improvement comes with increased computational time. The YOLOv5n exhibits slight deficiencies in detection and segmentation accuracy, but it achieves the highest recall rate. Furthermore, YOLOv5n demonstrates the best overall segmentation quality and fastest speed.

Conclusion

To achieve rapid and accurate instance segmentation of urine formed elements, this paper efficiently combines the YOLOv5n detection model with the FCN segmentation network. To validate this method, a dataset consisting of 500 images was created. By comparing five variants of YOLOv5 models, it was found that the YOLOv5n model yielded the best results. When compared with Mask RCNN, Mask Scoring RCNN, MaskDis R-CNN, and YOLOv8, YOLOv5n demonstrated superior results in terms of speed and segmentation quality.

Additionally, comparisons with SOLOv2, BoxInst, and ConvNeXt V2 highlighted YOLOv5n's superior balance of accuracy and speed, further solidifying its effectiveness for this application. The study also explored the impact of lightweight methods and attention mechanisms tailored for small objects on YOLOv5n, conducting experiments that integrated four state-of-the-art methods. While these methods enhanced object detection accuracy, further improvements are needed in instance segmentation.

Regarding dataset annotation, manual annotation of 500 images, each containing nearly 20 urine formed cells, proved to be time-consuming. In the future, efforts will be directed toward weakly supervised or unsupervised automatic annotation, the creation of larger datasets, and the adoption of more advanced algorithms to find efficient and rapid segmentation methods tailored for urine formed elements. This research marks a significant step toward automating urine formed element analysis, offering the potential for improved diagnostic capabilities in clinical settings.

Data availability

The dataset is available for access at https://terabox.com/s/1afaHP7g_Yd1Dn0qLNM8AhQ.

Received: 21 March 2024; Accepted: 13 November 2024
Published online: 19 November 2024

References

1. Almadhoun, M. D. Automated recognition of urinary epithelial cells. The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE): IEEE; 2013. pp. 568 – 72. (2013).
2. Jiang, X. & Nie, S. Urine sediment image segmentation based on level set and Mumford-Shah model. 1st International Conference on Bioinformatics and Biomedical Engineering: IEEE; 2007. pp. 1028-30. (2007).
3. Khalid, Z. M. A Comparative Study of Different Deep Learning Algorithms for Urinalysis Recognition System. (2022).
4. Lamchiagdhase, P. et al. Urine sediment examination: a comparison between the manual method and the iQ200 automated urine microscopy analyzer. *Clin. Chim. Acta*. **358** (1–2), 167–174. https://doi.org/10.1016/j.cccn.2005.02.021 (2005).
5. Lakatos, J., Bodor, T., Zidarics, Z. & Nagy, J. Data processing of digital recordings of microscopic examination of urinary sediment. *Clin. Chim. Acta*. **297** (1–2), 225–237. https://doi.org/10.1016/s0009-8981(00)00249-7 (2000).
6. , S. K & B, D. A review on various methods for recognition of urine particles using digital microscopic images of urine sediments. *Biomed. Signal Process. Control*. **68** https://doi.org/10.1016/j.bspc.2021.102806 (2021).
7. Cavanaugh, C. & Perazella, M. A. Urine sediment examination in the diagnosis and management of kidney disease: core curriculum 2019. *Am. J. Kidney Dis.* **73** (2), 258–272. https://doi.org/10.1053/j.ajkd.2018.07.012 (2019).
8. Wang, P. et al. Study of risk factor of urinary calculi according to the association between stone composition with urine component. *Sci. Rep.* **11** (1). https://doi.org/10.1038/s41598-021-87733-7 (2021).
9. Vu, Q. D. et al. Methods for segmentation and classification of Digital Microscopy Tissue Images. *Front. Bioeng. Biotechnol.* **7** https://doi.org/10.3389/fbioe.2019.00053 (2019).
10. Kutlu, H., Avci, E. & Özyurt, F. White blood cells detection and classification based on regional convolutional neural networks. *Med. Hypotheses*. **135** https://doi.org/10.1016/j.mehy.2019.109472 (2020).
11. Wang, C-W. et al. Deep learning for bone marrow cell detection and classification on whole-slide images. *Med. Image. Anal.* **75**, 102270 (2022).
12. Zheng, X. et al. White blood cell detection using saliency detection and CenterNet: a two-stage approach. *J. Biophotonics*. **16** (3), e202200174 (2023).
13. Gu, W., Bai, S. & Kong, L. A review on 2D instance segmentation based on deep neural networks. *Image Vis. Comput.* **120** https://doi.org/10.1016/j.imavis.2022.104401 (2022).
14. Yi, J. et al. Attentive neural cell instance segmentation. *Med. Image. Anal.* **55**, 228–240. https://doi.org/10.1016/j.media.2019.05.004 (2019).
15. Lin, S. & Norouzi, N. An effective Deep Learning Framework for Cell Segmentation in Microscopy images. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2021**, 3201–3204. https://doi.org/10.1109/EMBC46164.2021.9629863 (2021).
16. Atıcı H, Koçer HE. Mask R-CNN Based Segmentation and Classification of Blood Smear Images. *Gazi Journal of Engineering Sciences. (GJES).* 2023;**9**(1): 128-143.
17. Luo, Y. et al. A lightweight network based on dual-stream feature fusion and dual-domain attention for white blood cells segmentation. *Front. Oncol.* **13**, 1223353 (2023).
18. Wang, J. et al. *Improved BlendMask: Nuclei Instance Segmentation for Medical Microscopy Images* (IET Image Processing, 2023).
19. Ren, X., Zhou, S., Shen, D. & Wang, Q. Mask-RCNN for cell instance segmentation. *IEEE Trans. Med. Imaging* (2020).
20. Loh, D. R., Yong, W. X., Yapeter, J., Subburaj, K. & Chandramohanadas, R. A deep learning approach to the screening of malaria infection: automated and rapid cell counting, object detection and instance segmentation using Mask R-CNN. *Comput. Med. Imaging Graph.* **88** https://doi.org/10.1016/j.compmedimag.2020.101845 (2021).
21. Qiu, X., Lei, H., Xie, H. & Lei, B. Segmentation of Multiple Myeloma Cells Using Feature Selection Pyramid Network and Semantic Cascade Mask RCNN. IEEE 19th International Symposium on Biomedical Imaging (ISBI): IEEE; 2022. pp. 1–4. (2022).
22. Mitate, E. et al. Application of the sliding window method and Mask-RCNN method to nuclear recognition in oral cytology. *Diagn. Pathol.* **17** (1), 1–8 (2022).
23. Bai, B., Tian, J., Luo, S., Wang, T. & Lyu, S. YUSEG: Yolo and Unet is all you need for cell instance segmentation. Competitions in Neural Information Processing Systems: PMLR; pp. 1–15. (2023).
24. Prangemeier, T., Reich, C. & Koeppl, H. Attention-based transformers for instance segmentation of cells in microstructures. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM): IEEE; pp. 700-7. (2020).
25. Priego-Torres, B. M., Lobato-Delgado, B., Atienza-Cuevas, L. & Sanchez-Morillo, D. Deep learning-based instance segmentation for the precise automated quantification of digital breast cancer immunohistochemistry images. *Expert Syst. Appl.* **193**, 116471 (2022).
26. Zhao, M. et al. MSS-WISN: Multiscale multistaining WBCs instance segmentation network. *IEEE Access.* **10**, 65598–65610 (2022).
27. Liu, Y., Wang, C., Wen, Y., Huo, Y. & Liu, J. *Efficient Segmentation Algorithm for Complex Cellular Image Analysis System* (IET Control Theory & Applications, 2023).
28. Choi, H., Lee, H-J., You, H-J., Rhee, S-Y. & Jeon, W. Comparative Analysis of Generalized Intersection over Union and Error Matrix for Vegetation Cover classification Assessment. *Sens. Mater.* **31** (11). https://doi.org/10.18494/sam.2019.2584 (2019).
29. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. Proceedings of the IEEE international conference on computer vision pp. 2961-9. (2017).
30. Huang, Z., Huang, L., Gong, Y., Huang, C. & Wang, X. Mask scoring r-cnn. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp. 6409-18. (2019).

31. Tu, S. et al. *MaskDis R-CNN: An Instance Segmentation Algorithm with Adversarial Network for herd pigs* (IET Image Processing, 2023).
32. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition pp. 779 – 88. (2016).
33. Wang, X., Zhang, R., Kong, T., Li, L. & Shen, C. Solov2: dynamic and fast instance segmentation. *Adv. Neural. Inf. Process. Syst.* **33**, 17721–17732 (2020).
34. Tian, Z., Shen, C., Wang, X., Chen, H. & Boxinst High-performance instance segmentation with box annotations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 5443-52. (2021).
35. Woo, S. et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp. 16133-42. (2023).
36. Li, Y. et al. Rethinking vision transformers for mobilenet size and speed. Proceedings of the IEEE/CVF International Conference on Computer Vision pp. 16889-900. (2023).
37. Cai, H. et al. Lightweight multi-scale attention for high-resolution dense prediction. Proceedings of the IEEE/CVF International Conference on Computer Vision pp. 17302-13. (2023).
38. Liu, Y. et al. Learning to Aggregate Multi-scale Context for Instance Segmentation in Remote sensing images. (2021). arXiv preprint arXiv:211111057.
39. Li, C., Zhou, A. & Yao, A. Omni-dimensional dynamic convolution. *arXiv Preprint arXiv* :220907947. (2022).

Author contributions

HL: Methodology, Data Curation, Formal analysis, Writing – Original Draft and Visualization. ST: Conceptualization, Writing – Review & Editing, Validation and Supervision. CT: Project administration and Resources. WY：Project administration and Resources.HY: Data Curation. WC: Fromal analysis and Investigation. LM: Funding acquisition, model optimization.

Declarations

Competing interests

The authors declare no competing interests.

Ethical and informed consent

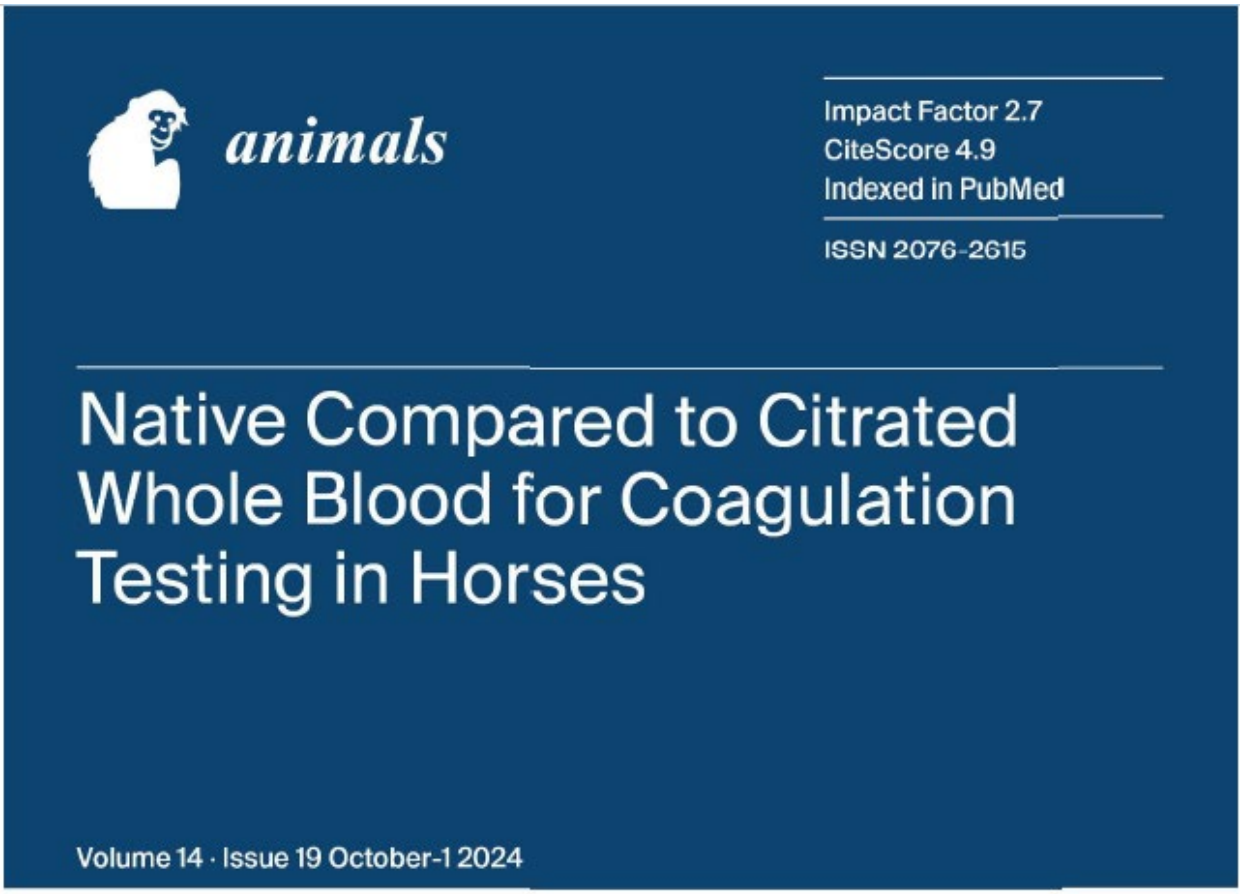
Written informed consent for publication of this paper was obtained from the South China.
Agricultural University, Songshan Lake Central Hospital of Dongguan, Science and all authors.

Additional information

Correspondence and requests for materials should be addressed to L.M. or C.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.


animals
Impact Factor 2.7
CiteScore 4.9
Indexed in PubMed
ISSN 2076-2615
Native Compared to Citrated Whole Blood for Coagulation Testing in Horses
Volume 14 · Issue 19 October-1 2024
MDPI

目录

目录

目录

目录

目录

animals

MDPI

Article

Tracking and Behavior Analysis of Group-Housed Pigs Based on a Multi-Object Tracking Approach

Shuqin Tu [1,2], Jiaying Du [1], Yun Liang [1,2,*], Yuefei Cao [1], Weidian Chen [1], Deqin Xiao [1] and Qiong Huang [1,2,*]

[1] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China; tsq5_6@scau.edu.cn (S.T.); dujiaying2001@163.com (J.D.); cyf_1208@163.com (Y.C.); chenwd2543@163.com (W.C.); deqinx@scau.edu.cn (D.X.)

[2] Guangzhou Key Laboratory of Intelligent Agriculture, Guangzhou 510000, China

* Correspondence: liang_yun168@163.com (Y.L.); qhuang@scau.edu.cn (Q.H.)

Simple Summary: This study developed a new method for automatically tracking and analyzing pig behavior in complex environments. We use the YOLOv8 algorithm for real-time detection and behavior classification, and employ the OC-SORT algorithm to address tracking issues caused by lighting changes, occlusions, and collisions between pigs. This method enables the automatic analysis of behavior, helping farm managers promptly detect abnormalities and health issues in pigs. Experimental results show that the method performs excellently in behavior recognition and tracking, accurately recording pig behavior to provide technical support for monitoring the health and welfare of pig herds.

Abstract: Smart farming technologies to track and analyze pig behaviors in natural environments are critical for monitoring the health status and welfare of pigs. This study aimed to develop a robust multi-object tracking (MOT) approach named YOLOv8 + OC-SORT(V8-Sort) for the automatic monitoring of the different behaviors of group-housed pigs. We addressed common challenges such as variable lighting, occlusion, and clustering between pigs, which often lead to significant errors in long-term behavioral monitoring. Our approach offers a reliable solution for real-time behavior tracking, contributing to improved health and welfare management in smart farming systems. First, the YOLOv8 is employed for the real-time detection and behavior classification of pigs under variable light and occlusion scenes. Second, the OC-SORT is utilized to track each pig to reduce the impact of pigs clustering together and occlusion on tracking. And, when a target is lost during tracking, the OC-SORT can recover the lost trajectory and re-track the target. Finally, to implement the automatic long-time monitoring of behaviors for each pig, we created an automatic behavior analysis algorithm that integrates the behavioral information from detection and the tracking results from OC-SORT. On the one-minute video datasets for pig tracking, the proposed MOT method outperforms JDE, Trackformer, and TransTrack, achieving the highest HOTA, MOTA, and IDF1 scores of 82.0%, 96.3%, and 96.8%, respectively. And, it achieved scores of 69.0% for HOTA, 99.7% for MOTA, and 75.1% for IDF1 on sixty-minute video datasets. In terms of pig behavior analysis, the proposed automatic behavior analysis algorithm can record the duration of four types of behaviors for each pig in each pen based on behavior classification and ID information to represent the pigs' health status and welfare. These results demonstrate that the proposed method exhibits excellent performance in behavior recognition and tracking, providing technical support for prompt anomaly detection and health status monitoring for pig farming managers.

Keywords: pig behavior tracking; OC-SORT; group-housed pigs; YOLOv8; behavior analysis

Citation: Tu, S.; Du, J.; Liang, Y.; Cao, Y.; Chen, W.; Xiao, D.; Huang, Q. Tracking and Behavior Analysis of Group-Housed Pigs Based on a Multi-Object Tracking Approach. *Animals* **2024**, *14*, 2828. https://doi.org/10.3390/ani14192828

Academic Editor: Christa Thöne-Reineke

Received: 3 September 2024
Revised: 25 September 2024
Accepted: 29 September 2024
Published: 30 September 2024

1. Introduction

With the development of intelligent and efficient animal farming, the automatic monitoring of pig health plays a crucial role in the modern livestock industry. Precision livestock

farming technology uses advanced sensors and data analytic technology to monitor individual animals in real-time, which provides great potential to improve management efficiency and production. The specific activity patterns of farmed animals, including standing, lying, and eating, evidently reflect the consequences of compromised health and welfare. Due to the large animal-to-staff ratio, conducting manual observation or recording animal behavior is labor-intensive, subjective, and inefficient. Additionally, human presence can change animal behavior, making monitoring without humans in the barn necessary for accurate observations. Therefore, there is an urgent need for effective automatic monitoring methods to accurately track and analyze pig behaviors under different environmental conditions. This not only helps in the timely detection of abnormalities such as diseases, stress, or environmental factors, but also enhances the efficiency of pig farming [1,2].

Regarding existing automated animal monitoring systems, there are two categories. Firstly, there are contact-based methods with attached sensors, such as utilizing a high-frequency radio frequency identification (HF RFID) system to register drinking behavior [3], employing accelerometers for the detection of abnormal gait patterns [4], using pressure pads to identify limping behavior [5], and using accelerometry and GNSS data to classify animal behavior [6]. Considering the cost of hardware and maintenance on large-scale commercial farms, contact-based solutions for automated tracking are not preferable. Secondly, there is contactless monitoring using computer vision and deep learning technologies [7–9], which has enjoyed growing popularity due to its low cost and sustainability compared to contact-based sensors. For example, Duc Duong Tran et al. proposed a method using deep learning for automatically monitoring and detecting abnormalities in pig behaviors [10]. Zhang et al. introduced a transformer-based neural network (TNN) model for recognizing pig feeding behavior and detecting potential dangers [11]. Alameer et al. utilized deep learning technology to identify pig postures and drinking behaviors and used drinking behavior as an indicator of pig health status [12]. Computer vision technology enables the real-time monitoring and identification of pig behavior through image processing, achieving accurate health monitoring. Further research is needed to extract behavioral information from surveillance videos in pig farms for improved monitoring [13].

In video surveillance, tracking and behavior recognition are crucial components for group-housed pig monitoring applications. In the field of pig tracking research, various methods have been proposed to achieve an accurate tracking performance. For instance, Chen et al. introduced a real-time detection and tracking method based on YOLACT, successfully detecting and tracking the various major body parts of pigs, thereby providing robust support for pig behavior analysis [14]. Aggaluck et al. achieved accurate pig tracking by training a faster region-based convolutional neural network to identify the body and head of pigs [15]. Gong et al. proposed an improved IOU-Tracker pig tracking algorithm, incorporating the YOLOv5s network for the real-time detection and tracking of pigs [16]. Zhang et al. employed a tracking method based on discriminative correlation filters for the rapid tracking of multiple pigs [17]. Tu et al. proposed an enhanced DeepSORT target-tracking algorithm that integrates the YOLOX-S and YOLOv5s detectors. By focusing on trajectory processing and data association tailored to pig-specific scenarios, they achieved significant improvements in tracking stability [18]. In the realm of pig behavior recognition, researchers have also made significant strides. Zhou et al. introduced a method for individual pig identification using the three-dimensional (3D) point clouds of the pig's back surface, successfully addressing the challenge of difficult sample collection in pig face recognition [19]. Hao et al. proposed a novel deep mutual learning enhanced two-stream pig behavior recognition approach. Their model integrates two mutual learning networks utilizing RGB and optical flow streams, with each branch comprising collaboratively learning student networks to capture a robust appearance and motion features. This innovative method effectively improves the performance of pig behavior recognition [20]. To investigate aggressive behaviors in group-housed pigs, Gao et al. presented a hybrid model that combines convolutional neural network (CNN) and gated recurrent unit (GRU) to distinguish between aggressive behaviors and other behaviors

in monitored videos [21]. Ji et al. proposed a method that employs the temporal shift module (TSM) for the automatic recognition of pig aggression. This approach enhances the accuracy of identifying aggressive behaviors in pigs by allowing the model to effectively process both spatial and temporal features [22]. These research methods have allowed significant progress in pig health monitoring. However, there are still considerable limitations when dealing with complex behavior analysis in pigs, especially in specific scenarios. For instance, collisions between pigs can easily result in target loss, making it challenging to effectively track pig targets for long periods. In commercial automated farming settings, variations in lighting conditions and dense occlusion can lead to frequent changes in target IDs during the tracking process.

Existing research primarily focuses on the identification or automatic tracking of pig behaviors, lacking specific behaviors' statistics and analysis. And, the Kalman filter (KF) of the current mainstream MOT methods is susceptible to noise when utilizing linear motion for target position estimation, resulting in the inaccurate estimation of the motion direction. The deviations in KF parameters also result in frequent target loss due to occlusions. Additionally, there are few studies discussing the work of recovering the lost trajectory for re-tracking, which has played an important role in long-time tracking.

To address these issues, we propose a pig tracking and behavior recognition algorithm based on V8-Sort to analyze the different behavioral statuses of group-housed pigs. Firstly, the algorithm employs the YOLOv8 detection algorithm for pig detection and behavior classification. Secondly, the OC-SORT algorithm is utilized for pig tracking under challenging scenarios. To reduce the impact of noise and occlusion during pig tracking tasks, OC-SORT employs observation-centered momentum (OCM), re-observation update (ORU), and recovery (OCR) strategies. Especially, when a target is lost for re-tracking, OC-SORT can employ the ORU strategy for target adjustment through virtual trajectories and utilize a second correlation to recover the lost trajectory. Finally, we create an automated behavior analysis algorithm, which records the duration of each behavior for each pig in each pen. To validate the effectiveness of V8-Sort, we compare it with current mainstream tracking algorithms, including Trackformer [23], JDE [24], and TransTrack [25], on the same dataset. Furthermore, a comprehensive evaluation of the proposed tracking algorithm is conducted on four ten-minute video datasets.

The main contributions of this work are as follows:

(1) A pig behavior tracking algorithm based on V8-Sort is proposed to decrease noise and improve the robustness to occlusions and nonlinear motions.

(2) The V8-Sort is validated on four ten-minute and one sixty-minute video datasets, exploring the work of recovering the lost trajectory for re-tracking.

(3) An automatic behavior analysis algorithm is designed to record the duration of four types of behaviors for each pig in each pen based on behavior information from tracking results.

(4) The effectiveness of the OC-SORT tracker is demonstrated through comparative experiments compared with other mainstream trackers.

2. Materials

The study utilizes two distinct datasets. The first dataset, provided by Posta et al. [26], is a publicly available dataset comprising 19 one-minute video clips and 4 ten-minute video clips. The videos cover a variety of complex environments, including different pig quantities, different pig ages, and lighting conditions. For training, 6 one-minute video clips are randomly selected, while the remaining clips are used for testing. The second dataset is a private dataset captured from a commercial pig farming facility in Foshan City. It consists of 19 one-minute video clips and 1 one-hour video clip, with 10 used for training and 10 used for testing. The resolution of these videos is 2688 $\times$ 1520, and each video is captured and annotated at a rate of 5 frames per second (fps).

Both datasets encompass scenes occurring during day and night, in crowded and sparse conditions, as well as in scenarios with varying levels of pig activity, ranging from

frequent(H) to moderate(M) and less(L) active situations. Simultaneously, based on manual observation, videos with a higher quantity of pigs and significant occlusion due to the pigs clustering together are categorized as dense videos, whereas videos having less occlusion are categorized as sparse videos. The pigs are of different ages, categorized into a nursery period (3–10 weeks), growing period (11–18 weeks), and finishing period (19–26 weeks). Additionally, according to growth curves, the weight varies across different age stages. Pigs in the nursery period (3–10 weeks) typically weigh between 7 and 25 kg. During the growing period (11–18 weeks), they weigh around 25–70 kg. And in the finishing period (19–26 weeks), their weight usually ranges from 70–to 130 kg. The specific description of the test dataset is outlined in Table 1.

Table 1. Description of the test dataset.

Dataset	Video Number	Day	Night	Sparse	Time	Activity Level	Number of Pigs
Public dataset	0102	√	—	√	1 min	H	7
	0402	√	—	×	1 min	M	15
	0502	—	√	√	1 min	M	8
	0602	√	—	×	1 min	H	16
	0702	√	—	×	1 min	M	12
	0802	—	√	×	1 min	L	13
	0902	√	—	×	1 min	M	14
	1002	—	√	×	1 min	M	14
	1102	√	—	×	1 min	H	16
	1202	√	—	×	1 min	L	15
	1502	—	√	×	1 min	M	16
	2001	√	—	√	10 min	L	7
	2002	—	√	√	10 min	H	8
	2003	√	—	×	10 min	L	16
	2004	—	√	×	10 min	M	15
Private dataset	3001	√	—	√	1 min	L	10
	3002	√	—	×	1 min	M	11
	3003	—	√	×	1 min	M	11
	3004	√	—	√	1 min	H	6
	3005	√	—	√	1 min	M	6
	3006	√	—	√	1 min	M	6
	3007	√	—	√	1 min	M	6
	3008	—	√	√	1 min	L	6
	3009	—	√	√	1 min	H	6
	3010	√	—	√	60 min	H	6

Some example images of group-housed pigs are illustrated in Figure 1. To ensure the diversity of pig behavior data, we extracted key frames using the Ffmpeg6.0 [27] software, and used the Darklabel [28] software to annotate pig behavior classifications, including standing, lying, eating, and other behaviors. The pigs are categorized into two breeds: white pigs and black-spotted pigs. Each pig is allocated approximately 1.2 square meters of pen space, and temperature and humidity within the barn are controlled to ensure suitable environmental conditions for the animals. Access to both public and private datasets facilitates a comprehensive analysis and validation of pig behavior across different environments. By analyzing various pig behaviors, we can gain deeper insights into their behavioral status in different scenarios, which is crucial for improving livestock management and enhancing production efficiency.

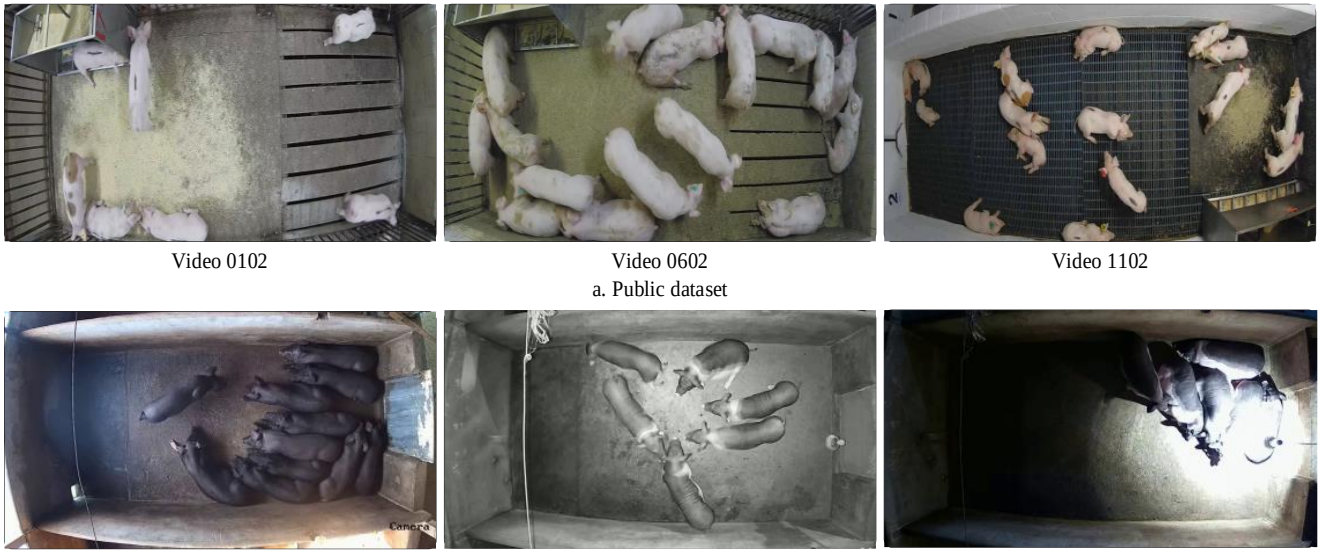


Figure 1. Part of group-housed pig images.

3. Methods

This study introduces the V8-Sort for the target tracking and behavior analysis of group-housed pigs. The overall structure is illustrated in Figure 2. First, YOLOv8 is employed as the detector to complete four tasks including target detection, displaying the target's bounding box positions, behavior categories, and confidence scores. Secondly, the OC-SORT algorithm is utilized to track each pig to reduce the impact of pigs' adhesion and occlusion on tracking. It contains three key tasks: motion prediction, data association, and track management. Finally, we create the behavior analysis method to calculate the frequency of each behavior for each pig based on the pig's ID and categories information. Therefore, the behavioral statistics information for all pigs is yield.

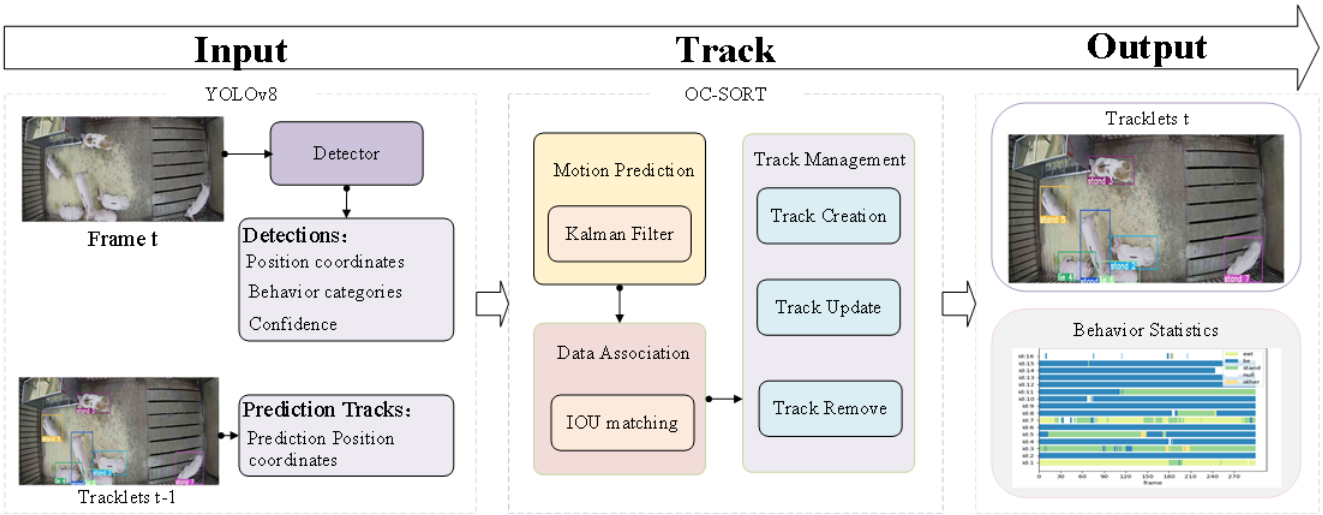


Figure 2. The overall structure of V8-Sort.

3.1. Pig Detection Based on the YOLOv8n Model

The YOLOv8n algorithm is employed as the target detector for object detection, and its model structure is primarily divided into three components: (i) Backbone for extracting image features; (ii) Neck for fusing multi-scale feature information; (iii) Prediction for predicting confidence, category, and anchor boxes. The pipeline of the YOLOv8n is shown

in Figure 3. The backbone of YOLOv8n consists of 10 layers and is organized into 5 sections, namely Stem layer, Stage layer1, Stage layer2, Stage layer3, and Stage layer4. The Neck part consists of 7 layers, which are divided into 6 parts, including TopDown layer1, TopDown layer2, Down Sample0, Bottom Up layer0, Down Sample1, and Bottom Up layer1. The Head part is composed of 6 layers and divided into 3 parts, each of which consists of ConvModule, Con2d, Bbox.Loss, and Cls.Loss. The functional descriptions of each layer can be found in Table 2. YOLOv8n is trained with specific parameters, including 80 epochs, an intersection over union (IOU) threshold of 0.7, and a default confidence threshold of 0.25. Compared to older versions of the YOLO algorithm, YOLOv8n maintains high speed and accuracy while also offering greater versatility, flexibility, and simplicity.

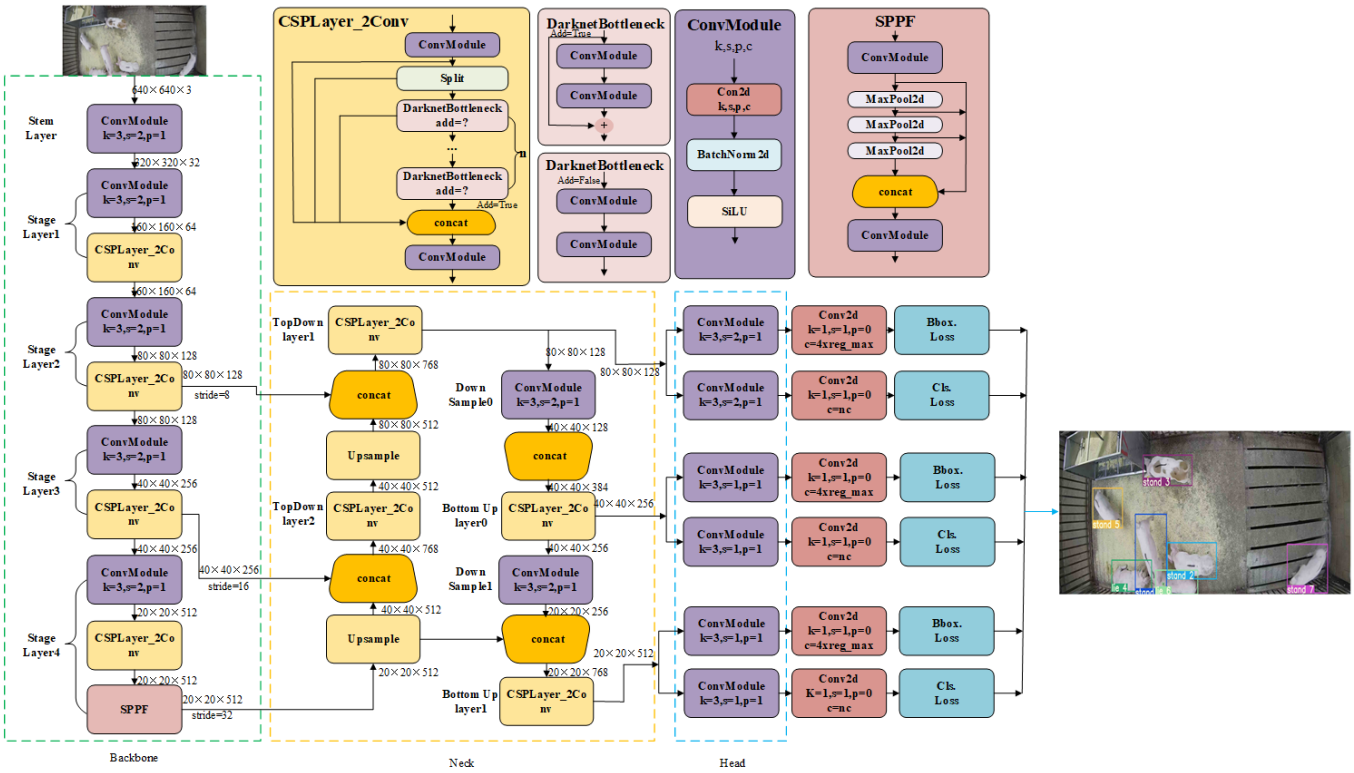


Figure 3. The pipeline of the YOLOv8n algorithm.

Table 2. Overview of the YOLOv8n layer functions.

Layer Name	Description
Stem layer	Initial layer for feature extraction and input processing.
Stage layer1	Processes the input with convolutional layers for feature refinement.
Stage layer2	Further refines features, capturing more complex patterns.
Stage layer3	Continues to extract and enhance feature representations.
Stage layer4	The final stage of the backbone, preparing features for the neck module.
TopDown layer1	Upsamples features for better spatial resolution.
TopDown layer2	Continues to upsample and merge features from different scales.
Down Sampl0	Reduces spatial dimensions for processing efficiency.
Bottom Up layer0	Integrates features from previous layers for enhanced information.
Down Sample1	Further downsampling to balance speed and accuracy.
Bottom Up layer1	Merges features, ensuring a comprehensive representation.
ConvModule	Applies convolution operations for feature extraction.
Con2d	Standard convolution layer for additional feature processing.
Bbox.Loss	Computes the loss related to bounding box predictions.
Cls.Loss	Computes the loss for classification accuracy.

3.2. Pig Tracking Based on the OC-SORT Algorithm

To reduce the impact of noise, including visual disturbances such as lighting changes and shadows, environmental interferences from surrounding movements, and data inaccuracies during the pig tracking task, we employ the OC-SORT to complete the pig's real-time tracking and behavior analysis tasks in video streams. Figure 4 shows the flowchart of the OC-SORT algorithm. The algorithm consists of three key modules: motion prediction, data association, and trajectory management. First, OC-SORT uses the KF to achieve the dynamic prediction of target trajectories in the motion prediction module. Then, it employs a data association algorithm to associate the detection targets and tracks across different frames of data. Finally, the trajectory management module of OC-SORT performs the real-time updating of trajectory information and status maintenance of the associated targets to ensure the accurate and coherent tracking and behavioral analysis of the pigs. The specific tracking process is as follows.

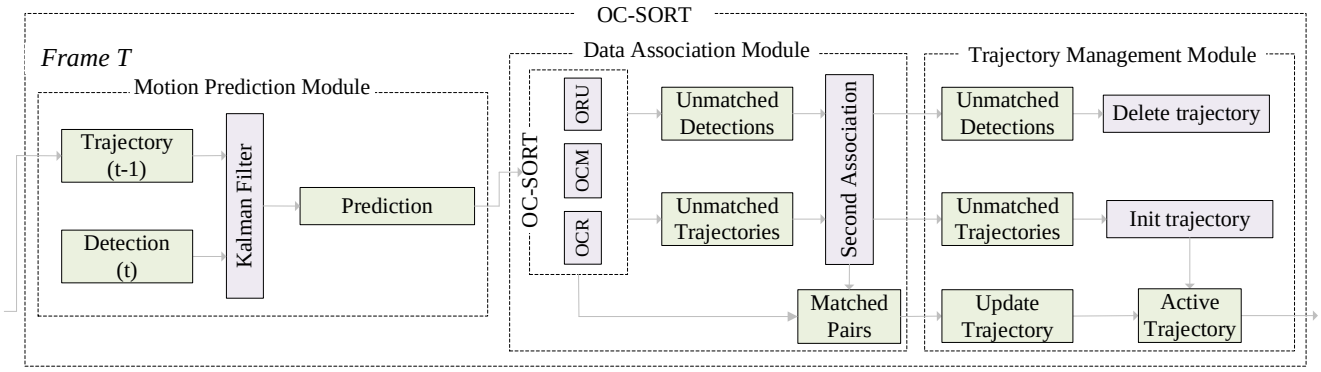


Figure 4. The flowchart of OC-SORT.

3.2.1. Pig Motion Prediction

The core KF within the motion prediction module is utilized to achieve the state prediction of pig trajectories. First, we define a seven-dimensional state vector x, as shown Equation (1) to represent the trajectory state of the pigs.

$$x = (u, v, s, r, \hat{u}, \hat{v}, \hat{s}) \tag{1}$$

where (u, v) represents the center position coordinates of the target, s is the bounding box scale (area), and r is the bounding box aspect ratio. The aspect ratio r is assumed to be a constant value. The other three variables, $\hat{u}$, $\hat{v}$, and $\hat{s}$ are the corresponding time derivatives.

Then, the previous frame's state estimation and the current frame's detection results are input to the KF prediction module to obtain an accurate estimation of the target's state in the current frame. And, it is obtained using the state estimation ($x_{t|t-1}$) and the covariance matrix ($P_{t|t-1}$) by Equations (2) and (3). Equations (2) and (3) are shown as follows:

$$x_{t|t-1} = F_t x_{t-1|t-1} \tag{2}$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t \tag{3}$$

where $x_{t|t-1}$ is the posterior state estimate at the current frame, $x_{t-1|t-1}$ is the posterior state estimate in the previous frame. P is the posterior estimate covariance matrix, F is the state transition model, and Q is the process noise. After predicting the mean x and covariance P of the target in the next frame using KF, the matching operation is performed between trajectories with the newly detected targets in the next frame.

3.2.2. Data Association

Data association enables the continuity of association for the same pig in continuous frames, ensuring consistency in pig ID value. The Hungarian algorithm [29] is employed

for data association matching between detected pigs and tracks. We utilize the OCM strategy of OC-SORT to incorporate the direction consistency of trajectories in the associated cost matrix operation, facilitating improved matching between tracking trajectories and observation results. After the usual association stage, the OCR technique of OC-SORT will initiate a second attempt to associate the final observation results of unmatched tracks with the unmatched observation values. Given N existing tracks and M detections, the association cost matrix is formulated as in Equation (4).

$$C\left(\hat{X}, Z\right) = C_{IoU}(\hat{X}, Z) + \lambda C_v(\mathbb{Z}, Z) \tag{4}$$

where $\hat{X} \in R^{N\times 7}$ is the set of object state estimations and $Z \in R^{M\times 5}$ is the set of observations on the new time step. λ is a weighting factor. $\mathbb{Z}$ contains the trajectory of observations of all existing tracks. $C_{IoU}(\cdot,\cdot)$ calculates the negative pairwise IoU (intersection over union) and $C_v(\cdot,\cdot)$ calculates the consistency between the directions of (i) linking two observations on an existing track and (ii) linking a track's historical observation and a new observation.

3.2.3. Trajectory Management

The management of trajectories involves three tasks, including trajectory creation, updating, and deletion. After completing pig data association matching, trajectories need to be updated for tracking pigs in the next frame. The management of trajectories includes the three following steps.

(1) For target pigs that fail to match, a new tracker is created for them, a new ID is assigned, and the information from the current detected pig is utilized for prediction in the next frame. For successfully matched tracking trajectories, we assign the detected pig to the ID of the successfully matched tracking target and update its trajectory information. The formula for updating the trajectory is as follows in Equations (5)–(7).

$$K_t = P_{t|t-1}H_t^T\left(H_tP_{t|t-1}H_t^T + R_t\right)^{-1} \tag{5}$$

$$x_{t|t} = x_{t|t-1} + K\left(z_t - H_tx_{t|t-1}\right) \tag{6}$$

$$P_{t|t} = (I - K_tH_t)P_{t|t-1} \tag{7}$$

where K is the output values of the KF operation, R is the observation noise, H is the observation model, z is the observation value, t represents the current step, and $t-1$ represents the previous time step. Equation (5) is used to calculate the Kalman gain to estimate the significance of errors, while Equations (6) and (7) are used to compute the mean and covariance values for updating trajectories.

(2) If a trajectory can be associated again after a period of being untracked, we employ the ORU module to reduce the accumulated error during the tracking loss process. First, a virtual trajectory is constructed for the object, starting from the last detection before tracking loss and ending at the newly matched detection. By using the last-seen observation before being untracked as z^{t1} and the observation triggering the re-association as z_{t2}, the virtual trajectory ($(\overline{z}_t)$) is denoted as in Equation (8).

$$\overline{z}_t = Traj_{virtual}(z_{t1}, z_{t2}, t), t1 < t < t2 \tag{8}$$

Then, along the trajectory of $\overline{z}_t(t1 < t < t2)$, we run the loop of predicting and re-updating. The re-update operation is as in Equation (9).

$$re-update\begin{cases} K_t = P_{t|t-1}H_t^T\left(H_tP_{t|t-1}H_t^T + R_t\right)^{-1} \\ \hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t\left(\overline{z}_t - H_t\hat{x}_{t|t-1}\right) \\ P_{t|t} = (I - K_tH_t)P_{t|t-1} \end{cases} \tag{9}$$

(3) For tracking trajectories that fail to match, we temporarily retain the trajectory without updating its status and associate this trajectory with the detected targets in the next frame. If the trajectory exceeds the predefined frame count (set as a maximum of 30 frames in this paper) without successfully matching any detection targets, we delete this trajectory.

When a target is lost for re-tracking, OC-SORT can employ the OCR, OCM, and ORU strategy to recover the lost trajectory through virtual trajectories. Figure 5 illustrates the process of pig re-tracking using OC-SORT. The red boxes are detections, yellow boxes are active tracks, blue boxes are untracked tracks, and dashed boxes are the estimates from KF. During association, OCM is used to add the velocity consistency cost. Target #1 is lost on the frame t + 1 because of occlusions. But, on the next frame, it is recovered by referring to its observation of the frame t by OCR. It being re-tracked triggers ORU from t to t + 2 for the parameters of its KF. Through the OCR, OCM, and ORU modules in OC-SORT, the robustness during pig occlusion and nonlinear motion has been enhanced.

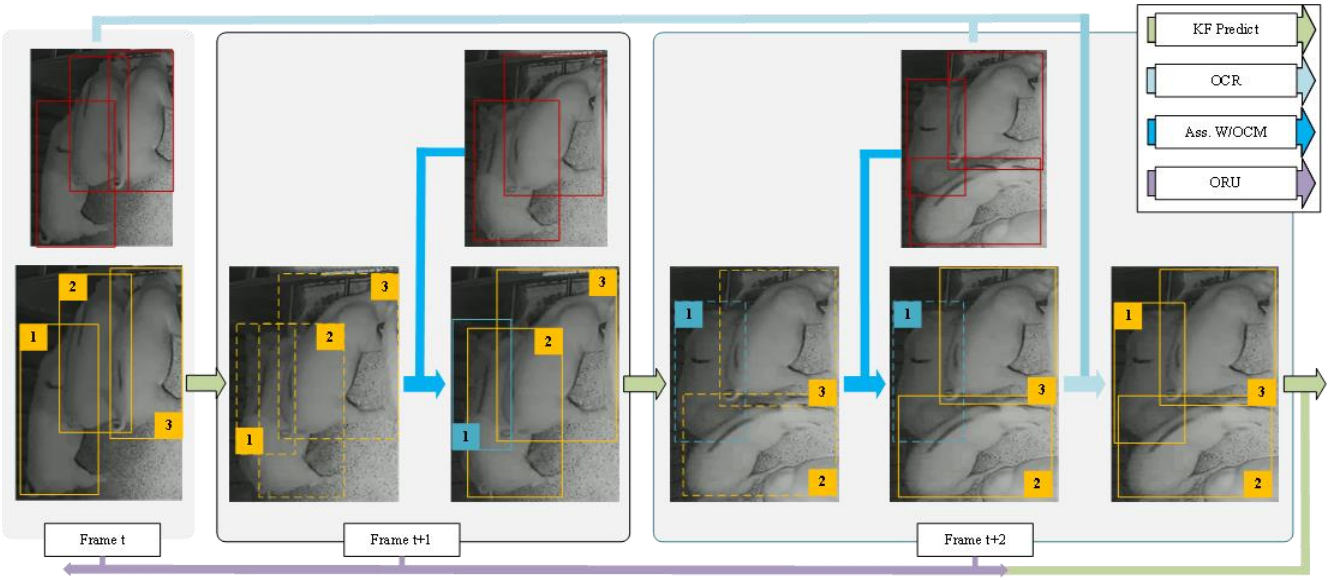


Figure 5. OC-SORT tracking process for pigs.

3.3. Pig Behavior Analysis Algorithm

We create and implement the pig's quadruple behavioral time calculation algorithm (shown in Algorithm 1) based on the video sequence tracking results. The specific implementation steps in this algorithm are as follows:

(1) A behavioral statistics array named $[A_1, A_2, A_3, A_4]$ for each trajectory is constructed to save the four behavioral classifications information. And A_1, A_2, A_3, and A_4, respectively, represent the cumulative frame counts for the four categories of lying, standing, eating, and other behaviors.

(2) Based on the categories (stand, lie, eat, and other behaviors) for each pig ID, we create a frame counting array named $[a_1, a_2, a_3, a_4]$. If the current detection box is identified as a "lie" behavior, the "lie" parameter (a_1) is set to 1, and the others are set to 0, and so on.

(3) After we associate detection boxes with tracking trajectories, we revise the values of the array $[A_1, A_2, A_3, A_4]$ if the detection box and trajectory match successfully. The formula for the operation is as follows in Equation (10):

$$\begin{bmatrix} A_{1new} \\ A_{2new} \\ A_{3new} \\ A_{4new} \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \quad (10)$$

The detection box is initialized as a new trajectory if it does not match the trajectory and its confidence score exceeds the threshold (equal to 0.7) [30], and the statistical array $[a_1, a_2, a_3, a_4]$ parameters are set to 0. Finally, by accumulating the behavioral frame counts of all trajectories and dividing by the frame rate, we can obtain the duration of each pig behavior.

Algorithm 1: Pseudo-code of pig behavior analysis

Input: A video sequence V; object detector Det; tracking score threshold η is set 0.75;
Frames per second Fps;
Output: Tracks T of the video

```
Initialization: T ← ∅
for frame f_k in V do
    D ← Det(f_k)
    Initialize time-count array including four elements for time statistics a ← [0, 0, 0, 0]
    Set variable category_index ← D{category}
    a[category_index] ← 1
    Associate T with D using OC-SORT:
        if succeed to match then
            Call the update or re-activate function to update the status of tracks
            Set variable A ← T[category_time_array] + a
            T[category_time_array] ← A
        end
        if D failed to match and D > η then
            Call the function to create a new track.
            Initialize time-count array A ← [0, 0, 0, 0]
            T[category_time_array] ← A
        end
End
T[category_time_array] = T[category_time_array]/Fps
```

Return T

3.4. Evaluation Metrics for MOT

We select HOTA, MOTA, and IDF1 as evaluation metrics for pig MOT. The calculation of HOTA is shown in Equation (11), where $DetA$ denotes the detection accuracy score, and $AssA$ represents the association accuracy score. c is a point belonging to TP, from which a unique ground truth trajectory can be determined. $A(c)$ represents the association accuracy. TP refers to the number of true positive samples, FN refers to positive samples incorrectly predicted as negative, and FP refers to negative samples incorrectly predicted as positive.

$$HOTA = \sqrt{DetA \cdot AssA} = \sqrt{\frac{\sum_{c \in TP} A(c)}{TP + FN + FP}} \tag{11}$$

The MOTA calculation is shown as Equation (12), where FP represents the total number of false detections in frame t. FN denotes the total number of missed detections in frame t, IDS refers to the number of ID switches that occur during tracking in frame t, and g_t indicates the number of targets observed at frame t.

$$MOTA = 1 - \frac{\sum_t (FP + FN + IDS)}{\sum_t g_t} \tag{12}$$

The IDF1 calculation is shown as Equation (13), where $IDTP$ represents the total number of targets correctly tracked with an unchanged ID. $IDFP$ represents the total number of targets incorrectly tracked with an unchanged ID. $IDFN$ represents the total number of targets lost in tracking with an unchanged ID.

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \tag{13}$$

Additionally, the model performance is evaluated with the number of identity switches (IDS). Higher values of HOTA, MOTA, and IDF1, and a lower value of IDs indicate a better model performance.

4. Results and Analysis

4.1. Results Comparison of V8-Sort and Other MOT Methods

The comparative experimental results on the public and private datasets are shown in Table 3 using V8-Sort and three widely used object tracking algorithms, namely Trackformer, JDE, and TransTrack. In the public dataset, V8-Sort achieved a HOTA of 82.0%, a MOTA of 96.3%, and 22 IDs. Compared with the other three methods, the proposed method has the highest accuracy and the least number of IDs. The HOTA value of V8-Sort is 11.2%, 19.4%, and 18.2% higher than those of Trackformer, JDE, and TransTrack, respectively; MOTA is higher by 7.8%, 12.9%, and 17%, respectively; and there are 261, 451, and 501 fewer IDs, respectively. Moreover, in the private dataset, V8-Sort has a HOTA of 74.8%, a MOTA of 96.7%, and 17 IDs. Comparing Trackformer and TransTrack, the HOTA of V8-Sort is 1.3 and 17.3 percentage points higher, and the IDs are fewer by 24 and 378, respectively.

Table 3. Comparison of V8-Sort with other MOT methods.

Video Sequence	Algorithm	HOTA/%↑	IDs↓	MOTA/%↑	IDF1/%↑	FP↓	FN↓
Public datasets	Trackformer	70.8	283	88.5	79.5	1048	3719
	JDE	62.6	473	83.4	71.2	3323	3455
	TransTrack	63.8	523	79.3	71.2	3627	4910
	V8-Sort	**82.0**	**22**	**96.3**	**96.8**	**658**	**953**
Private datasets	Trackformer	73.5	41	95.7	86.9	426	463
	TransTrack	57.5	395	82.1	67.2	1292	2279
	V8-Sort	**74.8**	**17**	**93.7**	**93.1**	**143**	**1232**

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable. and bolded results represent the method used in this study.

The results show that the tracking performance of V8-Sort outperforms those of Trackformer, JDE, and TransTrack. This is attributed to the fact that the observation-centered OC-SORT tracker introduces the momentum of object movement into the correlation phase and develops a pipeline that is less noisy and more robust to occlusions and no-linear motions as a means of enhancing the robustness and accuracy of tracking. Therefore, the V8-Sort method outperforms the other three algorithms, which are able to effectively handle issues such as target occlusion, intersection, and scale variations in complex scenes, thereby providing high-quality tracking results for pig behavior.

Figure 6 illustrates the comparison of V8-Sort with other MOT methods on public datasets. In Figure 6, the yellow arrows indicate the maximum ID for each frame. It can be observed that JDE (a), Trackformer (b), and TransTrack (c) have the highest tracking IDs values of 93, 419, and 31 in frame 201 of video 0602. These numbers of IDs differ from the actual number of pigs in the pigsty (16), leading to frequent pig ID errors. In contrast, V8-Sort (d) consistently maintained the maximum number of IDs of 16 without any erroneous ID changes.

Figure 7 illustrates the result of the comparison of V8-Sort with other MOT methods on private datasets. As depicted in Figure 7, Trackformer reaches a maximum tracking ID of 12 in frame 278 of video 3004, while TransTrack exhibits frequent erroneous IDs with a maximum number of tracked IDs of 12 and 11 in frames 199 and 278 of video 3004, respectively. V8-Sort maintains a stable maximum number of IDs of six. This stability is achieved because OC-SORT incorporates the motion trends of the targets into the similarity matrix, which helps to maintain the targets' IDs across different frames and thus avoids ID error switches during the long-time tracking process.

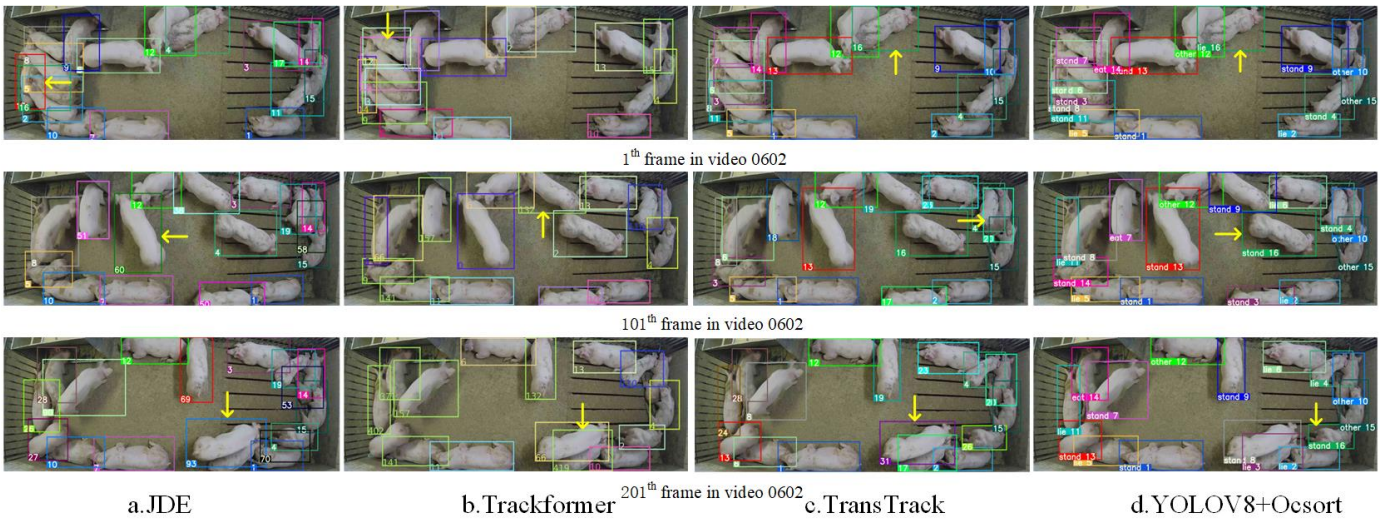


Figure 6. Comparison between V8-Sort and other tracking methods on public datasets.

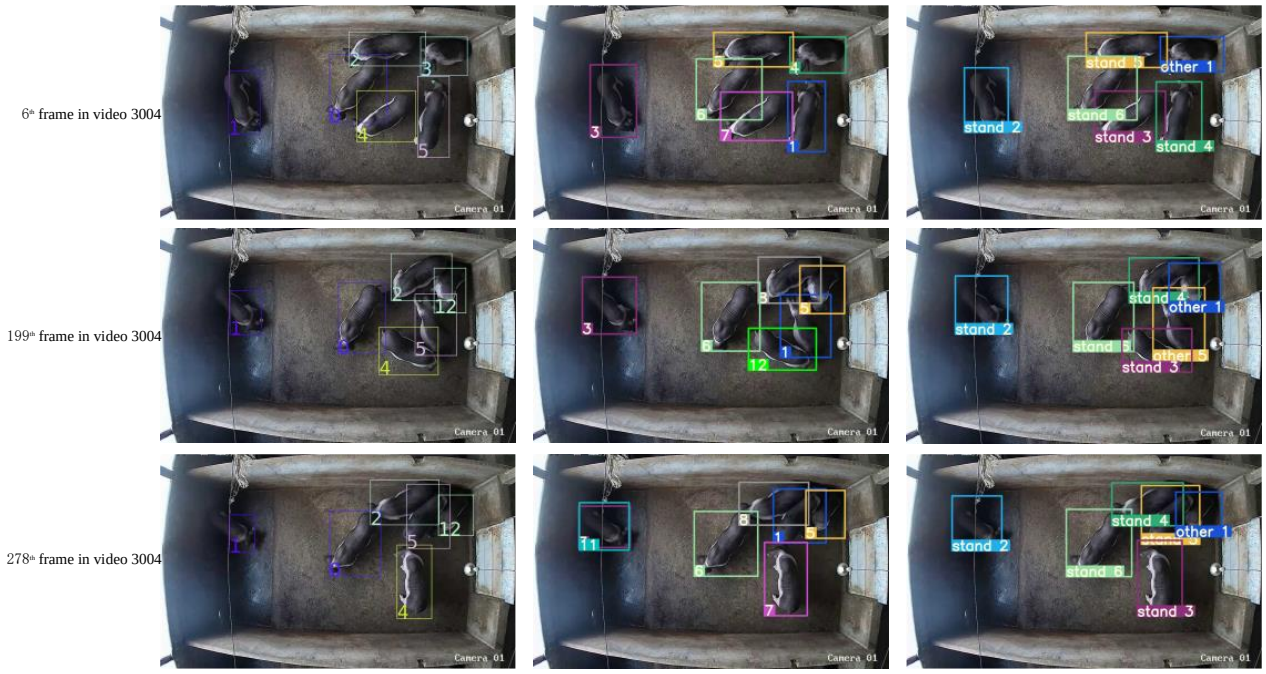


Figure 7. Comparison between V8-Sort and other tracking methods on private datasets.

4.2. Tracking Results of V8-Sort on One-Minute Videos Dataset

The results of the V8-Sort are shown in Tables 4 and 5 on one-minute videos from the public and private datasets. It can be observed that, in the public dataset, the average MOTA, HOTA, and number of IDs for V8-Sort are 96.3%, 82%, and 22, respectively. In the private dataset, the average MOTA, HOTA, and number of IDs for V8-Sort are 93.7%, 74.8%, and 17, respectively. Among them, the highest MOTA is achieved in video 0802 and 3005, reaching 99.9%, while the lowest MOTA is in video 3009, at 71.2%. The main reason for the difference in MOTA among the different videos is due to the complex environmental conditions, such as video background, daytime or night time, sparse or dense scenarios, and the activity status of the pigs. In the daytime video 3009, the pigs are more active, resulting in a lower MOTA. In the night time video 0802, pig activity status is low, and the background is simple, leading to the highest MOTA. And, V8-Sort detects and tracks each pig in both daytime and night time scenarios.

Table 4. The tracking results of V8-Sort on the public dataset.

Video Squences	HOTA/%↑	IDs↓	MOTA/%↑	IDF1/%↑	FP↓	FN↓
0102	85.4	2	98.2	97.1	1	35
0402	84.3	0	95.4	97.7	137	68
0502	84.0	10	99.4	99.7	10	4
0602	73.8	0	90.6	91.9	275	165
0702	83.8	0	98.1	99.0	13	55
0802	91.9	0	99.9	99.9	0	3
0902	85.8	0	97.5	98.7	45	61
1002	77.2	0	96.0	96.0	74	196
1102	83.1	5	97.6	92.8	29	83
1202	81.6	5	95.0	95.6	26	193
1502	77.3	0	98.4	99.2	29	50
Total/average	82.0	22	96.8	96.8	658	953

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable.

Table 5. The tracking results of V8-Sort on the private dataset.

Video Squences	HOTA/%↑	IDs↓	MOTA/%↑	IDF1/%↑	FP↓	FN↓
3001	78.0	2	97.0	97.1	11	87
3002	72.0	2	95.4	89.8	22	129
3003	70.1	3	93.0	93.0	27	202
3004	71.6	5	94.3	91.6	33	149
3005	82.0	0	99.9	99.9	0	2
3006	81.0	0	91.3	95.5	2	154
3007	82.6	0	97.6	98.8	18	25
3008	84.1	0	99.9	99.9	1	0
3009	52.3	5	71.2	70.6	29	484
Total/average	74.8	17	93.7	93.1	143	1232

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable.

To validate the feasibility and reliability of V8-Sort, we selected several segments from validation videos that were not used for training. From each of these segments, we chose three frames for presentation, and the tracking results are shown in Figures 8 and 9, which, respectively, depict the visualization of tracking results for V8-Sort on public and private datasets. In the video segments 0402, 0802, 3004, and 3008, the maximum numbers of IDs of pigs in each frame are maintained at 15, 18, 6, and 6, respectively, which match the actual numbers of pigs. In conclusion, the aforementioned tracking results comprehensively demonstrate the outstanding accuracy and stability of the method proposed in this study within the complex environment of pig activity areas. By visualizing the tracking results, we can intuitively observe changes in pig behavior and trajectory evolution, enabling a deeper understanding of pig behavioral patterns and activity habits. This can provide valuable technological support for precision farming management of pigs.

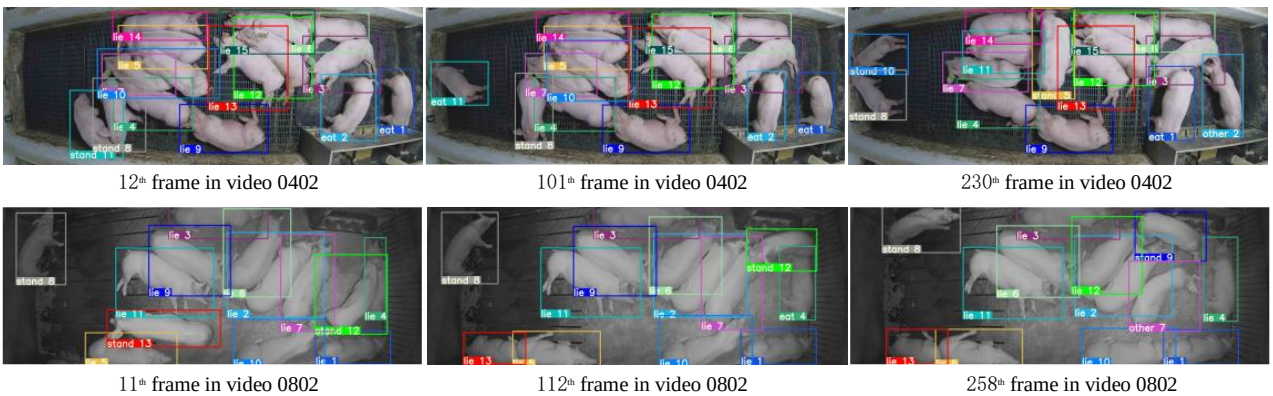
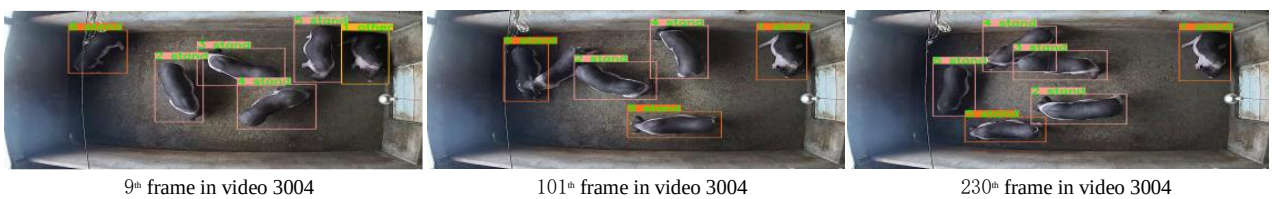


Figure 8. The visual results of V8-Sort on the public dataset.



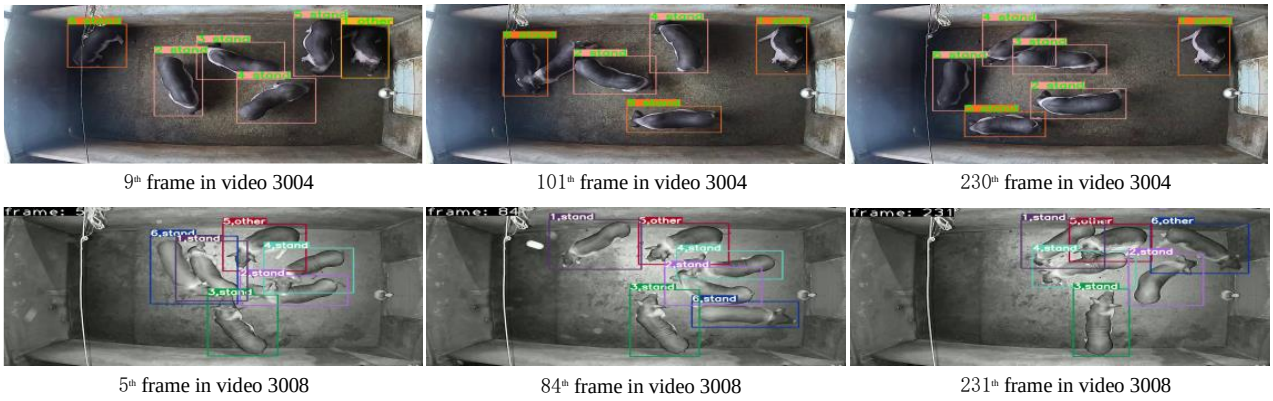


Figure 9. The tracking results visualization of V8-Sort on the private dataset.

4.3. The Long-Term Tracking Results of V8-Sort

The long-term tracking results based on V8-Sort are shown in Table 6. It can be observed that the average HOTA, number of IDs, MOTA, and IDF1 for videos 2001, 2002, 2003, 2004 and 3010 are 62.6%, 44, 93.1%, and 75.1%, respectively. There are significant differences between the tracking results of long-term and short-term videos. In the four 10 min videos, we achieve the highest values of HOTA, number of IDs, MOTA, and IDF1 (video 2002#) with 67.3%, 19, 97.1%, and 85.1%, respectively. In the 60 min video of 3010#, V8-Sort performs good, with HOTA of 69.0%, nine IDs, MOTA of 99.7%, and IDF1 of 75.1%. These disparities are primarily attributed to the duration and environmental conditions of the videos. Compared to 10 min videos, V8-Sort also makes the 60 min video achieve a better detection performance with a MOTA of 99.7%, and track the results with HOTA of 69%. The key reason is that, if a target is lost during long-term tracking, OC-SORT can employ the OCR, OCM, and ORU strategy to recover the lost trajectory through virtual trajectories and complete the re-tracking task. Therefore, V8-Sort achieves the more stable tracking of pig behaviors in long-term videos.

Table 6. The long-term tracking results based on the V8-Sort.

Video Squences	HOTA/%↑	IDs↓	MOTA/%↑	IDF1/%↑	FP↓	FN↓
2001	56.7	19	97.1	67.6	189	391
2002	67.3	29	90.4	85.1	291	1986
2003	61.2	73	93.6	73.0	1235	1741
2004	59.0	91	84.6	74.5	1737	5436
3010	69.0	9	99.7	75.1	75	201
Total/average	62.6	44	93.1	75.1	705	1951

The "↑" indicates that a higher value is better, while "↓" indicates that a lower value is preferable.

The visual results of long-term tracking in different scenarios are shown in Figure 10. In videos 2001 and 3010, the pigs are sparsely distributed with minimal occlusion, and each pig's behavior is accurately recognized. And, the maximum number of IDs consistently remains at eight and six, respectively, which both correspond to the actual number of pigs, reflecting stable tracking. However, the night time environment in video 2002 has led to occurrences of pig omissions, as indicated by the dashed boxes in the figure. Compared to videos 2001 and 2002, videos 2003 and 2004 exhibit a significant rise in the number of pigs. It can be observed that there are ID switches in frames 827 and 914 of video 2003, as indicated by the yellow arrows in the figure. In video 2004, substantial occlusion is present, resulting in a considerable number of omissions and ID switching issues. In these complex scenarios, the performance of the tracking system faces challenges, necessitating further algorithm optimization to enhance stability and accuracy.

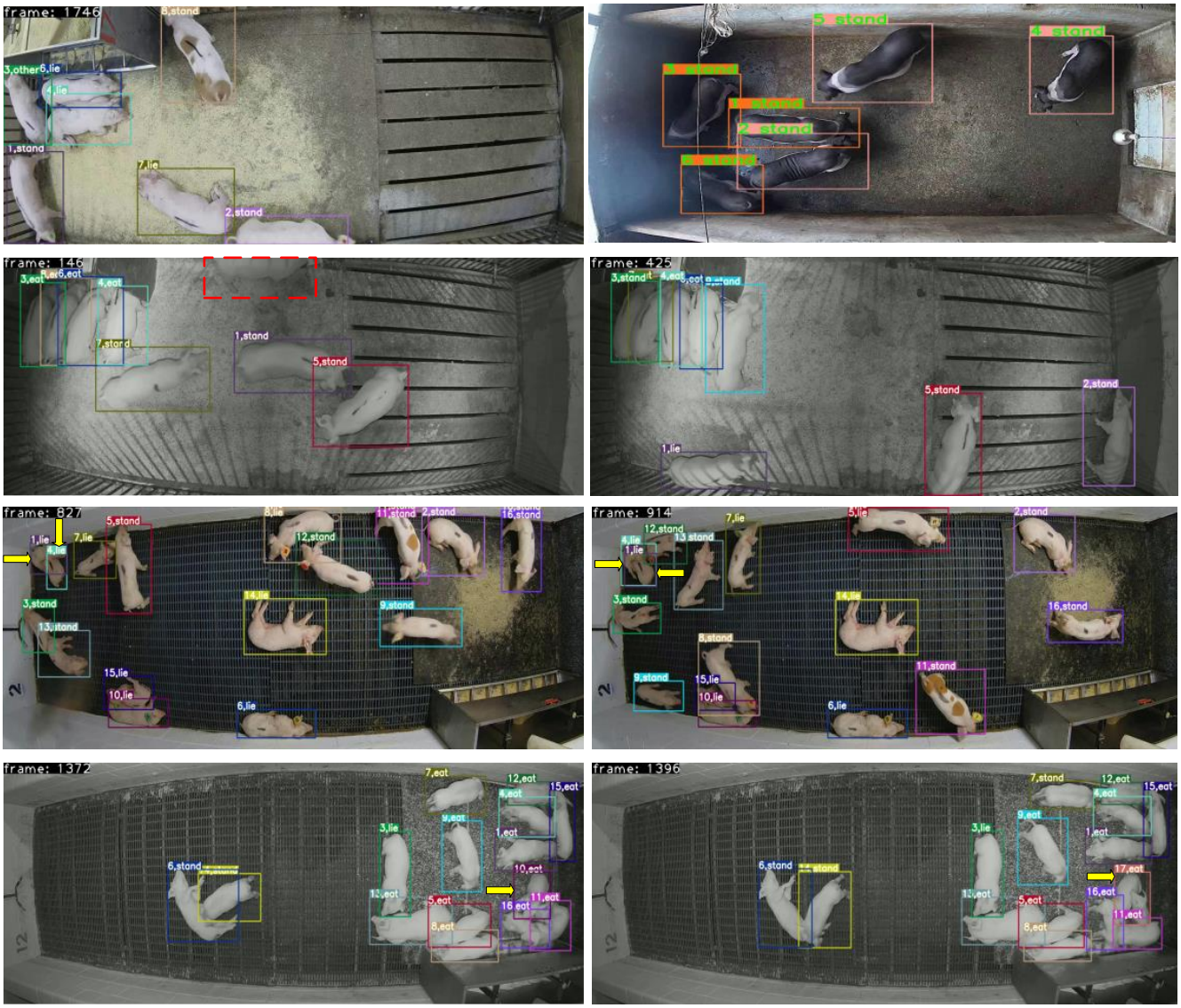


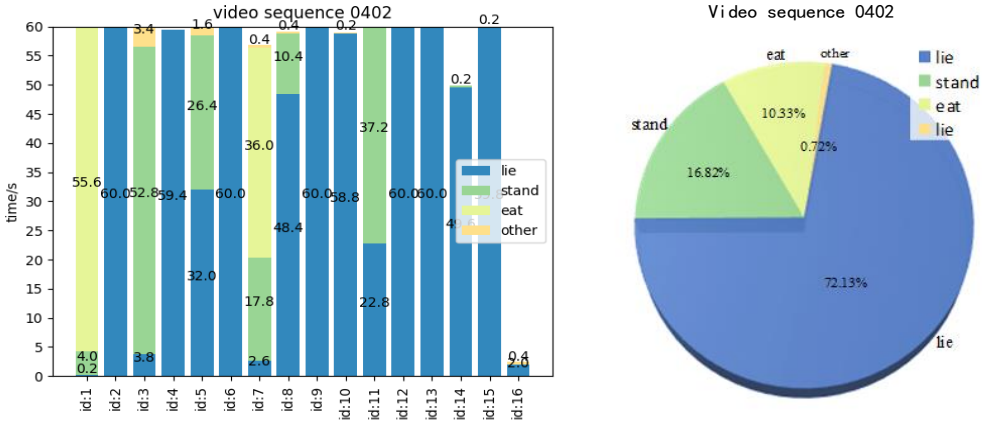
Figure 10. The visualization of long-term tracking results. (The first row shows the tracking results for videos 2001 and 3010, and the second, third, and fourth rows, respectively, depict the tracking results of two frames from videos 2002, 2003 and 2004).

4.4. Results of Behavior Analysis

Our proposed behavior analysis algorithm can record the duration of four types of behaviors for each pig and each pen based on behaviors classification and ID information to represent the pigs' health status and welfare. For example, the analysis results of pig behaviors on test video 0402# are shown in Figures 11 and 12.

In Figure 11a, the time allocation of each pig's ID for four behaviors (lie, stand, eat, and other) is depicted. It can be observed that pigs with a different ID value exhibit variations in time allocation for each behavior, with "lie" and "stand" behaviors occupying a larger portion of the time, while "eat" and "other" behaviors are less prominent. Figure 11b illustrates the percentage of four behaviors of all pigs throughout the entire video segment, with different colors indicating each behavior. It is noticeable that the "lie", "stand", and "eat" behaviors occurred 72.13%, 16.81%, and 10.33% of the time for all pigs, and the 'other' behavior occurred the least at 0.72%. This indicates that the entire herd is in a healthy state. Figure 12 depicts the number of frames occupied by each pig ID, where the horizontal coordinate is the number of frames and the vertical coordinate is the pig's ID value. It is worth noting that the behavior of some pigs varies frequently, such as those pigs with the

IDs with 3, 5, 7, and there are also cases where several pigs engage in a single behavior for a long time. For example, pigs with the IDs 2, 6, 9, 12, 13, and 15 are consistently engaged in "lie" behavior.



(a) The time allocation for each pig's ID across four different behaviors. (b) The behavioral changes of each ID-tagged pig.

Figure 11. Time allocation and proportion of pig behaviors.

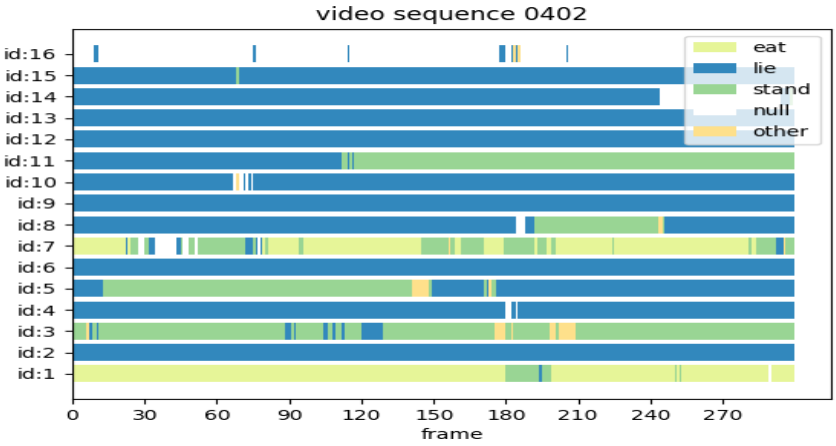


Figure 12. The proportional occurrence of the four behaviors.

Overall, the V8-Sort method can achieve the accurate recognition of pigs' basic behaviors, playing a positive role in optimizing breeding environments and enhancing pig welfare, as well as contributing to the economic efficiency of the breeding industry.

5. Conclusions and Future Work

This paper proposes an algorithm for the identification and tracking of pig behaviors based on V8-Sort. The algorithm aims to decrease noisy interference and improve robustness to occlusions and nonlinear motions and explore the work of recovering the lost trajectory for long-term tracking. On the public dataset, V8-Sort achieved a HOTA of 82.0%, MOTA of 96.3%, and 22 IDs. Compared to Trackformer, JDE, and TransTrack, V8-Sort has shown improvements of 11.2, 19.4, and 18.2 percentage points in HOTA, and improvements of 7.8, 12.9, and 17 percentage points in MOTA, respectively. On the private dataset, V8-Sort has a HOTA of 74.8%, and 17 IDs. Comparing Trackformer and TransTrack, the HOTA is 1.3 and 17.3 percentage points higher, and the IDs are fewer by 24 and 378, respectively. In the long-term videos' dataset, the average HOTA, number of IDs, MOTA, and IDF1 of V8-Sort are 62.6%, 44, 93.1%, and 75.1%, respectively. In conclusion, the V8-Sort can obtain higher accuracy and performance in pig behavior recognition and tracking tasks, which recover the lost trajectory during long-term tracking. This improves the accuracy and efficiency of the automatic monitoring of group-reared pigs under conditions of noisy

interference and scenes with occlusions, providing a more intelligent and reliable solution for pig farming management.

However, the behaviors analyzed in the current study are relatively simple and basic, primarily focusing on obvious behavior patterns such as eating and resting. Important behaviors such as more complex social interactions, stress responses, and environmental adaptability have not yet been addressed. Through obtaining the group-housed behaviors analysis using the MOT method in this paper, we can consider in future research to complete the accurate discrimination of pig health status and welfare by analyzing herd behaviors, including daily and abnormal pig behaviors. In addition, by integrating advanced sensors, data analysis and machine learning techniques, it is possible to establish predictive models for correlating the health status of pigs with each other based on their behaviors. By monitoring key information such as the pigs' activity levels, feeding patterns, and movement trajectories, combined with physiological parameters like body temperature and heart rate, we can obtain the real-time tracking of the health conditions of the pigs to facilitate production management in the modern livestock industry.

Author Contributions: Conceptualization, S.T. and Q.H.; methodology, J.D.; software, Y.L.; validation, D.X., W.C. and Y.C.; formal analysis, W.C.; investigation, Q.H.; resources, Y.L.; data curation, Y.C.; writing—original draft preparation, J.D.; writing—review and editing, S.T.; visualization, J.D.; supervision, S.T.; project administration, Y.L.; funding acquisition, Q.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by key R&D project of Guangzhou (202206010091, 2024B03J1358, 2023B03J1363), Nation key Research and Development Program of China (No. 2023YFF0725005 and No. 2023YFC3905800), and the horizontal project (No. H240794).

Institutional Review Board Statement: The animal study protocol was approved by the Animal Ethics Committee of South China Agricultural University (protocol code 2024F213 and date of approval: 14 March 2024).

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are included in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Matthews, S.G.; Miller, A.L.; Clapp, J.; Plötz, T.; Kyriazakis, I.J.T.V.J. Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. *Vet. J.* **2016**, *217*, 43–51. [CrossRef] [PubMed]
2. Matthews, S.G.; Miller, A.L.; Plötz, T.; Kyriazakis, I. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Sci. Rep.* **2017**, *7*, 17582. [CrossRef] [PubMed]
3. Maselyne, J.; Adriaens, I.; Huybrechts, T.; De Ketelaere, B.; Millet, S.; Vangeyte, J.; Nuffel, A.N.; Saeys, W.J.A. Measuring the drinking behaviour of individual pigs housed in group using radio frequency identification (RFID). *Animal* **2016**, *10*, 1557–1566. [CrossRef] [PubMed]
4. Chapa, J.M.; Maschat, K.; Iwersen, M.; Baumgartner, J.; Drillich, M.J.B.P. Accelerometer systems as tools for health and welfare assessment in cattle and pigs–a review. *Behav. Process.* **2020**, *181*, 104262. [CrossRef] [PubMed]
5. Meijer, E.; Oosterlinck, M.; van Nes, A.; Back, W.; van der Staay, F.J. Pressure mat analysis of naturally occurring lameness in young pigs after weaning. *BMC Vet. Res.* **2014**, *10*, 193. [CrossRef]
6. Arablouei, R.; Wang, Z.; Bishop-Hurley, G.J.; Liu, J. Multimodal sensor data fusion for in-situ classification of animal behavior using accelerometry and GNSS data. *Smart Agric. Technol.* **2023**, *4*, 100163. [CrossRef]
7. Chen, Z.; Lu, J.; Wang, H. A Review of Posture Detection Methods for Pigs Using Deep Learning. *Appl. Sci.* **2023**, *13*, 6997. [CrossRef]
8. Guo, J.; Wu, X.; Liu, J.; Wei, T.; Yang, X.; Yang, X.; He, B.; Zhang, W. Non-contact vibration sensor using deep learning and image processing. *Measurement* **2021**, *183*, 109823. [CrossRef]
9. Dhanya, V.G.; Subeesh, A.; Kushwaha, N.L.; Vishwakarma, D.K.; Nagesh Kumar, T.; Ritika, G.; Singh, A.N. Deep learning-based computer vision approaches for smart agricultural applications. *Artif. Intell. Agric.* **2022**, *6*, 211–229. [CrossRef]
10. Tran, D.; Thanh, N. Pig Health Abnormality Detection Based on Behavior Patterns in Activity Periods usingDeep Learning. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*, 603–610. [CrossRef]
11. Zhang, Y.; Yang, X.; Liu, Y.; Zhou, J.; Huang, Y.; Li, J.; Zhang, L.; Ma, Q. A time-series neural network for pig feeding behavior recognition and dangerous detection from videos. *Comput. Electron. Agric.* **2024**, *218*, 108710. [CrossRef]

12. Alameer, A.; Kyriazakis, I.; Bacardit, J. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs. *Sci. Rep.* **2020**, *10*, 13665. [CrossRef] [PubMed]
13. Yang, Q.; Xiao, D. A review of video-based pig behavior recognition. *Appl. Anim. Behav. Sci.* **2020**, *233*, 105146. [CrossRef]
14. Chen, F.; Liang, X.; Chen, L.; Liu, B.; Lan, Y. Novel method for real-time detection and tracking of pig body and its different parts. *Agric. Food Sci.* **2020**, *13*, 144–149. [CrossRef]
15. Jaoukaew, A.; Suwansantisuk, W.; Kumhom, P. Robust individual pig tracking. *Int. J. Electr. Comput. Eng. (IJECE)* **2024**, *14*, 279–293. [CrossRef]
16. Gong, W.C.; Wang, J.; Mao, L.; Lu, L. A Pig Tracking Algorithm with Improved IOU-Tracker. In Proceedings of the International Conference on Agri-Photonics and Smart Agricultural Sensing Technologies, Zhengzhou, China, 10–12 June 2022.
17. Zhang, L.; Gray, H.; Ye, X.; Collins, L.; Allinson, N. Automatic Individual Pig Detection and Tracking in Pig Farms. *Sensors* **2019**, *19*, 1188. [CrossRef]
18. Tu, S.; Zeng, Q.; Liang, Y.; Liu, X.; Huang, L.; Weng, S.; Huang, Q. Automated Behavior Recognition and Tracking of Group-Housed Pigs with an Improved DeepSORT Method. *Agriculture* **2022**, *12*, 1907. [CrossRef]
19. Zhou, H.; Li, Q.; Xie, Q. Individual Pig Identification Using Back Surface Point Clouds in 3D Vision. *Sensors* **2023**, *23*, 5156. [CrossRef]
20. Hao, W.; Zhang, K.; Zhang, L.; Han, M.; Hao, W.; Li, F.; Yang, G. TSML: A New Pig Behavior Recognition Method Based on Two-Stream Mutual Learning Network. *Sensors* **2023**, *23*, 5092. [CrossRef]
21. Gao, Y.; Yan, K.; Dai, B.; Sun, H.; Yin, Y.; Liu, R.; Shen, W. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatio-temporal attention mechanism. *Comput. Electron. Agric.* **2023**, *205*, 107606. [CrossRef]
22. Ji, H.; Teng, G.; Yu, J.; Wen, Y.; Deng, H.; Zhuang, Y. Efficient Aggressive Behavior Recognition of Pigs Based on Temporal Shift Module. *Animals* **2023**, *13*, 2078. [CrossRef] [PubMed]
23. Meinhardt, T.; Kirillov, A.; Leal-Taixe, L.; Feichtenhofer, C. Trackformer: Multi-object tracking with transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8844–8854.
24. Wang, Z.; Zheng, L.; Liu, Y.; Wang, S. Towards Real-Time Multi-Object Tracking. Computer Vision and Pattern Recognition. *arXiv* **2019**, arXiv:1909.12605.
25. Sun, P.; Cao, J.; Jiang, Y.; Zhang, R.; Xie, E.; Yuan, Z.; Wang, C.; Luo, P.J. Transtrack: Multiple object-tracking with transformer. In Proceedings of the Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
26. Psota, E.T.; Schmidt, T.; Mote, B.; Pérez, L.C. Long-Term Tracking of Group-Housed Livestock Using Keypoint Detection and MAP Estimation for Individual Animal Identification. *Sensors* **2020**, *20*, 3670. [CrossRef] [PubMed]
27. FFmpeg. Available online: https://ffmpeg.org/ (accessed on 1 September 2023).
28. Darklabel. Darklabel Software for Annotating Images and Videos. Available online: https://github.com/darkpgmr/DarkLabel (accessed on 28 July 2023).
29. Seo, Y.W.; Lee, S.G.; Chang, W.D.; Cha, E.Y. A Study on Multi-Object Tracking, Using The Hungarian Algorithm. In Proceedings of the Korea Information Processing Society Conference, Seoul, Republic of Korea, 14–15 May 2004; pp. 777–780.
30. Pujara, A.; Bhamare, M. DeepSORT: Real Time & Multi-Object Detection and Tracking with YOLO and TensorFlow. In Proceedings of the 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 24–26 November 2022; pp. 456–460.

7. Institutional review board statement

The animal study protocol was approved by the Animal Ethics Committee of South China Agricultural University (protocol code 2024F213 and date of approval: 14 March 2024).

CRediT authorship contribution statement

Shuqin Tu: Visualization, Supervision. **Yuefei Cao:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Conceptualization. **Liang Mao:** Data curation. **Yun Liang:** Data curation. **Hairan Yang:** Data curation. **Baiyang Tang:** Data curation. **Fang Yuan:** Data curation.

Funding

The work was supported by Shenzhen Polytechnic University Smart Agriculture Innovation Application R&D Center (No. 602431001PQ), Huizhou Municipal Key Areas Research and Development Project (No. 2024BQ010007), National Natural Science Foundation of China (No. 62272320), Shenzhen Science and Technology Innovation Commission Foundation (No. 20220812222043002), and Shenzhen Polytechnic University Research Fund (No. 6025310045K).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Liang Mao reports financial support was provided by Shenzhen Polytechnic University Smart Agriculture Innovation Application R&D Center. Has patent pending to. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

i have shared the link to my code at the manuscript

References

Meese, G., Ewbank, R., 1973. The establishment and nature of the dominance hierarchy in the domesticated pig. Anim. Behav. 21 (2), 326–334.
Stookey, J.M., Gonyou, H.W., 1994. The effects of regrouping on behavioral and production parameters in finishing swine. J. Anim. Sci. 72 (11), 2804–2811.
Jensen, P., 1982. An analysis of agonistic interaction patterns in group-housed dry sows—aggression regulation through an “avoidance order”. Appl. Anim. Ethol. 9 (1), 47–61.
Fraser, J., Aricibasi, H., Tulpan, D., & Bergeron, R. (2023). A computer vision image differential approach for automatic detection of aggressive behavior in pigs using deep learning. *Journal of Animal Science, 101*, skad347.
Stukenborg, A., Traulsen, I., Puppe, B., Presuhn, U., Krieter, J., 2011. Agonistic behaviour after mixing in pigs under commercial farm conditions. Appl. Anim. Behav. Sci. 129 (1), 28–35.
Xia, X., Zhang, N., Guan, Z., Chai, X., Ma, S., Chai, X., Sun, T., 2025. PAB-Mamba-YOLO: VSSM assists in YOLO for aggressive behavior detection among weaned piglets. Artif. Intell. Agric. 15 (1), 52–66.
Yan, K., Dai, B., Liu, H., Yin, Y., Li, X., Wu, R., Shen, W., 2024. Deep neural network with adaptive dual-modality fusion for temporal aggressive behavior detection of group-housed pigs. Comput. Electron. Agric. 224, 109243.
Gao, Y., Yan, K., Dai, B., Sun, H., Yin, Y., Liu, R., Shen, W., 2023. Recognition of aggressive behavior of group-housed pigs based on CNN-GRU hybrid model with spatio-temporal attention mechanism. Comput. Electron. Agric. 205, 107606.
Ji, H., Teng, G., Yu, J., Wen, Y., Deng, H., Zhuang, Y., 2023. Efficient aggressive behavior recognition of pigs based on temporal shift module. Animals 13 (13), 2078.
Chen, C., Zhu, W., Steibel, J., Siegford, J., Wurtz, K., Han, J., Norton, T., 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. Comput. Electron. Agric. 169, 105166.
Fu, Y., Wang, Z., Zheng, H., Yin, X., Fu, W., Gu, Y., 2025. Integrated detection of coconut clusters and oriented leaves using improved YOLOv8n-obb for robotic harvesting. Comput. Electron. Agric. 231, 109979.
Li, P., Chen, J., Chen, Q., Huang, L., Jiang, Z., Hua, W., Li, Y., 2025. Detection and picking point localization of grape bunches and stems based on oriented bounding box. Comput. Electron. Agric. 233, 110168.
Lu, J., Chen, Z., Li, X., Fu, Y., Xiong, X., Liu, X., Wang, H., 2024. ORP-Byte: a multi-object tracking method of pigs that combines Oriented RepPoints and improved Byte. Comput. Electron. Agric. 219, 108782.
Wei, J., Tang, X., Liu, J., Zhang, Z., 2023. Detection of pig movement and aggression using deep learning approaches. Animals 13 (19), 3074.
Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., Norton, T., 2020. A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. Biosyst. Eng. 195, 27–41.
McGlone, J.J., 1985. A quantitative ethogram of aggressive and submissive behaviors in recently regrouped pigs. J. Anim. Sci. 61 (3), 556–566.
D'Eath, R.B., Turner, S.P., 2009. The natural behaviour of the pig. In: *The Welfare of Pigs*. Springer, pp. 13–45.
Wang, W., 2023. Advanced Auto labeling solution with added Features. Retrieved from, Github repository https://github.com/CVHub520/X-AnyLabeling.
Chen, C., Zhu, W., Guo, Y., Ma, C., Huang, W., Ruan, C., 2018. A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs. Livest. Sci. 218, 70–78.
Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.
Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., . . . Wang, X. (2022). ByteTrack: Multi-object Tracking by Associating Every Detection Box. In *Computer Vision – ECCV 2022* (pp. 1-21).
Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. *Simple online and realtime tracking*. Paper Presented at the 2016 IEEE International Conference on Image Processing (ICIP).
Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., Leibe, B., 2021. Hota: a higher order metric for evaluating multi-object tracking. Int. J. Comput. Vis. 129 (2), 548–578.
Bernardin, K., Stiefelhagen, R., 2008. Evaluating multiple object tracking performance: the clear mot metrics. EURASIP J. Image Video Processing 2008 (1), 246309.
Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*.
Cao, J., Pang, J., Weng, X., Khirodkar, R., Kitani, K., 2023. *Observation-centric sort: Rethinking sort for robust multi-object tracking*. Paper Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
Maggiolino, G., Ahmad, A., Cao, J., Kitani, K., 2023. *Deep oc-sort: Multi-pedestrian tracking by adaptive re-identification*. Paper Presented at the 2023 IEEE International Conference on Image Processing (ICIP).
Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*.
Tu, S., Cai, Y., Liang, Y., Lei, H., Huang, Y., Liu, H., Xiao, D., 2024a. Tracking and monitoring of individual pig behavior based on YOLOv5-Byte. Comput. Electron. Agric. 221, 108997.
Tu, S., Cao, Y., Liang, Y., Zeng, Z., Ou, H., Du, J., Chen, W., 2024b. Tracking and automatic behavioral analysis of group-housed pigs based on YOLOX+ BoT-SORT-slim. Smart Agric. Technol. 9, 100566.

agriculture

MDPI

Article

RpTrack: Robust Pig Tracking with Irregular Movement Processing and Behavioral Statistics

Shuqin Tu [1], Hua Lei [1], Yun Liang [1,*], Enli Lyu [2] and Hongxing Liu [1]

[1] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China; tsq5_6@scau.edu.cn (S.T.); 20223170042@stu.scau.edu.cn (H.L.); 20223170060@stu.scau.edu.cn (H.L.)
[2] College of Engineering, South China Agricultural University, Guangzhou 510642, China; enlilv@scau.edu.cn
* Correspondence: liang_yun168@163.com

Abstract: Pig behavioral analysis based on multi-object tracking (MOT) technology of surveillance videos is vital for precision livestock farming. To address the challenges posed by uneven lighting scenes and irregular pig movements in the MOT task, we proposed a pig MOT method named RpTrack. Firstly, RpTrack addresses the issue of lost tracking caused by irregular pig movements by using an appropriate Kalman Filter and improved trajectory management. Then, RpTrack utilizes BIoU for the second matching strategy to alleviate the influence of missed detections on the tracking performance. Finally, the method utilizes post-processing on the tracking results to generate behavioral statistics and activity trajectories for each pig. The experimental results under conditions of uneven lighting and irregular pig movements show that RpTrack significantly outperforms four other state-of-the-art MOT methods, including SORT, OC-SORT, ByteTrack, and Bot-SORT, on both public and private datasets. The experimental results demonstrate that RpTrack not only has the best tracking performance but also has high-speed processing capabilities. In conclusion, RpTrack effectively addresses the challenges of uneven scene lighting and irregular pig movements, enabling accurate pig tracking and monitoring of different behaviors, such as eating, standing, and lying. This research supports the advancement and application of intelligent pig farming.

Keywords: multi-object tracking; pigs; RpTrack; behavioral statistics; uneven lighting scenes; irregular pig movements

Citation: Tu, S.; Lei, H.; Liang, Y.; Lyu, E.; Liu, H. RpTrack: Robust Pig Tracking with Irregular Movement Processing and Behavioral Statistics. *Agriculture* **2024**, *14*, 1158. https://doi.org/10.3390/agriculture14071158

Academic Editors: Claudia Arcidiacono and Sabina Angrecka

Received: 23 May 2024
Revised: 2 July 2024
Accepted: 10 July 2024
Published: 16 July 2024

1. Introduction

Pig farming holds a significant position in the livestock industry. In the field of pig farming, the application of precision livestock farming can reduce production costs, enhance productivity, improve animal welfare, meet food demand, and boost economic benefits [1,2]. To achieve precision livestock farming for pigs, MOT technology is crucial for the pig industry. It enables timely monitoring of the health and behavior of pigs via individual pig re-identification and behavioral analysis in video surveillance [3]. This technology provides the foundation for the realization of precision pig farming.

In recent research, several MOT methods have been applied and expanded to the livestock industry. For example, Zhang et al. introduced an online method for detecting and tracking multiple pigs, which removed the need for manual annotations or actual pig IDs and operated effectively both in daytime and nighttime conditions [4]. Cowton et al. [5] integrated Faster R-CNN [6] with two online multi-object tracking techniques, namely SORT and DeepSORT [7], creating a comprehensive system for individual pig localization and tracking. This system also extracted behavior-related metrics from RGB camera data, achieving impressive performance with MOTA and IDF1 scores reaching 92% and 73.4%, respectively. Guo et al. [8] proposed a weighted association algorithm for three multi-object tracking methods, namely JDE [9], FairMOT [10], and YOLOv5s-DeepSORT, to optimize pig re-identification, improve tracking performance, and reduce ID switches.

Tu et al. [11] combined YOLOv5 [12,13] with an improved DeepSORT algorithm to achieve highly accurate pig tracking, with a re-identification accuracy rate of up to 99.9%. Kim et al. improved YOLOv4 and DeepSORT to reduce computational expense while maintaining high accuracy in pig counting, which enabled real-time execution with an accuracy rate of 99.44% [14]. Odo et al. [15] utilized YOLOv4 [16] and YOLOv7 [17] detectors, combined with the DeepSORT and centroid tracking algorithms, to quantify ear-biting behavior in pigs. This approach achieved a detection accuracy of 98% and a tracking false positive rate of 14%. Han et al. proposed a detection-based tracking approach that utilized a YOLOv5 detector trained specifically for cattle detection to generate detection results. This method was designed to overcome challenges posed by scale variations, random movements, and occlusions in farm conditions when tracking cattle [18]. These algorithms all utilized the appearance information of targets during the association phase. However, due to the high similarity in appearance among the pigs' targets in the scenes and the uneven lighting conditions, the appearance information of the targets might be unreliable. Additionally, using appearance information is time-consuming and may not yield significant benefits in such scenarios.

Therefore, there are many studies that do not use the appearance information of targets in the data association phase for pig MOT under video surveillance monitoring. Yigui et al. proposed an improved pig-counting algorithm based on the YOLOv5+DeepSORT model, which achieves stable pig counting in a breeding environment with a 98.4% pig-counting correlation coefficient [19]. Zheng et al. [20] proposed a MOT method that effectively addressed the issues of false negatives and false positives resulting from complex environmental conditions in individual cow detection and tracking. Van der Zande et al. [21] combined YOLOv3 [22] with the SORT algorithm to achieve pig detection and tracking to monitor individual pig activity. These algorithms only used the target's motion information during the association phase, combined with the Hungarian algorithm, to achieve cross-frame target association. Due to the irregular movement patterns of targets in livestock tracking scenarios, which include variations in speed and sudden changes in direction, using Kalman Filters (KFs) in the traditional manner may not yield accurate bounding box positions, leading to the loss of target tracking. Additionally, the above research primarily focuses on the identification or automatic tracking of pig behaviors, lacking analysis of the time of the pig's different behaviors, such as eating, standing, and lying, according to the post-processing of the tracking results.

To address the aforementioned issues, based on the SORT algorithm [23], this paper introduces a robust pig-tracking method named RpTrack. Firstly, in the detection phase, this method utilizes the robust YOLOX [24] detector to generate detection results for each video frame. Then, in the association phase, RpTrack effectively deals with ID switches caused by irregular pig movement in pig tracking scenarios. This is achieved through improved trajectory management combined with a KF appropriate for pig tracking. Furthermore, it incorporates BIoU to mitigate the influence of missed detections due to uneven lighting conditions on tracking performance. Finally, we monitor and analyze the time of the pig's different behaviors, such as eating, standing, and lying, according to the post-processing of the tracking results throughout the video.

2. Materials and Methods

2.1. Materials

The video data were obtained in two parts: one part was provided by T. Psota et al. [25], and the other part was captured at the Lejiazhuang Breeding Base in Sanshui District, Foshan City, Guangdong Province, China. The cameras were installed directly above the central area of the pig breeding zone, capturing the entire area from a downward perspective. The resolution of the camera was 2560 × 960 pixels, the frame rate was 25 fps, and the video data were stored in MP4 format. In the subsequent data processing, we selected 27 videos, with 23 1-min videos and 4 10-min videos cropped from the one part for the public dataset and 18 1-min videos chosen from the other part for the private

dataset. Notably, the annotated dataset from T. Psota et al. was labeled with each pig's shoulder, tail, and ID for all frames of each video and could not be used for mainstream MOT tasks. Therefore, we annotated all video segments, including the public and private datasets, using DarkLabel1.3 software at 5 frames per second (fps) to evaluate the tracking performance of pig behaviors. The video scenes are illustrated in Figure 1. It includes daytime sparse and daytime dense scenes, as shown in Figure 1a,b, and nighttime even light and uneven light scenes, as shown in Figure 1c,d.

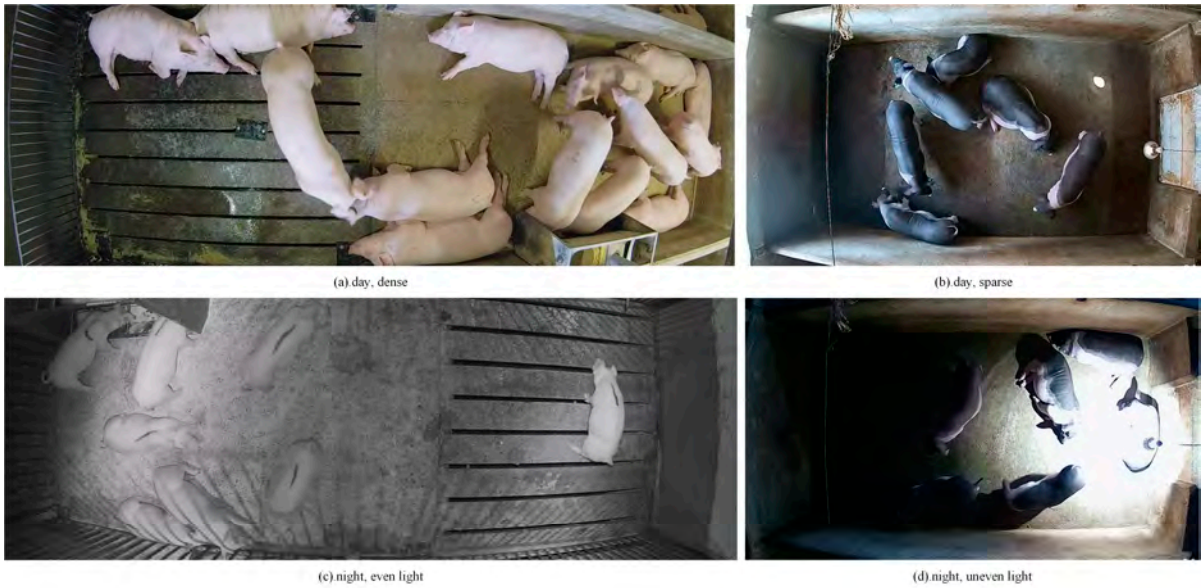


Figure 1. Dataset video scenes.

In this study, all videos in both datasets were manually annotated using the DarkLabel software. The annotation includes the identity, position, and behavior categories of the pigs. The behaviors are categorized into four types ("stand", "lie", "eat," and "other"). Examples of behavioral classification for some pigs are shown in Figure 2. To compare the tracking performance of the proposed method in different scenarios, this paper selected 15 videos from the public dataset as test videos. Additionally, 9 videos from the private dataset were chosen as test videos. Meanwhile, based on manual observations, videos with a higher number of pigs and more occlusions were categorized as pig-dense videos and vice versa as pig-sparse videos. Additionally, videos with relatively even lighting conditions were labeled as even lighting scenes, and videos with uneven lighting conditions were labeled as uneven lighting scenes, as shown in Figure 1. Table 1 presents a detailed description of the test video environments in both the public and private datasets. Note: No. 01, 05, 11, and 15 are 10 min videos; the rest are 1 min videos.

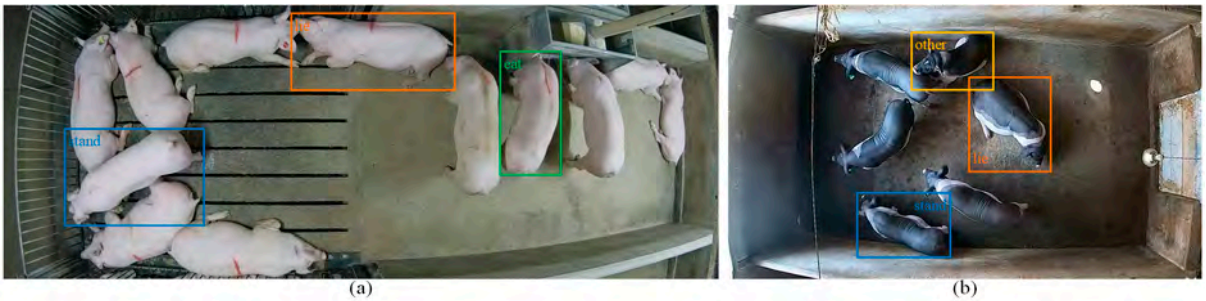


Figure 2. Examples of pig behavior classification. (**a**) Examples of "stand", "lie," and "eat" behavioral categories. (**b**) Examples of "stand", "lie," and "other" behavioral categories.

Table 1. Test video environments.

Dataset	No.	Sparse	Dense	Day	Night	Light
Public	0102	√	—	√	—	uniform
	0402	—	√	√	—	uniform
	0502	√	—	—	√	uniform
	0602	—	√	√	—	uniform
	0702	√	—	√	—	uniform
	0802	—	√	—	√	uniform
	0902	—	√	√	—	uniform
	1002	—	√	—	√	uneven
	1102	√	—	√	—	uniform
	1202	√	—	√	—	uniform
	1502	—	√	—	√	uniform
	01	√	—	√	—	uniform
	05	√	—	—	√	uniform
	11	—	√	√	—	uniform
	15	—	√	—	√	uniform
Private	0010	—	√	√	—	uniform
	0011	—	√	√	—	uniform
	0012	—	√	—	√	uneven
	0013	—	√	√	—	uniform
	0014	√	—	√	—	uniform
	0015	√	—	√	—	uniform
	0016	√	—	√	—	uniform
	0017	√	—	—	√	uniform
	0018	√	—	—	√	uneven

2.2. Methods

The robust pig tracking method named RpTrack is illustrated in Figure 3. Firstly, during the input phase, the YOLOX detector is used to obtain the detection results, including the bounding box positions, behavior categories, and confidence scores. Meanwhile, the improved trajectory prediction is applied to predict the current position for each trajectory in the trajectory set (except in the case of the first frame). Then, during the tracking phase, the current detection results and the trajectory prediction results are taken as inputs. The first matching is based on Intersection over Union (IoU). Unmatched detections and trajectories in the first matching are performed in a second matching based on BIoU. Finally, after completing the entire video tracking process, behavioral statistic information can be obtained on the behavioral states of the pigs and their activity trajectories within the video. Compared to the SORT method, RpTrack made improvements in three key components: the Kalman Filter, the trajectory management, and the BIoU. The following sections describe each of these components in detail.

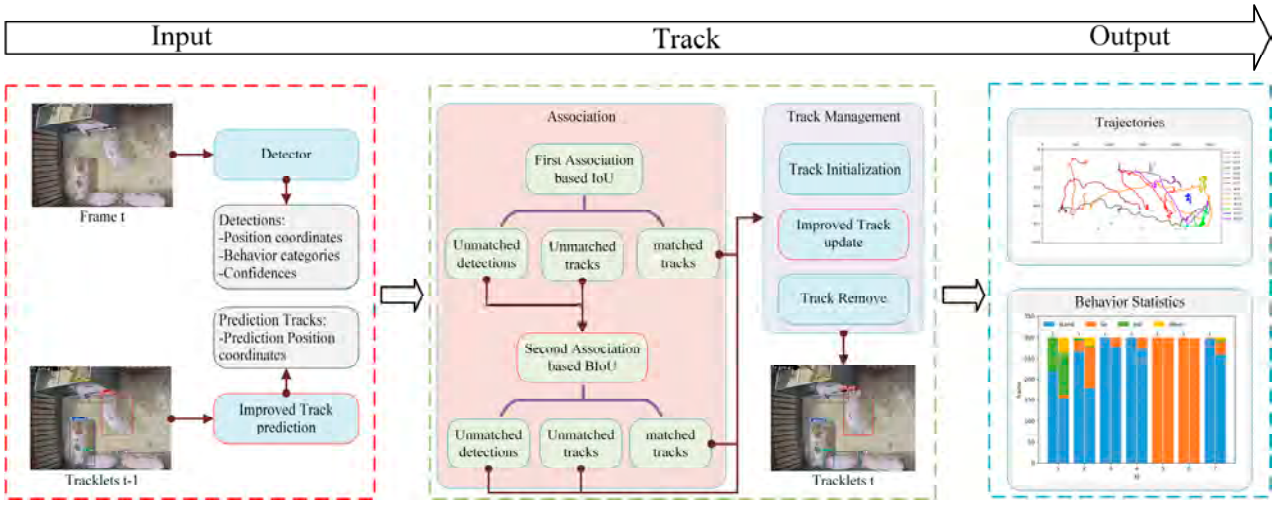


Figure 3. RpTrack tracking pipeline.

2.2.1. Improved Kalman Filter

In the SORT algorithm, the KF [26] state vector used is denoted as $x = [x_c, y_c, s, a, \hat{x}_c, \hat{y}_c, \hat{s}]^T$. Here, (x_c, y_c) represents the center coordinates of the target, s and a, respectively, denote the area and aspect ratio of the bounding box, and $[\hat{x}_c, \hat{y}_c, \hat{s}]$ is the velocity corresponding to these parameters.

In the pig tracking environment, this type of state vector may struggle to obtain accurate bounding box shapes. The predicted bounding boxes might not completely and accurately surround the pigs, thereby affecting the overall performance of the tracker. The KF state vector used in the RpTrack algorithm is denoted by $x = [x_c, y_c, w, h, \hat{x}_c, \hat{y}_c, \hat{w}, \hat{h}]^T$ as in the Bot-SORT [27] algorithm. Here, (x_c, y_c) represents the center coordinates of the target, w and h, respectively, denote the width and height of the bounding box, and $[\hat{x}_c, \hat{y}_c, \hat{w}, \hat{h}]$ is the velocity corresponding to these parameters. This state vector can provide more accurate bounding box shapes.

The visualization results of the KF state vector bounding boxes used in SORT and RpTrack are compared in Figure 4. The red dashed line and the green solid line represent the visualization results of the KF state vector bounding boxes in SORT and RpTrack, respectively. It can be observed that the green solid line can more completely and accurately surround the pigs. Therefore, RpTrack can enhance tracking performance compared with SORT.

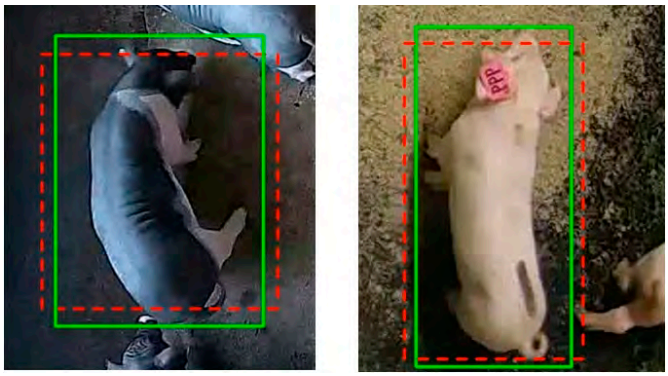


Figure 4. Comparison of visualization results of KF in SORT and RpTrack.

2.2.2. Improved Trajectory Management

The tracking scenarios for pedestrians typically exhibit different movement patterns compared to those for pigs. In pig tracking, the irregular movements of pigs make it challenging to apply pedestrian tracking methods directly. This often results in unsatisfactory tracking results. Most current motion-based MOT methods rely on utilizing position information observed across the entire trajectory for prediction. However, this approach is not effective in handling the irregular movement patterns of pigs. Therefore, in this study, we employ the position information observed in the last K frames of a trajectory (K is a hyperparameter) to predict the trajectory's position in the next frame. This approach allows for more accurate position predictions and addresses the issue of irregular pig movements. Improvements in the storage of trajectory information are necessary.

The improved trajectory information storage is illustrated in Figure 5b. Each long block represents a list for storing motion information, where K is the maximum length of the list. Each colored block in a long block represents the motion information stored in a trajectory. The list representing the storage of motion information for a trajectory can be denoted as $T = \left\{ [(x_i^1, P_i^1), (x_i^2, P_i^2), \ldots, (x_i^k, P_i^k)], s_i, c_i \right\}_{i=1}^{N^t}$, where $k \leq K$. Here, i denotes the trajectory index, x_i^k represents the KF state recording the position information of trajectory i for the last k frames, P_i^k indicates the corresponding covariance matrix for x_i^k, s_i represents the state of trajectory i (tracked or lost), c_i denotes the behavior category of trajectory i ("stand", "lie", "eat", and "other"), N^t represents the number of trajectories in frame t, and (x_i^k, P_i^k) denotes the motion information of trajectory i recording the position information of the last k frames.

The storage of trajectory information before improvement is depicted in Figure 5a, where each trajectory stores only one motion information, recording all observed position information for the trajectory. Additionally, each trajectory does not store the behavior state.

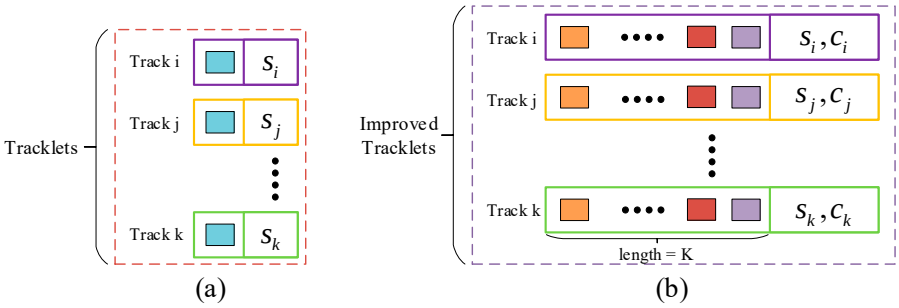


Figure 5. Improved before and after trajectory information storage: (**a**) the trajectory information storage before improvement. (**b**) The improved trajectory information storage.

To achieve accurate trajectory predictions, we made improvements to the trajectory prediction module. These improvements allowed the position information observed in the last K frames to be used to predict the position of the trajectory in the next frame. Figure 6 depicts the trajectory prediction procedure. If the trajectory state is "lost" or the behavior state is "lie", KF prediction processing is not applied to the trajectory, which is subsequently directly used in the association phase. If not, all the motion information stored for the trajectory is processed with the KF prediction, and the results are used in the association phase. Note: For each trajectory, only the prediction result of the motion information recording the most frames of position information is used for the association phase.

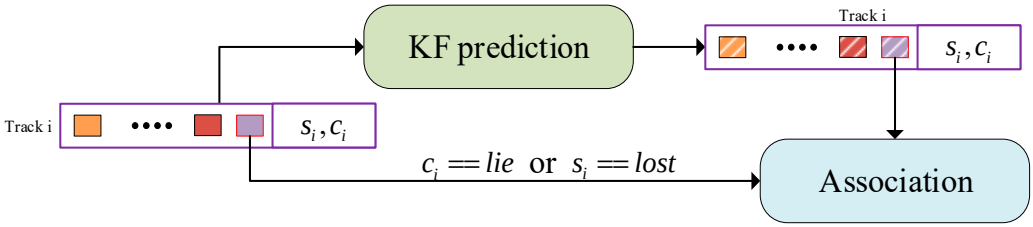


Figure 6. Trajectory prediction process. The red-bordered blocks indicate the motion information recording the most positional information.

The comparison between the improved and unimproved trajectory prediction results is depicted in Figure 7. In Figure 7, the green solid line represents the results obtained with the improved trajectory prediction, while the red dashed line represents the results without improvement. The more accurately the bounding box surrounds the pigs, the more accurate the prediction result. Figure 7a–c represent three motion patterns: slow movement, sudden turning, and fast movement. It can be observed that under all three different motion patterns, the improved trajectory prediction consistently yields more accurate results.

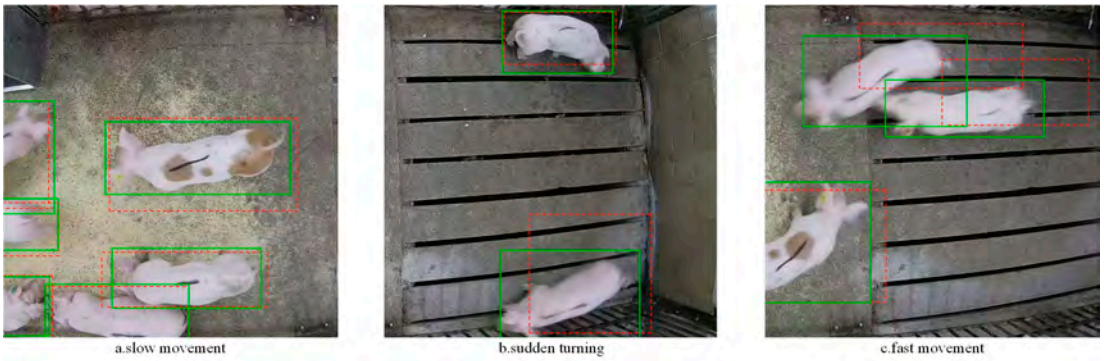


Figure 7. Comparison of trajectory prediction results with and without the improvement.

To maintain the accuracy of the stored motion information for trajectories, improvements are required in the trajectory update module. The specific implementation details are illustrated in Algorithm 1.

Algorithm 1: Improved trajectory update

Input: $t = \{[(x^1, P^1), (x^2, P^2), \ldots, (x^k, P^k)], s, c\}$: represents the storage information for trajectory t, x^k represents the KF state recording the position information of trajectory t over the last k frames, with P^k as the corresponding covariance matrix, s indicates the trajectory state (lost or tracked), and c represents the behavior category of trajectory t.
$d = \{l, c^d\}$: l indicates the position of detection d matched with trajectory t, where c^d denotes the behavior category of the detected target.
K: indicates the maximum length of the trajectory information storage list.
Output: $t = \{[(x^1, P^1), (x^2, P^2), \ldots, (x^k, P^k)], s, c\}$: represents the updated storage information for trajectory t, x^k represents the updated Kalman filter state, P^k represents the corresponding updated covariance matrix for x^k, c denotes the behavior category of the trajectory, and s reflects the trajectory state.

/*update historical KF state vectors and KF covariance matrixes*/
Step 1: Initialize an empty set
$t' \leftarrow \emptyset$
Step 2: Initialize KF for detection d
$x, P \leftarrow$ KF initialization d
$t' \leftarrow t' \cup \{x, P\}$ /*store current motion information*/
Step 3:update motion information
  if s == Tracked **then**
    for i $\leftarrow$ 1 **to** k **do**
      if i < K **then** /*ensure that maximum K frames of position information are recorded*/
        $x^i, P^i \leftarrow$ KF update x^i, P^i, x /*conduct KF update*/
        $t' \leftarrow t' \cup \{x^i, P^i\}$
$t \leftarrow t'$
$c \leftarrow c^d$ /*update trajectory behavior category*/
Return t

2.2.3. BIoU (Buffered IoU)

In the MOT task, motion-based target association algorithms commonly employ the IoU metric. The calculation of IoU is depicted in the left part of Figure 8, where A $= (w_A, h_A)$ and B $= (w_B, h_B)$, respectively, denote the bounding box information for detection and trajectory, w_A and h_A, respectively, denote the width and height of detection A, w_B and h_B, respectively, represent the width and height of trajectory B, and S_1 corresponds to the area of overlap between A and B, while S_2 represents the area of their union.

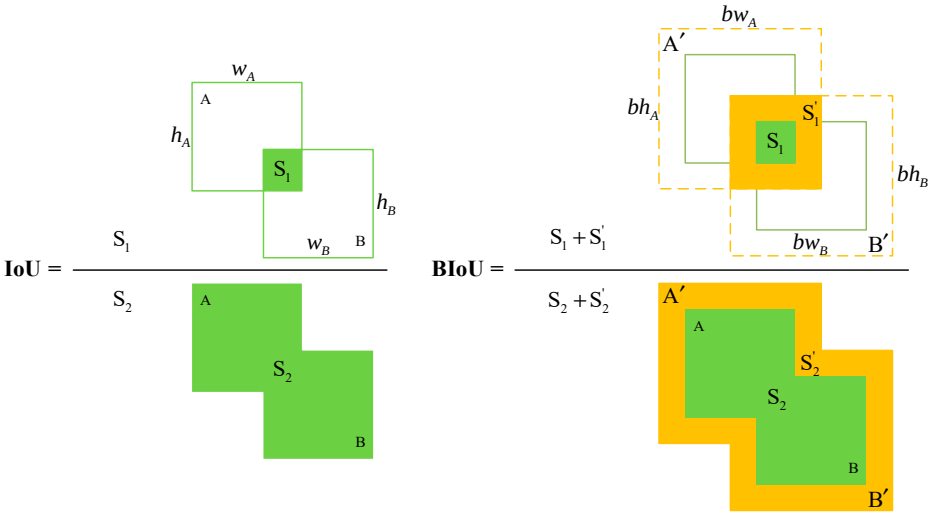


Figure 8. Calculation of IoU and BIoU.

To accurately associate targets with irregular motion patterns and similar appearances, Yang et al. introduced the BIoU (Buffered Intersection over Union) metric [28]. The computation of BIoU is illustrated in the right part of Figure 8, where A′ $= (bw_A, bh_A)$

and $B\prime = (bw_B, bh_B)$ denote the extended bounding box information and b represents a hyperparameter expansion factor. S_1' represents the extended area of overlap and S_2' corresponds to the extended area of their union. (Note that, as in [28], the value of b defaults to 1.5.)

3. Experiments

3.1. Experimental Platform and Parameter Settings

To validate the performance of the proposed RpTrack method in indoor pig tracking and behavioral statistics, three experiments were conducted: (1) a pig tracking experiment to analyze the performance of the RpTrack method; (2) a pig behavior statistics experiment to measure the duration of various behaviors for each pig in the videos; (3) an ablation experiment to evaluate the influences of the improved Kalman Filter, the improved trajectory management, and the BIoU on the tracking performance.

All experiments in this paper were conducted on the same computer, using Linux as the experiment platform with the Ubuntu 20.04 operating system. The hardware configuration included a 12th Gen Intel(R) i9-12900KF CPU, NVIDIA (Santa Clara, CA, USA) GeForce RTX 3090 GPU, 32GB of RAM, PyTorch version 1.11.1, Python version 3.7, and CUDA version 11.3.

3.2. Evaluation Metrics for Multi-Objective Tracking

We selected High Order Tracking Accuracy (HOTA) [29], Multiple Object Tracking Accuracy (MOTA) [30], and Identification F1 (IDF1) as the evaluation metrics for MOT of pigs. HOTA introduces a higher-dimensional tracking accuracy metric, which comprehensively assesses the performance of trackers. MOTA is used to measure the performance of the detector in detecting targets and the tracker in maintaining trajectories. IDF1 is employed to assess the stability of the tracker.

Additionally, in this study, the evaluation of algorithm performance also used two other metrics: the total number of identity switches (IDSW) and the frames per second (FPS) processed by the algorithm.

3.3. Tracking Results

To validate the performance of the proposed RpTrack algorithm, we used the public dataset, consisting of 11 videos, and the private dataset, containing nine videos, as the test videos. Due to the different farming environments and large differences in pig breeds and appearance between the public and private datasets, we used two different YOLOX-X models to complete the tracking experiments. The tracking experiment results of RpTrack are presented in Table 2.

As shown in Table 2, in the public dataset, video 0802 achieved the highest HOTA (92.1%). For all test videos, RpTrack achieved MOTA and IDF1 values of over 96% and 98%, respectively. It also maintained a frame rate (FPS) of 65 or higher. This indicates that RpTrack performs well in terms of tracking pigs in the public dataset while maintaining a fast-tracking speed. However, the tracking performance of videos 11 and 15 is much lower than the other videos, which is due to the existence of a large number of occlusions in videos 11 and 15, resulting in missed and false detections. This, in turn, results in low tracking performance, as shown by the red dashed boxes in Figure 9a,b. In the private dataset, video 0015 achieved the highest HOTA (85.5%). Except for video 0018, all other videos exhibited MOTA and IDF1 values of 97% and above. Except for videos 0010 and 0018, all videos had an ID Switch (IDSW) count of 0. The average FPS for the test videos was 71 or higher, demonstrating that RpTrack excels in terms of tracking accuracy and speed in the private dataset environment. Furthermore, video 0018 has a lower performance compared to the other test videos. This is mainly due to false and missed detections caused by low lighting conditions, as depicted by the red dashed bounding boxes in Figure 9c. The results of the RpTrack method in the public and private datasets indicate that the method has excellent tracking accuracy and real-time performance and can be applied to video surveillance pig tracking in complex scenarios.

Table 2. Results of tracking experiments on the test video.

Dataset	Video	HOTA/%	MOTA/%	IDF1/%	IDSW/%	FPS/(f/s)
Public dataset	0102	90.6	99.9	100.0	0	73.8
	0402	90.1	99.8	100.0	0	70.1
	0502	84.8	99.8	100.0	0	73.6
	0602	80.0	98.6	99.3	0	69.0
	0702	87.5	99.9	100.0	0	71.7
	0802	92.1	100.0	100.0	0	71.3
	0902	88.8	96.6	98.3	0	70.2
	1002	73.8	97.1	98.6	0	70.3
	1102	88.8	97.1	98.6	0	69.1
	1202	86.0	97.1	98.5	2	69.6
	1502	77.6	98.4	99.2	0	68.5
	01	77.0	96.6	93.9	8	75.4
	05	79.0	98.8	94.4	12	73.6
	11	61.8	94.2	71.8	54	68.7
	15	66.4	91.8	78.9	69	69.2
Private dataset	0010	80.0	99.0	97.8	2	65.7
	0011	81.4	97.1	98.5	0	69.7
	0012	79.0	98.2	99.1	0	69.6
	0013	81.9	97.6	98.8	0	69.6
	0014	84.1	99.9	100.0	0	74.7
	0015	85.5	99.6	99.8	0	74.9
	0016	86.1	99.8	99.9	0	74.3
	0017	81.2	99.9	100.0	0	73.6
	0018	66.2	88.2	90.7	4	74.8

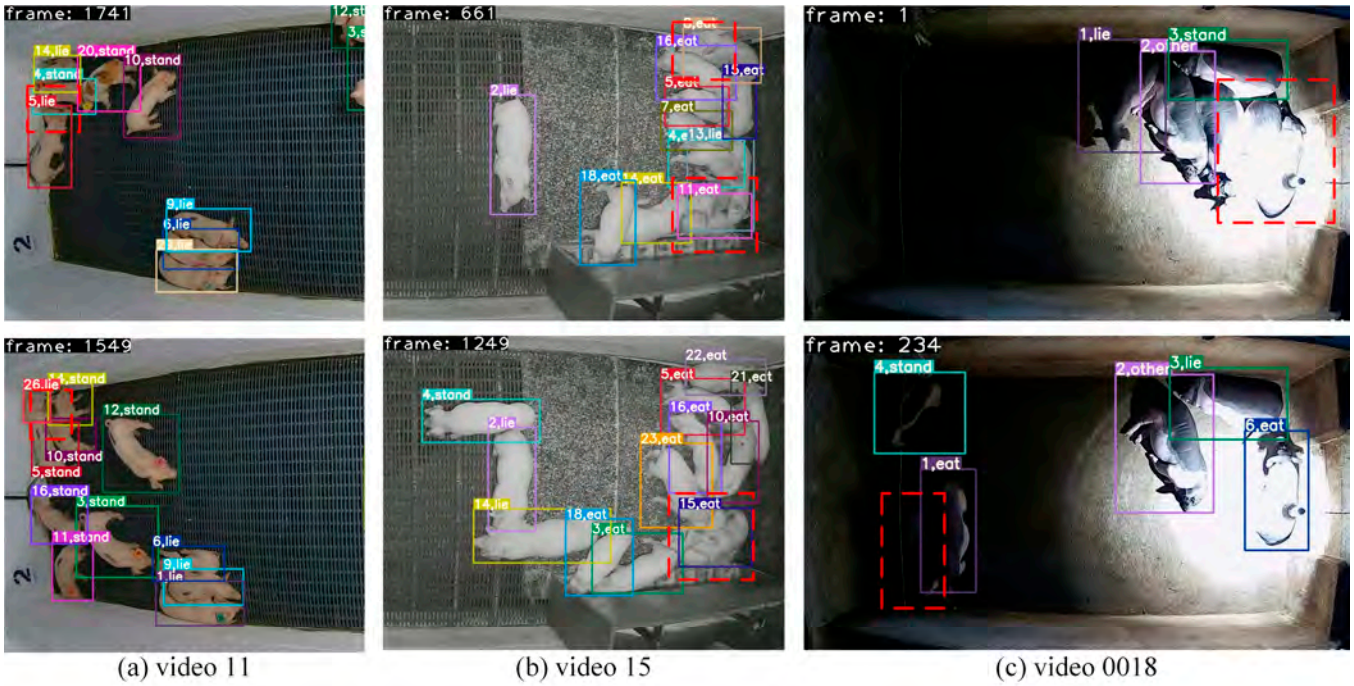


Figure 9. Tracking video environments. (**a**) Examples of video 11 missed detections. (**b**) Examples of video 15 missed detections. (**c**) Examples of video 0018 missed detections.

3.4. Comparison of Different MOT Algorithms

For both the public and private datasets, the experimental results for SORT, C-BIoU [28], ByteTrack [31], OC-SORT [32], Bot-SORT, and the proposed RpTrack are presented in Table 3. In both datasets, RpTrack outperforms other methods in terms of HOTA, MOTA, IDF1, and IDSW while maintaining a high FPS.

Table 3. Comparison of RpTrack results with other MOT methods.

Dataset	Method	HOTA/%	MOTA/%	IDF1/%	IDSW	FPS/(f/s)
Public	SORT	65.2	95.0	72.8	242	75.3
	ByteTrack	61.6	92.8	72.6	229	73.8
	C-BIoU	70.1	95.2	79.7	369	**79.1**
	OC-SORT	70.2	95.2	81.0	161	73.1
	Bot-SORT	69.1	95.1	78.8	317	19.2
	RpTrack	**73.2**	**95.5**	**85.6**	**146**	70.9
Private	SORT	77.7	97.4	93.0	29	78.6
	ByteTrack	73.3	93.2	90.1	41	79.3
	C-BIoU	76.8	95.4	91.7	45	**80.7**
	OC-SORT	78.6	97.4	94.3	18	78.1
	Bot-SORT	78.8	97.0	93.4	35	37.4
	RpTrack	**80.8**	**97.8**	**98.4**	**6**	72.9

The results demonstrate that the proposed RpTrack outperforms SORT, C-BIoU, ByteTrack, OC-SORT, and Bot-SORT in the pig tracking scenarios. This is attributed to the improved Kalman Filter, the improved trajectory management, and BIoU in RpTrack. These improvements effectively handle irregular pig movement, false detections, and missed detections, reducing ID switches and enhancing tracking performance. Therefore, the RpTrack algorithm's performance metrics are superior to other algorithms, indicating the effectiveness of RpTrack for MOT of pigs in complex scenarios.

The visualization results of Bot-SORT, ByteTrack, OC-SORT, and RpTrack methods in both the public and private datasets are presented in Figures 10 and 11. In these figures, the red-dashed bounding boxes indicate pigs that have lost track, and the yellow arrows represent pigs with ID switches. In Figure 10, in the 90th frame of video 0102, two pigs in the upper right corner suddenly move rapidly. In the subsequent three frames, Bot-SORT, ByteTrack, and OC-SORT lose track of these pigs and exhibit ID switches, while RpTrack can accurately track both pigs. In Figure 11, in the 152nd frame of video 0018, a pig in the lower left corner was lost due to uneven light from the 153rd frame to the 163rd frame. When the pig reappears in the 164th frame, only RpTrack can correctly track it, while the other algorithms fail to track the pig.

Figure 10. Comparison of RpTrack's tracking results with other MOT algorithms in the public dataset.

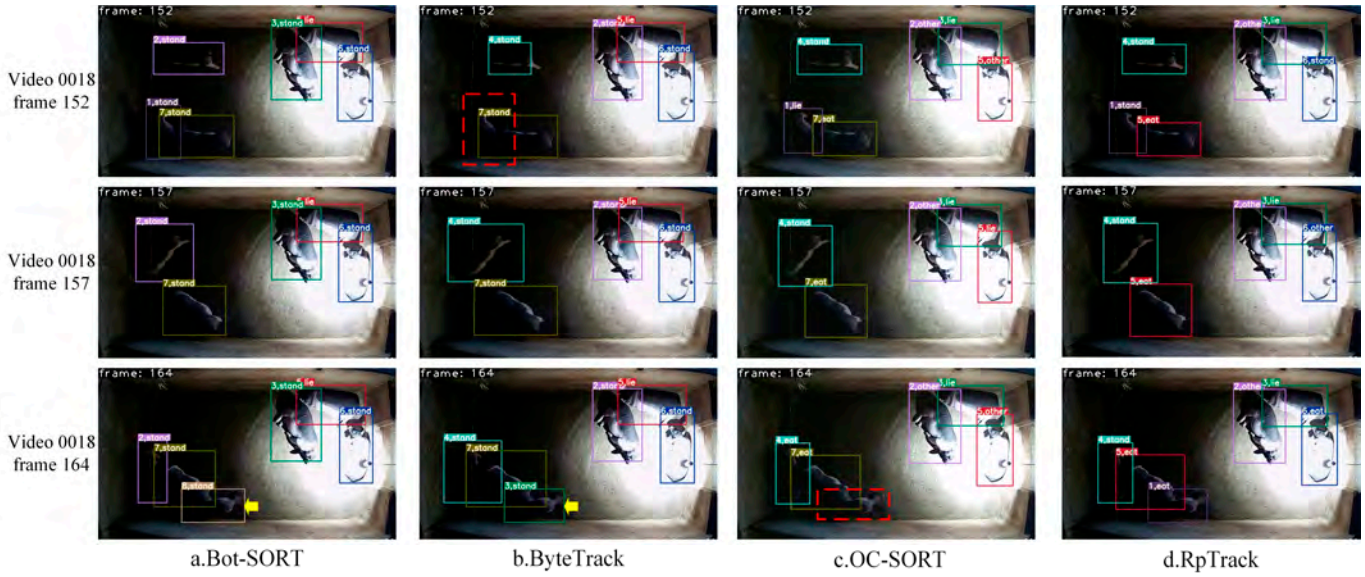


Figure 11. Comparison of RpTrack's tracking results with other MOT algorithms in the private dataset.

To further compare the overall tracking performance among different trackers, Figure 12 displays the pig tracking trajectories in selected videos. Figure 11a,d show the real activity trajectories and tracking trajectories of all individual pigs in video 0102 and video 0013, respectively, with labels on the right indicating the trajectory colors and their corresponding IDs. Figure 12b,c,e depict the real activity trajectories of pigs with IDs 3 and 6 in video 0102 and ID 10 in video 0013. These figures also show the tracking trajectories of each tracker. Taking video 0102 as an example, in Figure 12a, RpTrack produces tracking trajectories most similar to the real trajectories (GT). The number of trajectories matches that in the GT, indicating that RpTrack maintains consistent pig IDs throughout the tracking process. Additionally, Figure 11b shows the actual trajectory of a pig in video 0102 and the tracking trajectory of each tracker. It can be observed that Bot-SORT, OC-SORT, and ByteTrack all show ID switches when tracking a pig with GT ID 3, while RpTrack can correctly track this pig. A comprehensive analysis combining Table 3 and Figures 10–12 demonstrates that RpTrack exhibits the best tracking performance, accurately tracking pigs in complex scenarios.

3.5. Behavioral Statistics

To validate the accuracy of individual pig behavior identification and statistical analysis based on the RpTrack tracking method, this study employs the training dataset labeled with four distinct behavior categories ("stand", "lie", "eat," and "other") to train a YOLOX detector. The combination of the YOLOX detector and RpTrack tracker accomplishes simultaneous tracking and behavior statistics of pigs. The real behavior statistics of individual pigs in certain videos and the behavior statistics generated by RpTrack are presented in Figure 13. The horizontal axis represents pig identities, while the vertical axis indicates the number of frames in which the behavior duration. Bars with arrows represent the real behavior statistics, followed by adjacent bars illustrating RpTrack's behavior statistics. Different colors, such as blue, orange-red, green, and orange, correspond to the four behavior categories ("stand", "lie", "eat," and "other"), respectively. The greater the similarity between the actual behavior results and RpTrack's behavior results, the more accurate RpTrack's behavior statistics were considered. For pig ID 5 in video 0902, the real statistics for the blue, orange-red, and orange parts closely match the RpTrack behavioral statistics, indicating that RpTrack's results are accurate. Similar results were observed for other individual pigs. The experiment results indicate that the tracking method based on the YOLOX and RpTrack achieved relatively accurate behavior identification and statistical analysis of pig behaviors in the videos.

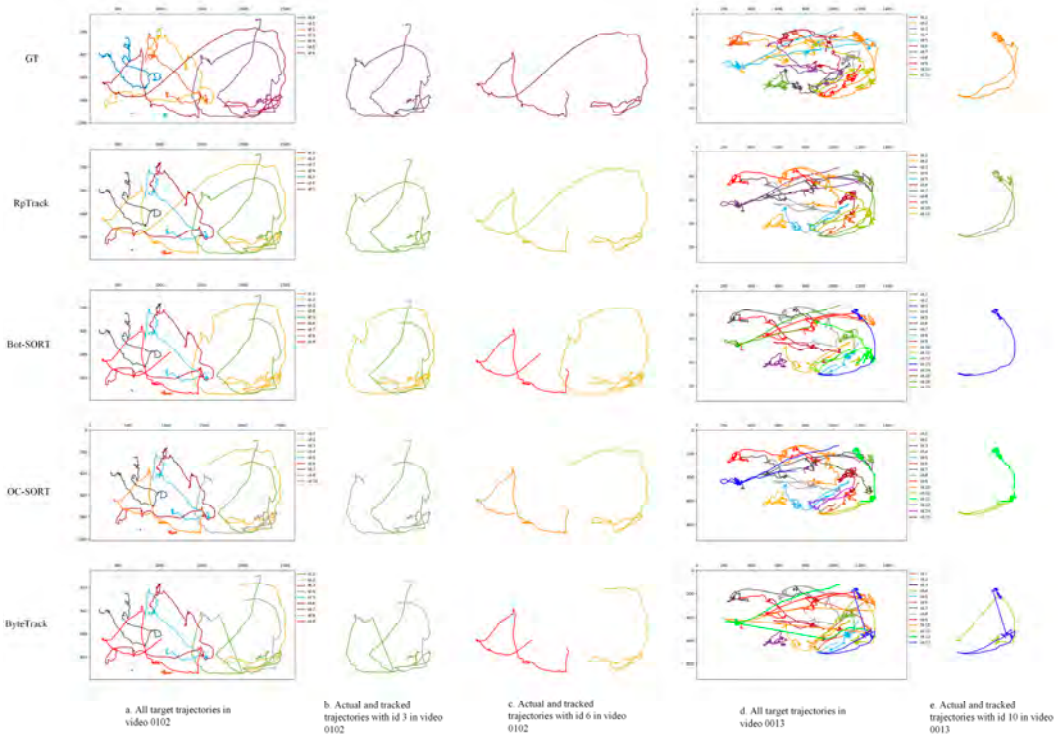


Figure 12. Comparison of target tracking trajectories for different methods.

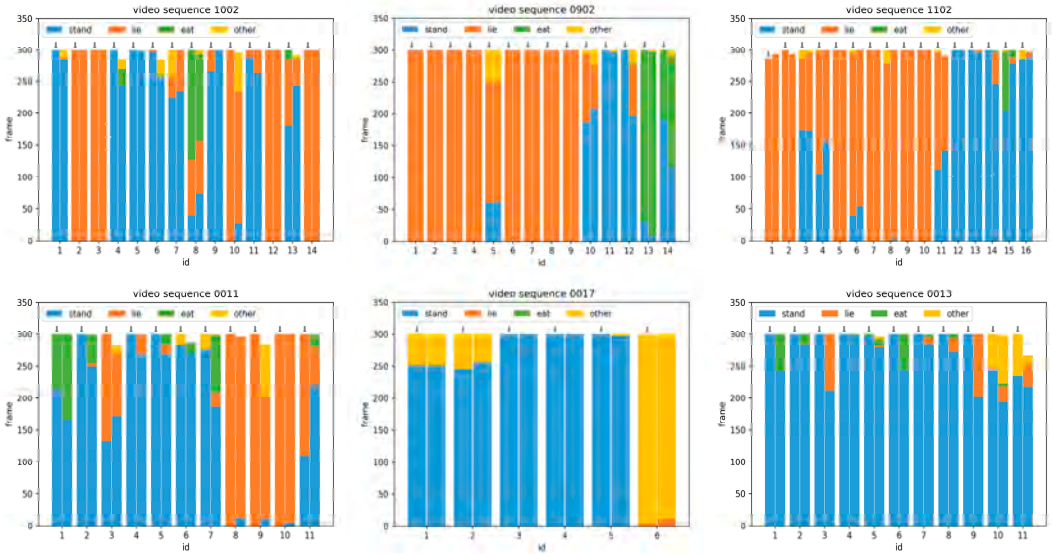


Figure 13. Comparison of actual behavioral statistics and tracking behavioral statistics for some videos.

3.6. Ablation Experiments and Analysis

3.6.1. Effect of Different K Values in Improved Trajectory Management

In the improved trajectory management, in order to obtain a suitable value for the hyperparameter K, we performed an ablation study on both the public and private datasets to analyze the effect of different values of K on the tracking performance and the results are shown in Table 4.

According to the results in Table 4, the best tracking performance can be obtained with $K = 2$ in the public dataset and with $K = 4$ in the private dataset. The tracking performance can be improved in all cases after using the improved trajectory management. Figure 14 shows the visualization results obtained with and without the improved trajectory management. In the figure, the red-dashed bounding boxes indicate pigs cannot be tracked,

and the yellow arrows represent pigs with ID switches. In this case, the pig with ID 2 was severely occluded at frame 1755, which led to a switch from ID 6 to 2. When the occluded pig reappeared at frame 1756, the pig failed to be retrieved correctly without the improved trajectory management, while it was able to be retrieved correctly after adopting the improved trajectory management. This demonstrates that the improved trajectory management is more accurate in predicting the target location.

Table 4. Effect of different values of *K* on tracking performance, where "-" indicates that the improved trajectory management is not used.

	Public				Private			
K Values	**HOTA/%**	**MOTA/%**	**IDF1/%**	**IDSW**	**HOTA/%**	**MOTA/%**	**IDF1/%**	**IDSW**
-	71.0	95.5	82.6	152	79.4	97.8	96.7	11
1	72.1	95.4	84.2	158	80.1	97.5	97.9	7
2	**73.2**	**95.5**	**85.6**	**148**	79.4	97.6	96.6	10
3	72.1	95.5	84.6	151	79.9	97.6	97.3	8
4	72.2	95.5	84.8	154	**80.7**	**97.7**	**98.3**	**6**
5	72.3	95.5	85.1	148	79.5	97.6	96.6	12

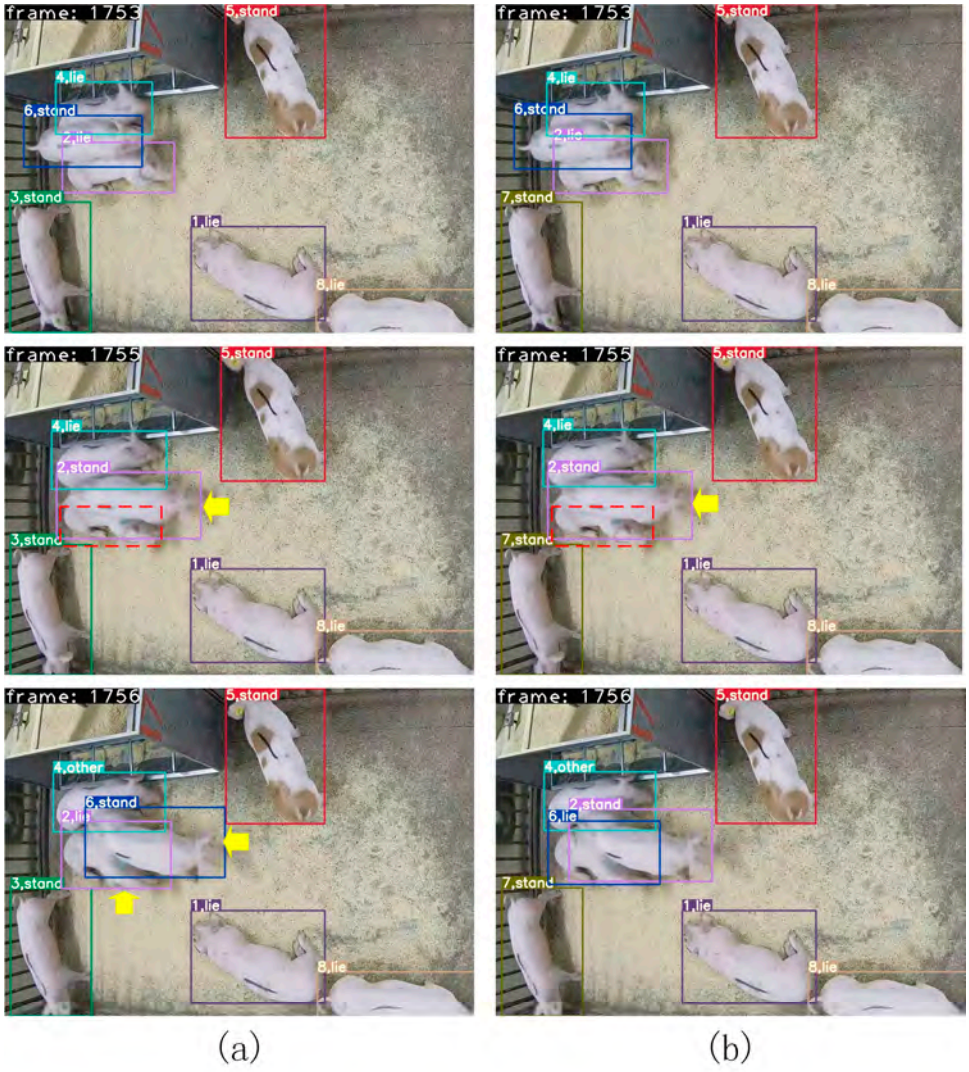


Figure 14. Comparison of visualization results on video 01 with and without the improved trajectory management, where (**a**) the results without the improved trajectory management and (**b**) the results with the improved trajectory management.

3.6.2. Effect of Each Module in the RpTrack

To validate the performance of the improved Kalman Filter, the improved trajectory management, and BIoU in tracking, we conducted tests on both public and private datasets before and after the improvements, as indicated in Table 5. It can be observed that, compared to the original SORT, the improved Kalman Filter, the improved trajectory management, and the BIoU all lead to performance improvements.

Table 5. Ablation experiments on the public and private datasets. IKF denotes the improved Kalman Filter, ITM denotes the improved trajectory management, and BIoU denotes Buffered IoU.

			Public				Private			
IKF	**ITM**	**BIoU**	**HOTA/%**	**MOTA/%**	**IDF1/%**	**IDSW**	**HOTA/%**	**MOTA/%**	**IDF1/%**	**IDSW**
			65.2	95.0	72.8	242	77.7	97.4	93.0	29
✔			70.5	95.5	81.8	146	79.7	97.7	96.6	11
✔	✔		70.7	95.5	82.0	**142**	80.2	97.8	97.2	8
✔	✔	✔	**73.2**	95.5	**85.6**	148	**80.8**	97.8	**98.4**	**6**

Figure 15 shows a comparison of tracking results with and without ITM in video 0102 and with and without BIoU in video 0018. In video 0102 (green-dashed box), in the 90th frame, pigs with IDs 2 and 4 suddenly exhibit rapid movements. Subsequently, in frames 92 and 93, ITM loses tracking (as indicated by the red-dashed box), while ITM maintains accurate tracking. In video 0018 (blue-dashed box), after frame 152, pig ID 1 is lost due to a missed detection. When the pig reappears at frame 164, not using BIoU ensures tracking is lost (as depicted by the red-dashed box), while using BIoU maintains accurate tracking. This indicates that the improvements are effective in dealing with irregular pig movements and missed detections, which in turn improves tracking performance.

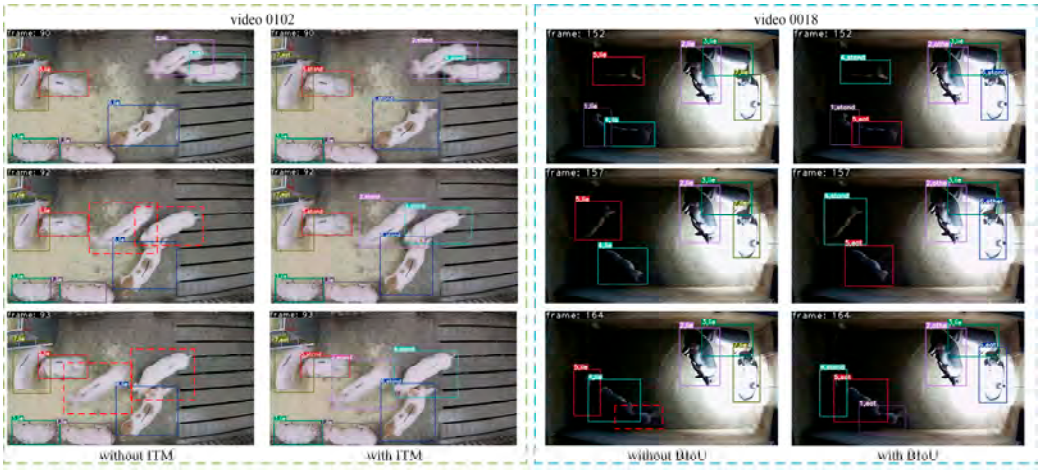


Figure 15. Comparison of tracking results with and without ITM or BIoU.

4. Conclusions and Limitation Discussion

The RpTrack method proposed in this paper utilizes YOLOX for pig detection and behavioral classification. It also leverages the improved KF and the improved trajectory management to address irregular pig movements. Additionally, it introduces BIoU to mitigate the issue of decreased tracking performance caused by missed detections resulting from uneven lighting. The summary is as follows:

1. Experiment results on both public and private datasets demonstrate that RpTrack achieves a competitive execution speed like the SORT method. In the public dataset, RpTrack achieves a HOTA of 73.2%, MOTA of 95.5, IDF1 of 85.6%, and IDSW of 148. In the private datasets, RpTrack's performance achieves a HOTA of 80.8%, MOTA of

97.8%, IDF1 of 98.4, and IDSW of 6, surpassing other leading tracking methods in all performance metrics.
2. Visualization results comparisons confirm that the improved KF and the improved trajectory management effectively address the issue of ID switches caused by irregular pig movements. Furthermore, It is also demonstrated that BIoU can alleviate the problem of ID switches resulting from missed detections due to uneven lighting. These improvements enhance tracking stability.
3. Pig behaviors are categorized into "stand", "lie", "eat," and "other". Based on RpTrack's more precise tracking results, it achieves more accurate pig behavior statistics.

RpTrack still has some limitations, as follows:

1. The improved trajectory management is sensitive to the hyperparameter K. Different values of K are used in different tracking environments to achieve better tracking results, and as K is larger, more computation is required (Section 3.6.1).
2. BIoU still has mismatch problems due to the bounding box extension.
3. The work in this paper focuses on improvements to the tracker, and the problems that would be faced by behavioral statistics are not studied in depth. Therefore, RpTrack fails to deal with the effect of ID switches on the results of behavioral statistics, but this is the direction of our future work.

The RpTrack method proposed in this paper effectively handles the challenges presented by irregular pig movements and uneven lighting in MOT of pigs. It demonstrates outstanding performance in tracking stability, accuracy, and behavioral statistics, providing support for research and applications in intelligent pig farming.

Author Contributions: Methodology, H.L. (Hua Lei); software, H.L. (Hua Lei); formal analysis, S.T. and E.L.; resources, S.T. and Y.L.; writing—original draft, H.L. (Hua Lei); writing—review and editing, S.T. and H.L. (Hua Lei); visualization, H.L. (Hua Lei) and H.L. (Hongxing Liu); supervision, S.T., Y.L., E.L., and H.L. (Hongxing Liu); project administration, S.T., Y.L., E.L., and H.L. (Hongxing Liu); funding acquisition, Y.L., E.L., and H.L. (Hongxing Liu). All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by key R&D project of Guangzhou (202206010091, 2024B03J1358, 2023B03J1363).

Institutional Review Board Statement: The animal experiment protocol was approved by the Ethics Committee for Research Animals of South China Agricultural University (Protocol number 2024F213).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

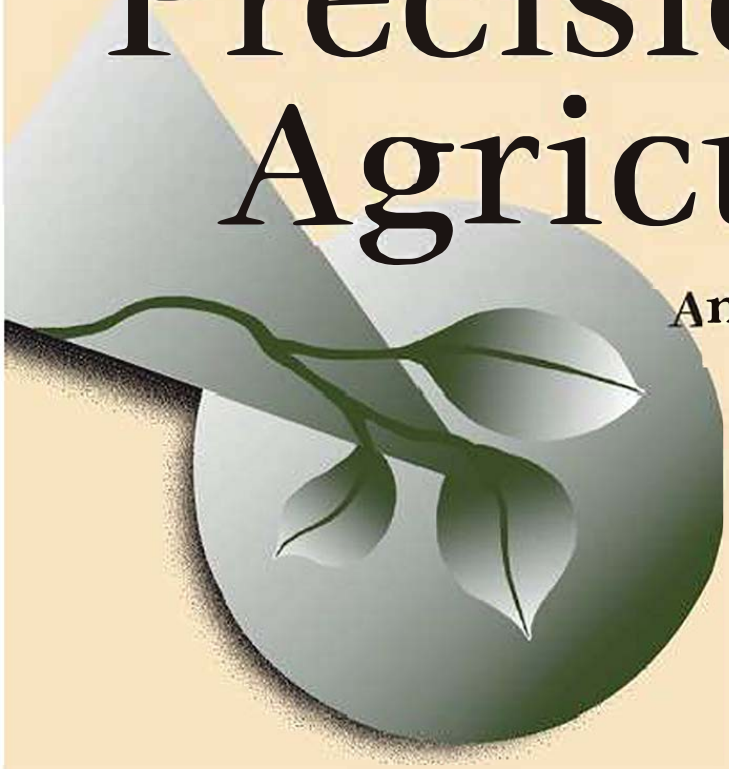
1. Tzanidakis, C.; Simitzis, P.; Arvanitis, K.; Panagakis, P. An overview of the current trends in precision pig farming technologies. *Livest. Sci.* **2021**, *249*, 104530. [CrossRef]
2. Yin, M.; Ma, R.; Luo, H.; Li, J.; Zhao, Q.; Zhang, M. Non-contact sensing technology enables precision livestock farming in smart farms. *Comput. Electron. Agric.* **2023**, *212*, 108171. [CrossRef]
3. Matthews, S.G.; Miller, A.L.; Plötz, T.; Kyriazakis, I. Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. *Sci. Rep.* **2017**, *7*, 17582. [CrossRef] [PubMed]
4. Zhang, L.; Gray, H.; Ye, X.; Collins, L.; Allinson, N. Automatic Individual Pig Detection and Tracking in Pig Farms. *Sensors* **2019**, *19*, 1188. [CrossRef] [PubMed]
5. Cowton, J.; Kyriazakis, I.; Bacardit, J. Automated Individual Pig Localisation, Tracking and Behaviour Metric Extraction Using Deep Learning. *IEEE Access* **2019**, *7*, 108049–108060. [CrossRef]
6. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
7. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In Proceedings of the: 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017. [CrossRef]
8. Guo, Q.; Sun, Y.; Orsini, C.; Bolhuis, J.E.; de Vlieg, J.; Bijma, P.; de With, P.H. Enhanced camera- based individual pig detection and tracking for smart pig farms. *Comput. Electron. Agric.* **2023**, *211*, 108009. [CrossRef]

9. Wang, Z.; Zheng, L.; Liu, Y.; Li, Y.; Wang, S. Towards real-time multi-object tracking. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 107–122.
10. Zhang, Y.; Wang, C.; Wang, X.; Zeng, W.; Liu, W. Fairmot: On the fairness of detection and re–identification in multiple object tracking. *Int. J. Comput. Vis.* **2021**, *129*, 3069–3087. [CrossRef]
11. Tu, S.; Liu, X.; Liang, Y.; Zhang, Y.; Huang, L.; Tang, Y. Behavior Recognition and Tracking Method of Group housed Pigs Based on Improved DeepSORT Algorithm. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 345–352. [CrossRef]
12. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
13. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Fang, J.; Michael, K.; Montes, D.; Nadar, J.; Skalski, P. Ultralytics/yolov5: v6.1–TensorRT, TensorFlow edge TPU and OpenVINO export and inference. *Zenodo* **2022**. [CrossRef]
14. Kim, J.; Suh, Y.; Lee, J.; Chae, H.; Ahn, H.; Chung, Y.; Park, D. EmbeddedPigCount: Pig Counting with Video Object Detection and Tracking on an Embedded Board. *Sensors* **2022**, *22*, 2689. [CrossRef]
15. Odo, A.; Muns, R.; Boyle, L.; Kyriazakis, I. Video Analysis Using Deep Learning for Automated Quantification of Ear Biting in Pigs. *IEEE Access* **2023**, *11*, 59744–59757. [CrossRef]
16. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-YOLOv4: Scaling cross stage partial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13024–13033.
17. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207. 02696.
18. Han, S.; Fuentes, A.; Yoon, S.; Jeong, Y.; Kim, H.; Sun Park, D. Deep learning-based multi–cattle tracking in crowded livestock farming using video. *Electron. Agric.* **2023**, *212*, 108044. [CrossRef]
19. Yigui, H.; Deqin, X.; Junbin, L.; Zhujie, T.; Kejian, L.; Miaobin, C. An Improved Pig Counting Algorithm Based on YOLOv5 and DeepSORT Model. *Sensors* **2023**, *23*, 6309. [CrossRef] [PubMed]
20. Zheng, Z.; Li, J.; Qin, L. YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Comput. Electron. Agric.* **2023**, *209*, 107857. [CrossRef]
21. Van der Zande, L.E.; Guzhva, O.; Rodenburg, T.B. Individual detection and tracking of group housed pigs in their home pen using computer vision. *Front. Animal Sci.* **2021**, *2*, 669312. [CrossRef]
22. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
23. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016. [CrossRef]
24. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
25. Psota, T.; Schmidt, E.; Mote, T.B.; Pérez, C.L. Long-Term Tracking of Group-Housed Livestock Using Keypoint Detection and MAP Estimation for Individual Animal Identification. *Sensors* **2020**, *20*, 3670. [CrossRef]
26. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]
27. Aharon, N.; Orfaig, R.; Bobrovsky, B.-Z. BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *arXiv* **2022**, arXiv:2206.14651. [CrossRef]
28. Yang, F.; Odashima, S.; Masui, S.; Jiang, S. Hard to Track. In Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–7 January 2023. [CrossRef]
29. Luiten, J.; Osep, A.; Dendorfer, P.; Torr, P.; Geiger, A.; Leal-Taixé, L.; Leibe, B. Hota: A Higher Order Metric for Evaluating Multi-object Tracking. *Int. J. Comput. Vis.* **2020**, *129*, 548–578. [CrossRef] [PubMed]
30. Bernardin, K.; Stiefelhagen, R. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *J. Image Video Proc.* **2008**, *2008*, 246309. [CrossRef]
31. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-object Tracking by Associating Every Detection Box. In *Computer Vision–ECCV*; Springer Nature: Cham, Switzerland, 2022. [CrossRef]
32. Cao, J.; Pang, J.; Weng, X.; Khirodkar, R.; Kitani, K. Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023.

2. 7 A passion fruit counting method based on the lightweight YOLOv5s and improved DeepSORT

Volume 25, Number 3, June 2024 ISSN: 1385-2256

Precision Agriculture



An International Journal on Advances in Precision Agriculture

Springer

Report Information from ProQuest

July 14 2025 03:47

目录

目录

Precision Agriculture
https://doi.org/10.1007/s11119-024-10132-1

A passion fruit counting method based on the lightweight YOLOv5s and improved DeepSORT

Shuqin Tu[1,2] · Yufei Huang[1] · Yun Liang[1] · Hongxing Liu[1] · Yifan Cai[1] · Hua Lei[1]

Accepted: 28 February 2024

Abstract
Accurate yield estimation of passion fruits is essential for planning acreage and harvest timing. However, due to the complexity of the natural environment and tracking instability, the existing yield estimation methods suffer from excessively large models that are difficult to deploy or repetitive counting of fruit. Therefore, an improved approach for efficient passion fruit yield estimation was proposed using the lightweight YOLOv5s and improved DeepSORT. First, the video is fed into the proposed lightweight YOLOv5s called YOLOv5s-little to obtain coordinates and confidence information about the fruits within each frame. Then, the information obtained from the detection model is input into improved DeepSORT for continuous frame tracking of passion fruit. Considering the frequent error IDs (ID switching), two improvements based on DeepSORT are proposed: delaying the creation of tracks and adding a second round of IoU matching. Finally, to overcome the problem of repetitive counting, a specific tracking counting method based on the track information and state is used for accurate passion fruit counting. Our method achieved a competitive result in tests. YOLOv5s-little detector achieved precision of 98.9%, 98.3% recall, 99.5% mAP, and only 0.9MB model size. The improved DeepSORT algorithm achieved higher order tracking accuracy (HOTA) of 79.6%, multi-object tracking accuracy (MOTA) of 92.58%, identification F1 (IDF1) of 95.02%, and ID switch (IDSW) of 11 respectively. Compared with DeepSORT, it improved by 4.66%, 1.8%, and 9.16% in HOTA, MOTA and IDF1, respectively, and IDSW improved the most with 85%. Compared with FairMOT and TransTrack, the HOTA of YOLOv5s-little + improved DeepSORT achieved improvements of 11.56% and 25.24%, respectively. The statistical average counting accuracy of our proposed counting method reaches 95.1%, which is a 7.09% improvement over the maximum ID value counting method. The counting results from test videos are highly correlated with the manual counting results ($R^2 = 0.96$), indicating that the counting method has high accuracy and effectiveness. These results show that YOLOv5s-little + improved DeepSORT can meet the practical needs of passion fruit yield estimation in real scenarios.

✉ Yun Liang
Liang_Yun168@163.com

[1] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

[2] Guangzhou Key Laboratory of Intelligent Agriculture, Guangzhou, China

Published online: 26 March 2024

Keywords Lightweight · Multiple object tracking · YOLOv5s · DeepSORT

Introduction

Passion fruit is rich in nutritional and medicinal value, which is mainly cultivated in tropical and subtropical regions. It is an important local crop in South America with high economic value (Bezerra et al., 2019; Thokchom & Mandal, 2017). However, manual yield estimation of passion fruit is a laborious and time-consuming task. This presents challenges for farmers in planning cultivation areas, determining optimal harvest times, and devising sales strategies. Fortunately, recent advancements in computer vision technology have provided effective support in addressing this issue (He et al., 2022a, 2022b; Vasconez et al., 2020). An automatic counting system for passion fruits can be developed, enabling farmers to plan their production and marketing strategies more accurately and maximizing their profitability.

With the development of deep learning methods, many researchers have implemented fruit detection and counting with the help of deep learning-based object detection (Assunção et al., 2022; Liu et al., 2017; Wang et al., 2022b). For example, Koirala et al. (2019) developed a new detection model called MangoYOLO based on the features of YOLOv2 (tiny) and YOLOv3. Tu et al. (2020) were able to robustly detect small fruits of passion fruit by feeding RGB-D images into an improved multiple scale faster region-based convolutional neural networks (MS-FRCNN) based model. Qi et al. (2022) used YOLOv5 as a target detection model to detect the main stem of litchi in litchi images and obtain the accurate location information of the main stem picking point. Liu et al. (2023) proposed a method based on binocular stereo vision and an improved YOLOv3 model to detect pineapples, improved by the optimization of the backbone network and the use of the binocular camera to get the left and right images. The above methods are used to detect fruits in capturing images for counting. However, the rapid development and widespread use of robots, mobile terminals, and smart devices in modern agriculture. Many detection models cannot be deployed on these devices due to limitations in calculating ability and storage capacity (He et al., 2022a, 2022b).

Therefore, model lightweight has become a trend in fruit detection. In particular, Zhou et al. (2020) used SSD with two lightweight backbones MobileNetV2 and InceptionV3 for kiwi detection in the field. Wang and He (2021) developed a small model size based on a channel-pruned YOLOv5s deep learning algorithm to accurately detect apple small fruit. Wang et al., (2022a, 2022b) proposed an improved YOLOv5 model to detect litchi, the improved backbone network uses ShuffleNetV2, introduces the CBAM module in the feature fusion stage, and adjusts the input size to 1280×1280. Shen et al. (2023) obtained a lighter-weight YOLOv5s cluster detection model based on the channel pruning algorithm, and soft non-maximum suppression is introduced in the prediction stage to improve the detection performance of grape overlapping clusters. Inspired by the above studies, a lightweight network based on YOLOv5 was designed according to the characteristics of passion fruit.

The input from object detection for counting fruits is images. These counting methods can only detect a portion of the fruits in that image, and each image has overlapping parts. Therefore, these methods will inevitably double count, leading to inaccurate results. To address the issue, most researchers use the multi-object tracking (MOT) technique to perform fruit counting (Egi et al., 2022; He et al., 2022a, 2022b; Liu et al.,

2019). MOT locates targets in a video and assigns a unique ID to each object (Guo et al., 2022; Luo et al., 2021). For example, Gao et al. (2022) used a YOLOv4-tiny network integrated with the CSR-DCF algorithm to count apples, which solved the problem of losing targets during the tracking process. Tan et al. (2022) used an optical flow-based tracking method to estimate camera motions. The number of cotton seedlings was updated by comparing the positions of bounding boxes predicted by optical flow and detected by the YOLOv4 network in the same frame. Yang et al. (2022) used CenterNet to detect cotton seedlings, extract their identity embedding, and associate the data based on DeepSORT fused localization and identity information.

In addition, there have been studies combining segmentation and multi-objective tracking for fruit tracking. De Jong et al. (2022) used MOTS architectures called TrackR-CNN and PointTrack and demonstrated the applicability of these methods for the joint detection of apples. Ariza-Sentís et al. (2023) applied the PointTrack algorithm for the detection and tracking of grape bunches and compared two instance segmentation algorithms (YOLACT and spatial embedding) to find the most suitable method for detecting grape bunches.

However, due to the complexity of the natural environment, there are frequent error IDs during the tracking process, leading to double counting. Regarding this issue, many researchers have already conducted studies on it. For example, Zhang et al. (2022) proposed OrangeSort to establish a specific tracking region counting strategy and tracking algorithm based on motion displacement estimation, aiming to alleviate the double counting problem associated with shaded fruits. Ge et al. (2022) obtained tracking results and used OpenCV to create a virtual counting line to count tomatoes. Rong et al. (2023) designed a specific tracking region counting method to overcome the problem of tracked tomato cluster IDs. Since passion fruits are often partially obscured by leaves or other fruits as they grow, and some of them are small, this can cause them to be easily missed. If the fruit is missed for a long time, it may lose its original ID and be re-identified as a new ID. This will cause duplicate counting problems in subsequent counting tasks. However, Current solutions are complex and not well suited to the problem of tracking passion fruit. Therefore, we design a method to alleviate frequent error IDs.

In addition to the challenges of models that are too large to deploy and frequent error IDs, we found that the currently proposed counting methods suffer from significant errors or cumbersome processes. Presently, the widely adopted approach is to use the maximum ID value generated by DeepSORT as the predicted value. While easily obtainable, the maximum ID often deviates significantly from manual counts. To address this issue, some studies have introduced alternative methods. Parico and Ahamed (2021) compared the ROI method with the maximum ID way. The ROI method counts fruits by crossing a line horizontally. Tests have shown that this method brings predicted counts closer to manual counts. Although the count line method has a high accuracy rate, it requires additional steps to be performed after tracking.

To address these challenges mentioned above, we have proposed an improved passion fruit yield estimation method based on YOLOv5s and DeepSORT. The specific improvements we have made are as follows:

(1) A lightweight network based on YOLOv5s is designed.
(2) We have introduced a delaying the creation of tracks method to avoid frequent error IDs and added a second round of IoU matching to improve tracking accuracy.

(3) We have proposed a convenient counting method based on track state and track information. We ensure that the estimated values are closer to the actual values and simplify the counting process by counting the fruits during the process of track creation and deletion.

Materials and methods

Dataset description

The videos used in the experiment were obtained in passion fruit plantations in Heyuan City, Guangdong Province, China, and Huadu District, Guangzhou City, China. The reason for choosing these two locations is that they represent typical characteristics of passion fruit cultivation. Both selected data collection sites are large passion fruit plantations using the widely used pergola cultivation technique. The video capture device was a cell phone (Huawei Mate30), and the horizontal shot was used. The distance between the capture device and the passion fruit was kept at 5–8 m in order to effectively capture multiple passion fruit plants and to ensure the passion fruit in the video is at the right size. In this acquisition, a total of 12 videos were captured, each video was about 15 s in length. The videos were stored in MP4 format with a frame rate of 25 fps and a resolution of 1280×720. Figure 1 illustrates a partial dataset.

The dataset processing mainly includes video annotation, video cropping into images, and data augmentation. The following are the specific steps:

(1) The videos of 15s with dense fruit content were chosen as the dataset. The length of the videos ensures that each video shows a portion of the characteristics of the passion fruit orchard and meets the criteria for fruit-densely (videos with 10 or more fruits). Then, each video was cut using FFmpeg software, and the captured videos were labeled

Fig. 1 Part of the dataset

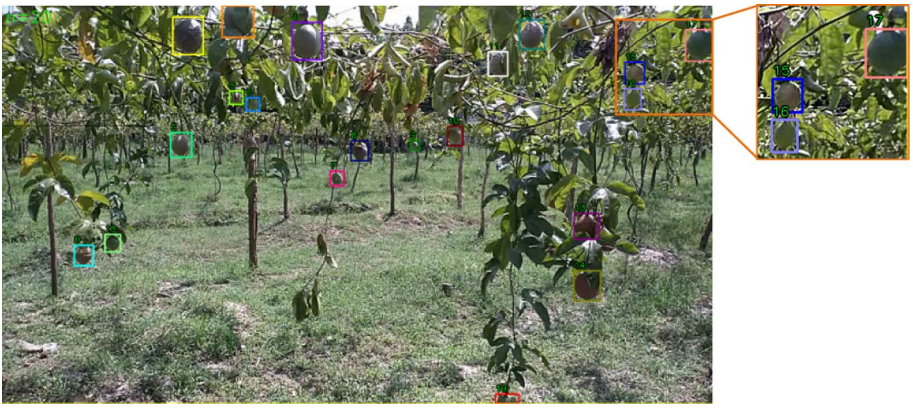


Fig. 2 Annotated images of video sequence

using DrakLabel software. All the passion fruits in each frame image are annotated and each passion fruit is given a unique ID. Figure 2 shows an annotated image of the video.

(2) After labeling the video, the videos were converted into image frames. A total of 1 287 image frames were obtained. These images were used for the training of YOLOv5 with DeepSORT. Additionally, a Python script was employed to transform the annotated files into txt files suitable for detection (COCO dataset) and reidentification (ReID) data suitable for DeepSORT. All image frames are randomly divided into training set, validation, and test set in the ratio of 7:1:2(900:130:257).

(3) When performing data augmentation on the original image for the training set, we perform the following steps, including randomly adjusting brightness and contrast, adding Gaussian noise, and cropping the image. Finally, the original 900 images used for YOLOv5 training were expanded to 1800 images.

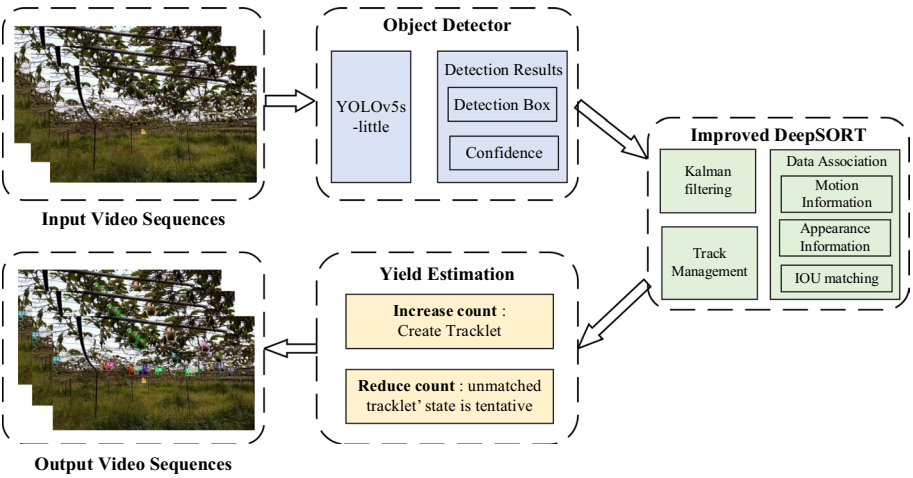


Fig. 3 Passion fruit yield estimation tracking flowchart

Method process and flow chart

Figure 3 illustrates the flowchart detailing the passion fruit yield estimation algorithm. The algorithm encompasses three main steps, each serving a crucial role in the overall process.

Object Detection using YOLOv5s-little: A lightweight detector named YOLOv5s-little is used to detect each passion fruit and obtain its detection box and confidence in each frame of the video.

Tracking with Improved DeepSORT: Information from the detector is imported into the improved DeepSORT algorithm to achieve accurate tracking of each passion fruit in the current frame and the next frame. To avoid error frequent IDs and improve tracking accuracy, two strategies in the improved DeepSORT are introduced: delaying the creation of track and adding the second round of IoU matching.

Passion Fruit Yield Estimation: The counting module completes passion fruit yield estimation according to track information and track state. The output includes the results of the MOT sequences and the passion fruit yield estimation.

YOLOv5s, YOLOv5s-2level and YOLOv5s-little detection network

As one of the most widely used one-stage object detection algorithms, YOLO has been widely applied in industry and agriculture due to its outstanding performance and fast detection speed (Bochkovskiy et al., 2020; Redmon & Farhadi, 2018). YOLOv5 was developed by Ultralytics and released in June 2020 as an upgrade to the YOLOv4 model. Compared to the structures of YOLOv3 and YOLOv4, YOLOv5 has been further optimized to enable fast and accurate object detection. In addition, YOLOv5's tractable network architecture allows for further optimization and enhancement, while also being conducive to use with other multi-target tracking algorithms. The YOLOv5 series includes four versions, YOLOv5s, YOLOv5m, YOLOv5x, and YOLOv5xl, which vary in network depth and width. YOLOv5s was chosen as the detector for improvement because it can balance the requirements of speed and accuracy.

Two improved YOLOv5 detection networks for passion fruit targets are proposed, the two-layer (YOLOv5s-2Level) and the lightweight network (YOLOv5s-little). YOLOv5s-little adopts the same structure as YOLOv5s-2Level. The network architecture of YOLOv5s and YOLOv5s-2Level is shown in Fig. 4. As the targets of passion fruit in video tracking are generally small, the YOLOv5s model has little impact on the accuracy of fruit detection in high-level feature maps for large targets, and it consumes more detection time. Therefore, YOLOv5s-2Level mainly removes the P5 layer in the YOLOv5s model. Since passion fruits are usually round and have little shape variation, based on YOLOv5s-2Level, YOLOv5s-little

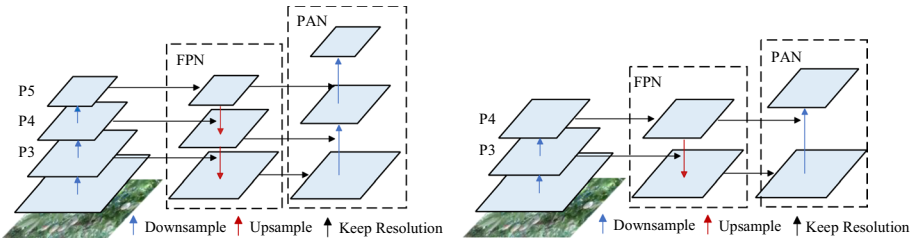


Fig. 4 Network structure of YOLOv5s and YOLOv5-2level

further decreases model parameters, compresses network depth and width, and reduces the number of anchor boxes to shrink the model size, making it less than one-tenth the size of the YOLOv5s model while maintaining fast detection speed and good detection accuracy.

Improved deepSORT for passion fruit yield estimation model

The DeepSORT algorithm is an improved algorithm based on the SORT algorithm (Wojke et al., 2017). The SORT algorithm consists of three parts, Kalman filtering, the Hungarian matching algorithm, and track management (Bewley et al., 2016). The algorithm first uses Kalman filtering (KF) to predict the position of each detected target. Then a data association algorithm is used to associate the tracked targets with their detection results in consecutive frames. Finally, tracks on matches are updated, tracks that are no longer detected are deleted, and new tracks are created for unmatched objects. On top of the SORT, DeepSORT uses a convolutional neural network to extract the features of the target and employs appearance similarity for data association. This enhances the precision and robustness of multi-object tracking, particularly in complex scenarios.

Figure 5 shows the flowchart of the improved DeepSORT algorithm. The specific tracking process is as follows:

(1) First, cascade matching is performed between the detection results and the confirmed state track predicted by the KF. The Hungarian matching algorithm is used to match the previous tracks with the current detection boxes by solving the cost matrix of motion and appearance information.
(2) For unmatched detection boxes, unmatched tracks, and unconfirmed tracks, the algorithm uses first round of IoU matching to complete trajectory tracking using the Hungarian algorithm.
(3) The second round of IoU matching is performed to match the unmatched tracks and detection boxes from the first round of IoU matching. If an unmatched detection satisfies the track creation condition, a track is created. If an unmatched track satisfies the deletion condition, the track is deleted. If an unmatched track does not satisfy the deletion condition, it continues to participate in the next round of matching.
(4) After all matching in the current frame is completed, the KF update is performed on the matched tracks.

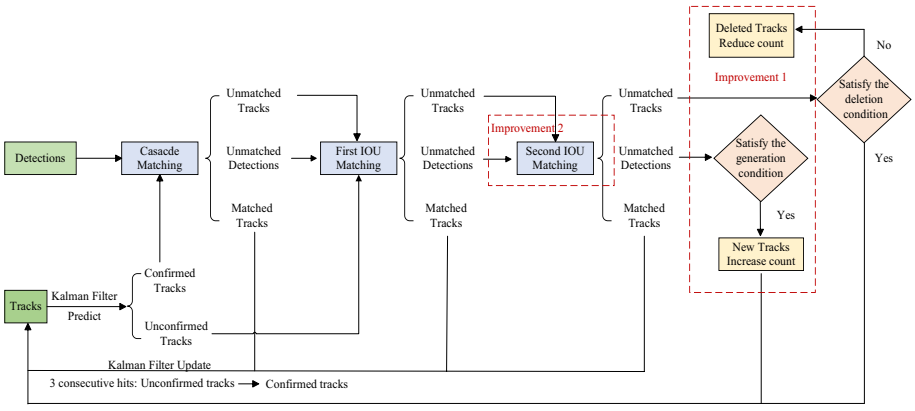


Fig. 5 Flow chart of the tracking process of the improved DeepSORT

(a) (b)

Fig. 6 Illustration of passion fruit frequent IDs

Fig. 7 Area map allowing creation of tracks

Two improvements to DeepSORT (as shown in Fig. 5) are proposed to address the problems encountered in passion fruit tracking, namely delaying the creation of tracks and the second round of IoU matching. They are described below:

(1) Delaying the creation of tracks

Due to the mutual occlusion between fruits and the similarity in color between immature fruits and leaves, it is easy to cause false detection of fruits, leading to frequent IDs. Figure 6 displays the frequent IDs of passion fruit in real-life scenarios. In Fig. 6a, the highest ID shown is 47, whereas the manual annotation for this frame indicates 27. Similarly, in Fig. 6b, the highest ID is 21, but the manual annotation indicates 19. After many experiments, we find that frequent error IDs occur mainly in the central region and the border region. The reasons for this include two points. First, fruits in the central region lose their original ID due to feature significant changes (e.g., illumination, shape, occlusion, etc.) and are assigned a new ID. Second, leaves and stems in the border region are easily error detected as fruits, leading to error frequent IDs.

Therefore, we propose a method for the creation of tracks to limit the increase in fruit ID in the central and boundary regions. This method is shown as improvement 1 in Fig. 5. Specifically, in Fig. 7, the inner red area is the central area, and the outer red area is the boundary area. New tracks can only be created in the red area, not in the center area or the boundary area. During the experiment, we use the two values of coordinates of the bounding box (BB) and the resolution of the current frame to determine whether the bounding box is in the central region or the boundary region. After the third frame, if the unmatched detection is in the central area or the boundary area, no track is allowed to be created,

otherwise a track is created and given an ID. The boundary value is set according to the size of the fruit (set 25 in this paper). The detailed steps are described in Algorithm 1.

Algorithm 1 The creation of track

Input: **[x, y, w, h]:** bounding box information

frame_idx: current frame index

img_width: current frame width

img_height: current frame height

next_id: next ID number

total: the total number of passion fruits

Output: next_id; total

1: **if** frame_idx > 3 **then**

2: **if** w <25 and (0 <x <30 or img_width – 30 < x < img_width) **then**

3: **return**

4: **end if**

5: **else if** 70 < x < img_width - 70 - w and 70 < y < (img_height – 70 – h) **then**

6: **return**

7: **end if**

8: **end if**

9: create_track (x, y, w, h, next_id)

10: next_id←next_id + 1

11: total← total+ 1

12: **return** next_id, total

(2) The second round of IoU matching

Due to delaying the creation of tracks, new tracks cannot be created in either the central region or the boundary region. If a track in these regions fails to match new detection boxes multiple times, the track is then deleted. In addition, no new tracks could be created for fruits in these regions, leading to missed detections. Therefore, we add the second round of IoU matching to achieve better results in associating unmatched detections with tracks that were not matched in the first two rounds of matching.

Yield estimation algorithm

The maximum ID value generated based on DeepSORT can be used to predict the number of passion fruits, and the maximum ID value counting method (MAX-IDVC) is commonly used. However, the predicted count from MAX-IDVC exceeds the actual count. The single line counting method, which counts fruits by tracking them as they cross a line horizontally, can solve the repetition counting problem. But it needs to be counted after each frame of tracking, which is a rather cumbersome process. Our proposed method estimates the yield of fruits based on the track information and state during the track management process. This method not only resolves duplicate counting problems but also simplifies operations. This yield estimation algorithm is shown as improvement 1 in Fig. 6.

The passion fruit yield count of each video is set as Total. The yield count is estimated in three cases based on track information and track state, specifically as shown in Eq. 1. Total is added by 1 If there is a creation of the track. The total is subtracted by 1 for the unmatched track state is tentative. The total remains unchanged if the number of recent KF update (time-since-update) parameters of the track is greater than 30.

$$\text{Total} = \begin{cases} \text{Total} + 1 & \text{if a creation of track} \\ \text{Total} - 1 & \text{if unmatched track state is tentative} \\ \text{Total} & \text{if track's time-since-update} > 30 \end{cases} \tag{1}$$

Experiment and result analysis

Implementation details

The experiments for estimating the passion fruit yield in this study are divided into three parts:

- YOLOv5s, YOLOv5s-2level, and YOLOv5s-little models were used to detect passion fruit.
- Based on the YOLOv5s-little model, passion fruit was tracked using DeepSORT and the improved DeepSORT.
- Based on YOLOv5s-little and the improved DeepSORT, MAX-IDVC, and our counting method were used to count passion fruits.

For the detection experiment, 4000 images with data augmentation were used for model training. The original resolution of the images was 1280 × 720 pixels, but it was compressed to 640×640 pixels after input preprocessing. In the training process, a pre-trained

model was not used, and the optimizer was set to SGD. The learning rate, batch size, and epoch were set to 0.01, 32, and 200, respectively.

Twelve annotated videos were used, with three videos used for tracking testing and the remaining nine videos used for ReID training. The target dataset for ReID training contained 198 passion fruits. The number of images per passion fruit was increased to around 250 through data augmentation. The ratio of the training set to the test set was set at 4:1.

For the tracking experiment, three videos with 28, 42, and 19 passion fruits, respectively, were used, covering both sparse and dense scenes. The max_IoU_distance threshold was set to 0.7, while other parameters were kept at their default values.

Evaluation metrics

In this section, we have chosen metrics because of their widespread use and comprehensive nature, which allows for effective comparisons with other studies and ensures the credibility and comparability of our research. The evaluation metrics used in each of the three experiments are described below.

The object detection section was evaluated using precision (P), recall (R), mean average precision (mAP), run time, and model size for YOLOv5s, YOLOv5s-2Level, and YOLOv5s-little (Padilla et al., 2020). Among them, P, R and, mAP reflect the overall performance of the detector, and the higher the value, the better the performance. The running time and model size reflect the lightness of the detector, and the lower the value, the higher the lightness.

The multi-object tracking part uses higher order tracking accuracy (HOTA), multiple object tracking accuracy (MOTA), identification F1 (IDF1), ID switch (IDSW) for DeepSORT and the improved DeepSORT is evaluated comprehensively (Luiten et al., 2021). Among them, HOTA, MOTA, IDF1, and IDSW are important to reflect the performance of the tracking algorithm, the higher the value of HOTA, MOTA, and IDF1, the better the performance, and the smaller the value of IDSW, the better the performance.

The counting part uses counting accuracy and Statistical average counting accuracy (Rong et al., 2023) to evaluate the counting method. The higher the value of these two metrics, the closer the predicted value of the counting method is to the true value.

The detection results of YOLOv5s, YOLOv5s-2Level, and YOLOv5s-little

Table 1 displayed the evaluation results of the YOLOv5s, YOLOv5s-2Level, and YOLOv5s-little detectors. The main difference in precision between the three models is reflected in the values of mAP@0.5:0.95. And the mAP@0.5:0.95 values for YOLOv5s-2Level and YOLOv5s-little were 0.856 and 0.782, respectively, representing a decrease of 1.4% and 8.8% compared to the original YOLOv5s model. Furthermore, the detection time

Table 1 Detection results of YOLOv5s, YOLOv5s-2Level, and YOLOv5s-little

	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	Time/s	Size/MB
YOLOv5s	99.1	98.8	99.8	87	0.0127	14.4
YOLOv5s-2Level	98.7	99.6	99.8	85.6	0.0097	3.4
YOLOv5s-little	98.9	98.3	99.5	78.2	0.0079	0.9

per image for YOLOv5s-2Level and YOLOv5s-little models were 0.0097 s and 0.0079 s, respectively, resulting in a speed improvement of 30% and 60%. With a size of less than 1MB, the YOLOv5s-little model was more suitable for online, real-time applications and portable devices than the YOLOv5s and YOLOv5s-2Level. Considering detection speed, precision, and model size, the YOLOv5s-little model was chosen as detector for the later tracking experiments.

Figure 8 showed the detection results for YOLOv5s and YOLOv5s-little under sparse (in the left column of Fig. 8) and dense scene (in the right column of Fig. 8). In the former, nine fruits were present in the original image, and both YOLOv5s and YOLOv5s-little successfully detected them. In the latter, the original image contained twenty fruits, and both models also detected all of them with equal accuracy. Taken together, these results demonstrated the robust detection capability of both YOLOv5s and YOLOv5s-little across diverse scene densities.

(a) Original images in sparsely-fruited scenes.
Ground labeled result : 9

(b) Original images in densely-fruited scenes.
Ground labeled result : 20

(c) Detection map of YOLOv5s in sparsely-fruited scenes.
Detection result : 9

(d) Detection map of YOLOv5s in densely-fruited scenes.
Detection result : 20

(e) Detection map of YOLOv5s-little in sparsely-fruited scenes.
Detection result : 9

(f) Detection map of YOLOv5s-little in densely-fruited scenes.
Detection result : 20

Fig. 8 The detection results of YOLOv5s and YOLOv5s-little in different scenes

The tracking results of improved and original DeepSORT Based on YOLOv5s-little

The comparison results between the improved and original DeepSORT tracking algorithms based on YOLOv5s-little were presented in Table 2. The average or total of each evaluation metric in all test videos is labeled in bold. Five metrics mentioned in Sect. 3.2 were used to evaluate the tracking performance. The values of the HOTA, MOTA, IDF1, and IDSW indicators were 74.94%, 90.78%, 85.86%, and 76, respectively using the original DeepSORT, while they were 79.6%, 92.58%, 95.02%, and 11, respectively using the improved DeepSORT. The improved DeepSORT showed excellent performance in the first four indicators, with improvements of 4.66%, 1.8%, 9.16%, and 85% in HOTA, MOTA, IDF1, and IDSW, respectively, compared to the original algorithm. The improvement in IDSW was most significant, effectively reducing the problem of frequent IDs. This was very helpful for subsequent fruit-counting tasks and would lead to more accurate fruit-counting results, thus improving the efficiency and accuracy of crop management. Despite using IOU matching one more time than the original, the tracking speed of the improved DeepSORT did not decrease significantly. In conclusion, the improved DeepSORT showed significant improvements in both tracking stability and accuracy.

Figures 9 and 10 showed the tracking results of test videos 02 and 03 using the original and improved DeepSORT. Both videos had 100 frames. Test video 02 contained 19 fruits which were sparsely distributed, and video sequences were captured from right to left. Test video 3 contained 42 fruits which are densely distributed, while video 03 sequences were captured from right to left.

In Fig. 9a and c, the highest ID value by the original DeepSORT was 14, while the maximum ID value by the improved DeepSORT was 10. The original DeepSORT showed discontinuity in ID in the 50th frame, while the ID value by the improved DeepSORT remained continuous. In Fig. 9b and d, the maximum ID value by the improved DeepSORT was 19, while the maximum ID value by the original DeepSORT was 26. The maximum ID value by the original DeepSORT differed from the actual number of fruits by 7 units, while the improved DeepSORT was closer to the actual situation.

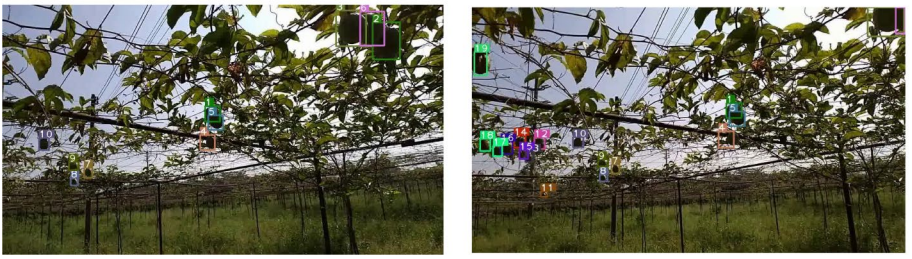
In Fig. 10a and c, the maximum ID value of the original DeepSORT reached 65, which exceeded the ground truth count significantly. However, the maximum ID value of the improved DeepSORT was 40, which was close to the ground truth count (equal to 42). In Fig. 10b and d, the maximum ID value of the original DeepSORT was 89, which was more than twice the ground truth count, while the maximum ID of the improved

Table 2 Tracking results of improved and original DeepSORT Based on YOLOv5s-little

Method	Test video	HOTA/%↑	MOTA/%↑	IDF1/%↑	IDSW↓	FPS $(f \cdot s^{-1})$↑
Original DeepSORT	01	83.42	96.24	97.06	4	35
	02	66.53	87.58	75.05	71	52
	03	79.72	88.52	94.08	1	27
	All	**74.94**	**90.78**	**85.86**	**76**	**38**
Improved DeepSORT	01	84.25	96.73	98.27	2	33
	02	75.57	90.29	92.58	9	52
	03	80.69	90.46	95.25	0	28
	All	**79.6**	**92.58**	**95.02**	**11**	**38**

(a) Original DeepSORT, Frame 50

(b) Original DeepSORT, Frame 100



(c) Improved DeepSORT, Frame 50

(d) Improved DeepSORT, Frame 100

Fig. 9 The tracking results of original and improved DeepSORT of test 02

(a) Original DeepSORT, Frame 50

(b) Original DeepSORT, Frame 100

(c) Improved DeepSORT, Frame 50

(d) Improved DeepSORT, Frame 100

Fig. 10 The tracking results of original and improved DeepSORT of test 03

DeepSORT was 43, which was also close to the ground truth count. In conclusion, the improved DeepSORT can effectively solve the problem of frequent IDs in fruit sparse and dense scenarios.

Evaluation of counting method

The comparison experiments between the MAX-IDVC approach and our counting method were shown in Fig. 11. The x-axis in Fig. 11 represented the video number, and the y-axis represented the fruit counts. The blue bars represented the ground truth of the fruit count value, while the red bars represented the fruit count values between the two methods including the MAX-IDVC approach and our counting method. The counting accuracy percentage values were shown above each bar.

According to the results of Fig. 11, our counting method achieved higher counting accuracy than that of the MAX-IDVC method for most videos. In Fig. 11a, the lowest counting accuracy achieved by the MAX-IDVC method is 52.4%, while the highest counting accuracy was 94.7%. In Fig. 11b, our counting method achieved the lowest counting accuracy of 76.2% and the highest counting accuracy of 100%. Statistical average counting accuracy using the MAX-IDVC method was 86.19%, while statistical average counting accuracy using our counting method was 93.28%. Therefore, our counting method can achieve better result than that of the MAX-IDVC method.

The regression curves between the predicted values and the ground truth count between the maximum ID value counting and the proposed methods were shown in Fig. 12. Figure 12a displayed the regression curve between the maximum ID value and the ground truth count, while Fig. 12b showed the regression curve between the proposed method count value and the ground truth count. The R^2 value of the regression curve between the maximum ID value and the ground truth count was 0.9419, while the R^2 value of the proposed method was 0.9633, which was higher than that of the maximum ID value regression curve. In addition, the slope of the regression curve of the proposed method was closer to 1 than that of the maximum ID value. In conclusion, the proposed method was more accurate in predicting the passion fruit yield.

Comparison with other MOT method

FairMOT and TransTrack were chosen for comparison with lightweight YOLOv5s-little and improved DeepSORT. FairMOT (Zhang et al., 2021) consists of two homogeneous branches—one for predicting pixel-level object scores and the other for re-ID features. This

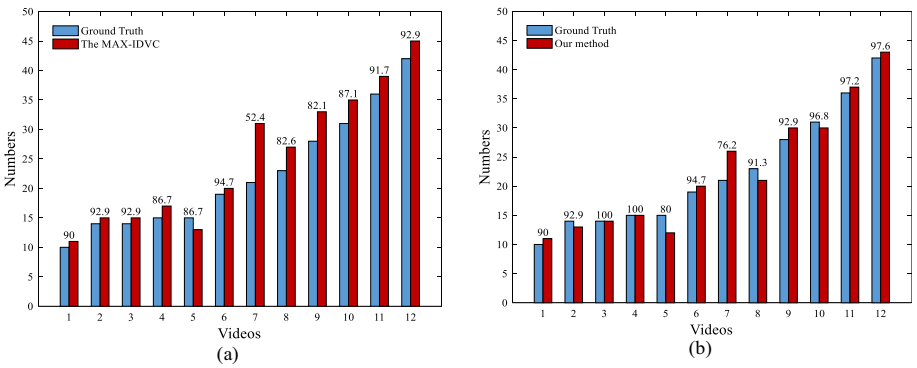


Fig. 11 The counting results of the ground truth and the predicted value. **a** The MAX-IDVC method. **b** Our counting method

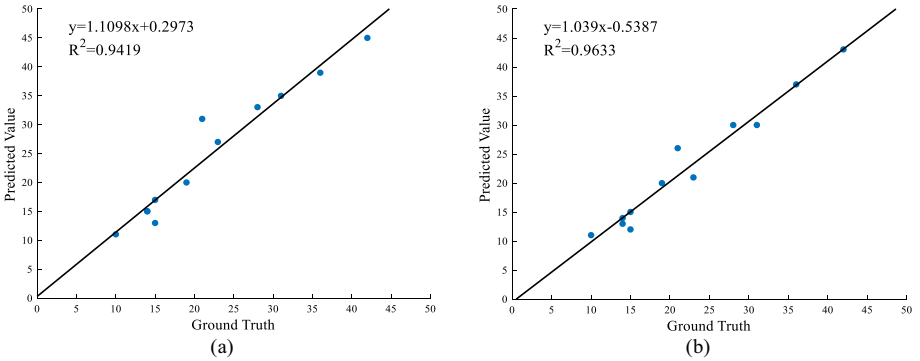


Fig. 12 The regression curve between the ground truth and the predicted value. **a** The MAX-IDVC method. **b** Our counting method

algorithm addresses the fairness issue between these two tasks and achieves high levels of detection and tracking accuracy. On the other hand, TransTrack (Sun et al., 2020) establishes a new joint detection and tracking paradigm. TransTrack fully leverages the advantages of transformer architecture to achieve object detection and object association in a single shot.

Table 3 shows the comparison of experimental results using our method with FairMOT and TransTrack tracking methods. The best value for each metric is shown in bold. Our method performs the best on four metrics of MOT. Specifically, the HOTA, MOTA, IDF1, and IDSW values of our approach were 79.6%, 92.58%, 95.02%, and 11, respectively. Compared to the FairMOT method, our method improved by 11.56% for HOTA, 7.91% for MOTA, and 19.29% for IDF1, while reducing IDSW by 269. Compared to TransTrack, our method achieved significant improvements of 25.24% for HOTA, 37.22% for MOTA, and 23.67% for IDF1. Additionally, the IDSW values of TransTrack were 351 higher than that of our approach. These comparative results further validated the effectiveness of our method in reducing frequent errors in fruit IDs and improving the accuracy of yield estimation algorithms.

Discussion

While existing fruit counting methods focus on clustered fruits (e.g., tomatoes, etc.), our study focuses on addressing the challenge of counting passion fruit individually. The characteristics of passion fruit make it similar in color to leaves and occlusion by leaves. To

Table 3 The experimental results of comparison with other tracking methods

Method	HOTA/%↑	MOTA/%↑	IDF1/%↑	IDSW↓
YOLOv5s-little + Improved DeepSORT	**79.60**	**92.58**	**95.02**	**11**
FairMOT	68.04	84.67	75.73	280
TransTrack	54.36	55.36	71.35	351

address these challenges, three key improvements for passion fruit counting are proposed: model lightweighting, delaying the creation of track and adding a second round of IoU matching to enhance tracking stability, and a counting method based on track information and track state. Our study shows that with these improvements, our proposed method makes significant progress in terms of model lightweight, tracking stability, and counting accuracy.

First, the YOLOv5s detection model is replaced by a lightweight detector named YOLOv5s-little. In the test set, YOLOv5s-little can effectively detect passion fruit in dense or sparse scenes, and the size of the model is only 0.9 MB. Compared to the lightweight networks such as MobileNet or ShuffleNet (Wang et al.,) based on YOLOv5s, the model of the proposed YOLOv5s-little is smaller in size. While the YOLOv5s-little performs well in that scene, it may not perform well when faced with different lighting or angle changes in different scenes. This may pose a challenge to its application in a wide range of scenes.

Second, we delay the creation of tracks and add the second round of IoU matching during passion fruit tracking, effectively reducing the frequent error IDs. Table 4 shows the results of the ablation study for the tracking improvement method. The best value for each metric is shown in bold. The original tracker achieved the HOTA of 74.94%, MOTA of 90.78%, IDF1 of 85.86%, and IDSW of 76. The value of HOTA and MOTA using the second round of IoU matching increases by 2.89% and 1.68% compared with the original tracker, which indicates that adopting it can get better tracking performance. Delaying the creation of tracks and the second round of IoU matching can further decrease the IDSW to 11. At the same time, the values of HOTA, MOTA, and IDF1 can also be enhanced by 4.66%,1.8%, and 9.16% compared with the original one (Table 4 first row). Therefore, by reducing IDs and improving tracking performance, the improved DeepSORT can consistently track passion fruit in videos. These enhancements significantly contribute to the accuracy improvement in subsequent counting results. However, when the restricted area is set unreasonably, it limits the generation of new tracks within that area, which might lead to missed detections. Therefore, considering the characteristics of different scenarios, it's necessary to adjust the restricted area based on the characteristics of different scenes.

Finally, when using DeepSORT to track each passion fruit, calculating the total number of passion fruits directly based on the maximum ID value leads to inaccurate counting. To solve this problem, a counting method based on track information and track state is proposed to overcome the double counting problem, and the number of passion fruits using our method is close to manual counting.

The different classes of MOT algorithms currently available all have their own characteristics. For example, FairMOT has higher computational efficiency and is able to perform target tracking in real time, but its performance may be poor in complex scenarios. TransTrack is capable of handling large-scale target sets based on Transformer, it may face computational complexity and speed challenges and is not applicable in real

Table 4 Ablation study of tracking improvement

Delaying the creation of tracks	The second round of IoU matching	HOTA↑	MOTA↑	IDF1↑	IDSW↓
		74.94	90.78	85.86	76
√		74.87	86.52	87.94	29
	√	77.83	92.46	91.30	33
√	√	**79.60**	**92.58**	**95.02**	**11**

production environments. Compared to these two tracking algorithms, the proposed improved DeepSORT offers an improvement in tracking accuracy while satisfying better real-time performance and can be adapted for use in practical applications.

When estimating the passion fruit yield, there are two key challenges in video data collection. Firstly, rapid movement of the camera can result in blurred images, making it difficult to capture and detect the fruits. Hence, stabilizing the camera is necessary to ensure accurate fruit detection. Secondly, shooting toward sunlight can result in excessively dark footage, affecting image quality. To better track the fruits, video capture should avoid shooting toward sunlight.

In passion fruit yield estimation, it is difficult to fully consider agricultural expertise and specific environmental factors using only machine learning models. In subsequent studies, incorporating human expertise into machine learning processes, such as integrating the experience of agricultural experts, can significantly improve the accuracy and interpretability of prediction results (Holzinger et al., 2022). This approach not only provides yield-related data, but also delves into the factors that influence plant growth stages, including soil conditions, climate change, and plant pests and diseases. Such insights help to identify the true causality behind yield fluctuations, providing farmers with more targeted support and decision-making advice.

Conclusion

This study presents a method for estimating passion fruit yield by combining the lightweight YOLOv5s-little detector and an improved DeepSORT algorithm. By adopting a lightweight detection network and multi-object tracking techniques, our method achieves accurate passion fruit counting in natural environments. The YOLOv5s-little detector demonstrates excellent performance with 98.9% precision, 98.3% recall, and 99.5% mAP@0.5, while the model size is only 0.9MB. The improved DeepSORT algorithm achieves HOTA of 79.6%, MOTA of 92.58%, IDF1 of 95.02%, and IDSW of 11 respectively. A specific counting method based on track information and track state is proposed to overcome the issue of double counting, achieving a statistical average counting accuracy of 95.1%. And the counting results from test videos are highly correlated with the manual counting results ($R^2 = 0.96$).

The passion fruit yield estimation method proposed in this study will have a profound impact on the agricultural industry. Issues of model oversizing and double counting in existing methods have been successfully addressed, providing a more stable and real-time means of yield estimation for agricultural production. In the future, further optimization of the model's performance will be undertaken by considering additional factors in real-world planting scenarios, such as changes in light and fruit ripeness. Meanwhile, the integration of human expertise into the machine learning process is planned to enhance the accuracy and interpretability of prediction results.

Funding Funding was provided by National Natural Science Foundation of China (Grant Nos. 61772209, 31600591), Science and Technology Planning Project of Guangdong Province (Grant No. 2019A050510034), National College Students Innovation and Entrepreneurship Training Program (Grant No. 202110564025), Key Research and Development Program of Guangzhou (No. 2024B03J1358).

References

Ariza-Sentís, M., Baja, H., Vélez, S., & Valente, J. (2023). Object detection and tracking on UAV RGB videos for early extraction of grape phenotypic traits. *Computers and Electronics in Agriculture, 211*, 108051. https://doi.org/10.1016/j.compag.2023.108051

Assunção, E. T., Gaspar, P. D., Mesquita, R. J. M., Simões, M. P., Ramos, A., Proença, H., & Inacio, P. R. M. (2022). Peaches detection using a deep learning technique—a contribution to yield estimation, resources management, and circular economy. *Climate, 10*(2), 11. https://doi.org/10.3390/cli10020011

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. *2016 IEEE international conference on image processing (ICIP)* (pp. 3464–3468). IEEE.

Bezerra, A. D. M., Pacheco Filho, A. J. S., Bomfim, I. G. A., Smagghe, G., & Freitas, B. M. (2019). Agricultural area losses and pollinator mismatch due to climate changes endanger passion fruit production in the Neotropics. *Agricultural Systems, 169*, 49–57. https://doi.org/10.1016/j.agsy.2018.12.002

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. Preprint retrieved from https://arxiv.org/abs/2004.10934

De Jong, S., Baja, H., Tamminga, K., & Valente, J. (2022). APPLE MOTS: Detection, segmentation and tracking of homogeneous objects using MOTS. *IEEE Robotics and Automation Letter, 7*(4), 11418–11425. https://doi.org/10.1109/LRA.2022.3199026

Egi, Y., Hajyzadeh, M., & Eyceyurt, E. (2022). Drone-computer communication based tomato generative organ counting model using YOLO V5 and deep-sort. *Agriculture, 12*(9), 1290. https://doi.org/10.3390/agriculture12091290

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., Li, R., Fu, L., & Zhang, Q. (2022). A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. *Computers and Electronics in Agriculture, 197*, 107000. https://doi.org/10.1016/j.compag.2022.107000

Ge, Y., Lin, S., Zhang, Y., Li, Z., Cheng, H., Dong, J., Shao, S., Zhang, J., Qi, X., & Wu, Z. (2022). Tracking and counting of tomato at different growth period using an improving YOLO-deepsort network for inspection robot. *Machines, 10*(6), 489. https://doi.org/10.3390/machines10060489

Guo, S., Wang, S., Yang, Z., Wang, L., Zhang, H., Guo, P., Gao, Y., & Guo, J. (2022). A review of deep learning-based visual multi-object tracking algorithms for autonomous driving. *Applied Sciences, 12*(21), 10741. https://doi.org/10.3390/app122110741

He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., Majeed, Y., & Dhupia, J. (2022a). Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. *Computers and Electronics in Agriculture, 195*, 106812. https://doi.org/10.1016/j.compag.2022.106812

He, L., Wu, F., Du, X., & Zhang, G. (2022b). Cascade-SORT: A robust fruit counting approach using multiple features cascade matching. *Computers and Electronics in Agriculture, 200*, 107223. https://doi.org/10.1016/j.compag.2022.107223

Holzinger, A., Saranti, A., Angerschmid, A., Retzlaff, C. O., Gronauer, A., Pejakovic, V., Medel-Jimenez, F., Krexner, T., Gollob, C., & Stampfer, K. (2022). Digital transformation in smart farm and forest operations needs human-centered AI: Challenges and future directions. *Sensors, 22*(8), 3043. https://doi.org/10.3390/s22083043

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO.' *Precision Agriculture, 20*(6), 1107–1135. https://doi.org/10.1007/s11119-019-09642-0

Liu, S., Cossell, S., Tang, J., Dunn, G., & Whitty, M. (2017). A computer vision system for early stage grape yield estimation based on shoot detection. *Computers and Electronics in Agriculture, 137*, 88–101. https://doi.org/10.1016/j.compag.2017.03.013

Liu, T. H., Nie, X. N., Wu, J. M., Zhang, D., Liu, W., Cheng, Y. F., Zheng, Y., Qiu, J., & Qi, L. (2023). Pineapple (*Ananas comosus*) fruit detection and localization in natural environment based on binocular stereo vision and improved YOLOv3 model. *Precision Agriculture, 24*(1), 139–160. https://doi.org/10.1007/s11119-022-09935-x

Liu, X., Chen, S. W., Liu, C., Shivakumar, S. S., Das, J., Taylor, C. J., Underwood, J., & Kumar, V. (2019). Monocular camera based fruit counting and mapping with semantic data association. *IEEE Robotics and Automation Letter, 4*(3), 2296–2303. https://doi.org/10.1109/LRA.2019.2901987

Luiten, J., Osep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, L., & Leibe, B. (2021). HOTA: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision, 129*(2), 548–578. https://doi.org/10.1007/s11263-020-01375-2

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. K. (2021). Multiple object tracking: A literature review. *Artificial Intelligence, 293*, 103448. https://doi.org/10.1016/j.artint.2020.103448

Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. *2020 international conference on systems, signals and image processing (IWSSIP)* (pp. 237–242). Niterói: IEEE.

Qi, X., Dong, J., Lan, Y., & Zhu, H. (2022). Method for identifying litchi picking position based on YOLOv5 and PSPNet. *Remote Sensing, 14*(9), 2004. https://doi.org/10.3390/rs14092004

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. Preprint retrieved from https://arxiv.org/abs/1804.02767

Rong, J., Zhou, H., Zhang, F., Yuan, T., & Wang, P. (2023). Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion. *Computers and Electronics in Agriculture, 207*, 107741. https://doi.org/10.1016/j.compag.2023.107741

Shen, L., Su, J., He, R., Song, L., Huang, R., Fang, Y., Song, Y., & Su, B. (2023). Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLOv5s. *Computers and Electronics in Agriculture, 206*, 107662. https://doi.org/10.1016/j.compag.2023.107662

Sun, P., Cao, J., Jiang, Y., Zhang, R., Xie, E., Yuan, Z., Wang, C., & Luo, P. (2020). Transtrack: Multiple object tracking with transformer. Preprint retrieved form https://arxiv.org/abs/2012.15460

Tan, C., Li, C., He, D., & Song, H. (2022). Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. *Computers and Electronics in Agriculture, 193*, 106683. https://doi.org/10.1016/j.compag.2021.106683

Thokchom, R., & Mandal, G. (2017). Production preference and importance of passion fruit (*Passiflora edulis*): A review. *Journal of Agricultural Engineering and Food Technology, 4*(1), 27–30.

Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., Wan, H., & Xue, Y. (2020). Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. *Precision Agriculture, 21*(5), 1072–1091. https://doi.org/10.1007/s11119-020-09709-3

Vasconez, J. P., Delpiano, J., Vougioukas, S., & Auat Cheein, F. (2020). Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. *Computers and Electronics in Agriculture, 173*, 105348. https://doi.org/10.1016/j.compag.2020.105348

Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering, 210*, 271–281. https://doi.org/10.1016/j.biosystemseng.2021.08.015

Wang, L., Zhao, Y., Xiong, Z., Wang, S., Li, Y., & Lan, Y. (2022a). Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model. *Frontiers in Plant Science, 13*, 965425. https://doi.org/10.3389/fpls.2022.965425

Wang, Z., Jin, L., Wang, S., & Xu, H. (2022b). Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. *Postharvest Biology and Technology, 185*, 111808. https://doi.org/10.1016/j.postharvbio.2021.111808

Wang, X., Wu, Z., Jia, M., Xu, T., Pan, C., Qi, X., & Zhao, M. (2023). Lightweight SM-YOLOv5 tomato fruit detection algorithm for plant factory. *Sensors, 23*(6), 3336. https://doi.org/10.3390/s23063336

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. *2017 IEEE international conference on image processing (ICIP)* (pp. 3645–3649). IEEE.

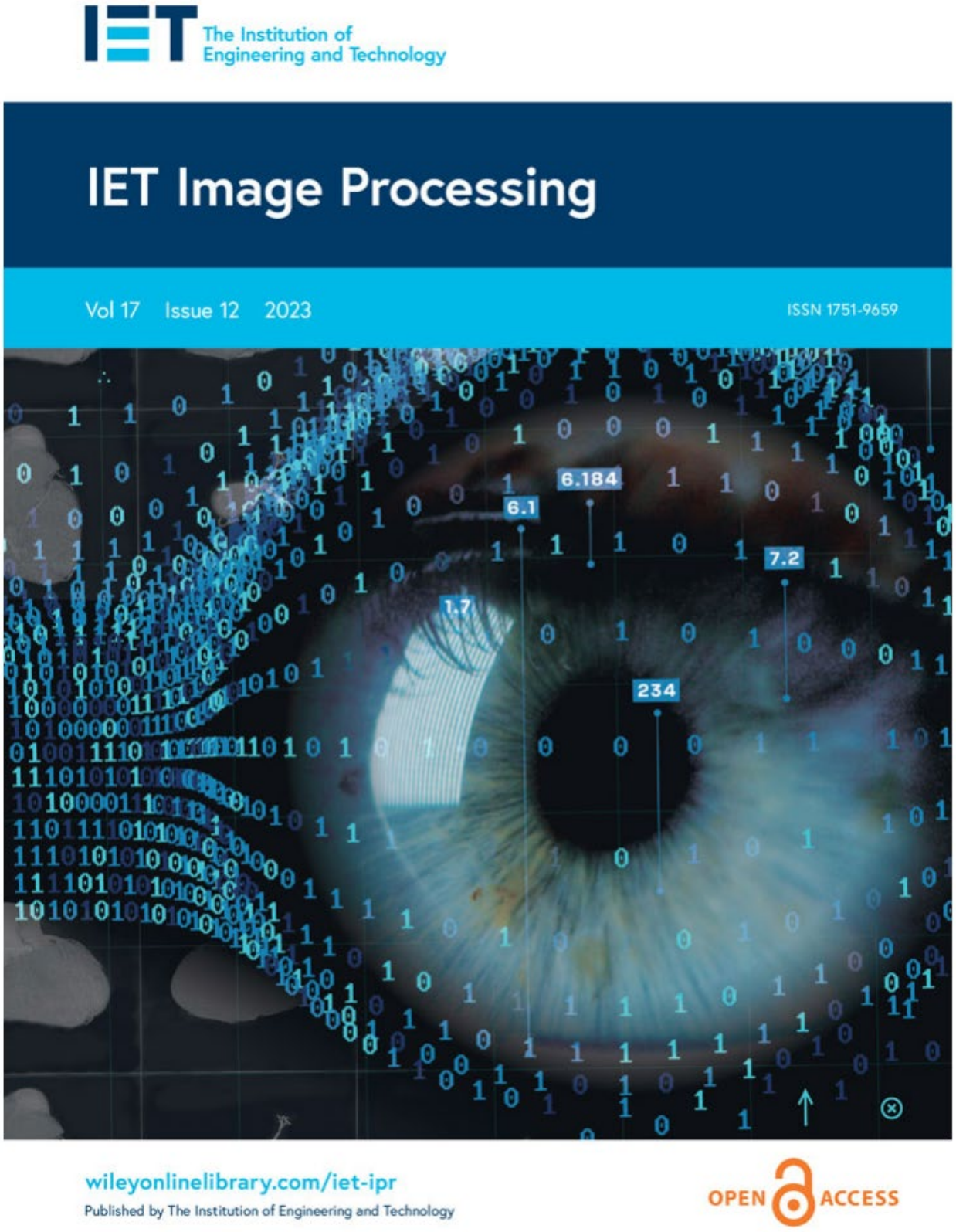
Yang, H., Chang, F., Huang, Y., Xu, M., Zhao, Y., Ma, L., & Su, H. (2022). Multi-object tracking using Deep SORT and modified CenterNet in cotton seedling counting. *Computers and Electronics in Agriculture, 202*, 107339. https://doi.org/10.1016/j.compag.2022.107339

Zhang, W., Wang, J., Liu, Y., Chen, K., Li, H., Duan, Y., Wu, W., Shi, Y., & Guo, W. (2022). Deep-learning-based in-field citrus fruit detection and tracking. *Horticulture Research, 9*, 1–10.

Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2021). FairMOT: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision, 129*(11), 3069–3087. https://doi.org/10.1007/s11263-021-01513-4

Zhou, Z., Song, Z., Fu, L., Gao, F., Li, R., & Cui, Y. (2020). Real-time kiwifruit detection in orchard using deep learning on Android™ smartphones for yield estimation. *Computers and Electronics in Agriculture, 179*, 105856. https://doi.org/10.1016/j.compag.2020.105856

2.8 MaskDis R-CNN: An instance segmentation algorithm with adversarial network for herd pigs



Received: 1 April 2023 | Revised: 2 July 2023 | Accepted: 11 July 2023

IET Image Processing

DOI: 10.1049/ipr2.12880

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

MaskDis R-CNN: An instance segmentation algorithm with adversarial network for herd pigs

Shuqin Tu[1] | Qiantao Zeng[1] | Haofeng Liu[2] | Yun Liang[1] | Xiaolong Liu[1] | Lei Huang[1] | Zhengxin Huang[1]

[1]College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China

[2]Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

Correspondence
Yun Liang, College of Mathematics and Informatics, South China Agricultural University, No. 483, Wushan Road, Tianhe District, Guangzhou City, Guangdong Province 510642, China.
Email: Liang_Yun168@163.com

[Correction added on 8 September 2023, after first online publication: Funding information has been updated in this version].

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 61772209, 31600591; the Science and Technology Planning Project of Guangdong Province, Grant/Award Number: 2019A050510034; Guangzhou Key R&D Program Project, Grant/Award Numbers: 2023B03J1363, 202206010091; College Students' Innovation and Entrepreneurship Competition, Grant/Award Number: 202110564025

Abstract

The current instance segmentation method can achieve satisfactory results in common scenarios. However, under the overlap or partial occlusion between targets caused by the complex scenes, accurate segmentation of pigs remains a challenging task. To address the problem, the authors propose an instance segmentation method based on Mask Scoring region-based convolutional neural networks (R-CNN) (MS R-CNN), which creates the adversarial network called MaskDis in the head branch of MS R-CNN. The MaskDis is trained as a discriminator using a generative adversarial network, and the MS R-CNN model is used as a generator during model training. The adversarial training enables the generator to learn context information and features at the pixel level, which effectively improves the segmentation quality under pigs' overlapping or dense occlusions scenes. Experimental conducted on the pig object segmentation dataset show that the proposed approach achieves a precision of 92.03%, a recall of 92.18%, and an F1 score of 0.9210. Compared with the basic MS R-CNN model, the approach achieved a 2.25% improvement in precision and 1.18% improvement in F1 score. Furthermore, the improved approach outperformed advanced instance segmentation methods such as YOLACT, Swin Transformer, YOLOv5-seg, and SOLOv2 on COCO evaluation metrics. These experimental results demonstrate the effectiveness of the proposed approach in instance segmentation of pigs in complex scenes, providing technical support for non-contact pig automatic management.

1 | INTRODUCTION

The pig farming industry plays an important role in the national economy, and the main direction of the pig farming industry is large-scale and intelligent pig farming. In large-scale intelligent pig breeding, accurate and timely collection of phenotype information for controlling pig growth is the key technology for precision farming [1–3]. Observing animals on an individual level to assess their health and welfare is necessary. However, on a real-world commercial farm, observing pig contour information by humans is impractical, and subjective human observation can lead to errors [4]. Techniques based on deep learning have been adopted in the last few years, achieving excellent performance in many fields, image inpainting, natural language processing, and so on [5–7]. There have been successful uses of deep learning algorithms for acquiring pig information, providing an efficient, contactless, and non-destructive intelligent method [8–10]. However, objective factors in the pig farm environment, such as light fluctuations, high similarity, and pig adhesion, result in decreased accuracy of detection and segmentation, which fails to meet practical applications in pig farming. Therefore, designing and developing an accurate and efficient segmentation algorithm for large-scale and intelligent pig farming is important. The ability to automatically detect and segment the contour of individual pigs can assist in the early detection of potential health or welfare problems without the need for human observation.

Due to the development of deep learning technology, the performance of instance segmentation has been significantly improved. The classical methods rely on object detectors, such

as Faster region-based convolutional neural networks (R-CNN) [11] by detecting objects, obtaining bounding boxes, and then extracting region features for pixel-wise segmentation using techniques like RoIPooling or RoI-Align. Mask R-CNN [12] is a representative instance segmentation algorithm that adds a mask branch on the Faster R-CNN to predict the mask of the object. Mask Scoring R-CNN [13] modifies the scoring strategy for mask segmentation based on Mask R-CNN and Cascade R-CNN [14] progressively improves object localization using cascade structure to achieve more accurate mask prediction. In addition, a series of instance segmentation algorithms are derived from the single-stage fully convolutional one-stage object detection (FCOS) [15] as the target detection framework. MEInst [16] and CondInst [17] extend FCOS by predicting the encoding mask vector or mask kernel for dynamic convolution [18]. These methods have achieved satisfactory performance in instance segmentation and effectively promote the development of instance segmentation. However, these methods encountered boundary leakage issues, where they failed to properly segment instances of individuals that overlapped within the same class, and the resulting mask segmentation lacked smoothness in its details.

Instance segmentation methods have been widely used in livestock welfare analyses. For example, cattle instances were segmented from real animal farms using these methods, achieving an average accuracy of 92% [19]. Mask R-CNN based on a dual attention guided feature pyramid network was introduced for instance segmentation of group-housed pigs [20], effectively segmenting individual pigs and achieved an average precision (AP) of 93.1%. In addition, our previous study applied Soft-NMS to Mask R-CNN for instance segmentation of pigs with complex backgrounds [21], achieving a harmonic average (F1) of 93.74%. Mask R-CNN combined with a support vector machine (SVM) classifier to identify individual cows and achieved an accuracy of 98.67% for cows [22]. However, the above methods do not consider the diversity of samples and objective factors such as complex light variations, occlusion, adhesion of pigs, and complex backgrounds. Moreover, adversarial networks can be used to significantly improve the performance of the model during training using automatic annotation of samples, which can solve the problems of instance segmentation mentioned above.

Adversarial networks [23] have become popular algorithm because it is capable of learning data distributions without relying on annotations. And its performance can be significantly improved if annotations are used in the training. Adversarial networks were applied to semantic segmentation [24], which detects and corrects higher-order inconsistencies between segmentation maps generated by segmentation networks and the ground truth (GT) segmentation maps. It had also been used for the segmentation of medical images [25], overcoming the limitation of classical adversarial network discriminators, which provide a single scalar true/false output, by generating stable and sufficient gradient feedback for the network. In addition, adversarial networks have been applied to image in painting. Repair network and optimization network (RNON) is an efficient image in painting method consisting of two mutually independent generative adversarial networks, with one network functioning as an image in painting network and the other as an image optimization network [26]. Therefore, it can be widely used in many fields such as image segmentation [27, 28], image classification [29, 30], and so on [31–33].

To address the issue of unsatisfactory segmentation performance in scenarios involving pig overlapping and occlusions, this paper proposed an instance segment network model combining adversarial network named MaskDis with the basic MS R-CNN model. Firstly, the MaskDis is used as a discriminator and the mask head of MS R-CNN is used as a generator; they are trained using an adversarial training approach. After adversarial learning between the generator and the discriminator, the generator can learn pixel-level, low-level, and mid-level features, as well as context information for better segmentation performance. Finally, adversarial training makes the prediction mask close to the GT, resulting in improved segmentation quality in complex scenarios.

For this paper, the main contributions are as follows: (1) we proposed an improved instance segmentation algorithm to enhance the segmentation quality by fusing the adversarial network in the MS R-CNN model. (2) We designed an adversarial network (MaskDis) model achieving better performance of instance segmentation under pig overlapping and occluded scenarios. (3) We completed experimental validation with a variety of advanced instance segmentation algorithms on the pig segmentation dataset and proved that our method has better segmentation performance.

2 | METHODS

2.1 | Overall framework of the instance segmentation algorithm

The structure of MaskDis R-CNN (shown in Figure 1) includes two components: MS R-CNN and MaskDis Head. The method firstly extracted the feature maps from the input images using a backbone network of ResNet-101and Feature Pyramid Networks (FPN), and the resulting feature maps are then fed to the Region Proposal Network (RPN) to generate Region of Interests (RoIs). Secondly, the RoiAlign layer fixed the ROIs to the same size and then fed them to the Fully Connected layers (FC layers) used for classification and detection and Fully Convolutional Network (FCN) used for segmentation. The prediction mask, the image of individual pig, and the GT are simultaneously fed into the adversarial network head (MaskDis head) for adversarial training, and then back-propagated to the generator(MS R-CNN) by the multiscale loss function.

The following subsections illustrate the process of the MS R-CNN and MaskDis head model.

2.2 | The basic framework of MS R-CNN

The basic framework of MS R-CNN consists of the following four components:

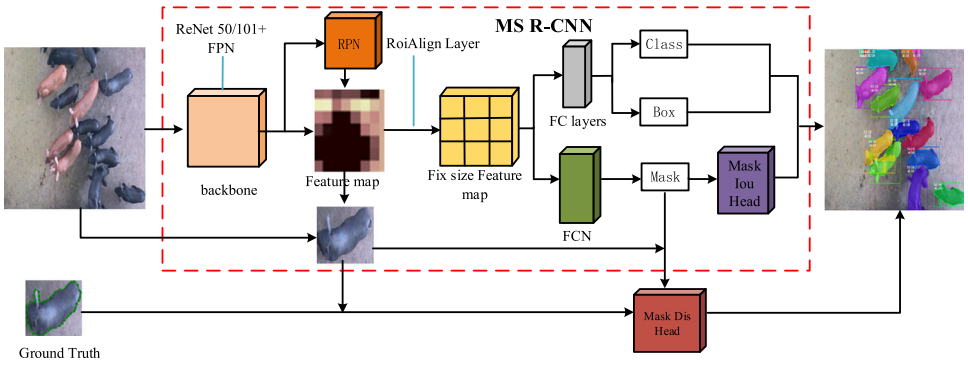


FIGURE 1 The structure of MaskDis R-CNN.

Backbone. ResNet-101 and FPN are a network structure designed for multi-scale feature extraction. The ResNet-101 consists of multiple residual blocks, which effectively reduces the number of parameters in the convolution process and prevents the degradation of the network due to the increase in network depth. ResNet-101 generates five different scale feature maps, which are fed into the FPN. In FPN, the pre-processed images are firstly extracted by the bottom-up forward process to get feature maps of different scales, and then the feature maps P6 are firstly obtained by down-sampling, and then the corresponding feature maps are fused by top-down up-sampling to get multi-scale fused. The features obtained by fusion have more robust semantic information, and also effectively improve the speed and the accuracy of detection.

Region Proposal Networks. RPN are efficient in generating regions of interest because of the advantages of fast candidate region generation and low computational cost, the input of RPN networks is the feature map extracted from the backbone network, and the output is a batch of candidate frames and region scores, enabling end-to-end training. The anchor mechanism was used in the RPN network. The anchor is in the form of a 3×3 sliding window ($n = 3$) over each layer of the input feature map, generating k region boxes of different sizes and proportions at the centre of the sliding window.

Mask Head. The output of Mask Head includes classification prediction, bounding-box regression, and instance segmentation mask prediction. The classification branch and bounding-box regression branch share the features extracted in the first stage, including pooling the RoIs by RoIAlign, followed by extracting the deep features of the RoI feature map by two FC layers, after which they start to divide into two branches and perform one full connection each and then output the results. The principles of classification regression and bounding-box regression are the same as those of classification and border regression in RPN. The RoI feature map used in the segmentation branch of the example is independent of the two branches mentioned above. First, RoI is processed by RoIAlign to obtain a $14 \times 14 \times 256$ RoI feature map, where 14×14 represents the pixels of the feature map and 256 represents the number of channels of the feature map. Then after passing through four convolutional layers comprising the FCN, one layer of deconvolution, and one layer of convolution, the final result of instance segmentation with a scale size of 28×28 is generated.

MaskIoU Head. The input features of the head branch of MaskIoU Head are obtained from Mask Head. Then, the MaskIoU values are obtained through the calculation of four convolutional layers and three FC layers. In the four convolutional layers, the first layer uses a convolutional kernel with a size of $3 \times 3 \times 257$, and the remaining three layers use convolutional kernels with a size of $3 \times 3 \times 256$. In the three FC layers, the output of the first two layers is 1024, and the output of the last layer is the number of categories (set to 2 in the experiment).

2.3 | MaskDis Head

Figure 2 shows the structure of the adversarial network of the herd pigs instance segmentation model. The adversarial network head branch, called MaskDis Head, is added to the MS R-CNN model. And Mask Head is used as the generator and MaskDis Head as the discriminator during model training.

Firstly, True/False samples are obtained by dot-multiplying the true mask or the predicted mask with the RoI of the original image, respectively. Then, these True/False samples are used as input data and fed into the network for computation. The input data consists of pixel-level information with dimensions of $28 \times 28 \times 3$. The output of the first layer has dimensions of $14 \times 14 \times 64$, and the output of the second layer has dimensions of $7 \times 7 \times 128$. The network structure comprises two convolutional layers, each with a 5×5 convolutional kernel size, 64 and 128 channels, and a stride of 2. Finally, the pixel-level features extracted from the first convolutional layer output, and the features from the second convolutional layer are combined to form a one-dimensional vector. Hierarchical features are extracted from multiple layers of the MaskDis Head to compute the multi-scale L1 loss. This loss effectively captures both

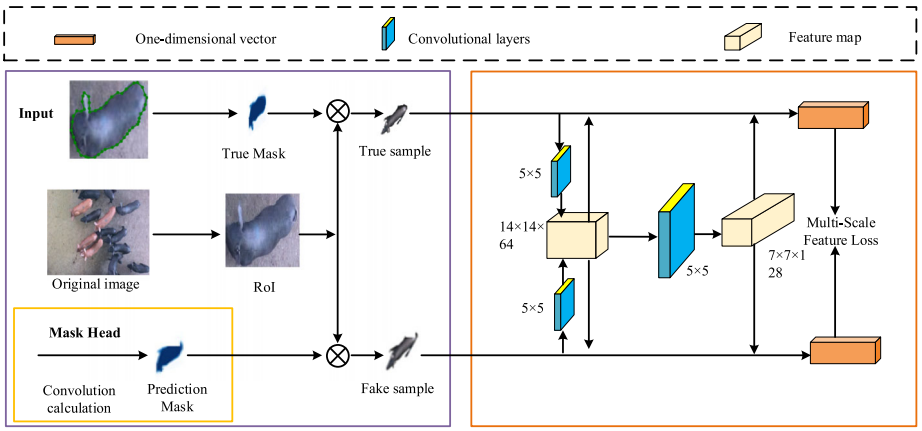


FIGURE 2 The architecture of the MaskDis head.

ALGORITHM 1

Input: *Truemask*,*RoI*,*Predictionmask*; Output: *Loss*

1. $Truesample \leftarrow Truemask \otimes RoI$
2. $Featuremap1\_T \leftarrow Convolution\_layer1(Truesample, (5,5), 64, 2)$
3. $Featuremap2\_T \leftarrow Convolution\_layer2(featuremap1\_T, (5,5), 128, 2)$
4. $Fakesample \leftarrow Predictionmask \otimes RoI$
5. $Featuremap1\_P \leftarrow Convolution\_layer1(Fakesample, (5,5), 64, 2)$
6. $Featuremap2\_P \leftarrow Convolution\_layer2(featuremap1\_P, (5,5), 128, 2)$
7. $Loss \leftarrow \frac{1}{N}\sum_{n=1}^{N} smoothL1_{loss}(fc(Truesample, Featuremap1\_T, Featuremap2\_T), fc(Fakesample, Featuremap1\_T, Featuremap2\_T))$

Return: *Loss*

long- and short-range spatial relations between pixels by utilizing hierarchical features, including pixel-level, low-level, and mid-level features.

The generator and discriminator networks are trained alternately in an adversarial manner: the Mask Head is trained to minimize the multi-scale L1 loss, while the MaskDis Head is trained to maximize the same loss function. This adversarial training improves the quality of segmentation as the generator and discriminator learn pixel-level features, low-level features, and mid-level features. The relevant pseudo-code is presented in the table, and the improvements are described in Algorithm 1.

More details including number of feature maps used in each convolutional layer can be found in Figure 2.

In MaskDis Head, the generator generates n predicted masks denoted as $x_n$and the corresponding original map RoI and true masks denoted as r_n and y_n, respectively, and the Multi-Scale Feature Loss function L_{Dis} is defined as

$$\min_{\theta_G}\max_{\theta_D} L_{Dis}(\theta_G, \theta_D) = \frac{1}{N}\sum_{n=1}^{N} smoothL1_{loss}\left(fc(r_n \cdot x_n), fc(r_n \cdot y_n)\right) \quad (1)$$

θ_G and θ_D represent the parameters for the generator and the discriminator. $r_n \cdot x_n$is the result of the original map RoI and the predicted mask dot product, and $r_n \cdot y_n$ is the result of the original map RoI and the GT mask dot product. The formula $smoothL1_{loss}(T_{pred}, T_{gt})$ is defined as follows:

$$smoothL1_{loss}\left(T_{pred}, T_{gt}\right) = \begin{cases} (T_{pred} - T_{gt})^2/2, if \left|T_{pred} - T_{gt}\right| < 1 \\ \left|T_{pred} - T_{gt}\right| - 1/2, otherwise \end{cases} \quad (2)$$

In the pig instance segmentation algorithm for fusion adversarial networks, the RoI head branch consists of three parts, Mask Head, MaskIoU Head, and MaskDis Head. L_{cls} is the loss value for classification regression, L_{box} is the loss value for border regression, L_{mask} is the loss value for instance segmentation generator, L_{IoU} is the loss value for MaskIoU regression, and L_{Dis} is the loss value for MaskDis Head regression. When training the mask generator, the loss function of the RoI branch is defined as follows:

$$L_{RoI} = L_{\text{cls}} + L_{box} + L_{mask} + L_{IoU} + L_{Dis} \quad (3)$$

3 | RESULTS AND DISCUSSION

In this section, MaskDis R-CNN is compared with Mask R-CNN and MS R-CNN. In addition, it is evaluated in the same experimental configuration: (1) comparison of experimental results in front and top views; (2) comparison of results on COCO evaluation metrics; (3) comparison in segmentation quality and scoring; and (4) comparison with other advanced instance segmentation methods.

3.1 | Experimental parameters and evaluation indicators

A device configuration is established for the implementation of the proposed approach, which consists of Python 3.7

FIGURE 3 Part of the dataset.

TABLE 1 distribution of the dataset.

Dataset	Front view	Top view	Total
Train	105	75	180
Test	65	70	135
Total	170	145	315

software PyTorch running on a PC with AMD Ryzen5 2600X and 3.00 GHz processor, 64 GB RAM and NVIDIA GeForce RTX TITAN X GPU with 12GB GPU VRAM.

Experimental data were collected randomly over 10 days of videos, containing 7 h of video per day, from 9:00 to 16:00 in the 'Lejiazhuang Pig Farm', Foshan City, Guangdong Province. And they were recorded by FL3-U3-88S2C-C cameras. The experimental data were saved in the audio video interleaved format with a video frame rate of 25 fps. To obtain adequate and better images, we focused on the five pens from the top and front view angles. And the size of the pens was 7 m × 5 m × 3 m (m represents the unit of length), respectively, and the number of pigs in each pen ranged from 3 to 20. Part of the training set and test set data are shown in Figure 3.

After obtaining the video data, a total of 315 images were selected according to a certain ratio, and these images were divided into the train dataset and test dataset. The detailed information of the dataset is shown in Table 1. In total, 180 images were selected as the train dataset, including 105 images collected by top view and 75 images by front view. And 135 images were selected as the test dataset, including 65 images collected by front view and 70 images by top view. Labelling the 315 images with including 3423 pig objects cost about 135 person-hours. Moreover, all animal experiments were conducted following the guidelines provided by the Guangdong Provincial Laboratory Animal Welfare and Ethical Review Guidelines and were approved by the Animal Welfare Committee of South China Agricultural University (No: 2021F129).

Data pre-processing uses data augmentation, which includes random horizontal flip, random brightness adjustment, random contrast adjustment, random saturation adjustment, and random hue adjustment. The Resnet-101 and FPN is used as Backbone, where FPN uses layers 2 to 5 of the Resnet-101 network, and the output of FPN has 256 channels. The number of foregrounds retained after post-processing NMS is adjusted in RPN, and its size is set to 1000. The learning rate is set to 0.0025, the epoch is set to 90, the batch size is set to 2, the scale of degradation is 0.1, and the value of weight decay is set to 0.0001.

To analyze the quality of segmentation results, we used recall, precision, F1 scores, and Precision-Recall (P-R) curves as evaluation metrics. AP is the area under the P-R curve and IoU is the degree of overlap between the predicted bounding box and the GT, which can be used to summarize the performance of an object detection model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\text{AP} = \int_0^1 \text{Precision} \cdot \text{Recalldr} \tag{7}$$

$$\text{IoU} = \frac{\text{detection result} \cap \text{ground truth}}{\text{detection result} \cup \text{ground truth}} \tag{8}$$

where True Positive (TP) is the number of pixels correctly predicted to be pig category, False Positive (FP) is the number of pixels incorrectly predicted to be pig category, False Negative (FN) is the number of pixels predicted to the pig category as the background, and F1 is a comprehensive evaluation

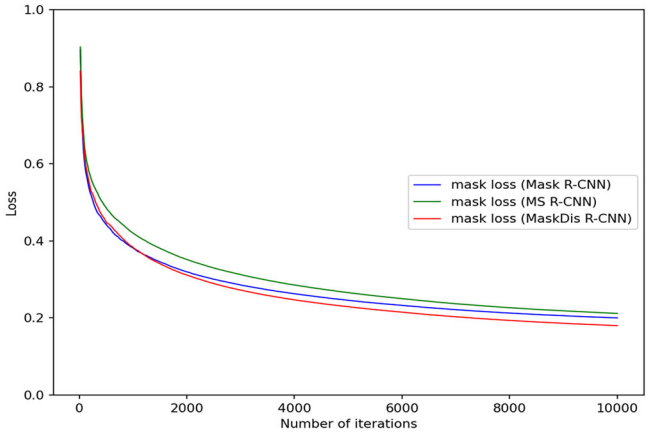


FIGURE 4 Comparison of training loss iteration curves of the model.

metrics of precision and recall rate. In our study, we adopted the standard COCO style AP0.5:0.95 metric, which computes the AP across various IoU thresholds ranging from 0.5 to 0.95 with an interval of 0.05. Additionally, we calculated the AP0.5 and AP0.75 metrics, which provide the AP values for different IoU thresholds. Moreover, we computed AR0.5 and AR0.75, representing the average recall (AR) values for different IoU thresholds, and obtained the mean average recall (mAR) by averaging the AR values at each IoU threshold. The IoU metric serves to assess the accuracy of object detection by measuring the overlap between the predicted bounding boxes and the GT. It is computed as the area of overlap between the two boxes divided by the area of their union.

3.2 | Experimental results

The comparative training loss iteration curves of the three instance segmentation models are shown in Figure 4. The blue, green, and red curves are the segmentation mask loss values of Mask R-CNN, MS R-CNN, and MaskDis R-CNN, respectively. According to Figure 4, each curve starts to converge at approximately 2000 iterations, after which the gap between the loss values of the Mask R-CNN model and the MS R-CNN model gradually decreases, while the gap between the loss values of both and MaskDis R-CNN model gradually increases. In general, the loss value of the MaskDis R-CNN is relatively small, indicating that the segmentation model of the fusion the adversarial network convergence more easily.

Table 2 shows the statistics of image detection results and comparisons. The number of pigs is the total of individual pigs in the test dataset, the test number is the total of individual pigs detected by the model, and the correct number is the total of individual pigs correctly detected by the model in the test number. The MaskDis R-CNN model detected a total of 1204 group-housed pigs, of which 1108 pigs are correctly detected, with a slight improvement in recall to 92.18%, precision increased by 2.25% to 92.03%, and F1 score increased by 0.0118 to 0.9210 when compared with the MS R-CNN model. The model shows a larger improvement on the test set from the front view, where the number of detections is 494, the number of correctly detected is 413. The F1 score increased by 2.56% to 84.63%. In the test set from the top view, the number of pigs detected is 710, the number of correct detections is 695, the recall rate is 96.53%, the precision rate increased by 1.46% to 97.89%, and the F1 score increased only slightly. The MaskDis R-CNN performs better in Precision and F1 values than the MS R-CNN. Thus, the MaskDis R-CNN demonstrates improvements in recall and precision, indicating an improved segmentation quality compared to the Mask R-CNN and MS R-CNN models.

The P-R curve (PR curve) of the three models is shown in Figure 5, with an IoU threshold of 0.75 for the PR curve evaluation criterion. The blue line represents the PR curve of Mask R-CNN, the red line represents the PR curve of MS R-CNN, and the green line represents the PR curve of the MaskDis R-CNN. Figure 5 shows that the green line is closest to the right and covers the largest area, indicating better segmentation performance. Based on the COCO evaluation metric, higher scores for correctly detected results lead to a larger area under the PR curve and higher average accuracy rates. The blue triangle in Figure 5 represents the intersection point of the three curves. To the left of the intersection point, the distance between the

TABLE 2 Statistics of image detection results and comparison.

Model	Image type	Number of pigs	Number of tests	Number of corrects	Recall (%)	Precision (%)	F1	Time (s)
	Front view	482	608	423	87.76	69.57	0.7761	
Mask R-CNN	Top view	720	707	690	96.83	97.60	0.9671	0.284
	Total	1202	1315	1113	**92.60**	84.64	0.8844	
	Front view	482	505	405	84.02	80.20	0.8207	
MS R-CNN	Top view	720	728	702	97.50	96.43	0.9696	0.286
	Total	1202	1233	1107	92.10	89.78	0.9092	
	Front view	482	494	413	85.68	83.60	0.8463	
MaskDis R-CNN	Top view	720	710	695	96.53	97.89	0.9720	0.288
	Total	1202	1204	1108	92.18	**92.03**	**0.9210**	

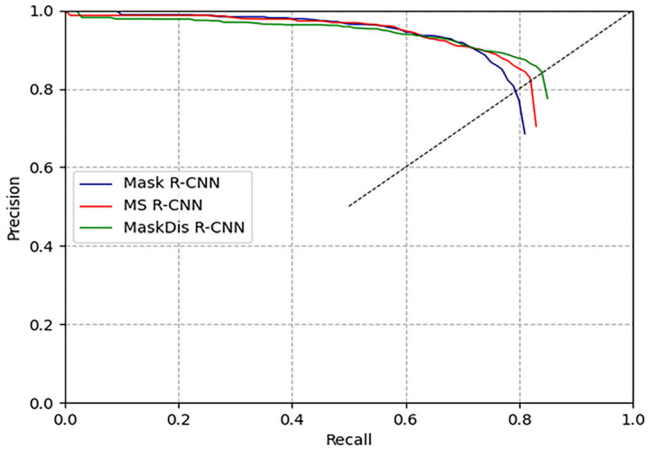


FIGURE 5 Precision-recall curve of herd pig instance segmentation model with fusion adversarial network.

three curves is relatively similar, but to the right of the intersection point, the distance between the curves gradually increases. Due to the fusion of adversarial networks, the quality of detection and segmentation results is improved, resulting in higher recall and precision values.

Based on Figure 5, MaskDis R-CNN demonstrates superior segmentation effectiveness compared to the other two models. Figure 6 illustrates the segmentation results achieved by MaskDis R-CNN. Each detection box in the upper left corner is labelled with classification (CLS) and MS, representing the classification and segmentation quality scores, respectively. Based on the classification scores in Figure 5, the pigs exhibit a CLS of 1.00, with an average MS score exceeding 0.9. When dealing with densely packed and closely connected pigs, MaskDis R-CNN achieves more comprehensive pig segmentation, displaying smooth segmentation boundaries without fragmentation or missed segments.

3.3 | Comparison of results on the COCO evaluation metrics

The results of the three models' segmentation tasks on the COCO evaluation metrics are shown in Table 3. We compare the three models in the front view, the top view, and the total of front view and top view. According to Table 3, the proposed MaskDis R-CNN method performs better in instance segmentation of objects. The performance of MaskDis R-CNN reaches 96.09(AR_{50}), 85.36(AR_{75}), 73.36(mAR), 92.76(AP_{50}), 80.76(AP_{75}), and 68.55(mAP), has a significant promotion compared to MS R-CNN. Also, compared with MS R-CNN, the MaskDis R-CNN increases by 2.25%, 2.28%, 1.37%, and 1.9% in the metrics of AR_{75}, mAR, AP_{75}, and mAP, respectively. Therefore, our method is validated in improving the segmentation performance.

The ablation experiments of our method are shown in Table 4 under the COCO evaluation metrics. The best performance is achieved using Resnet-101 as the backbone network and adopting Stochastic Gradient Descent (SGD) for the optimizer, as shown in Table 4. Our approach's AP50, AP75, mAP, and mAR values were 92.8%, 80.8%, 68.6%, and 73.4%, respectively. Therefore, we use this configuration to compare it with other advanced instance segmentation methods.

3.4 | Comparison of segmentation quality between MaskDis R-CNN and MS R-CNN

The segmentation result of the MS R-CNN model is shown at the top of Figure 7. The segmentation result of the MaskDis R-CNN model is shown at the bottom of Figure 7. According to Figures 7a and 7b, the segmentation quality of the MS R-CNN model is flawed in the case of dense overlap of pigs, resulting in the segmentation target's MS score falling below 0.7, and

FIGURE 6 The segmentation result of the MaskDis R-CNN model.

TABLE 3 Results of the model under the COCO evaluation metrics.

Model	Image type	AR_{50} (%)	AR_{75} (%)	*mAR* (%)	AP_{50} (%)	AP_{75} (%)	*mAP* (%)
Mask R-CNN	Front view	89.00	64.73	56.64	84.05	55.03	49.86
	Top view	97.78	93.06	78.40	96.60	92.17	75.79
	Total	94.26	81.7	69.68	92.08	77.68	65.67
MS R-CNN	Front view	91.49	65.35	58.42	82.68	55.43	50.35
	Top view	99.03	95.00	79.56	98.31	93.34	76.80
	Total	96.01	83.11	71.08	**93.09**	79.39	66.65
MaskDis R-CNN	Front view	92.12	70.12	61.76	82.84	58.61	52.76
	Top view	98.75	95.56	81.13	97.24	93.98	77.95
	Total	**96.09**	**85.36**	**73.36**	92.76	**80.76**	**68.55**

TABLE 4 ablation experiments under the COCO evaluation metrics.

Model	Backbone	Optimizer	Image type	AP_{50} (%)	AP_{75} (%)	*mAP* (%)	*mAR* (%)
MaskDis R-CNN	R50	Adam	Front view	84.8	47.4	47.1	57.0
			Top view	97.3	91.0	76.3	78.9
			Total	91.5	75.1	65.4	69.7
		SGD	Front view	79.2	49.6	52.4	55.8
			Top view	97.8	91.6	76.9	78.4
			Total	91.2	75.7	68.0	69.3
	R101	Adam	Front view	78.4	49.7	46.1	55.0
			Top view	97.6	93.5	77.2	79.5
			Total	91.9	77.1	65.9	69.7
		SGD	Front view	82.9	58.6	52.8	61.8
			Top view	97.2	94.0	78.0	81.1
			Total	**92.8**	**80.8**	**68.6**	**73.4**

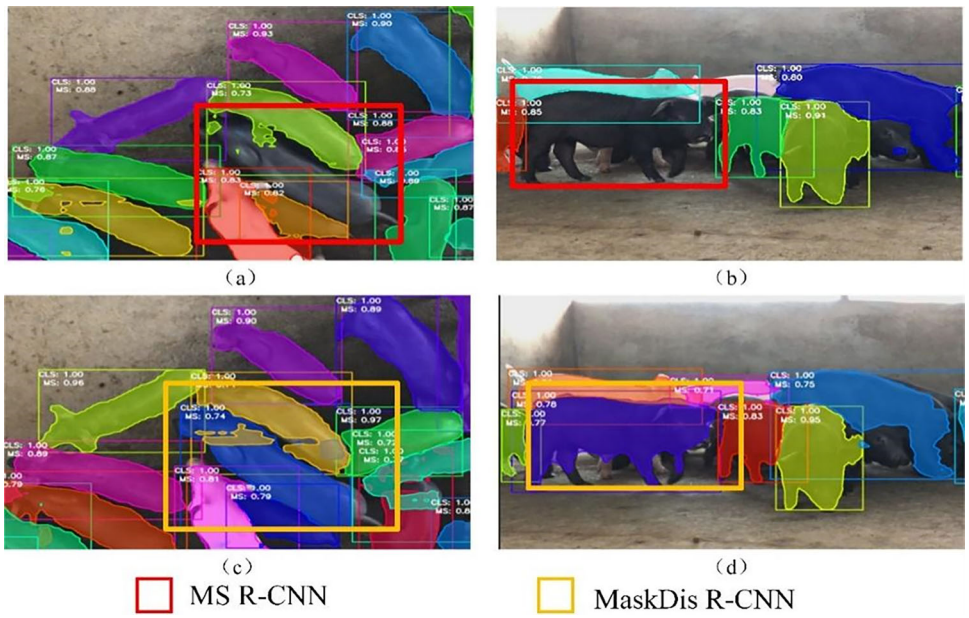


FIGURE 7 Comparison of the segmentation quality score between MaskDis R-CNN model and MS R-CNN model. MS R-CNN, mask scoring R-CNN.

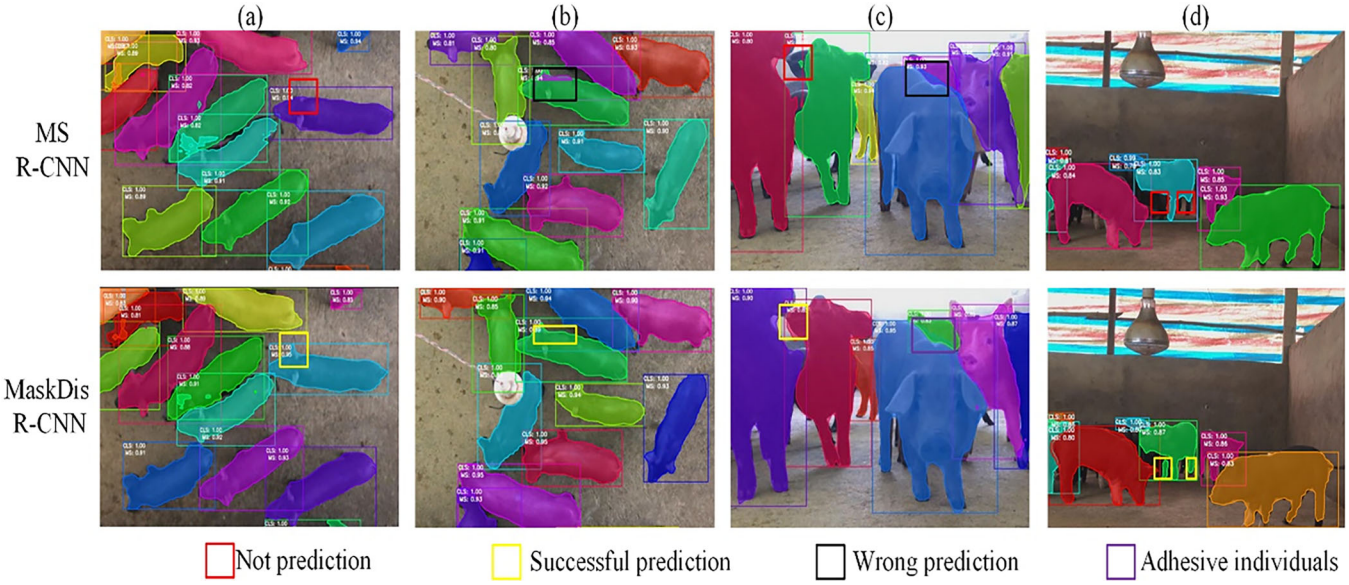


FIGURE 8 Comparison of front view and top view segmentation quality between MaskDis R-CNN model and MS R-CNN model. MS R-CNN, mask scoring R-CNN.

therefore the model has missed detection. According to Figures 7c and 7d, the model with the fused adversarial network has improved segmentation quality, increasing the MS score of the segmentation results, thus avoiding missed detections to some extent and improving the recall rate.

To further demonstrate the effectiveness of our proposed MaskDis R-CNN, we compare the segmentation quality of the MaskDis R-CNN model and the MS R-CNN model from the top and front views. The segmentation result of the MS R-CNN model is shown at the top of Figure 8, and the segmentation result of the MaskDis R-CNN model is shown at the bottom of Figure 8. According to Figure 8, the main problem of MS R-CNN is the incomplete segmentation of pig bodies (Figures 8a, 8c, 8d), and a small number of segmented fragments (Figure 8b) appear on other pig bodies in the case of dense and sticky pig populations. And the segmentation boundary is not smooth enough (Figure 8c). However, there are improvements in the segmentation results of the MaskDis Mask method with the addition of the adversarial network, such as improved quality of both segmentation details and segmentation boundaries, more complete segmented pigs, and fewer segmented fragments.

3.5 | Results comparison with other advanced instance segmentation methods

We used the proposed model to conduct comparison experiments with the four advanced instance segmentation methods, including YOLACT [34], Swin Transformer [35],YOLOv5-seg, and SOLOv2 [36] on the same dataset. The YOLACT is a real-time instance segmentation model that combines detection and segmentation, achieving 33.5 fps on the MS COCO dataset. The Swin Transformer is a transform-based instance segmentation method that introduces a hierarchical Transformer whose representation is computed with Shifted windows to better capture global contextual information. The YOLOv5-seg is a high-speed instance segmentation algorithm that keeps accuracy while achieving real-time performance. The SOLOv2 is a one-stage instance segmentation method that combines detection and segmentation, which uses deformable convolution and a feature selection module with an attention mechanism that can better adapt to the shape and scale variations of the instances.

The comparison results of our approach with other advanced instance segmentation methods are shown in Table 5. Our approach achieved the best performance on the metrics of COCO. The AP_{50}, AP_{75}, mAP, and mAR values of our approach were 92.8%, 80.8%, 68.6%, and 73.4%, respectively. Compared with the YOLACT method, our approach improved by 4.9%, 12.1%, 8.9%, and 5.9% in AP_{50}, AP_{75}, mAP, and mAR. Compared with the Swin Transformer method, our approach improved by 2.7%, 20.7%, 16.3%, and 10.6% in AP_{50}, AP_{75}, mAP, and mAR. Compared with the SOLOv2 method, our approach improved by 0.5%, 6.9%, and 2.9% in AP_{50}, AP_{75}, mAP. Compared with the YOLOv5-seg method, our approach improved by 2.7%, 2.5%, and 5.5% in AP_{50}, AP_{75}, and mAP. These comparison results demonstrate that our approach can effectively improve the pig segmentation performance.

4 | DISCUSSION

In this paper, we propose MaskDis R-CNN, a deep learning algorithm by fusing adversarial networks, for the detection and instance segmentation of herd pigs in complex scenarios. The key innovations include designing adversarial networks, integrating them into MS R-CNN, and then proving their

TABLE 5 Results comparison with other advanced instance segmentation methods.

Model	Image type	AP_{50} (%)	AP_{75} (%)	*mAP* (%)	*mAR* (%)
YOLACT	Front view	77.8	47.0	44.7	55.9
	Top view	93.8	84.5	70.5	75.3
	Total	87.9	68.7	59.7	67.5
Swin transformer	Front view	83.9	40.8	42.4	54.6
	Top view	94.9	77.3	61.2	68.3
	Total	90.1	60.1	52.3	62.8
YOLOv5-seg	Front view	80.2	66.8	53.6	–
	Top view	96.9	86.7	73.1	–
	Total	90.1	78.3	63.1	–
SOLOv2	Front view	83.5	54.8	51.8	63.3
	Top view	97.7	86.7	75.3	80.6
	Total	92.3	73.9	65.7	**73.7**
MaskDis R-CNN	Front view	82.8	58.6	52.8	61.8
	Top view	97.24	94.0	78.0	81.1
	Total	**92.8**	**80.8**	**68.6**	73.4

effectiveness, for instance, segmentation tasks of group-housed pigs. The advantage of the MaskDis R-CNN method is the ability to detect and segment instances of heavily obscured and overlapping pigs, which can be further developed to perform tasks such as pig welfare monitoring [37–39].

Previous instance segmentation studies in pigs were challenged by occlusions, light variations, and background factors [21]. To address this issue, adding adversarial network can enhance the quality of instance segmentation, as the training process uses adversarial training to make the prediction mask closer to the GT in order to achieve enhanced segmentation quality. In our work, the adversarial network head branch is designed and added to the MS R-CNN, called MaskDis Head, which is used as a discriminator, and Mask Head is used as a generator during model training. Through the adversarial training of the generative network, the generator learns pixel-level, low-level, and middle-level features. The segmentation quality is improved as the function of the segmentation mask loss converges more easily during model training.

Mask R-CNN and MS R-CNN are efficient methods in both the top and front views. However, it does not achieve the expected results in the situation of the overlapped pigs. When there is a problem caused by dense overlap and severe occlusion, the problem is more severe in front views. To get better segmentation results, the MaskDis R-CNN instance segmentation model creates the adversarial network branch for guiding the segmentation mask training of the model. The improved model is compared with the MS R-CNN model for experiments and analysis, and the results show that MaskDis R-CNN improves the segmentation quality. In the comparative analysis of the segmentation results between MaskDis R-CNN and the MS R-CNN model, it is found that the MaskDis R-CNN improves the quality of instance segmentation on the situation of the overlapped pigs, which indicates that the adversarial network branch of MaskDis R-CNN improves the ability to handle the detail information of instance segmentation.

The final experimental results have shown that the MS R-CNN model can achieve detection accuracy with a recall of 92.10% and a precision of 89.78%. And this method sometimes misses the detection and does not work well for segmentation of herd pigs in the case of dense and adhesive. However, the MaskDis R-CNN can void this situation and improve the quality of intensive herd raised piglets with a precision of 92.03% and an F1 score of 92.10%. In conclusion, the MaskDis R-CNN significantly outperforms MS R-CNN for instance segmentation in a dense environment of pig pens.

5 | CONCLUSION

To solve the problem of unsatisfactory segmentation performance under overlapping or partial occlusion between pig's scenes, we propose an approach named MaskDis R-CNN by fusing the MS R-CNN model and the adversarial network. The new method solves the problem that MS R-CNN does not achieve the expected results in specific situations such as pig dense and occlusion situations. Mask Head is used as a generator, and MaskDis Head as a discriminator during model training. Adversarial training of the generated network keeps the prediction mask close to GT for improving segmentation quality during model testing.

We conducted comparative experiments between Mask R-CNN, MS R-CNN, and our proposed approach to demonstrate our method's effectiveness in the overlapped pigs and occlusion situation. Our method outperforms the other two approaches, which can achieve a recall of 92.18%, a precision of 92.03%, and an F1 score of 0.9210. In addition, by adopting the COCO evaluation metrics, our approach achieves an mAP of 73.36% and an mAR of 68.55%, which are higher than the other two algorithms. In addition, our method performance is better than other advanced instance segmentation methods. Considering these results, it is demonstrated that the proposed method improves the problem of poor segmentation quality due to the dense and occlusive herd of pigs.

Although this paper achieves accurate segmentation, there are still some constraints: (1) the proposed model cannot achieve the characteristics of lightweight, fast speed, and strong portability, which requires operations such as convolution and ROI pooling on each candidate region. This leads to higher computational and memory requirements. (2) In the case of severe occlusion, the proposed model still suffers from missed and false detections, resulting in segmentation performance that does not meet practical needs. Future work will optimize the network architecture of the proposed approach, which can reduce the computational requirements and increase the inference speed. The improved model also provides a theoretical basis for the intelligent development of pig farming and has great significance for improving pig welfare and guiding the production.

AUTHOR CONTRIBUTIONS

Shuqin Tu:Writing—Review & Editing, Methodology, Conceptualization, Funding Acquisition, Resources, Formal Analysis, **Qiantao Zeng**:Writing—Original Draft, Validation, Methodology, Visualization, Writing—Review & Editing, **Haofeng Liu**:Conceptualization, Formal Analysis, Funding Acquisition, **Yun Liang**:Supervision, Funding Acquisition, Project Administration, **Xiaolong Liu**: Resources, Visualization, Investigation, **Lei Huang**: Resources, Data Curation, Visualization, **Zhengxin Huang**: Resources, Data Curation.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (61772209 and 31600591), the Science and Technology Planning Project of Guangdong Province (Grant No. 2019A050510034), Guangzhou Key R&D Program Project (2023B03J1363, 202206010091), and College Students' Innovation and Entrepreneurship Competition (202110564025).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Yun Liang https://orcid.org/0000-0003-0799-0054

REFERENCES

1. Wongsriworaphon, A., Arnonkijpanich, B., Pathumnakul, S.: An approach based on digital image analysis to estimate the live weights of pigs in farm environments. Comput. Electron Agr. 115, 26–33 (2015)
2. Wang, K., Guo, H., Ma, Q., Su, W., Chen, L., Zhu, D.: A portable and automatic Xtion-based measurement system for pig body size. Comput. Electron. Agric. 148, 291–298 (2018)
3. Bhoj, S., Tarafdar, A., Chauhan, A., Singh, M., Gaur, G.K.: Image processing strategies for pig liveweight measurement: Updates and challenges. Comput. Electron. Agric. 193, 106693 (2022)
4. Nasirahmadi, A., Edwards, S.A., Sturm, B.: Implementation of machine vision for detecting behaviour of cattle and pigs. Livestock Sci. 202, 25–38 (2017)
5. Chen, Y., Xia, R., Zou, K., Yang, K.: FFTI: Image inpainting algorithm via features fusion and two-steps inpainting. J. Visual Commun. Image Represent. 91, 103776 (2023)
6. Chen, Y., Xia, R., Yang, K., Zou, K.: MFFN: Image super-resolution via multi-level features fusion network. Vis. Comput. (2023)
7. Patil, R., Boit, S., Gudivada, V.N., Nandigam, J.: A survey of text representation and embedding techniques in NLP. IEEE Access 11, 36120–36146 (2023)
8. Kim, J., Chung, Y., Choi, Y., Sa, J., Kim, H., Chung, Y., Park, D., Kim, H.: Depth-based detection of standing-pigs in moving noise environments. Sensors (Basel) 17(12), 2757 (2017)
9. Dominiak, K.N., Pedersen, L.J., Kristensen, A.R.: Spatial modeling of pigs' drinking patterns as an alarm reducing method I. Developing a multivariate dynamic linear model. Comput. Electron Agr. 161, 79–91 (2019)
10. Pan, X., Zhu, J., Tai, W., Fu, Y.: An automated method to quantify the composition of live pigs based on computed tomography segmentation using deep neural networks. Comput. Electron Agr. 183, 105987 (2021)
11. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39(6), 1137–1149 (2017)
12. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV 2017), Venice, Italy, pp. 2980–2988 (2017)
13. Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X.: Mask scoring R-CNN. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 6409–6418 (2019)
14. Cai, Z., Vasconcelos, N.: Cascade R-CNN: High quality object detection and instance segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 43(5), 1483–1498 (2021)
15. Tian, Z., Shen, C., Chen, H., He, T.: FCOS: Fully convolutional one-stage object detection. In: International Conference on Computer Vision (ICCV 2019), Seoul, South Korea, pp. 9626–9635 (2019)
16. Zhang, R., Tian, Z., Shen, C., You, M., Yan, Y.: Mask encoding for single shot instance segmentation. In: Conference on Computer Vision and Pattern Recognition (CVPR 2020), Seattle, WA, USA, pp. 10223–10232 (2020)
17. Tian, Z., Shen, C., Chen, H.: Conditional convolutions for instance segmentation. In: Conference on Computer Vision and Pattern Recognition (CVPR 2020), Seattle, WA, USA, pp. 282–298 (2020)
18. Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., Liu, Z.: Dynamic convolution: Attention over convolution kernels. In: Conference on Computer Vision and Pattern Recognition (CVPR 2020), Seattle, WA, USA, pp. 11027–11036 (2020)
19. Qiao, Y.L., Truman, M., Sukkarieh, S.: Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. Comput. Electron Agr. 165, 10458 (2019)
20. Hu, Z., Yang, H., Lou, T.: Dual attention-guided feature pyramid network for instance segmentation of group pigs. Comput. Electron Agr. 186, 106140 (2021)
21. Tu, S., Yuan, W., Liang, Y., Wang, F., Wan, H.: Automatic detection and segmentation for group-housed pigs based on PigMS R-CNN. Sensors (Basel) 21(9), 3251 (2021)
22. Xiao, J., Liu, G., Wang, K., Si, Y.: Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM. Comput. Electron Agr. 194(3), 106738 (2022)
23. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., David Warde-Farley, S., Ozair, A.C., CourvilleBengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems, pp. 2672–2680 (2014)
24. Luc, P., Couprie, C., Soumith ChintalaVerbeek, J.: Semantic segmentation using adversarial networks. CoRR 2016, abs/1611.08408, (2016)
25. Xue, Y., Xu, T., Zhang, H., Long, L.R., Huang, X.: SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation. Neuroinformatics 16(3), 383–392 (2018)
26. Chen, Y., Xia, R., Zou, K., Yang, K.: RNON: Image inpainting via repair network and optimization network. Int. J. Mach. Learn Cybern. 1–17 (2023)
27. Manohara Pai, M.M., Mehrotra, V., Ujjwal VermaPai, R.M.: Improved semantic segmentation of water bodies and land in SAR images using generative adversarial networks. Int. J. Semant Comput. 14, 55–69 (2020)
28. Khaled, A., Han, J.-J., Ghaleb, T.A.: Multi-model medical image segmentation using multi-stage generative adversarial networks. IEEE Access 10, 28590–28599 (2022)
29. Man, R., Yang, P., Xu, B.: Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks. IEEE Access 8, 155362–155377 (2020)
30. Minagi, A., Hokuto HiranoTakemoto, K.: Natural images allow universal adversarial attacks on medical image classification using deep neural networks with transfer learning. J. Imaging 8(2), 38 (2022)
31. Choi, S.H., Shin, J.-M., Peng LiuChoi, Y.-H.: ARGAN: Adversarially robust generative adversarial networks for deep neural networks against adversarial examples. IEEE Access 10, 33602–33615 (2022)

32. Guo, X., Hiroyuki OkamuraDohi, T.: Automated software test data generation with generative adversarial networks. IEEE Access 10, 20690–20700 (2022)
33. Nistal, J.: Exploring generative adversarial networks for controllable musical audio synthesis. (Synthèse audio musicale contrôlable à l'aide de réseaux adverses génératifs) (2022)
34. Bolya, D., Zhou, C., Xiao, F., Lee, Y.J.: YOLACT: Real-time instance segmentation. In: International Conference on Computer Vision (ICCV 2019), pp. 9156–9165, Seoul, South Korea (2019)
35. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: International Conference on Computer Vision (ICCV 2021), Montreal, QC, Canada, pp. 9992–10002 (2021)
36. Wang, X., Zhang, R., Kong, T., Li, L., Shen, C.: SOLOv2: Dynamic and fast instance segmentation. In: Annual Conference on Neural Information Processing Systems (NeurIPS 2020) (2020)
37. Valletta, J.J., Torney, C., Kings, M., Thornton, A., Madden, J.: Applications of machine learning in animal behaviour studies. Anim. Behav. 124, 203–220 (2017)
38. Wathes, C.M., Kristensen, H.H., Aerts, J.M., Berckmans, D.: Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? Comput. Electron Agr. 64(1), 2–10 (2008)
39. Alameer, A., Kyriazakis, I., Bacardit, J.: Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs. Sci. Rep. 10(1), 13665 (2020)

How to cite this article: Tu, S., Zeng, Q., Liu, H., Liang, Y., Liu, X., Huang, L., Huang, Z.: MaskDis R-CNN: An instance segmentation algorithm with adversarial network for herd pigs. IET Image Process. 17, 3488–3499 (2023).
https://doi.org/10.1049/ipr2.12880

2.9 基于改进DeepSORT的群养生猪行为识别与跟踪方法

美国《工程索引》（Ei）收录期刊
美国《化学文摘》（CA）收录期刊
Scopus数据库收录期刊
中文核心期刊　中国科技核心期刊
中国科学引文数据库来源期刊
RCCSE中国权威学术期刊

ISSN 1000-1298
CODEN NUYCA3

农业机械学报

NONGYE JIXIE XUEBAO

Transactions of the Chinese Society for Agricultural Machinery

第53卷

特约专稿

西北现代生态灌区建设理论与技术保障体系构建

邓铭江　陶汪海　王全九　苏李君　马昌坤　宁松瑞

2022 8

ISSN 1000-1298
08>
9 771000 129220

中国农业机械学会主办

农业机械学报
NONGYE JIXIE XUEBAO
2022 年　第 8 期(第 53 卷)

目　次

农业信息化工程

农 业 机 械 学 报

NONGYE JIXIE XUEBAO

2022 年　第 53 卷　第 8 期

（月刊，1957 年创刊）

2022 年 8 月 25 日出版

Transactions of the Chinese Society for Agricultural Machinery

No. 8　Vol. 53　2022

（Monthly, started in 1957）

Published on 25, August 2022

主　管	中国科学技术协会
主　办	中国农业机械学会 中国农业机械化科学研究院
编辑出版	中国农业机械学会 （地址：北京德外北沙滩 1 号 6 信箱） 邮政编码　100083
主　编	任露泉
印　刷	北京富泰印刷有限责任公司
国内发行	中国邮政集团公司北京报刊发行局
订 购 处	全国各地邮局
国外发行	中国国际图书贸易集团有限公司 （北京 399 信箱）

Responsible Department: China Association for Science and Technology

Sponsored by: Chinese Society for Agricultural Machinery, Chinese Academy of Agricultural Mechanization Sciences

Published by: Chinese Society for Agricultural Machinery

Editor in Chief: Ren Luquan

Editorial Office: No. 1 Beishatan, Deshengmen Wai, Beijing 100083, China

Tel: 86-10-64882610/64882231

Fax: 86-10-64867367

http://www.j-csam.org

E-mail: njxb@caams.org.cn

Overseas Distributor: China International Book Trading Corporation (P. O. Box 399, Beijing, China)

ISSN 1000-1298　CN 11-1964/S　　国内邮发代号　2-363　　国外发行代号　M289　　国内定价　100.00 元

doi:10.6041/j.issn.1000-1298.2022.08.037

基于改进 DeepSORT 的群养生猪行为识别与跟踪方法

涂淑琴[1]　刘晓龙[1]　梁　云[1]　张　宇[2]　黄　磊[1]　汤寅杰[1]
(1. 华南农业大学数学与信息学院, 广州 510642; 2. 华南农业大学电子工程学院, 广州 510642)

摘要: 为改善猪只重叠与遮挡造成的猪只身份编号(Identity,ID)频繁跳变,在 YOLO v5s 检测算法基础上,提出了改进 DeepSORT 行为跟踪算法。该算法改进包括两方面:一针对特定场景下猪只数量稳定的特点,改进跟踪算法的轨迹生成与匹配过程,降低 ID 切换次数,提升跟踪稳定性;二将 YOLO v5s 检测算法中的行为类别信息引入跟踪算法中,在跟踪中实现准确的猪只行为识别。实验结果表明,在目标检测方面,YOLO v5s 的 mAP 为 99.3%,F1 值为 98.7%。在重识别方面,实验的 Top-1 准确率达到 99.88%。在跟踪方面,改进 DeepSORT 算法的 MOTA 为 91.9%,IDF1 为 89.2%,IDS 为 33;与 DeepSORT 算法对比,MOTA 和 IDF1 分别提升了 1.0、16.9 个百分点,IDS 下降了 83.8%。改进 DeepSORT 算法在群养环境下能够实现稳定 ID 的猪只行为跟踪,能够为无接触式的生猪自动监测提供技术支持。

关键词: 群养生猪; 目标检测; 行为识别; 多目标跟踪; DeepSORT
中图分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2022)08-0345-08　　**OSID**:

Behavior Recognition and Tracking Method of Group-housed Pigs Based on Improved DeepSORT Algorithm

TU Shuqin[1]　LIU Xiaolong[1]　LIANG Yun[1]　ZHANG Yu[2]　HUANG Lei[1]　TANG Yinjie[1]
(1. *College of Mathematics and Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*
2. *College of Electronic Engineering*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: Behavior recognition and tracking of group-housed pigs are an effective aid to monitor pigs' health status in smart farming. In real farming scenarios, it is still challenging to automatically track the behavior of group-housed pigs by using computer vision techniques due to the pigs' overlapping occlusion and illumination change, which cause the identity (ID) of pig to switch wrongly. To improve the situation, an improved DeepSORT algorithm of behavior tracking based on YOLO v5s was proposed. The improvement of the algorithm included two parts. One was that the trajectory processing and data association were improved in the scene where there was a fixed number of pigs. This reduced ID switch and enhanced tracking stability. The other was that the behavior information from YOLO v5s detection algorithm was introduced into the tracking algorithm, thereby achieving behavior recognition of pigs in tracking. The experimental results showed that YOLO v5s algorithm had a mAP of 99.3% and an F1 of 98.7% in object detection. In terms of re-identification, the Top-1 accuracy of the experiment was 99.88%. In terms of tracking, the method achieved a favorable performance with a MOTA of 91.9%, an IDF1 of 89.2% and an IDS of 33. Compared with the original DeepSORT algorithm, the proposed method improved 1.0 percentage points and 16.9 percentage points in MOTA and IDF1 respectively, and decreased 83.8% in IDS. This showed that the improved DeepSORT algorithm was able to achieve behavior tracking of group-housed pigs with stable ID. The method can provide technical support for no-contact automatic monitoring of pigs.

Key words: group-housed pigs; object detection; behavior recognition; multi-object tracking; DeepSORT

收稿日期: 2022-05-08　修回日期: 2022-05-31
基金项目: 广东省科技计划项目(2019A050510034)、广州市重点项目(202206010091)、广州市科技计划重点实验室建设项目(201902010081)和广东省企业特派员项目(GDKTP2021055700)
作者简介: 涂淑琴(1978—),女,讲师,主要从事计算机视觉研究,E-mail: tushuqin@163.com
通信作者: 张宇(1976—),女,讲师,主要从事农业信息化研究,E-mail: zhangyu@scau.edu.cn

0 引言

猪只的健康情况决定着生猪养殖业的发展与经济效益,多数生猪疾病的临床或亚临床体征表现之前常伴随行为异常[1-2],故对猪只的运动、饮食等行为的监测有助于判断猪只健康情况。目前人工监测是猪场的主要管理方式,但需要大量劳动力并且难以实现长期持续观察,借助无线射频技术(Radio frequency identification,RFID)的监测可避免人工监测的缺陷,但易对猪只造成刺激。信息技术手段已成为现代农业发展的重要引擎,通过借助计算机视觉技术,可以实现低成本、无接触地自动监测猪只[3],实现生猪智慧养殖。

国内外学者在生猪自动监测方面的研究可分为两类,分别为猪只行为识别和猪只跟踪。在猪只行为识别中,主要采用传统图像处理与深度学习的方法。ZHU 等[4]利用阈值分割和形态学处理等机器视觉技术,实现非接触地对猪只饮水行为进行识别。李菊霞等[5]针对猪舍环境下猪只饮食行为自动化检测程度较低的问题,提出了一种基于 YOLO v4 的猪只饮食行为检测模型。ALAMEER 等[6]基于 YOLO v2 与 Faster R-CNN 监测猪只的姿态与饮水行为,其无需传感器或对个体识别即可监测猪只行为。高云等[7]提出使用 3D CONV 对群养猪侵略性行为进行识别,可为猪场养殖环境中猪只侵略性行为检测提供参考。在猪只跟踪方面,近年来也有很多研究者开展相关工作。XIAO 等[8]通过颜色信息识别猪只,并通过分析二值图像中连接区域消除噪声,根据 DT-ACR 关联规则对猪进行跟踪。SUN 等[9]提出一种多通道彩色特征自适应融合算法,并利用目标猪的轮廓信息实时更新比例,提升在目标形变与尺度变化下的跟踪效果。ZHANG 等[10]利用分层数据关联算法将基于卷积神经网络的检测器与相关滤波器的跟踪器结合,实现单个猪只跟踪。张伟等[11]基于 CenterNet 设计断奶仔猪目标检测模型,结合 DeepSORT 算法[12]实现断奶仔猪的多目标跟踪,改善猪只外观高度相似与遮挡情况下的跟踪效果。

目前,视频监控技术由于其设备价格低廉和实现简单的优点,已广泛应用于生猪的行为识别与跟踪研究。但是,在真实养殖场景下,由于光照变化与猪群的密集遮挡,容易造成跟踪中目标 ID 频繁跳变;同时,很多研究只是对猪只的行为进行识别[13-15]或跟踪[16-17],算法获取的猪只信息难以满足现代化生猪养殖业的要求。基于此,本文将行为识别与多目标跟踪进行融合,并改进 DeepSORT 算法中的轨迹生成与匹配过程,提出基于 YOLO v5s 的改进 DeepSORT 算法。通过将 YOLO v5s 检测算法得到的行为类别信息引入 DeepSORT 跟踪算法,实现跟踪中的行为识别,使得在跟踪中算法可以获取到目标的行为信息;同时,针对特定场景下猪只数量稳定不变的特点,改进 DeepSORT 算法中的轨迹生成与匹配过程,提升跟踪中猪只 ID 的稳定性;最终在真实群养环境下实现稳定 ID 的猪只行为跟踪,以期为无接触式自动监测生猪提供技术支持。

1 实验数据与数据增广

1.1 数据集

实验数据从文献[18]的数据集中选择,筛选保留 15 段猪只移动较多的有效视频,每段视频为 60 s,每秒为 5 帧。将视频分辨率裁剪为 2 688 像素×1 012像素,在视频段中只保留同一猪舍下的猪只。使用 DarkLabel 软件对视频段进行标注,构建 3 个数据集分别用于训练目标检测模块、重识别模块和验证行为跟踪算法效果,这 3 个数据集中均包含白天与夜间场景、猪只拥挤与稀疏场景、猪只活动频繁与较少场景。

选取真实群养环境下 4 段不同条件的视频段作为行为跟踪数据集,用于验证算法效果,记为序号 01~04,如表 1 所示。

表 1 行为跟踪数据集

Tab. 1 Behavior tracking dataset

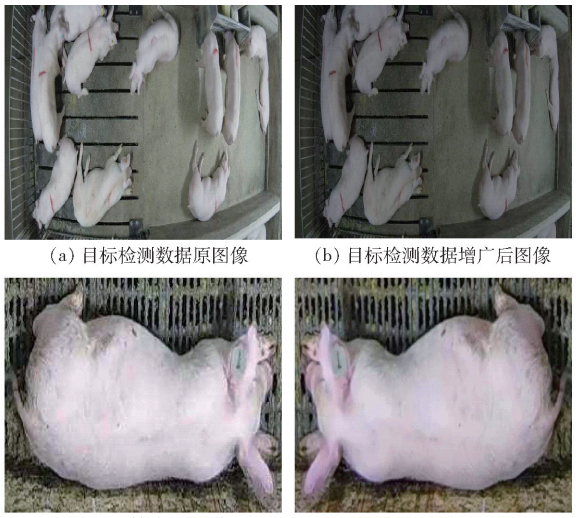
视频段序号	照明情况	密集程度	活动程度	猪只数量
01	白天	稀疏	频繁	7
02	白天	拥挤	较少	15
03	白天	拥挤	频繁	16
04	夜晚	拥挤	频繁	16

其余视频段利用 ffmpeg 工具分割成图像,使用脚本将图像与标注信息分别构建目标检测数据集和重识别数据集,其中目标检测数据集将猪只分为躺卧、站立、饮食和其他行为 4 种类别。目标检测数据集共 3 300 幅图像,按照比例 7∶2∶1随机划分为训练集、验证集和测试集。重识别数据集包含 137 头猪只,平均每头猪只 300 幅图像,按照比例 7∶3随机划分为训练集和测试集。

1.2 数据增广技术

为提升模型泛化能力,使用左右翻转、上下翻转、改变色彩属性的数据增广技术,扩充生猪的目标检测与重识别训练数据。其中,对图像进行左右翻转与上下翻转可以扩大数据集的规模,以获得理想的训练效果;改变图像色彩属性,如随机改变图像的色调、饱和度和明度可以模拟光照情况变化对图像

的干扰，在一定程度上消除光环境的影响[19]。最终目标检测训练数据集扩充到 4 620 幅图像，重识别训练数据集中每头猪只扩充到 420 幅图像。生猪的目标检测与重识别图像数据增广后样例见图 1。



(a) 目标检测数据原图像　(b) 目标检测数据增广后图像

(c) 重识别数据原图像　(d) 重识别数据增广后图像

图 1　数据增广后结果

Fig. 1　Results after data augmentation

2　行为跟踪算法原理

采用 YOLO v5s 检测器结合改进 DeepSORT 算法实现猪只行为跟踪，如图 2 所示，首先图像输入 YOLO v5s 目标检测器，得到检测结果；然后改进 DeepSORT 算法基于卡尔曼滤波与匈牙利算法对前后两帧之间目标进行匹配，生成跟踪轨迹；最后输出行为跟踪图像。

2.1　YOLO v5s 检测算法

在基于检测的跟踪算法中，目标检测的效果对跟踪算法的效果起着至关重要的作用。当前基于深度学习的目标检测算法中，主要可划分为 One－stage 和 Two－stage 两类。在 One－stage 算法中占主流的是 YOLO[20－23]系列等。Two－stage 算法中占主流的是 R－CNN[24]、Fast R－CNN[25]、Faster R－CNN[26]等。One－stage 与 Two－stage 算法各有优劣，前者在检测速度上有明显的优势，后者在检测精度有更好的效果。YOLO v5s 算法可以兼顾速度与精度的要求，本文将其作为目标检测器。YOLO v5s 网络结构如图 3 所示，该算法主要包括输入端、主干网络、颈部和输出端 4 部分。其中，在输入端对数据进行预处理，采用如 Mosaic 数据增强、自适应图像缩放和自适应锚框计算等技术实现；主干网络主要由 Focus 层、CONV 层、C3 层、SPP 层等结构组成；颈部包含特征金字塔网络(Feature pyramid networks，FPN)[27]、路径聚合网络(Path aggregation network，PAN)[28]；输出端主要是在 3 个不同大小的特征图上预测不同尺寸的目标。

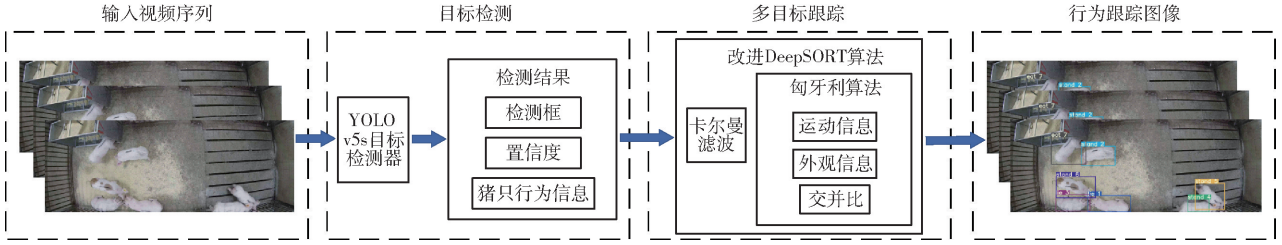


图 2　行为跟踪算法流程图

Fig. 2　Flow chart of behavior tracking algorithm

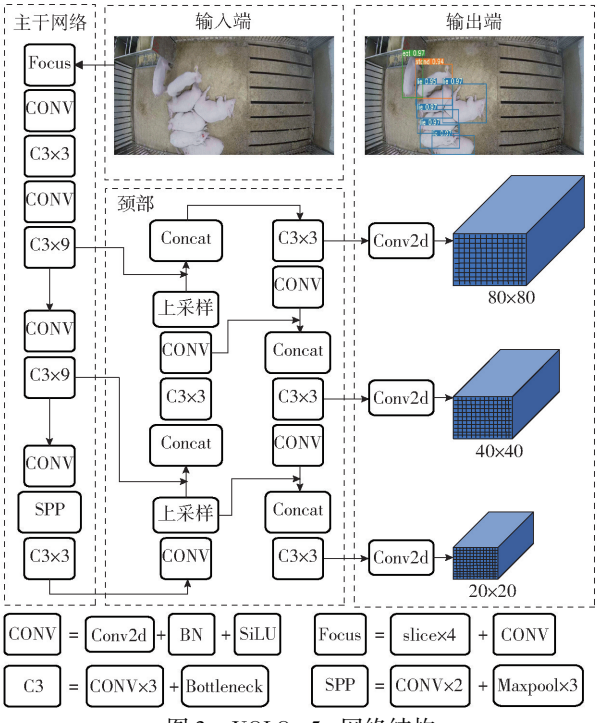


图 3　YOLO v5s 网络结构

Fig. 3　Network structure of YOLO v5s

2.2　DeepSORT 算法

DeepSORT 算法是在 SORT[29]基础上引入重识别模型，通过外观信息与运动信息增强匈牙利算法的匹配效果，减少 ID 切换的数量；其通过卡尔曼滤波器和匈牙利算法，分别处理跟踪问题的运动预测和数据关联部分。

2.2.1　轨迹的状态估计

轨迹中目标对象的状态和运动信息定义于 8 维状态空间$(u,v,\gamma,h,\dot{x},\dot{y},\dot{\gamma},\dot{h})$，其中$(u,v)$表示目标的中心位置坐标，$\gamma$ 为纵横比，h 为高度，$(\dot{x},\dot{y},\dot{\gamma},$

$\dot{h}$)表示(u,v,γ,h)对应参数在图像坐标中各自的速度。算法采用带有等速运动和线性观测模型的标准卡尔曼滤波器,将边界框坐标(u,v,γ,h)作为目标对象的直接观测值。

2.2.2 匹配问题

DeepSORT 算法结合运动信息度量与外观信息度量,使用匈牙利算法解决检测框与轨迹的匹配问题。对于运动信息度量,算法表示为

$$d^{(1)}(i,j)=(\boldsymbol{d}_j-\boldsymbol{y}_i)^{\mathrm{T}}\boldsymbol{S}_i^{-1}(\boldsymbol{d}_j-\boldsymbol{y}_i) \tag{1}$$

式中 $d^{(1)}(i,j)$——第 j 个检测框和第 i 个轨迹预测得到的边界框的马氏距离

$\boldsymbol{d}_j$——第 j 个检测框

$\boldsymbol{y}_i$——第 i 个轨迹预测后的边界框

$\boldsymbol{S}_i$——第 i 个轨迹预测得到的在当前测量空间的协方差矩阵

马氏距离用于过滤可能性极低的匹配,当 $d^{(1)}(i,j)$ 小于指定的阈值,认为匹配成功。

对于外观信息度量,引入重识别模型,用于提取目标的外观信息。重识别模型网络结构如表 2 所示,主要包括 1 个卷积层、1 个最大池化层、8 个残差层和 1 个平均池化层。

表 2 重识别模型网络结构

Tab. 2 Network structure of re-identification model

名称	卷积核尺寸、步幅	输出尺寸
卷积层 1	3×3、1	64×128×64
最大池化层 2	3×3、2	64×64×32
残差层 3	3×3、1	64×64×32
残差层 4	3×3、1	64×64×32
残差层 5	3×3、2	128×32×16
残差层 6	3×3、2	128×32×16
残差层 7	3×3、2	256×16×8
残差层 8	3×3、2	256×16×8
残差层 9	3×3、2	512×8×4
残差层 10	3×3、2	512×8×4
平均池化层 11	8×4、1	512×1×1

对于外观信息度量,算法表示为

$$d^{(2)}(i,j)=\min\{1-\boldsymbol{r}_j^{\mathrm{T}}\boldsymbol{r}_k^{(i)}\mid\boldsymbol{r}_k^{(i)}\in\boldsymbol{R}_i\} \tag{2}$$

式中 $d^{(2)}(i,j)$——第 j 个检测框与第 i 个轨迹的最小余弦距离

$\boldsymbol{r}_j$——第 j 个检测框 $\boldsymbol{d}_j$ 相应的外观描述符,设置 $\|\boldsymbol{r}_j\|=1$

$\boldsymbol{r}_k^{(i)}$——第 i 个轨迹相应的外观描述符

$\boldsymbol{R}_i$——第 i 个轨迹的外观信息仓库,保存最新 100 条目标成功匹配的外观描述符

当 $d^{(2)}(i,j)$ 小于指定阈值时,认为匹配成功。

马氏距离提供基于运动目标可能的位置信息,余弦距离通过考虑外观信息可以在目标发生遮挡重叠情况下恢复 ID,为提升跟踪效果,将马氏距离与余弦距离结合作为最终度量,公式为

$$c_{i,j}=\lambda d^{(1)}(i,j)+(1-\lambda)d^{(2)}(i,j) \tag{3}$$

式中 $c_{i,j}$——第 j 个检测框与第 i 个轨迹的关联程度

λ——权重系数

沿用原算法将 λ 设置为 0,即运动信息度量用于限制明显不可行的匹配,关联矩阵中只使用外观信息度量计算。

2.3 改进 DeepSORT 算法

2.3.1 跟踪中的行为识别

在原 DeepSORT 跟踪算法的基础上,将猪只行为类别添加到目标跟踪的轨迹中,其实现效果如图 4 所示,图 4a 中左上角为目标猪只的 ID 编号,是原算法效果;图 4b 中左上角为目标的行为类别和 ID 编号,是改进后效果。首先在 YOLO v5s 目标检测算法中将检测到的猪只行为类别分为躺卧、站立、饮食和其他 4 类;然后将行为类别作为 DeepSORT 算法的输入,存储在目标猪只轨迹的参数中,以此实现猪只跟踪过程中的行为识别。

(a) 原算法效果 (b) 添加类别信息后效果

图 4 行为识别效果

Fig. 4 Effect of behavior recognition

2.3.2 改进匹配过程与轨迹生成

DeepSORT 算法在跟踪实验中,随着视频帧增长,同一猪只目标易被分配不同的 ID,导致 ID 最大值大幅超出真实的猪只目标数量。主要原因是目标猪只发生运动或遮挡重叠现象时,检测结果无法与原轨迹匹配,导致未匹配的检测结果生成新轨迹。基于上述问题,针对猪舍特定场景,对算法的匹配过程与轨迹生成进行改进。

由于猪舍为封闭场景,无目标新增或减少,为提升匹配效果,增加第 2 轮交并比匹配对未匹配的检测框进行处理。改进后的匹配过程:第 1 轮交并比匹配完成初步匹配,第 2 轮放宽交并比最大距离的限制来尽可能使未匹配的检测框与轨迹匹配成功。

在该场景下通过改进轨迹生成的方式来限制 ID 增长。具体实现如下:封闭场景下目标总数是稳定不变的,则 ID 最高数量是已知的,动态存储检测结果中的目标总数,将其计为 ID 极大值,由于检测

结果有一定概率出现误检漏检现象，ID 极大值由最近 3 帧中检测目标数量的平均数决定，若当前生成轨迹 ID 超过 ID 极大值，不生成新轨迹；若未超过，则生成新轨迹。

改进后算法匹配过程如图 5 所示，其中虚线框部分为算法改进内容。轨迹初始化为未确认态，满足连续 3 帧都成功匹配，将未确认态转化为确认态。

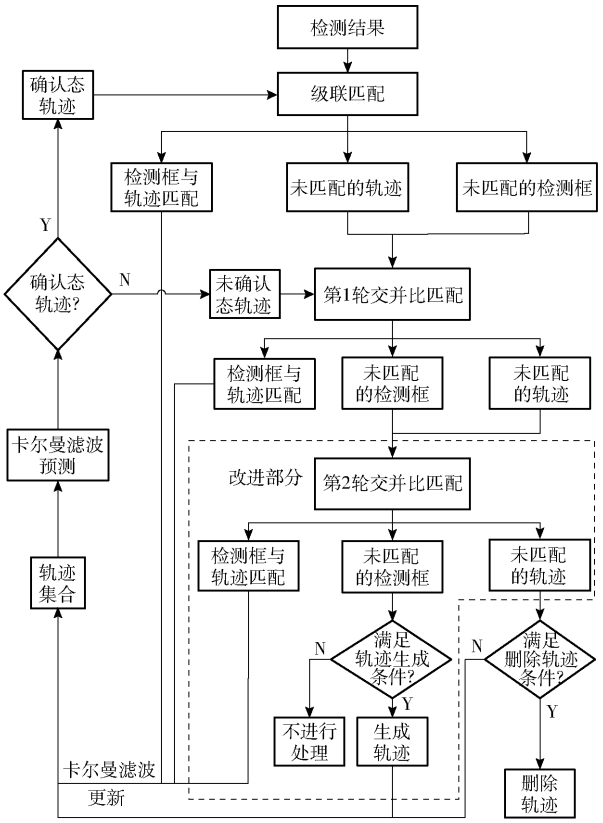


图 5　匹配过程流程图

Fig. 5　Flow chart of matching process

算法步骤如下：

(1) 首先将检测结果与卡尔曼滤波预测的确认态轨迹进行级联匹配；级联匹配采用匈牙利算法，对运动信息与外观信息的关联矩阵求解，从而匹配检测框与轨迹。

(2) 未确认态轨迹、级联匹配中未匹配的轨迹和未匹配的检测框进行第 1 轮交并比匹配；交并比匹配采用匈牙利算法，对交并比关联矩阵求解，从而匹配检测框与轨迹。

(3) 进行第 2 轮交并比匹配，将第 1 轮交并比匹配中未匹配的轨迹和未匹配检测框进行第 2 轮最大交并比距离匹配；判断未匹配的检测框是否满足轨迹生成条件，满足则生成轨迹，不满足则不做处理；最终判断未匹配的轨迹是否满足删除条件，满足则删除轨迹，不满足则继续参与匹配。

(4) 当前帧匹配过程结束后执行卡尔曼滤波更新。

3　实验与结果分析

针对猪舍中群养猪的行为跟踪进行 3 个实验，分别是：目标检测实验训练 YOLO v5s 检测器；重识别实验训练 DeepSORT 算法中重识别模块；在生猪行为跟踪数据集进行实验，测试与分析改进 DeepSORT 算法的性能。

实验采用 Windows 10 平台，编程语言为 Python 3.6，模型框架为 Pytorch 1.7.1，硬件环境为 AMD Ryzen5 2600X 处理器、64 GB 内存、NVDIA GeForce GTX TITAN X 显卡。

3.1　目标检测实验与结果分析

目标检测实验采用 YOLO v5s 模型，检测生猪躺卧、站立、饮食和其他行为，其他行为描述猪只行为转换过程的中间状态。实验不采用预训练模型，输入图像尺寸为 640 像素 ×640 像素，优化器为随机梯度下降法（Stochastic gradient descent，SGD），初始学习率为 0.01，批量大小为 64，模型迭代 200 次。

为验证检测效果，选取精确率（Precision）、召回率（Recall）、调和平均数（F1-score，F1）、平均精度均值（Mean average precision，mAP）对模型进行综合评价。

YOLO v5s 在测试集上的精确率-召回率曲线如图 6 所示，检测器对猪只的 4 种行为均有较好的检测效果。

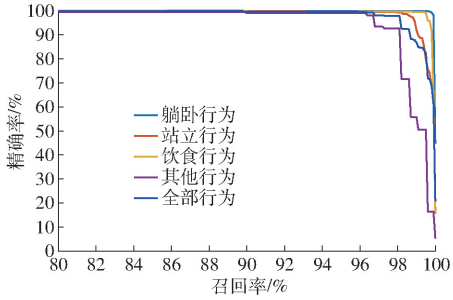


图 6　精确率-召回率曲线

Fig. 6　Precision－recall curves

算法在测试集上全部结果如表 3 所示。在猪只的躺卧、站立、饮食行为上检测效果较好，其平均精确率（AP）均达到 99% 以上，说明检测器对于猪舍场景中大多数行为的检测效果较好，而在其他行为上的召回率为 95.2%，说明模型在检测猪只行为转

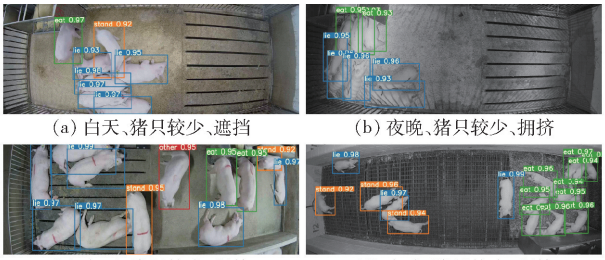
表 3　目标检测实验结果

Tab. 3　Experimental results of object detection　%

行为	精确率	召回率	F1 值	AP	mAP
躺卧	99.9	99.4	99.6	99.7	
站立	99.2	98.3	98.7	99.5	
饮食	99.1	99.6	99.3	99.5	
其他	99.0	95.2	97.1	98.6	
全部行为	99.3	98.1	98.7		99.3

换过程中存在少量漏检情况。综合猪只的各个行为结果,检测器在处理猪舍场景行为检测问题上效果良好,可为跟踪阶段建立最优的输入。

利用 YOLO v5s 算法对不同条件下群养生猪图像进行测试,结果如图 7 所示。图 7a 为白天、猪只较少和遮挡情况下检测效果,算法对于猪只较少的情况能达到优越的检测效果;图 7b 为夜晚、猪只较少和拥挤的情况下检测效果,在拥挤重叠情况下,算法仍保持着精准的检测效果;图 7c 为白天、猪只较多和拥挤的情况下检测效果,算法也具有精准的检测效果;图 7d 为夜晚、猪只较多和拥挤的情况下检测效果,算法在猪只严重拥挤情况下保持着较强的性能,无漏检。不同场景下 YOLO v5s 算法都能准确识别猪只行为。



(a) 白天、猪只较少、遮挡 (b) 夜晚、猪只较少、拥挤

(c) 白天、猪只较多、拥挤 (d) 夜晚、猪只较多、拥挤

图 7 YOLO v5s 算法的群养生猪检测结果

Fig. 7 Detection results of group-housed pigs based on YOLO v5s algorithm

3.2 重识别实验与结果分析

实验使用 Market－1501[30] 数据集上预训练得到的权重文件,为提升对生猪的重识别效果,在本文数据集上重新训练重识别模型,批量大小为 256,迭代 100 次,其余参数沿用原算法。

重识别模型可以提取出猪只具有区分度的特征,实现不同帧中同一猪只的重新识别,使用 Top－1 准确率评价模型效果,其表示模型预测概率最大结果的正确总数占所有样本的比率,值越接近 1 说明模型提取特征能力越强,即重识别效果越好。

图 8 为重识别模型的 Top－1 准确率曲线,在迭代 25 次后,曲线趋于平稳,此时模型基本达到收敛,迭代 100 次后 Top－1 准确率在测试集上结果为 99.88%,此时重识别模型能够较好地提取出目标的表观特征,从而准确地实现猪只的重识别。

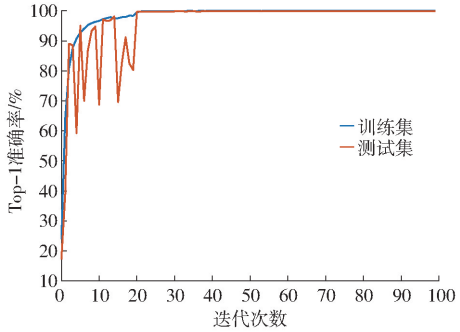


图 8 重识别模型的 Top－1 准确率曲线

Fig. 8 Top－1 accuracy curves of re-identification model

3.3 行为跟踪实验与结果分析

3.3.1 行为跟踪算法评价指标

选用 5 个指标评价行为跟踪算法的效果:身份编号切换次数(Identity switch,IDS),跟踪目标 ID 发生改变的次数,值越小表示跟踪稳定性越好。识别平均数比率(Identification F1,IDF1)是识别精确率与识别召回率的调和平均数,用于评价跟踪算法的稳定性,值越大说明算法越能长时间地对某个目标进行准确地跟踪。多目标跟踪准确率(Multiple object tracking accuracy,MOTA),同时考虑误报、漏报和 IDS,衡量跟踪算法在检测目标和保持轨迹时的性能,与目标检测精度无关,值越大表示算法的性能越好。多目标跟踪精确度(Multiple object tracking precision,MOTP),量化检测器的定位精度越大表示检测器的精度越高。帧率(Frames per second,FPS)是算法每秒处理的视频帧数,值越大表明处理速度越快。

3.3.2 行为跟踪实验结果与分析

为验证改进算法对性能的提升,沿用原算法的参数设置,利用生猪行为跟踪数据集测试改进 DeepSORT 算法,改进前后实验结果如表 4 所示。

表 4 改进前后实验结果

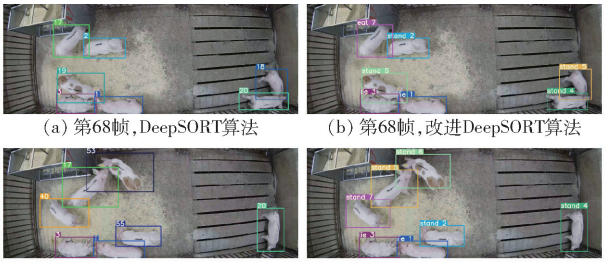
Tab. 4 Experimental results before and after improvement

算法	测试视频序号	IDS	IDF1/%	MOTA/%	MOTP/%	FPS/(f·s^{-1})
DeepSORT	01	47	52.7	91.8	92.9	11
	02	29	82.0	94.0	89.7	7
	03	71	66.9	85.2	88.0	7
	04	57	77.2	93.3	93.1	8
	总计/平均	204	72.3	90.9	90.6	8
改进 DeepSORT	01	0	97.2	94.5	92.5	11
	02	3	94.1	95.0	89.8	7
	03	13	83.2	88.8	87.9	7
	04	17	87.0	90.7	92.8	8
	总计/平均	33	89.2	91.9	90.5	8

在 IDS 方面,改进 DeepSORT 算法为 33,较改进前的 204,降低了 83.8%,特别是测试视频 01,改进算法对该测试视频的 IDS 为 0,即不发生 ID 切换,说明改进算法对跟踪场景条件良好(如猪只较少)的情况下效果显著。在 IDF1 方面,改进算法为 89.2%,较原算法提升了 16.9 个百分点,在所有测试视频段均有明显提升,说明改进部分在不同场景条件下可以明显提升跟踪算法的稳定性。在 MOTA

方面，改进算法为 91.9%，较改进前提升了 1.0 个百分点，改进算法对测试视频 01、02、03 的处理均优于原算法，说明在大多数场景条件下改进后算法均能提升跟踪准确率。在 MOTP 方面与 FPS 方面，改进前后算法基本维持不变。综上，改进算法在跟踪准确率与稳定性方面显著提升。

改进前后 DeepSORT 算法在白天、猪群稀疏、猪只活动频繁场景（视频段 01）下跟踪结果如图 9 所示，图 9a 中猪只最大 ID 为 20，而图 9b 中猪只最大 ID 稳定在 7，可以看出在猪只剧烈运动的情况下，原算法跟踪中 ID 频繁切换；图 9c 中原算法最大 ID 已经增长到 55，而图 9d 中改进算法最大 ID 依旧稳定在 7，无 ID 切换，跟踪性能优异，且可以准确识别全部猪只的行为。



（a）第68帧，DeepSORT算法　（b）第68帧，改进DeepSORT算法

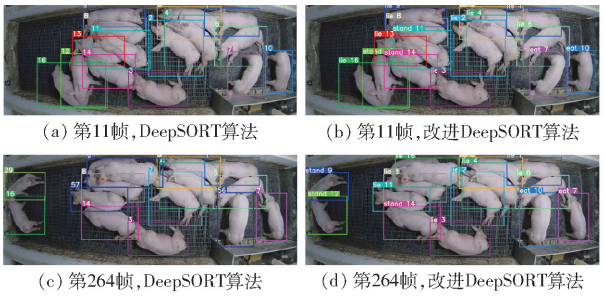
（c）第207帧，DeepSORT算法　（d）第207帧，改进DeepSORT算法

图 9　改进前后 DeepSORT 算法跟踪结果（视频段 01）

Fig. 9　Tracking results of DeepSORT algorithm before and after improvement (video segment 01)

改进前后 DeepSORT 算法在白天、猪群拥挤、猪只活动较少场景（视频段 02）下跟踪结果如图 10 所示，在猪只发生严重遮挡重叠情况下，改进前后算法均出现少量目标丢失，图 10a 与图 10b 中猪只最大 ID 均为 16；图 10c 中原算法最大 ID 已经增长到 57，而图 10d 中改进算法最大 ID 稳定在 16，可以看出改进算法可以大幅抑制 ID 切换，在猪只严重密集拥挤情况下仍可以实现良好跟踪，并且可以识别到大部分猪只的行为。

改进前后 DeepSORT 算法在夜晚、猪群拥挤、猪只活动频繁场景（视频段 04）下跟踪结果如图 11 所示，图 11a 与图 11b 中猪只最大 ID 均为 16，图 11c 与图 11d 为第 246 帧中算法对比结果，光照条件较弱与猪群发生严重遮挡导致原算法在跟踪中出现大量 ID 切换，而改进算法能有效抑制 ID 切换，表明在复杂环境下改进算法仍有良好的跟踪效果，并且可以准确识别到大部分猪只行为。

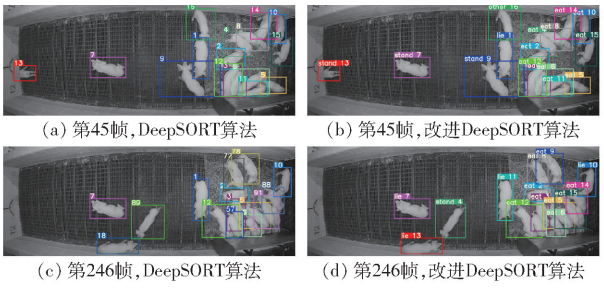


（a）第11帧，DeepSORT算法　（b）第11帧，改进DeepSORT算法

（c）第264帧，DeepSORT算法　（d）第264帧，改进DeepSORT算法

图 10　改进前后 DeepSORT 算法跟踪结果（视频段 02）

Fig. 10　Tracking results of DeepSORT algorithm before and after improvement (video segment 02)



（a）第45帧，DeepSORT算法　（b）第45帧，改进DeepSORT算法

（c）第246帧，DeepSORT算法　（d）第246帧，改进DeepSORT算法

图 11　改进前后 DeepSORT 算法跟踪结果（视频段 04）

Fig. 11　Tracking results of DeepSORT algorithm before and after improvement (video segment 04)

综上，改进 DeepSORT 算法可以有效抑制 ID 切换，在复杂环境中仍有效，可以良好稳定地跟踪群养生猪并准确识别其行为。

4　结论

（1）在 YOLO v5s 检测算法基础上，将检测结果中行为信息引入 DeepSORT 算法，并针对猪舍特定场景改进 DeepSORT 算法中的轨迹生成与匹配过程，提出了改进 DeepSORT 算法。

（2）实验结果表明，YOLO v5s 目标检测实验的 mAP 为 99.3%，F1 值为 98.7%；生猪重识别实验的 Top－1 准确率达到 99.88%；在跟踪方面，改进 DeepSORT 算法的 MOTA 为 91.9%，IDF1 为 89.2%，IDS 为 33，MOTA 和 IDF1 比原算法分别提升了 1.0、16.9 个百分点，IDS 下降了 83.8%。

（3）所构建群养生猪行为跟踪算法可以满足实际养殖环境中的需要，能够为无接触式的生猪自动监测提供技术支持。

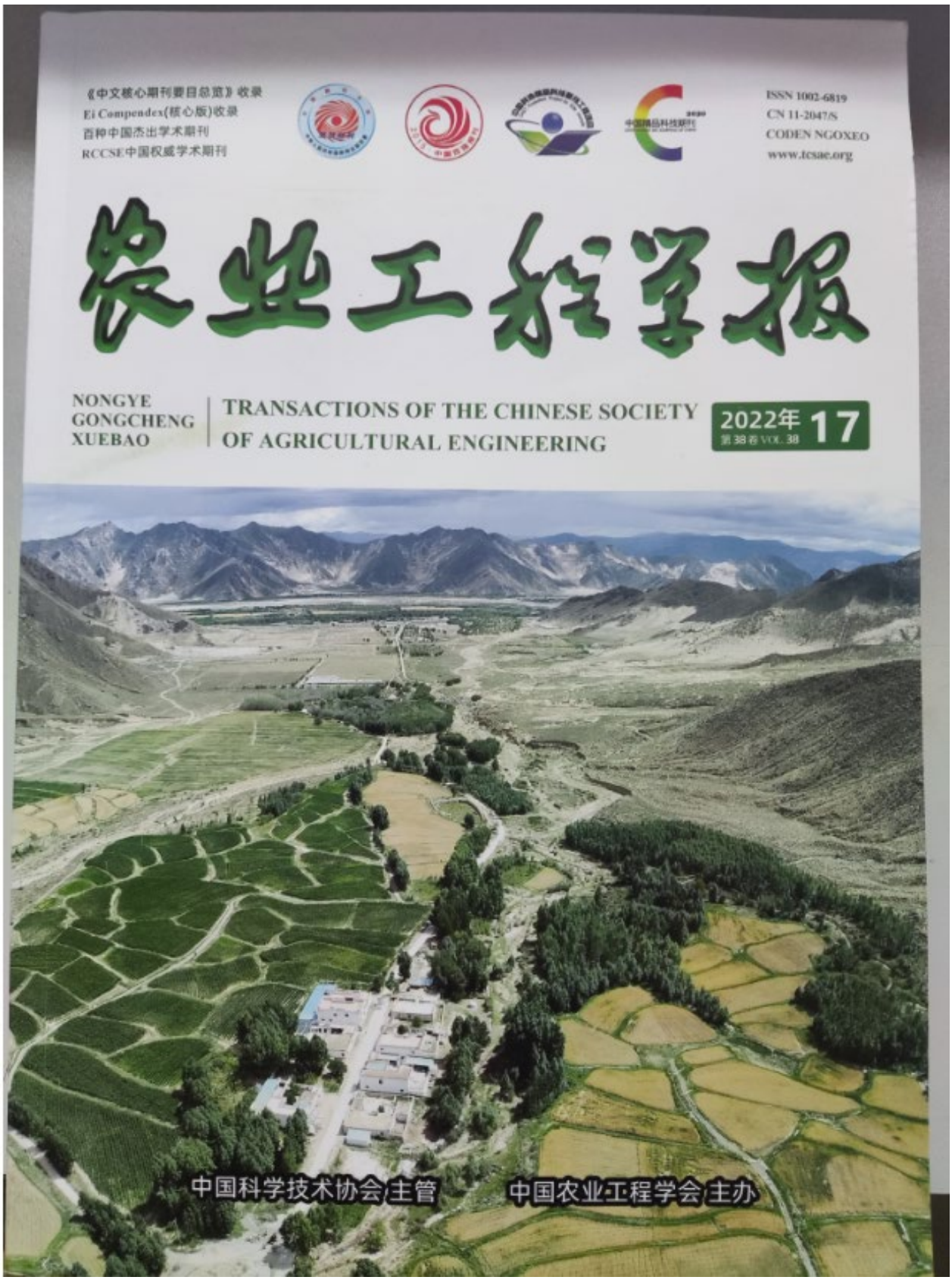
参　考　文　献

[1]　沈明霞，王梦雨，刘龙申，等. 基于深度神经网络的猪咳嗽声识别方法[J]. 农业机械学报，2022，53(5)：257－266.
SHEN Mingxia, WANG Mengyu, LIU Longshen, et al. Recognition method of pig cough based on deep neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 257－266. (in Chinese)

[2]　李丹，张凯锋，李行健，等. 基于 Mask R－CNN 的猪只爬跨行为识别[J]. 农业机械学报，2019，50(增刊)：261－266，275.
LI Dan, ZHANG Kaifeng, LI Xingjian, et al. Mounting behavior recognition for pigs based on Mask R－CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 261－266, 275. (in Chinese)

[3] CHEN C, ZHU W X, NORTON T. Behaviour recognition of pigs and cattle: journey from computer vision to deep learning[J]. Computers and Electronics in Agriculture, 2021, 187: 106255.
[4] ZHU W X, GUO Y Z, JIAO P P, et al. Recognition and drinking behaviour analysis of individual pigs based on machine vision [J]. Livestock Science, 2017, 205: 129 - 136.
[5] 李菊霞,李艳文,牛帆,等. 基于 YOLOv4 的猪只饮食行为检测方法[J]. 农业机械学报,2021,52(3):251 - 256.
LI Juxia, LI Yanwen, NIU Fan, et al. Pig diet behavior detection method based on YOLOv4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(3): 251 - 256. (in Chinese)
[6] ALAMEER A, KYRIAZAKIS I, BACARDIT J. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs[J]. Scientific Reports, 2020, 10(1): 13665.
[7] 高云,陈斌,廖慧敏,等. 群养猪侵略性行为的深度学习识别方法[J]. 农业工程学报,2019,35(23):192 - 200.
GAO Yun, CHEN Bin, LIAO Huimin, et al. Recognition method for aggressive behavior of group pigs based on deep learning [J]. Transactions of the CSAE, 2019, 35(23): 192 - 200. (in Chinese)
[8] XIAO D Q, FENG A J, LIU J. Detection and tracking of pigs in natural environments based on video analysis[J]. International Journal of Agricultural and Biological Engineering, 2019, 12(4): 116 - 126.
[9] SUN L Q, CHEN S H, LIU T, et al. Pig target tracking algorithm based on multi-channel color feature fusion[J]. International Journal of Agricultural and Biological Engineering, 2020, 13(3): 180 - 185.
[10] ZHANG L, GRAY H, YE X J, et al. Automatic individual pig detection and tracking in pig farms[J]. Sensors, 2019, 19(5): 1188.
[11] 张伟,沈明霞,刘龙申,等. 基于 CenterNet 搭配优化 DeepSORT 算法的断奶仔猪目标跟踪方法研究[J]. 南京农业大学学报,2021,44(5):973 - 981.
ZHANG Wei, SHEN Mingxia, LIU Longshen, et al. Research on weaned piglet target tracking method based on CenterNet collocation optimized DeepSORT algorithm[J]. Journal of Nanjing Agricultural University, 2021, 44(5): 973 - 981. (in Chinese)
[12] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing (ICIP), 2017:3645 - 3649.
[13] NASIRAHMADI A, EDWARDS S A, STURM B. Implementation of machine vision for detecting behaviour of cattle and pigs [J]. Livestock Science, 2017, 202: 25 - 38.
[14] LI D, ZHANG K F, LI Z B, et al. A spatiotemporal convolutional network for multi-behavior recognition of pigs[J]. Sensors, 2020, 20(8): 2381.
[15] 杨秋妹,肖德琴,张根兴. 猪只饮水行为机器视觉自动识别[J]. 农业机械学报,2018,49(6):232 - 238.
YANG Qiumei, XIAO Deqin, ZHANG Genxing. Automatic pig drinking behavior recognition with machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(6): 232 - 238. (in Chinese)
[16] JUNG W, KIM S H, HONG S P, et al. An AIoT monitoring system for multi-object tracking and alerting[J]. CMC-Computers Materials & Continua, 2021, 67(1): 337 - 348.
[17] GAN H M, OU M Q, ZHAO F Y, et al. Automated piglet tracking using a single convolutional neural network[J]. Biosystems Engineering, 2021, 205: 48 - 63.
[18] PSOTA E T, SCHMIDT T, MOTE B, et al. Long-term tracking of group-housed livestock using keypoint detection and MAP estimation for individual animal identification[J]. Sensors, 2020, 20(13): 3670.
[19] 魏贤哲,卢武,赵文彬,等. 基于改进 Mask R - CNN 的输电线路防外破目标检测方法研究[J]. 电力系统保护与控制,2021,49(23):155 - 162.
WEI Xianzhe, LU Wu, ZHAO Wenbin, et al. Target detection method for external damage of a transmission line based on an improved Mask R - CNN algorithm[J]. Power System Protection and Control, 2021, 49(23): 155 - 162. (in Chinese)
[20] REDMON J, DIVVALA S K, GIRSHICK R B, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779 - 788.
[21] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517 - 6525.
[22] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. ArXiv, 2018: 1804.02767.
[23] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. ArXiv, 2020: 2004.10934.
[24] GIRSHICK R B, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.
[25] GIRSHICK R B. Fast R - CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), 2015: 1440 - 1448.
[26] REN S, HE K, GIRSHICK R B, et al. Faster R - CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6):1137 - 1149.
[27] LIN T Y, DOLLÁR P, GIRSHICK R B, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936 - 944.
[28] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759 - 8768.
[29] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing (ICIP), 2016: 3464 - 3468.
[30] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV), 2015: 1116 - 1124.

2.10 基于JDE模型的群养生猪多目标跟踪



农 业 工 程 学 报

2022 年 9 月第 17 期 （总第 441 期） 第 38 卷

目 次

·农业生物环境与能源工程·

·土地保障与生态安全·

·农产品加工工程·

致谢：感谢西北农林科技大学焦菊英团队提供的封面图片“拉萨河畔农田航拍图”。

《农业工程学报》：对标一流，追求卓越

《农业工程学报》（以下简称《学报》）创刊于 1985 年，现为半月刊，全年 24 期，大 16 开面向国内外公开发行。《学报》是由中国科学技术协会主管、中国农业工程学会主办的全国性专业学术期刊。读者对象为农业工程学科及相关领域的科研、教学及生产科技人员、技术管理及推广人员和高等院校师生。刊稿内容涵盖了农业装备工程与机械化、农业航空工程、农业水土工程、农业信息与电气技术、农业生物环境与能源工程、土地保障与生态安全、农产品加工工程等学科专业领域。

《学报》始终坚持“双为”方向和“双百”方针的办刊宗旨及“内容为王，质量为本”的办刊理念。拥有国内外农业工程相关领域各专业的知名专家学者组成的编委会，其中两院院士 30 余人。本刊坚持专家办刊，编委在稿件同行评审把关中发挥重要作用。

《学报》是中国农业工程领域的领军期刊，在行业具有很高的学术影响力，被 EI Compendex（核心版）、Scopus、CA、CSA、CAB Abstracts、CSCD、《中文核心期刊要目总览》、《中国科技核心期刊目录》、《中国农林核心期刊概览 2020》等国内外多个权威数据库收录。中国科学技术信息研究所 2021 年最新影响因子 2.162，在农业工程类核心期刊中位列第一名。

我们愿与广大农业工程同仁携手并肩，共同奋斗，对标一流，追求卓越，创建国际知名品牌，引领学科发展，培养人才，激励创新，为全面推进乡村振兴，加快农业农村现代化，促进农业农村高质量发展作出新贡献！

期刊荣誉：多项位列农业工程类期刊榜首

- 国家新闻出版广电总局“双效期刊”
- 国家新闻出版广电总局“百强报刊”
- 中国科技期刊卓越行动计划-梯队期刊
- 百种中国杰出学术期刊
- 中国精品科技期刊
- TOP5%中国最具国际影响力学术期刊
- RCCSE 中国权威学术期刊（A+）
- 科技期刊数字影响力 100 强
- 世界学术影响力指数 WAJCI-Q1 区期刊（为本学科唯一入选中文刊）
- Google Scholar 学术期刊影响力排名位列高被引中文刊第四名
- 中国农林领域高质量科技期刊分级目录第一区（T1）
- 中国科协精品科技期刊工程项目资助期刊

對標一流
追求卓越

ISSN 1002-6819
CN 11-2047/S
国内邮发代号：18-57
海外发行代号：BM3335
国内定价：80元

ISSN 1002-6819
9 771002 681221 17>

基于 JDE 模型的群养生猪多目标跟踪

涂淑琴，黄 磊，梁 云※，黄正鑫，李承桀，刘晓龙
（华南农业大学数学与信息学院，广州 510642）

摘 要：为实现群养生猪在不同场景下（白天与黑夜，猪只稀疏与稠密）的猪只个体准确检测与实时跟踪，该研究提出一种联合检测与跟踪（Joint Detection and Embedding，JDE）模型。首先利用特征提取模块对输入视频序列提取不同尺度的图像特征，产生 3 个预测头，预测头通过多任务协同学习输出 3 个分支，分别为分类信息、边界框回归信息和外观信息。3 种信息在数据关联模块进行处理，其中分类信息和边界框回归信息输出检测框的位置，结合外观信息，通过包含卡尔曼滤波和匈牙利算法的数据关联算法输出视频序列。试验结果表明，本文 JDE 模型在公开数据集和自建数据集的总体检测平均精度均值（mean Average Precision，mAP）为 92.9%，多目标跟踪精度（Multiple Object Tracking Accuracy，MOTA）为 83.9%，IDF1 得分为 79.6%，每秒传输帧数（Frames Per Second，FPS）为 73.9 帧/s。在公开数据集中，对比目标检测和跟踪模块分离（Separate Detection and Embedding，SDE）模型，本文 JDE 模型在 MOTA 提升 0.5 个百分点的基础上，FPS 提升 340%，解决了采用 SDE 模型多目标跟踪实时性不足问题。对比 TransTrack 模型，本文 JDE 模型的 MOTA 和 IDF1 分别提升 10.4 个百分点和 6.6 个百分点，FPS 提升 324%。实现养殖环境下的群养生猪多目标实时跟踪，可为大规模生猪养殖的精准管理提供技术支持。

关键词：目标检测；目标跟踪；联合检测与跟踪；数据关联；群养生猪
doi：10.11975/j.issn.1002-6819.2022.17.020
中图分类号：TP391.4 **文献标志码**：A **文章编号**：1002-6819(2022)-17-0186-10

涂淑琴，黄磊，梁云，等. 基于 JDE 模型的群养生猪多目标跟踪[J]. 农业工程学报，2022，38(17)：186-195.
doi：10.11975/j.issn.1002-6819.2022.17.020 http://www.tcsae.org
Tu Shuqin, Huang Lei, Liang Yun, et al. Multiple object tracking of group-housed pigs based on JDE model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(17): 186-195. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2022.17.020 http://www.tcsae.org

0 引 言

生猪产业一直是国内畜牧业的支柱产业，其发展关系到国家食物安全、社会稳定及国民经济的协调发展。生猪养殖业正朝着规模化、专业化、智能化和精细化发展。目前，在劳动力短缺的情况下，智能与精准畜牧业对帮助农户实现畜牧业规模化生产具有重要作用[1]。通过视频摄像头，采用计算机视觉技术获取每头猪每天的体重变化、运动轨迹、饮食情况和行为变化等数据，监测猪只行为和健康，预测猪只个体异常情况，实现生猪生产过程的精确控制[2]，对提高生猪的福利具有重要价值[3]。因此，采用多目标跟踪技术，准确跟踪群养生猪中的个体，识别猪只行为变化，对提高农场的智能化管理水平和生产力具有重要意义。

目前，国内外研究者在禽畜跟踪的方面进行很多研究。有些研究者通过给禽畜穿戴自动跟踪设备实现跟踪禽畜。如 Zambelis 等[4]使用耳标加速计对饲养奶牛的喂养和活动行为进行观察。Giovanetti 等[5]将三轴加速度计传感器安装在羊的身体上，然后测量羊在牧场的行为。Krista 等[6]将运动能耗仪安装在母羊的项圈上，以此评估绵羊行为活动水平。这些方法在某些情况下对于禽畜的观察是可行的，但是，使用可穿戴自动跟踪设备会影响禽畜的行为，严重情况下会影响其自由活动，降低动物福利。另外，大量可穿戴自动跟踪设备会增加生产的成本。

近年来，使用计算机视觉技术进行猪只日常行为监控取得了多方面的研究成果，例如猪的攻击行为[7-10]、饮食饮水行为[11-15]、母猪行为检测[16]、攀爬和玩耍行为[17-18]，猪只姿态识别[11,19-22]，早期发现呼吸道疾病[23-24]。

多目标跟踪的性能在很大程度上取决于其检测目标的性能。传统的目标检测算法，如 Zhao 等[25]使用背景减法来检测移动奶牛目标，Zhang 等[26]提出了一种基于光流估计的运动目标检测方法，于欣等[27]提出一种基于光流法与特征统计的鱼群异常行为检测方法，这些算法在速度和准确性方面不能满足实际场景要求。目前，基于深度学习的目标检测算法不断完善，其准确性和速度都有显著提升，能够满足实际应用。深度学习的目标检测算法主要分为一阶段和二阶段算法。二阶段算法在检测时首先生成候选区域，之后对候选区域进行分类和校准，准确率相对较高，典型的有 R-CNN（Region Convolution Neural Network）算法[28]，Fast R-CNN 算法[29]，Faster R-CNN 算法[30]。如王浩等[31]利用改进的 Faster R-CNN 算

收稿日期：2022-04-19 修订日期：2022-08-16
基金项目：广东省省级科技计划项目（2019A050510034）；广州市重点科技计划项目（202206010091）；大学生创新创业大赛项目（202110564025）
作者简介：涂淑琴，博士，讲师，研究方向为图像处理与计算机视觉。
Email：tushuqin@163.com
※通信作者：梁云，博士，教授，研究方向为图像处理与计算机视觉。
Email：yliang@scau.edu.cn

法定位群养生猪的圈内位置，识别准确率可达 96.7%。一阶段算法在检测时无需生成候选区域，直接对目标类别和边界进行回归，如 YOLO 系列算法[32-35]。如金耀等[36]利用 YOLOv3 算法[32]对生猪个体进行识别，对母猪的识别精度均值达 95.16%。相较于二阶段算法，一阶段算法的检测速度更快。

在多目标跟踪方面，现有多目标跟踪算法的应用大多是基于检测跟踪（Tracking by Detection，TBD）范式，即 SDE（Separate Detection and Embedding）模型，先用检测器输出检测结果，再用基于卡尔曼滤波和匈牙利算法的后端追踪优化算法进行跟踪，如使用 SORT（Simple Online and Realtime Tracking）[37]、DeepSORT [38]算法来提取目标的表观特征进行多目标重识别进行跟踪，其中 DeepSORT 算法在 SORT 算法的基础上，通过提取深度表观特征提高了多目标的跟踪效果。如张宏鸣等[39]利用改进 YOLOv3 算法结合 DeepSORT 算法进行肉牛多目标跟踪，张伟等[40]利用基于 CenterNet 结合优化 DeepSORT 算法进行断奶仔猪目标跟踪。上述研究的算法是两阶段过程，先检测再跟踪，目标检测和跟踪模块分离导致跟踪速度慢，达不到实时跟踪效果。

本研究将目标检测与跟踪融合在一个过程中，提出一种实时、非接触的群养生猪多目标跟踪 JDE（Joint Detection and Embedding）算法，通过一个端对端网络同时输出多目标的分类信息、边界框回归信息和外观信息，以减少算法的运行时间，达到实时跟踪的效果。在相同的公开试验数据集中将 JDE 算法与 SDE 算法进行对比，以验证本文算法的速度，同时与 TransTrack 算法[41]对比，进一步验证本文算法的准确性与实时性。

1　基于 JDE 的群养生猪多目标跟踪算法

1.1　多目标跟踪算法概述

基于 JDE 的群养生猪多目标跟踪算法如图 1 所示。该算法以群养生猪视频序列为输入；采用特征提取模块提取不同尺度的图像特征，得到 3 个不同尺度特征图的预测头，输入数据关联模块；预测头的分类信息和边界框回归信息用于得到检测框的位置结果，在跟踪部分，利用外观信息结合检测框，通过包含卡尔曼滤波和匈牙利算法的数据关联算法，输出检测与跟踪的视频序列结果。

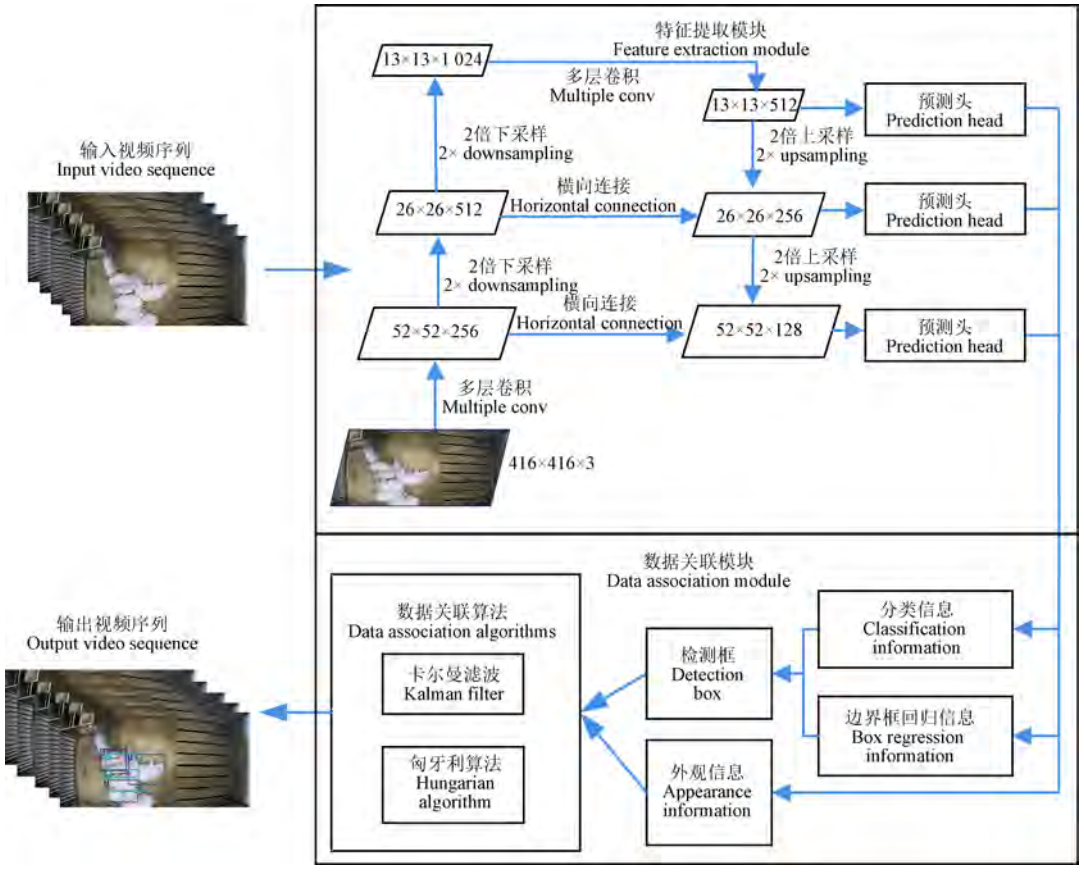


图 1　基于 JDE 的群养生猪多目标跟踪算法

Fig.1　Multiple object tracking algorithm for group-housed pigs based on Joint Detection and Embedding (JDE)

1.2　特征提取模块

特征提取模块由 Darknet-53 网络和多尺度模块特征金字塔构成，如图 2 所示。Darknet-53 网络包括 6 个卷积层和 5 个残差层，其中卷积层和残差层的大小和数量见表 1。卷积层由卷积层、批量归一化层和激活函数层共同构成，残差层由一个 1×1 大小的卷积层和 3×3 大小的卷积层构成。

特征金字塔采用同一图像的不同尺度来检测目标，有助于检测小目标。本文特征金字塔利用 Darknet-53 网络中的第 3、4 和 5 个残差块进行特征融合，产生 3 个输出预测头，分别输出分类信息、边界框回归信息和外观信息。

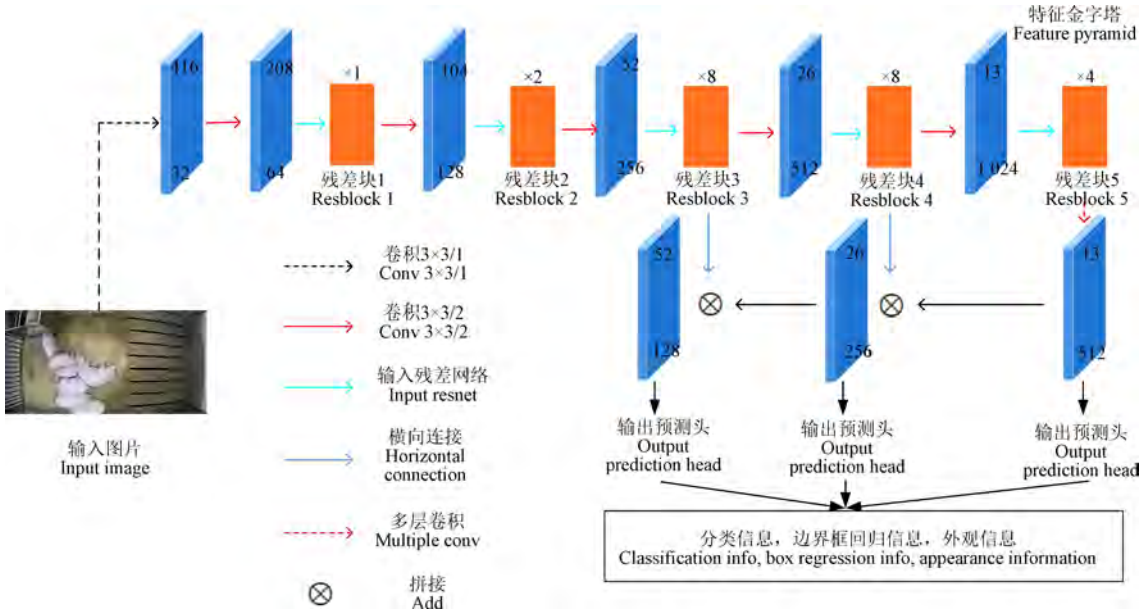


图 2 特征提取网络结构

Fig.2 Diagram of feature extraction network structure

表 1 Darknet-53 网络结构参数

Table 1 Darknet-53 network structure parameters

名称 Name	输出图片大小 Output image size/Pixel	滤波器个数 Number of filters	滤波器大小 Filters size/Pixel	数量 Numbers
卷积层 1 Conv layer 1	416×416	32	3×3	1
卷积层 2 Conv layer 2	208×208	64	3×3/2	1
残差层 1 Residual layer 1	208×208	32，64	1×1，3×3	1
卷积层 3 Conv layer 3	104×104	128	3×3/2	1
残差层 2 Residual layer 2	104×104	64，128	1×1，3×3	2
卷积层 4 Conv layer 4	52×52	256	3×3/2	1
残差层 3 Residual layer 3	52×52	128，256	1×1，3×3	8
卷积层 5 Conv layer 5	26×26	512	3×3/2	1
残差层 4 Residual layer 4	26×26	256，512	1×1，3×3	8
卷积层 6 Conv layer 6	13×13	1 024	3×3/2	1
残差层 5 Residual layer 5	13×13	512，1 024	1×1，3×3	4

1.3 数据关联模块

本文 JDE 算法的学习目标为多任务协同学习，其总体损失 L_{total} 为分类损失、边界框回归损失和外观信息学习损失之和，如式（1）所示。

$$L_{total}=\omega_\alpha L_\alpha+\omega_\beta L_\beta+\omega_\gamma L_\gamma \tag{1}$$

式中 ω_α、ω_β、ω_γ 分别为分类、边界框回归和外观信息学习的权重值，L_α 为分类损失，L_γ 为外观信息学习损失，其中损失均为交叉熵损失，计算公式如式（2）所示。

$$L_\alpha=L_\gamma=-\frac{1}{N}\sum_{i}^{M}\sum_{c=1}y_{ic}\lg p_{ic} \tag{2}$$

式中 M 为类别的数量，N 为样本数，y_{ic} 为符号函数（0 或 1），c 为类别数。如果样本 N 的真实类别等于 c，则 y_{ic}=1，否则 y_{ic}=0。p_{ic} 为观测样本 i 属于类别 c 的预测概率。

L_β 为边界框回归损失，为 smooth-L1 损失，计算公式如式（3）所示。

$$L_\beta=\begin{cases}0.5x^2, |x|<1\\ |x|-0.5, \text{其他}\end{cases} \tag{3}$$

式中 x 为输入样本。

算法采用基于任务的不确定性计算加权系数，最终自动加权的损失 L_{total} 如式（4）所示。

$$L_{total}=\frac{1}{2}\left(\frac{1}{e^{s_\alpha}}L_\alpha+s_\alpha\right)+\frac{1}{2}\left(\frac{1}{e^{s_\beta}}L_\beta+s_\beta\right)+\frac{1}{2}\left(\frac{1}{e^{s_\gamma}}L_\gamma+s_\gamma\right) \tag{4}$$

式中 s_α、s_β、s_γ 为每个个体损失的任务依赖的不确定性，为可学习参数。

模型通过分类损失和回归损失学习到的分类信息和回归信息生成检测框对视频帧中每个猪只进行定位，外观学习损失得到的外观信息包括每个猪只的外观特征，二者通过数据关联，对每头猪分配 ID，实现多目标跟踪。猪只多目标跟踪的具体实现流程如图 3 所示，具体步骤如下：

1）创建初始跟踪轨迹。对于给定的视频帧序列，第一帧将根据视频帧序列的检测结果利用卡尔曼滤波对轨迹进行初始化，并维护一个跟踪轨迹池，包含所有可能与预测值相关联的轨迹。

2）数据关联。对于下一帧的输出结果，利用卡尔曼滤波进行轨迹预测，计算出预测值与轨迹池之间的运动亲和信息和外观亲和信息，其中外观亲和信息采用余弦相似度计算，运动亲和信息采用马氏距离计算，然后利用匈牙利算法的代价矩阵进行轨迹分配。

3）更新轨迹。如果出现在 2 帧内的预测值没有被分配给任何一个轨迹池中的轨迹，那么这条轨迹将被初始

化为新的轨迹，然后根据卡尔曼滤波进行所有匹配轨迹状态的更新，如果某条轨迹在连续 30 帧内没有被更新，则终止该轨迹，所有视频帧处理完毕后，输出视频帧序列。

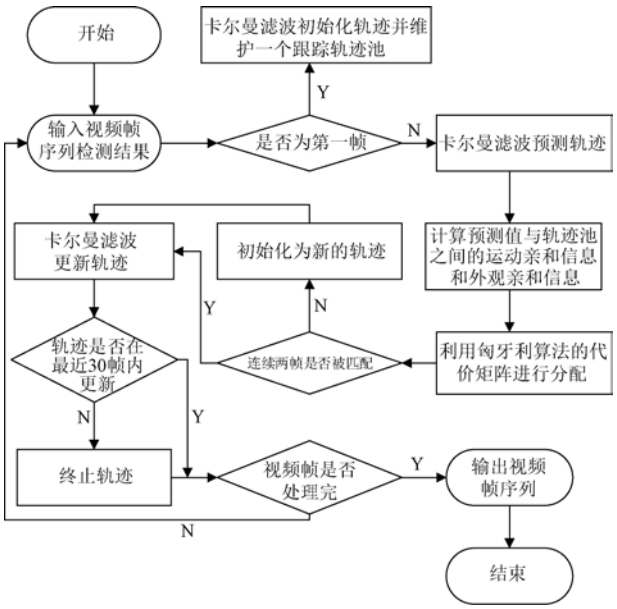


图 3　卡尔曼滤波结合匈牙利算法的猪只目标跟踪流程

Fig.3　Pig object tracking process of Kalman filter combined with Hungarian algorithm

2　数据准备与评价指标

2.1　数据集

本试验采用的数据集包括 2 部分：一部分为 Psota 等[42]提供的公开数据集，包含不同日龄、大小、数量和不同环境的猪只视频，其中，视频 1、2、4、5 为保育猪（3～10 周龄），视频 6、7、8、9、10 为早期育成猪（11～18 周龄），视频 12、15 为晚期育成猪（19～26 周龄）。根据时间段的不同将猪只的活动水平分为 3 类：白天的高活动、白天（或夜晚）的中等活动、白天（或夜晚）的低活动，详表 2。同时，根据人工观察，将猪只个数较多且黏连遮挡情况较为严重的视频定义为稠密视频，反之为稀疏视频，见表 2。另外一部分为自建数据集[43]。两部分数据集均为俯拍视频片段，由于摄像头高度及焦距的影响，不可避免拍摄到猪圈外的物品，因此，在试验中采用视频裁剪方法将视角固定为猪圈内，以减少外部环境的影响。

表 2　公开数据集

Table 2　Public dataset

视频 Videos	白天 Day	黑夜 Night	稀疏 Sparse	稠密 Dense	活动水平 Activity level	猪只个数 Number of pigs
1	√	—	√	—	高	7
2	√	—	√	—	低	7
4	√	—	—	√	中	15
5	—	√	√	—	中	8
6	√	—	—	√	高	16
7	√	—	—	√	中	12
8	—	√	—	√	低	13
9	√	—	—	√	中	14
10	—	√	—	√	中	14
12	√	—	—	√	低	15
15	—	√	—	√	中	16

首先，利用 FFmpeg 软件完成视频剪辑，从中截取稠密、稀疏、白天、黑夜的视频，2 部分数据集共 21 个视频。然后利用 DarkLabel 软件对数据进行标注，其中，公开数据集 11 个视频，共 3 300 张图像，自建数据集 10 个视频，共 1 000 张图像。部分数据集如图 4 所示。为对比不同场景下模型的检测和跟踪能力，选取不同的视频进行模型训练和测试，参与训练的视频不参与测试。本文共设计 3 个试验，其中试验 1 以视频 4、6、12 为测试集，这些视频均为白天稠密，其余视频为训练集。试验 2 以视频 2、5、8 为测试集，其中视频 5、8 分别为夜晚稀疏与夜晚稠密，视频 2 为白天稀疏，其余视频为训练集。试验 3 以自建数据集的 7 个视频为测试集（视频 3、11、14、16、18、19、21），另外 3 个视频为测试集（视频 13、17、20）。其中猪只活动水平定义如下：根据视频的人工观察结果，在白天（10:00－12:30）猪只的饮食和玩耍等行为较频繁，此时间段定义为猪只白天的高活动水平。在白天（12:30－17:00）或夜晚（17:00－20:00）猪只的饮食和玩耍等行为没有白天（10:00－12:30）高，此时间段定义为白天或夜晚的中等活动水平。在白天（7:00－10:00）或夜晚（20:00－7:00）猪只的饮食和玩耍等行为较少，躺卧行为较多，此时间段定义为白天或夜晚的低活动水平。

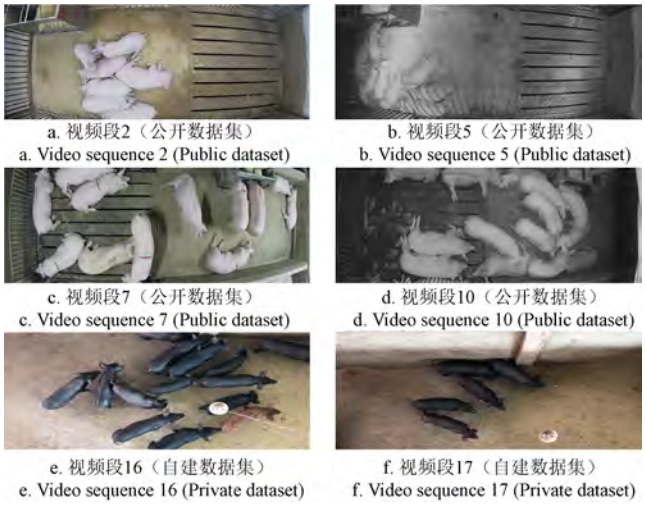


图 4　部分数据集

Fig.4　Part of the dataset

2.2　试验环境

本文所有试验在同一计算机上完成，硬件配置为 12th Gen Intel(R) i9-12900KF CPU，NVIDIA GeForce RTX 3090 GPU，32GB 内存，64 位 Linux 操作系统，Pytorch 版本 1.7.1，Python 版本 3.8，CUDA 版本 11.0。

训练过程中设置图片尺寸为 416×416（像素），批处理大小（Batchsize）设置为 32，初始学习率（Learning Rate）为 0.01，动量（Momentum）设置为 0.9，共训练 30 个时期（Epoch），使用随机梯度下降法（Stochastic Gradient Descent, SGD）进行优化，保存训练过程中精度最高的模型参数进行模型测试。

2.3 评价指标

选择精确率（Precision，P），召回率（Recall，R）和平均精度均值（mean Average Precision，mAP）3 个指标评判模型的检测性能。精确率衡量模型对猪只目标检测的精确程度，如式（5），其中 DTP 是检测正确的目标数量，DFP 是检测错误的目标数量。

$$P=\frac{\mathrm{DTP}}{\mathrm{DTP+DFP}}\times 100\% \tag{5}$$

召回率衡量模型对猪只目标检测的覆盖能力，如式（6），其中 DFN 是漏检的目标数量。

$$R=\frac{\mathrm{DTP}}{\mathrm{DTP+DFN}}\times 100\% \tag{6}$$

平均精度均值是对检测的类别对应的精度均值取平均，如式（7），其中 $P(R)$是以召回率 R 为自变量，精确率 P 为因变量的函数。

$$\mathrm{mAP}=\int_0^1 P(R)\mathrm{d}R \tag{7}$$

选择多目标跟踪精度（Multiple Object Tracking Accuracy，MOTA）和 IDF1 得分（ID F1 Score）作为多目标跟踪的主要评价指标。MOTA 衡量跟踪器检测目标和保持轨迹跟踪的性能。IDF1 为引入跟踪目标标号 ID 的 F1 值，由于引入了跟踪目标标号 ID，IDF1 更重视目标的轨迹跟踪能力。MOTA 计算公式如式（8）所示。

$$\mathrm{MOTA}=1-\frac{\sum_t(\mathrm{FP+FN+IDS})}{\sum_t g_t} \tag{8}$$

式中 FP 为在第 t 帧中目标误报总数（假阳性）；FN 为在第 t 帧目标丢失总数（假阴性）；IDS 为在第 t 帧中跟踪目标标号 ID 发生切换的次数；g_t 是 t 时刻观测到的目标数量。

IDF1 计算公式如式（9）所示。

$$\mathrm{IDF1}=\frac{2\mathrm{IDTP}}{2\mathrm{IDTP+IDFP+IDFN}} \tag{9}$$

式中 IDTP 为 ID 保持不变的情况下正确跟踪到的目标总数，IDFP 为 ID 保持不变的情况下跟踪错误的目标总数，IDFN 为 ID 保持不变的情况下跟踪目标丢失总数。

此外，其他相关指标还有碎片数（Fragmentation，FM）、主要跟踪到的目标（Mostly Tracked Target，MT）（被跟踪到的轨迹比例大于 80%）、主要丢失目标（Mostly Lost Target，ML）（被跟踪到的轨迹比例小于 20%）、部分跟踪到的目标（Partially Tracked Target，PT）（被跟踪到的轨迹比例不大于 80%且不小于 20%）、一条跟踪轨迹改变目标标号 ID 的次数（Identity Switches，IDS）以及平均每秒传输帧数（Frames Per Second，FPS）。

本文对群养生猪目标跟踪模型性能的分析选择 MOTA、IDF1 和 FPS 作为主要评价指标，辅助以 FP、FN、FM、IDS、MT、ML 等指标进行模型的性能评估。其中 MOTA、IDF1、MT 和 FPS 数值越高模型性能越好，FP、FN、FM、IDS 和 ML 数值越低模型性能越好。

3 结果与分析

3.1 JDE 模型试验结果

JDE 模型的检测结果见表 3。可以发现，本文算法在公开数据集中的 mAP 平均值达到 92.5%，测试集 2、4、6、8、12 视频的 mAP 分别为 96.2%、95.6%、96.1%、98.0%、92.2%。对于视频 5，其 mAP 为 77.0%，主要原因是该视频的场景与其他视频相比差异较大，增加了目标检测的难度；在自建数据集中的 mAP 平均值达到 93.8%，总体平均 mAP 达到 92.9%，表明本文 JDE 算法对于不同复杂场景具有较好的检测能力。

表 3 JDE 模型的目标检测试验结果

Table 3 Object detection experiment results of the Joint Detection and Embedding (JDE) model

%

测试集 Test set	视频 Video	精确率 Precision P	召回率 Recall R	平均精度均值 Mean Average Precision mAP
公开数据集 Public dataset	2	96.0	94.8	96.2
	4	84.7	96.1	95.6
	5	81.2	79.2	77.0
	6	96.2	95.1	96.1
	8	99.2	87.9	98.0
	12	83.2	93.6	92.2
自建数据集 Private dataset	13	99.5	90.9	94.8
	17	93.1	99.0	98.6
	20	94.5	80.1	88.0

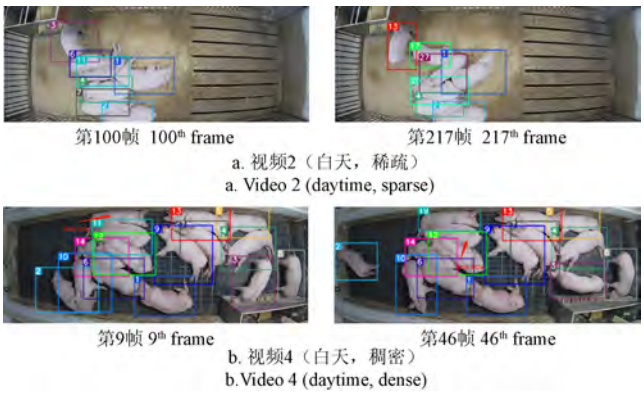
JDE 模型的跟踪结果如表 4 所示。可以发现，在公开数据集中，视频 2、4、5、6、8、12 的 MOTA 分别为 91.4%、82.5%、59.2%、90.8%、94.2%、74.4%，平均 MOTA 为 82.1%，在自建数据集中，视频 13、17、20 的 MOTA 分别为 84.4%、88.1%、90.2%，平均 MOTA 为 87.6%，总体平均 MOTA 为 83.9%。不同视频的 MOTA 产生差别的主要原因是每个视频的环境不同，如视频背景、白天、黑夜、稀疏、稠密和猪只的活动状态，在视频背景干扰严重、猪只活动较为频繁（如饮食，玩耍等行为）情况下，MOTA 相对较低，在夜晚视频 8 中，猪只活动较少且背景对猪只的干扰较小，MOTA 最高，为 94.2%。在夜晚视频 5 中，视频背景干扰严重，MOTA 较低，为 59.2%，根据 IDF1 和 FPS 可以看出，本文 JDE 模型在公开数据集中的 IDF1 平均值为 77.7%，FPS 平均值为 74.26 帧/s，在自建数据集中的 IDF1 平均值为 83.5%，FPS 平均值为 73.19 帧/s，总体平均 IDF1 值为 79.6%，总体平均 FPS 值为 73.9 帧/s。可以发现，本文 JDE 模型对猪只目标的 ID 跟踪精度和 FPS 均达到较高水平，能够实现实际养殖环境下的群养猪多目标快速实时跟踪，为实际群养猪养殖场的精准管理提供技术支持。

表 4　JDE 模型的多目标跟踪试验结果

Table 4　Multiple object tracking experiment results of the JDE model

视频 Video	猪只个数 Pig numbers	主要跟踪到的目标 Mostly tracked target	部分跟踪到的目标 Partially tracked target	主要丢失目标 Mostly lost target	假阳性 False positive	假阴性 False negative	ID 跳变 ID switch	碎片数 Fragmentation	多目标跟踪精度 Multiple object tracking accuracy/%	IDF1 得分 IDF1 score/%	每秒检测帧数 Frames Per Second FPS/(帧·s^{-1})
2	7	6	1	0	29	136	14	19	91.4	80.3	77.64
4	15	12	3	0	377	384	26	30	82.5	75.3	72.47
5	8	4	4	0	438	497	40	51	59.2	59.2	76.46
6	16	14	2	0	115	285	39	63	90.8	75.1	71.56
8	13	13	0	0	104	108	15	26	94.2	94.1	73.88
12	15	12	3	0	847	286	14	45	74.4	82.2	73.55
13	14	11	3	0	24	185	9	14	84.4	79.1	71.19
17	5	5	0	0	24	31	4	9	88.1	88.3	75.73
20	8	6	1	1	6	57	6	5	90.2	83.1	72.66

猪只白天稀疏和稠密 2 种分布情况的可视化分析结果如图 5 所示。



注：图中数字表示猪只 ID 号，算法中第一帧图像的检测会对每头猪只分配一个从 1 递增的 ID 号，例如（1、2、3…），对后续帧进行检测和跟踪时，由于猪只的移动，可能会对某个猪只的 ID 识别错误，此时把这个猪只识别为新的猪只，则该猪只的 ID 号就变为错误的 ID 号，直至所有视频帧处理完毕。下同。

Note: The number in the figure indicates the pig ID No., the first image detection frame of the algorithm will assign an incremental ID No. from 1 to each pig, for example (1, 2, 3...), when detecting and tracking the subsequent frames, due to the movement of the pig, the ID of a pig may be identified incorrectly, at this time to identify this pig as a new pig, the ID No. of the pig will change to the wrong ID No. until all video frames are processed. Same below.

图 5　猪只白天稀疏和稠密分布情况的可视化分析结果

Fig.5　Results of the visualization analysis of the sparse and dense distribution of pigs during the day

对于猪只白天稀疏的视频 2，本文算法可以准确地检测和跟踪每一只猪，如图 5a。但是，对猪只白天稠密且猪只粘连遮挡情况较为严重的视频 4 存在漏检，如图 5b 中箭头标识的猪。这说明在猪只白天稠密的环境下，由于猪只目标出现漏检，从而影响了算法的跟踪性能。

对猪只白天和夜晚情况下的可视化分析如图 6 所示，可以发现，在猪只白天稠密且有遮挡的情况下，本文 JDE 模型可以很好地跟踪到每一只猪，如图 6a。在夜晚视频背景比较黑暗且猪只密集有遮挡的情况下，JDE 模型也可以准确地跟踪每一只猪，如图 6b。但在猪只夜晚稀疏的视频 5 中，由于所有猪只都分布于猪圈的左方，且视频背景和猪只颜色相似，这使得检测器和跟踪器较难检测和跟踪这些猪只目标，出现猪只漏检的情况，如图 6c 所示。总体上，本文 JDE 模型对于不同场景下的群养生猪多目标跟踪达到较好水平。

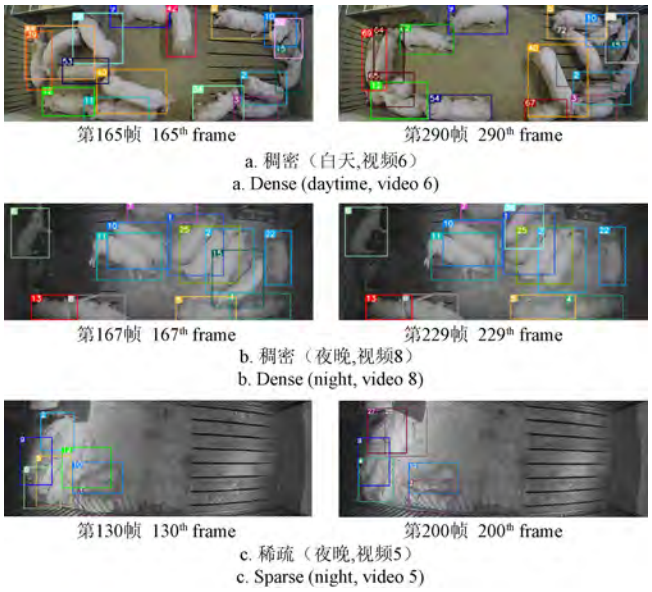


图 6　猪只白天和夜晚不同分布情况的的可视化分析结果

Fig.6　Results of the visualization analysis of the different distribution of pigs during the day and night

3.2　SDE 模型试验结果

为验证本文 JDE 模型的多目标跟踪性能，与经典的 SDE 模型进行对比试验。SDE 检测器与本文 JDE 模型相同，跟踪器使用 DeepSORT，采用相同的公开数据集进行训练和测试，试验结果如表 5 所示。可以发现，SDE 模型的 MOTA 和 IDF1 平均值分别为 81.6%和 78.2%，对比表 4，本文 JDE 模型的 MOTA 提升了 0.5 个百分点。从总体性能指标来看，本文 JDE 模型的 MT、PT、ML、FN、MOTA 和 FPS 指标均优于 SDE 模型。在速度方面，SDE 模型的 FPS 均值为 16.88 帧/s，本文 JDE 模型的 FPS 均值达到 74.26 帧/s。总体来说，二者在跟踪准确度和跟踪精度接近情况下，本文 JDE 模型的视频处理速度比 SDE 模型提升了 340%，这对于实现养殖场长时间群养生猪视频的实时多目标跟踪有重要意义。

表 5 SDE 模型的多目标跟踪试验结果

Table 5 Multiple object tracking experiment results of the Separate Detection and Embedding (SDE) model

视频 Video	猪只个数 Pig numbers	主要跟踪到的目标 Mostly tracked target	部分跟踪到的目标 Partially tracked target	主要丢失目标 Mostly lost target	假阳性 False positive	假阴性 False negative	ID 跳变 ID switch	碎片数 Fragmentation	多目标跟踪精度 Multiple object tracking accuracy/%	IDF1 得分 IDF1 score/%	FPS/(帧·s^{-1})
2	7	6	1	0	128	138	8	13	87.0	85.1	21.18
4	15	15	0	0	173	215	33	30	90.6	81.3	17.00
5	8	3	4	1	216	824	19	39	55.9	60.4	15.28
6	16	13	3	0	22	636	32	53	85.6	74.2	16.16
8	13	13	0	0	430	74	9	10	86.8	87.1	16.52
12	15	10	3	2	45	1 075	8	29	74.9	76.0	15.11

选取部分数据集进行可视化分析，结果如图 7 所示，在猪只夜晚稠密的视频 8 中，SDE 模型存在错检情况，如图 7b 左下角第二头猪出现 2 个跟踪框，而本文 JDE 模型没有错检情况，如图 7a 所示。在猪只白天稠密的视频 12 中，由于猪只密集躺在一起，检测器较容易发生漏检，如图 7a、7b，JDE 模型漏检 2 头猪，SDE 模型漏检 3 头猪，JDE 比 SDE 模型具有更好的检测跟踪结果。

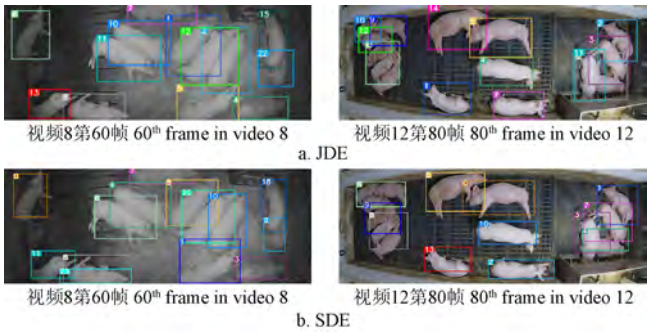


图 7 JDE 与 SDE 模型对猪只不同分布情况的可视化结果对比

Fig.7 Comparison of visualization results of JDE and SDE models for different distribution of pigs

此外，文献[40]采用基于 SDE 模型对猪只目标检测的平均精度均值达 99.0%，多目标跟踪精度 MOTA 为 96.8%，但文献[40]的数据场景单一，无法应对其他场景。尽管包括白天和黑夜（光照变化），但训练和测试场景相同。本文数据集包含不同情况下的场景，共有 11 个视频场景，各个场景环境不同，猪只大小也不同，训练和测试场景完全不相同。

3.3 TransTrack 试验结果

为进一步验证本文算法在群养猪多目标跟踪方面的性能，与 TransTrack 模型在相同的公开数据集上进行对比试验，试验结果如表 6 所示。TransTrack 模型的平均 MOTA、IDF1 和 FPS 分别为 71.7%、71.1%和 17.53 帧/s，与表 4 结果比较发现，本文 JDE 模型比 TransTrack 模型的 MOTA 和 IDF1 分别提升 10.4 和 6.6 个百分点，同时 FPS 提升 324%。从性能指标 MT、PT、ML、FP、FN、IDS、FM、MOTA、IDF1 和 FPS 的数值对比可以发现，本文 JDE 模型性能均优于 TransTrack 模型。

表 6 TransTrack 模型的试验结果

Table 6 Experimental results of the TransTrack model

视频 Video	猪只个数 Pig numbers	主要跟踪到的目标 Mostly tracked target	部分跟踪到的目标 Partially tracked target	主要丢失目标 Mostly lost target	假阳性 False positive	假阴性 False negative	ID 跳变 ID switch	碎片数 Fragmentation	多目标跟踪精度 Multiple object tracking accuracy/%	IDF1 得分 IDF1 score/%	FPS/(帧·s^{-1})
2	7	6	1	0	98	130	19	20	88.2	78.9	17.29
4	15	14	1	0	487	337	33	67	81.0	84.1	17.64
5	8	4	3	1	109	621	31	54	68.3	73.9	17.66
6	16	7	9	0	516	1 370	94	166	58.8	46.6	17.83
8	13	9	3	1	507	762	41	64	66.4	69.9	17.53
12	15	11	2	2	546	909	16	33	67.3	73.4	17.21

对 2 种模型的跟踪结果选取部分数据进行可视化分析，结果如图 8 所示。对比发现，相较于 TransTrack 模型，JDE 模型对猪只严重遮挡情况有更好的检测和跟踪能力，如图 8a。而 TransTrack 模型在猪只严重遮挡情况下，会出现猪只的漏检或者是猪只追踪的缺失，如图 8b。可以看出，本文算法在不同场景中，检测框更加贴合猪只目标，对于严重遮挡的猪只目标具有更强的检测跟踪能力。

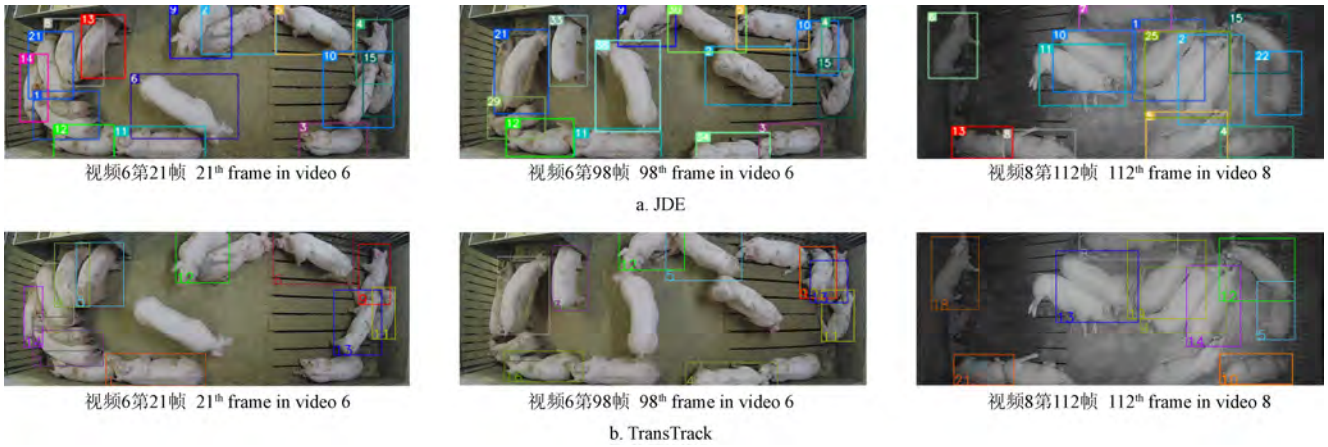


图 8　JDE 与 TransTrack 模型的可视化结果对比

Fig.8　Visualization comparison of JDE model and TransTrack model

4　结　论

1）本文 JDE 模型在二阶段目标检测和跟踪分离框架的基础上进行改进，在输出检测框的同时，给网络增加目标外观信息学习损失对应的输出分支，实现检测和跟踪的多任务协同学习，实现联合目标检测和跟踪。

2）本文制作了 2 个数据集，分别为公开数据集和自建数据集。其数据场景复杂多样，各个场景的猪只大小、数量、日龄和光照条件都不同，并在公开数据集中与 SDE 模型和 TransTrack 模型进行了对比。

3）试验结果表明，本文 JDE 模型在 2 个数据集的总体平均精度均值 mAP 为 92.9%，平均多目标跟踪精度 MOTA 为 83.9%，平均 IDF1 得分为 79.6%，平均每秒检测帧数 FPS 为 73.9。在公开数据集中与 TransTrack 模型进行对比，本文 JDE 模型的 MOTA 和 IDF1 分别提升 10.4 和 6.6 个百分点，FPS 提升 324%。在公开数据集中与 SDE 模型进行对比，本文 JDE 模型在 MOTA 和 IDF1 的数值接近下，FPS 提升 340%，解决了 SDE 模型目标检测和跟踪模块分离导致目标跟踪速度慢的问题，这对于养殖场群养生猪长时间视频的实时多目标跟踪具有重要意义。

[参　考　文　献]

[1] Rowe E, Dawkins M S, Gebhardt-Henrich S G A. Systematic review of precision livestock farming in the poultry sector: Is Technology focussed on improving bird welfare?[J]. Animals (Basel), 2019, 9(9): 614.

[2] Cowton J, Kyriazakis I, Plotz T , et al. A combined deep learning GRU-autoencoder for the early detection of respiratory disease in pigs using multiple environmental sensors[J]. Sensors (Basel), 2018, 18(8): 2521.

[3] Sébastien F, Alain N R, Benoit L. Rethinking environment control strategy of confined animal housing systems through precision livestock farming[J]. Biosystems Engineering, 2017, 155: 96-123.

[4] Zambelis A, Wolfe T, Vasseur E. Technical note: Validation of an ear-tag accelerometer to identify feeding and activity behaviors of tiestall-housed dairy cattle[J]. Journal of Dairy Science, 2019, 102(5): 4536-4540.

[5] Giovanetti V, Decandia M, Molle G, et al. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer[J]. Livestock Science, 2017, 196: 42-48.

[6] Krista M M, Elizabeth A S, Carlos J B R, et al. Technical note: Validation of an automatic recording system to assess behavioural activity level in sheep (Ovis aries)[J]. Small Ruminant Research, 2015, 127: 92-96.

[7] Chen C, Zhu W X, Ma C H, et al. Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs[J]. Computers and Electronics in Agriculture, 2017, 142: 380-387.

[8] Chen C, Zhu W X, Guo Y Z, et al. A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs[J]. Livestock Science, 2018, 218: 70-78.

[9] Chen C, Zhu W X, Liu D, et al. Detection of aggressive behaviours in pigs using a RealSence depth sensor[J]. Computers and Electronics in Agriculture, 2019, 166: 105003.

[10] Chen C, Zhu W X, Steibel J, et al. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory[J]. Computers and Electronics in Agriculture, 2020, 169: 105166.

[11] Alameer A, Kyriazakis I, Bacardit J. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs[J]. Scientific Reports, 2020, 10(1): 13665.

[12] Lao F, Brown B, Stinn J P, et al. Automatic recognition of lactating sow behaviors through depth image processing[J]. Computers and Electronics in Agriculture, 2016, 125: 56-62.

[13] Zhu W X, Guo Y Z, Jiao P P, et al. Recognition and drinking behaviour analysis of individual pigs based on machine vision[J]. Livestock Science, 2017, 205: 129-136.

[14] Leonard S M, Xin H, Brown-Brandl T M, et al. Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls[J]. Computers and Electronics in Agriculture, 2019, 163: 104866.

[15] Yang A Q, Huang H S, Zheng B, et al. An automatic recognition framework for sow daily behaviours based on

motion and image analyses[J]. Biosystems Engineering, 2020, 192: 56-71.

[16] Zhang Y Q, Cai J H, Xiao D Q, et al. Real-time sow behavior detection based on deep learning[J]. Computers and Electronics in Agriculture, 2019, 163: 104884.

[17] Nasirahmadi A, Hensel O, Edwards S, et al. Automatic detection of mounting behaviours among pigs using image analysis[J]. Computers and Electronics in Agriculture, 2016, 124: 295-302.

[18] Li D, Chen Y F, Zhang K F, et al. Mounting beaviour recognition for pigs based on deep learning[J]. Sensors (Basel), 2019, 19(22): 4924.

[19] Nasirahmadi A, Sturm B, Olsson A, et al. Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and support vector machine[J]. Computers and Electronics in Agriculture, 2019, 156: 475-481.

[20] Zheng C, Zhu X M, Yang X F, et al. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. Computers and Electronics in Agriculture, 2018, 147: 51-63.

[21] Zhu X M, Chen C X, Zheng B, et al. Automatic recognition of lactating sow postures by refined two-stream RGB-D faster R-CNN[J]. Biosystems Engineering, 2020, 189: 116-132.

[22] Zheng C, Yang X F, Zhu X M, et al. Automatic posture change analysis of lactating sows by action localisation and tube optimisation from untrimmed depth videos[J]. Biosystems Engineering, 2020, 194: 227-250.

[23] Jorquera-Chavez M, Fuentes S, Dunshea F R, et al. Remotely sensed imagery for early detection of respiratory disease in pigs: A pilot study[J]. Animals (Basel), 2020, 10(3): 451.

[24] Jorquera-Chavez M, Fuentes S, Dunshea F R, et al. Using imagery and computer vision as remote monitoring methods for early detection of respiratory disease in pigs[J]. Computers and Electronics in Agriculture, 2021, 187: 106283.

[25] Zhao K X, He D J. Target detection method for moving cows based on background subtraction[J]. International Journal of Agricultural and Biological Engineering, 2015, 8(1): 42-49.

[26] Zhang Y G, Zheng J, Zhang C, et al. An effective motion object detection method using optical flow estimation under a moving camera[J]. Journal of Visual Communication and Image Representation, 2018, 55: 215-228.

[27] 于欣，侯晓娇，卢焕达，等. 基于光流法与特征统计的鱼群异常行为检测[J]. 农业工程学报，2014，30(2)：162-168.
Yu Xin, Hou Xiaojiao, Lu Huanda, et al. Anomaly detection of fish school behavior based on features statistical and optical flow methods[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(2): 162-168. (in Chinese with English abstract)

[28] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Columbus, OH, USA, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 580-587.

[29] Girshick R. Fast R-CNN[C]// Santiago, Chile, IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.

[30] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[31] 王浩，曾雅琼，裴宏亮，等. 改进 Faster R-CNN 的群养猪只圈内位置识别与应用[J]. 农业工程学报，2020，36(21)：201-209.
Wang Hao, Zeng Yaqiong, Pei Hongliang, et al. Recognition and application of pigs'position in group pens based on improved Faster R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(21): 201-209. (in Chinese with English abstract)

[32] Redmon J, Farhadia A. YOLOv3: An incremental improvement [EB/OL]. 2018-04-08, https://pjreddie.com/media/files/papers/YOLOv3.pdf.

[33] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Las Vegas, NV, USA, Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.

[34] Redmon J, Farhadi A. YOLO9000: Better,faster,stronger[C]// Honolulu, HI, USA, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 7263-7271.

[35] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. 2020-04-23, https://arxiv.org/pdf/2004.10934.pdf.

[36] 金耀，何秀文，万世主，等. 基于 YOLO v3 的生猪个体识别方法[J]. 中国农机化学报，2021，42(2)：178-183.
Jin Yao, He Xiuwen, Wan Shizhu, et al.Individual pig identification method based on YOLOv3[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(2): 178-183. (in Chinese with English abstract)

[37] Bewley A, Ge Z Y, Ott L, et al. Simple online and realtime tracking[C]//Phoenix, Arizona, USA. IEEE International Conference on Image Processing (ICIP), 2016: 3464-3468.

[38] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//Beijing, China. IEEE International Conference on Image Processing (ICIP), 2017: 3645-3649.

[39] 张宏鸣，汪润，董佩杰，等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. 农业机械学报，2021，52(4)：249-256.
Zhang Hongming, Wang Run, Dong Peijie, et al. Multi-object tracking method for beef cattle based on DeepSORT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 249-256. (in Chinese with English abstract)

[40] 张伟，沈明霞，刘龙申，等. 基于 CenterNet 搭配优化 DeepSORT 算法的断奶仔猪目标跟踪方法研究[J]. 南京农业大学学报，2021，44(5)：973-981.
Zhang Wei, Shen Mingxia, Liu Longshen, et al. Research on weaned piglet target tracking method based on CenterNet collocation optimized DeepSORT algorithm[J]. Journal of Nanjing Agricultural University, 2021, 44(5): 973-981. (in Chinese with English abstract)

[41] Sun P Z, Cao J K, Jiang Y, et al. TransTrack: Multiple object tracking with transformer[EB/OL]. 2021-05-04, https://arxiv.org/abs/2012.15460v1.

[42] Psota E T, Schmidt T, Mote B, et al. Long-term tracking of

group-housed livestock using keypoint detection and MAP estimation for individual animal identification[J]. Sensors (Basel), 2020, 20(13): 3670.

[43] Tu S Q，Yuan W J，Liang Y，et al. Automatic detection and segmentation for group-housed pigs based on PigMS R-CNN[J]. Sensors (Basel), 2021, 21(9): 3251.

Multiple object tracking of group-housed pigs based on JDE model

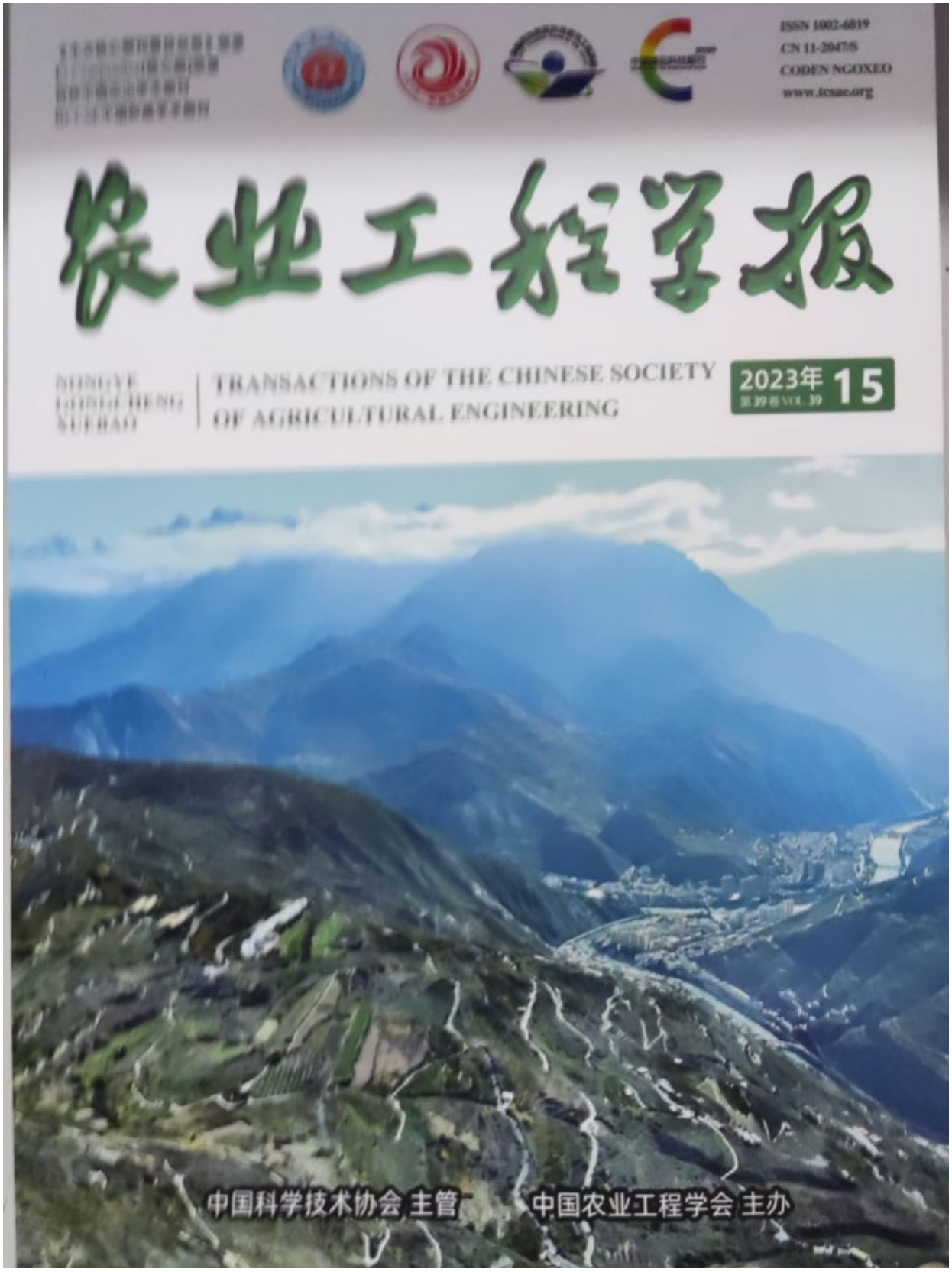
Tu Shuqin, Huang Lei, Liang Yun※, Huang Zhengxin, Li Chengjie, Liu Xiaolong

(*College of Mathematics and Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: Pig production has been always the pillar of the industrial livestock industry in China. Therefore, the pig industry is closely related to food safety, social stability, and the coordinated development of the national economy. An intelligent video surveillance can greatly contribute to the large-scale production of animal husbandry under labor shortage at present. It is very necessary to accurately track and identify the abnormal behavior of group-housed pigs in the breeding scene. Much effort has been focused on Multiple Object Tracking (MOT) for pig detection and tracking. Among them, two parts are included in the Tracking By Detection (TBD) paradigm, e.g., the Separate Detection and Embedding (SDE) model. Previously, the detector has been developed to detect pig objects. And then the tracking models have been selected for the pig tracking using Kalman filter and Hungarian (Sort or DeepSORT). The detection and association steps have been designed to increase the running and training time of the model in the dominant MOT strategy. Thus, real-time tracking cannot fully meet the requirement of the group-housed pigs. In this study, a Joint Detection and Embedding (JDE) model was proposed to automatically detect the pig objects and then track each one in the complex scenes (day or night, sparse or dense). The core of JDE model was to integrate the detector and the embedding model into a single network for the real-time MOT system. Specifically, the JDE model incorporated the appearance model into a single-shot detector. As such, the simultaneous output was performed on the corresponding appearance to improve the runtime and operational efficiency of the model. An overall loss of one multiple task learning loss was utilized in the JDE model. Three loss functions were included classification, box regression and appearance. Three merits were achieved after operations. Firstly, the multiple tasks learning loss was used to realize the object detection and appearance to be learned in a shared model, in order to reduce the amount of occupied memory. Secondly, the forward operation was computed using the multiple tasks loss at one time. The overall inference time was reduced to improve the efficiency of the MOT system. Thirdly, the performance of each prediction head was promoted to share the same set of low-level features and feature pyramid network architecture. Finally, the data association module was utilized to process the outputs of the detection and appearance head from the JDE, in order to produce the position prediction and ID tracking of multiple objects. The JDE model was validated on the special dataset under a variety of settings. The special dataset was also built with a total of 21 video segments and 4 300 images using the dark label video annotation software. Two types of datasets were obtained, where the public dataset contained 11 video sequences and 3 300 images, and the private dataset contained 10 video segments and 1 000 images. The experimental results show that the mean Average Precision (mAP), Multiple Object Tracking Accuracies (MOTA), IDF1 score, and FPS of the JDE on all test videos were 92.9%, 83.9%, 79.6%, and 73.9 frames/s, respectively. A comparison was also made with the SDE model and TransTrack method on the public dataset. The JDE model improved the FPS by 340%, and the MOTA by 0.5 percentage points in the same test dataset, compared with the SDE model. It infers the sufficient real-time performance of MOT using the JDE model. The MOTA, IDF1 metrics, and FPS of the JDE model was improved by 10.4 and 6.6 percentage points, and 324%, respectively, compared with the TransTrack model. The visual tracking demonstrated that the JDE model performed the best detection and tracking ability with the SDE and TransTrack models under the four scenarios, including the dense day, sparse day, dense night, and sparse night. The finding can also provide an effective and accurate detection for the rapid tracking of group-housed pigs in complex farming scenes.

Keywords: object detection; object tracking; joint detection and tracking; data association; group-housed pigs

2.11 改进TransTrack多目标生猪行为跟踪方法



农 业 工 程 学 报

2023 年 8 月第 15 期　（总第 464 期）　第 39 卷

目　次

期刊基本参数：CN 11-2047/S*1985*s*16*304*zh*P* ¥ 80.00*280*31*2023-08

✦✶✦✦✶✦✶✦✦✶✦✦✶✦✦✶✦✦✶✦✦✶✦✦✶

致谢：感谢中国科学院水利部成都山地灾害与环境研究所张继飞团队提供的封面图片“汶川县县城景观”

《农业工程学报》：对标一流，追求卓越

《农业工程学报》（以下简称《学报》）创刊于 1985 年，现为半月刊，全年 24 期，大 16 开面向国内外公开发行。《学报》是由中国科学技术协会主管、中国农业工程学会主办的全国性专业学术期刊。读者对象为农业工程学科及相关领域的科研、教学及生产科技人员、技术管理及推广人员和高等院校师生。刊稿内容涵盖了农业装备工程与机械化、农业航空工程、农业水土工程、农业信息与电气技术、农业生物环境与能源工程、土地保障与生态安全、农产品加工工程等学科专业领域。

《学报》始终坚持“双为”方向和“双百”方针的办刊宗旨及“内容为王，质量为本”的办刊理念。拥有国内外农业工程相关领域各专业的知名专家学者组成的编委会，其中两院院士 30 余人。本刊坚持专家办刊，编委在稿件同行评审把关中发挥重要作用。

《学报》是中国农业工程领域的领军期刊，在行业具有很高的学术影响力，被 EI Compendex（核心版）、Scopus、CA、CSA、CAB Abstracts、CSCD、《中文核心期刊要目总览》、《中国科技核心期刊目录》、《中国农林核心期刊概览 2020》等国内外多个权威数据库收录。中国科学技术信息研究所 2022 年最新影响因子 2.274。

我们愿与广大农业工程同仁携手并肩，共同奋斗，对标一流，追求卓越，创建国际知名品牌，引领学科发展，培养人才，激励创新，为全面推进乡村振兴，加快农业农村现代化，促进农业农村高质量发展作出新贡献！

期刊荣誉：多项位列农业工程类期刊榜首

- ✧ 国家新闻出版广电总局“双效期刊”
- ✧ 国家新闻出版广电总局“百强报刊”
- ✧ 中国科技期刊卓越行动计划-梯队期刊
- ✧ 百种中国杰出学术期刊
- ✧ 中国精品科技期刊
- ✧ TOP5%中国最具国际影响力学术期刊
- ✧ RCCSE 中国权威学术期刊（A+）
- ✧ 科技期刊数字影响力 100 强
- ✧ 世界学术影响力指数 WAJCI-Q1 区期刊（为本学科唯一入选中文刊）
- ✧ Google Scholar 学术期刊影响力排名位列高被引中文刊第四名
- ✧ 中国农林领域高质量科技期刊分级目录第一区（T1）
- ✧ 中国科协精品科技期刊工程项目资助期刊

對標一流
追求卓越

ISSN 1002-6819
CN 11-2047/S
国内邮发代号：18-57
海外发行代号：BM3335
国内定价：80元

ISSN 1002-6819

改进 TransTrack 多目标生猪行为跟踪方法

涂淑琴，黄正鑫，梁　云※，黄　磊，刘晓龙

（华南农业大学数学与信息学院，广州 510642）

摘　要：高效准确地监测群养生猪的行为变化以获取其生理、健康和福利状况，对于实现生猪智能精细化养殖具有重要意义。针对猪场自然场景下光照变化和猪只粘连遮挡等因素影响，使得猪只行为跟踪中存在误检、漏检和身份频繁错误变换问题，该研究提出一种改进的 TransTrack 多目标生猪行为跟踪方法。首先，在目标检测模块中，采用改进的并集交并比的匹配算法，去除猪只遮挡导致的目标误检检测框。然后，在跟踪模块中，根据高低匹配阈值进行 2 次数据关联，提高光照变化下漏检目标的跟踪准确性。最后，针对误检与漏检导致跟踪中猪只身份错误变换，根据猪栏中猪只数量信息，限制猪只身份编号值的错误增加，提高猪只身份准确识别率。在公开数据集和私有数据集上的试验结果表明，改进的 TransTrack 在多目标跟踪准确率（multiple object tracking accuracy，MOTA），高阶跟踪准确率（higher order tracking accuracy，HOTA）和身份变换（identity switches，IDs）分别为 92.0%、69.8% 和 210。在公开数据集中，对比 Trackformer，JDE 和 TransTrack 模型，改进的 TransTrack 方法在 MOTA 分别提高 3.9，9.0 和 13.1 个百分点，HOTA 分别提高 1.3，9.5 和 8.3 个百分点，IDs 分别降低 136，326 和 376。在私有数据集中，对比 Trackformer 和 TransTrack 模型，改进的 TransTrack 方法在 MOTA 分别提高 14.4 和 15.8 个百分点，HOTA 分别提高 1.8 和 9.5 个百分点。结果显示，改进的 TransTrack 方法能够更加稳定地实现对群养生猪的行为跟踪，为群养生猪行为识别与智能分析提供技术支持。

关键词：识别；多目标跟踪；生猪；TransTrack；数据关联

doi：10.11975/j.issn.1002-6819.202303189

中图分类号：TP391.4　　**文献标志码：**A　　**文章编号：**1002-6819(2023)-15-0172-09

涂淑琴，黄正鑫，梁云，等. 改进 TransTrack 多目标生猪行为跟踪方法[J]. 农业工程学报，2023，39(15)：172-180. doi: 10.11975/j.issn.1002-6819.202303189　http://www.tcsae.org

TU Shuqin, HUANG Zhengxin, LIANG Yun, et al. Methods for multi-target tracking of pig action using improved TransTrack[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(15): 172-180. (in Chinese with English abstract)　doi: 10.11975/j.issn.1002-6819.202303189　http://www.tcsae.org

0 引　言

生猪产业作为中国经济的重要组成部分，其降本增效对经济发展具有重要意义[1]。为了在大规模生猪养殖中降低成本、提高效益，数字化信息技术被广泛应用，并成为未来发展的方向。然而，在目前的生猪养殖中，群养生猪的行为监测主要依赖于人工，这导致管理成本高、存在人畜交叉感染的风险[2-3]。为应对这些挑战，深度学习技术已被应用于生猪行为监测中，实现非接触式和低应激的健康监测，提高猪只福利和经济效益[4-5]。

在猪只行为识别领域，基于深度学习的猪只行为与姿态识别研究已取得许多成果[6-10]。例如，在猪只行为识别方面，ZHENG 等[6] 研究猪只站立、坐、趴卧和侧卧识别，CHEN 等[7] 研究生猪的攻击性行为。SHAO 等[8] 研究生猪站立、俯卧、侧卧和探索行为。在姿态识别方面，RIEKERT 等[9] 研究生猪姿态检测，平均精度为 80.20%。王鲁等[10] 研究母猪 4 种姿态识别，包括跪立、站立、坐和躺卧姿态，取得平均精度为 98.20% 的性能。以上研究在特定条件下达到较高精度，猪只行为与姿态识别取得一定进展，但缺乏对不同光照场景下的充分试验，在不同光照场景及猪只粘连密集遮挡下，准确识别猪只姿态行为仍存在一定挑战，且上述研究仅对图像中猪只行为进行静态识别，无法实现对视频中每头猪只的行为进行动态的自动跟踪。

多目标跟踪 (multiple object tracking，MOT) 算法在给定视频中对多个感兴趣的目标进行定位，维持个体 ID 号，跟踪每个个体的运动轨迹[11]。常用的 MOT 算法包括 SORT[12]、DeepSORT[13]、JDE[14]、ByteTrack[15]、TrackFormer[16] 和 TransTrack[17] 等。近年来，MOT 算法也引入猪只视频监控应用中。例如，张伟等[18] 利用 CenterNet+DeepSORT 模型实现了断奶仔猪的检测和多目标跟踪，改善了猪只外观相似和遮挡情况下的跟踪效果。文献 [19-20] 提出基于 YOLOv5s 及 YOLOX-S 的改进 DeepSORT 算法，改善了猪只重叠与遮挡造成的猪只身份编号（ID）频繁跳变。文献 [21] 利用一种 JDE 群养生猪 MOT 方法，实现了实时的跟踪效果。文献 [22] 提出改进 ByteTrack 多目标跟踪算法用于群养生猪行为跟踪，该方法提高了跟踪的性能和效率。虽然上述研究在猪只目标跟踪领域取得了一定的进展，然而这些方法中的检测器、跟踪器和外观模型是分开训练，在目标检测、跟踪和重识别任务之间缺乏有效的信息交互利用，在复

收稿日期：2023-03-27　修订日期：2023-06-30

基金项目：广州市重点研发计划（202206010091）；广州市重点研发计划项目（2023B03J1363）

作者简介：涂淑琴，博士，讲师，研究方向为图像处理与计算机视觉。Email：tushuqin@163.com

※通信作者：梁云，博士，教授，研究方向为图像处理与计算机视觉。Email：yliang@scau.edu.cn

杂场景下，容易出现猪只漏检、错检和 ID 频繁跳变等问题。另外，上述多数研究仅完成了生猪目标跟踪，未将生猪行为信息与跟踪相融合，因此，需要进一步将猪只行为与跟踪任务进行融合，以实现对每头生猪行为的准确自动跟踪。

针对上述问题，本文提出一种改进的 TransTrack 多目标生猪行为跟踪方法。首先，该方法采用 TransFormer 网络实现目标检测、跟踪和重识别的信息交互，同时，在目标检测模块中采用改进的并集交并比匹配算法，消除猪只遮挡导致的错误检测框；然后，在跟踪模块中，引入不同的匹配阈值，进行 2 次数据关联，以确保在不同光照条件下准确跟踪漏检目标。为解决误检和漏检引起的猪只身份错误变换问题，利用猪栏猪只真实头数的先验信息，更正猪只身份错误变换，实现准确的猪只身份识别；最后，在跟踪轨迹中引入猪只行为信息，构建生猪多目标行为跟踪模型，为实现猪只行为分析提供可靠的技术支撑。

1　试验数据

本研究使用公开数据集和私有数据集进行试验。其中，公开数据集来自文献 [23]，包含 23 个猪舍监控视频段，每个猪舍中的猪只数量不同，范围为 7～16 头，涵盖白天和夜晚的不同时间段，共包含图像 6 900 张。其中，12 个视频段用于模型训练和验证，其余 11 个视频段用于测试。私有数据集为 8 个视频段，包含 2 400 张图像，在佛山商业猪场进行拍摄，包括 08:00－10:00 和 15:00－16:00 的晴天和阴天，每个猪舍有 6～11 头猪只，其中 4 个用于训练，4 个用于测试。每个视频段的时长为 1 min，帧率为 5 帧/s，每个视频都包含 300 张图像。这 2 个数据集中都包含具有不同日龄、大小和数量的猪只视频。根据时间段的不同将猪只的活动水平分为 3 类：白天的高活动、白天（或夜晚）的中等活动、白天（或夜晚）的低活动。其中，猪只活动水平[21] 定义如下：根据视频的人工观察结果，在白天（10:00－12:30）猪只的饮食和玩耍等行为较频繁，此时间段定义为猪只白天的高活动水平。在白天（12:30－17:00）或夜晚（17:00－20:00）猪只的饮食和玩耍等行为没有白天（10:00－12:30）高，此时间段定义为白天或夜晚的中等活动水平。在白天（07:00－10:00）或夜晚（20:00－07:00）猪只的饮食和玩耍等行为较少，躺卧行为较多，此时间段定义为白天或夜晚的低活动水平。详细的测试视频如表 1 所示。

表 1　测试数据集

Table 1　Test dataset

数据集 Dataset	视频编号 Vedio number	白天 Day	黑夜 Night	活动水平 Activity level	猪只个数 Number of pigs
公开数据集 Public dataset	0102	√	–	高	7
	0402	√	–	中	15
	0502	–	√	中	8
	0602	√	–	高	16
	0702	√	–	中	12
	0802	–	√	低	13
	0902	√	–	中	14
	1 002	–	√	中	14
	1 102	√	–	高	16
	1 202	√	–	低	15
	1 502	–	√	中	16
私有数据集 Private dataset	1 602	√	–	低	11
	1 604	√	–	中	11
	1 701	√	–	高	6
	1 703	√	–	低	6

部分数据样本如图 1 所示，其中图 1a 是公开数据集的示例，图 1b 是私有数据集的示例。在生猪各个生长阶段中, 躺卧、饮食和站立是猪只行为研究的基础需求，对猪只的这些行为进行追踪，能更好地了解猪只生长过程的心理和生理状态，满足猪场管理实际需求。因此，本研究将生猪的行为分为 4 个类别，分别是躺卧、饮食、站立和其他。

视频0102 Video 0102　　视频0702 Video 0702　　视频1502 Video 1502

a. 公开数据集
a. Public dataset

视频1602 Video 1602　　视频1604 Video 1604　　视频1701 Video 1701

b. 私有数据集
b. Private dataset

图 1　部分生猪图像

Fig.1　Part of group-housed pig images

2　改进的 TransTrack 多目标生猪行为跟踪方法

针对 TransTrack 方法在猪场环境下，由于遮挡和重复检测导致跟踪性能下降的问题，提出改进的 TransTrack 方法，其工作流程如图 2 所示。该方法分为特征提取和检测与跟踪模块两部分，首先，将视频输入到特征提取模块，利用基于残差网络的骨干网络提取连续 2 帧图像的特征（F_t，F_{t-1}），将其组合为全局特征图后，融合位置嵌入信息输入 TransFormer 编码器；然

后，将编码之后的连续 2 帧图像特征（F_t编码，F_{t-1}编码）输入到 2 个 TransFormer 解码器网络，分别输入前馈网络，形成目标检测网络和跟踪网络，产生目标检测框和跟踪框；最后，将检测框和跟踪框进行并集交并比（complete intersection over union, CIOU）数据关联匹配，输出多目标跟踪图像序列结果。

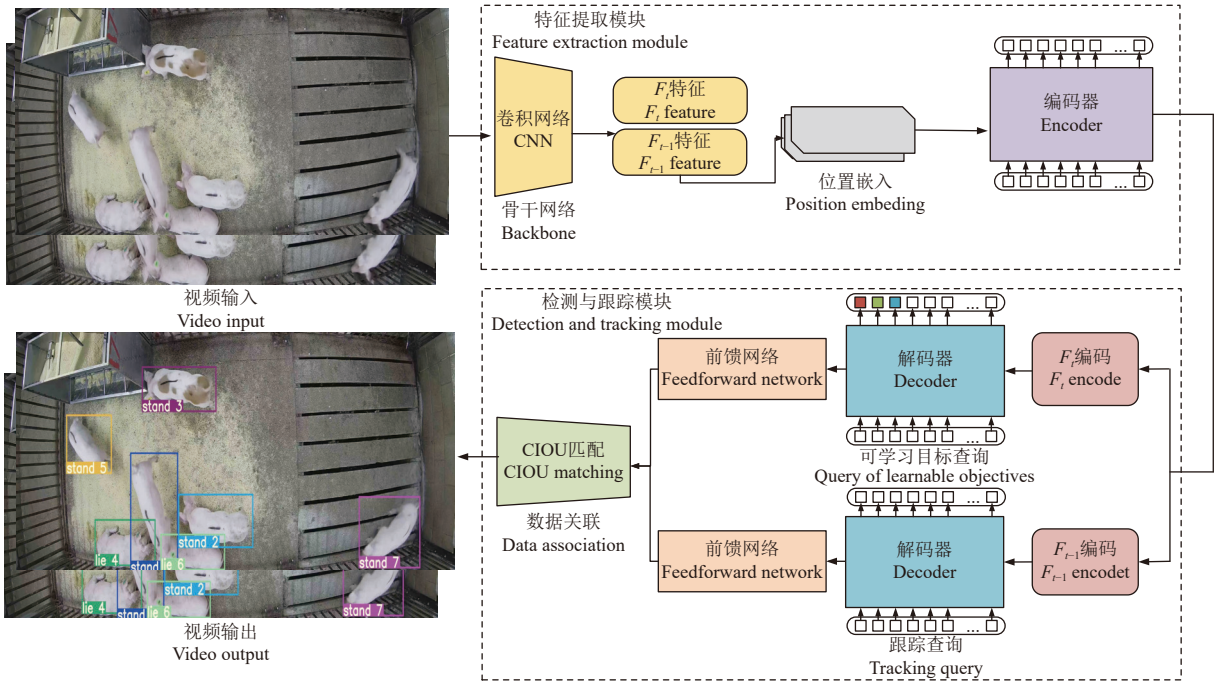


图 2 改进的 TransTrack 多目标生猪行为跟踪方法

Fig.2 Behavior tracking flow chart of improved TransTrack on group-housed pigs

2.1 TransFormer 编码器和解码器

TransTrack 的关键模块为 TransFormer 编码器和解码器。TransFormer 编码器结构如图 3a 所示，首先对输入的连续 2 帧图像特征进行自注意力处理，然后通过前馈网络和 2 次残差连接正则化处理，对特征进行增强处理，输出增强的特征，分别为F_t和F_{t-1}编码。TransFormer 解码器结构如图 3b 所示，以前一帧的目标特征和当前帧的可学习目标作为跟踪和目标查询，将目标查询和跟踪查询通过自注意力学习和 1 次残差连接正则化处理，分别与编码器输出的图像特征（F_t编码，F_{t-1}编码）进行交叉注意力，经过 2 次残差连接正则化和前馈网络处理，产生增强的目标特征和跟踪特征，获得跟踪目标和检测结果。

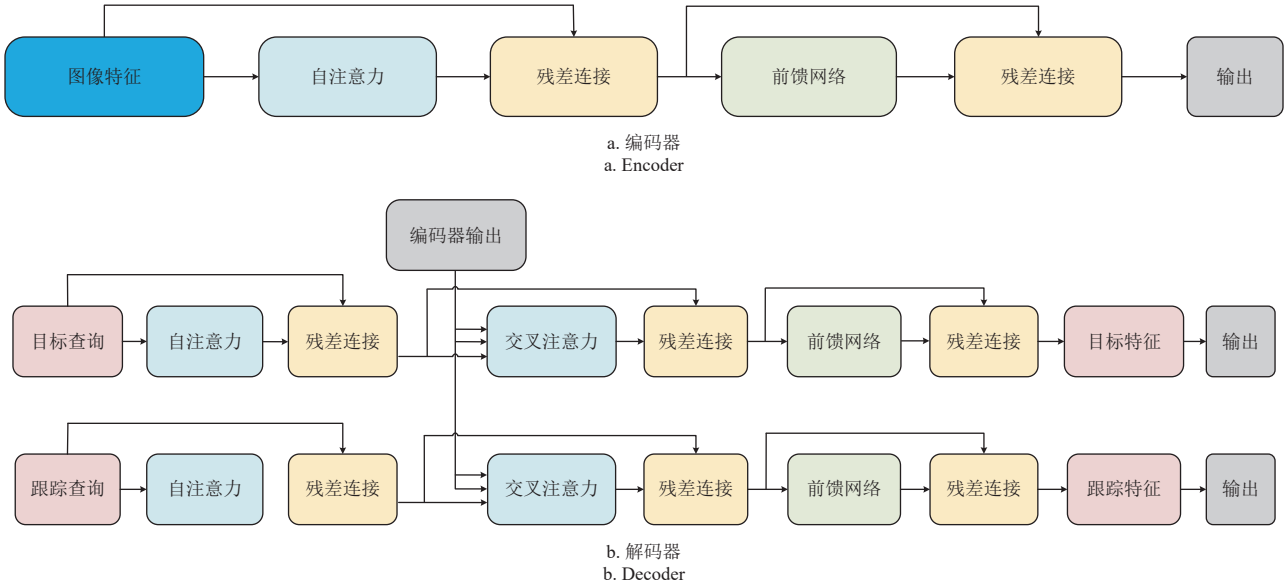


图 3 TransFormer 编码器和解码器

Fig.3 TransFormer encoder and decoder

TransFormer 编码器和解码器采用相同的训练损失，损失函数γ计算式为：

$$\gamma = \lambda_{\text{cls}} \cdot \gamma_{\text{cls}} + \lambda_{L1} \cdot \gamma_{L1} + \lambda_{\text{GIOU}} \cdot \gamma_{\text{GIOU}} \quad (1)$$

式中γ_{cls}为预测的类别和真实类别之间的 Focal loss[24]，

γ_{L1}为归一化后的预测框中心点坐标与真实坐标之间的 L_1 损失，γ_{GIOU}为归一化后的预测框宽度和高度与真实框宽度和高度之间的广义交并比（generalized intersection over union, GIOU）损失[25]，λ_{cls}，λ_{L1}和λ_{GIOU}都是固定系数。

2.2　数据关联方法

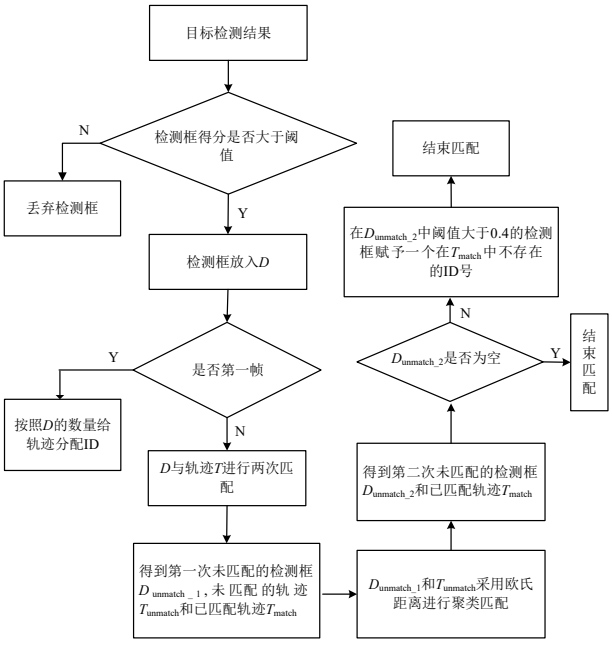
TransTrack 在 MOT 任务取得优秀的性能，然而，TransTrack 在数据关联匹配中采用 GIOU，无法应对遮挡和重复检测等问题造成的目标跟踪性能下降的挑战。因此，本文在基本的 TransTrack 方法基础上，对检测和跟踪模块两部分进行有效改进。

1）检测模块的重叠框修正和改进数据关联算法

首先，在目标检测中计算检测框之间的重叠程度，根据设定阈值去除重叠的多余检测框；然后，根据检测置信度将无重叠的检测框分为高分和低分的目标检测框。同时，将 4 种猪只行为类别信息引入目标分类中。最后，在轨迹跟踪中，使用 CIOU 分别对高分和低分目标检测框进行 2 次数据关联，并在轨迹框中增加 4 种行为类别参数，实现猪只行为跟踪。

2）跟踪模块的 ID 修正处理

在群养生猪自然场景下，由于目标误检或者密集遮挡等原因，TransTrack 存在猪只频繁的 ID 错误变换。针对该挑战，本文根据养猪场猪栏中猪只数量在一段时间内固定的特点，在对群养生猪进行跟踪时，通过优化轨迹 ID 的赋值方法，限制 ID 的增长，减少猪只 ID 频繁错误切换。对轨迹 ID 的修正处理流程如图 4 所示。



注：D 为置信度得分大于阈值的检测框，T 为轨迹，$D_{unmatch\_1}$ 为第一次未匹配的检测框，$T_{unmatch}$ 为未匹配的轨迹，T_{match} 为已匹配轨迹，$D_{unmatch\_2}$ 为第二次未匹配的检测框

Note: D is the detection box with confidence score greater than the threshold, T is the track, $D_{unmatch\_1}$ is the first unmatched detection box, $T_{unmatch}$ is the unmatched track, T_{match} is the matched track, and $D_{unmatch\_2}$ is the second unmatched detection box

图 4　跟踪模块的 ID 处理

Fig.4　ID processing for tracking modules

由图 4 可知，跟踪模块的 ID 处理过程中，将目标检测结果中，置信度得分大于阈值的检测框放入 D 中。若为第一帧，创建一个新的轨迹 $T=\{T_1,\cdots,T_n\}$，按 D 数量n从$\{1,\cdots,n\}$，给轨迹 T 的 ID 赋值。否则，将 D 与 T 进行两次匹配，得到第一次未匹配的检测框 $D_{unmatch\_1}$,未匹配的轨迹 $T_{unmatch}$ 和已匹配轨迹 T_{match}。对 $D_{unmatch\_1}$ 和 $T_{unmatch}$ 采用欧氏距离进行聚类匹配。产生第二次未匹配的检测框 $D_{unmatch\_2}$ 和匹配轨迹 T_{match}。若 $D_{unmatch\_2}$ 为空，说明检测框 D 和轨迹 T 匹配成功；否则，依次给 $D_{unmatch\_2}$ 里每一个阈值大于 0.4 的检测框，从小到大赋予一个在 T_{match} 中不存在的 ID 号，并更新 T_{match}。

2.3　试验平台

本文所有试验在同一台计算机上完成，使用 Linux 作为试验平台，采用 ubuntu20.04 操作系统，硬件配置为 12 th Gen Intel(R) i9-12900KF CPU，NVIDIA GeForce RTX 3 090 GPU，32G 内存，Pytorch 版本 1.11.1，Python 版本 3.8，CUDA 版本 11.3。

训练过程中设置图片尺寸为 1 920×1 080 像素大小，批处理大小设置为 2，初始学习率为 0.000 2，动量设置为 0.9，共训练 40 个迭代轮次，使用随机梯度下降法 (stochastic gradient descent, SGD) 进行优化，高低匹配阈值分别为 0.8 和 0.4，去除重叠框阈值为 0.7，保存训练过程中精度最高的模型参数进行试验测试。

多目标跟踪器的性能，选择高阶跟踪准确率（higher order tracking accuracy，HOTA）、多目标跟踪准确率（multiple object tracking accuracy，MOTA）、识别 F1 值（identification F1 Score，IDF1）和身份变换（ID switches，IDs）等作为主要评价指标[26-27]。HOTA 将精确的检测、关联跟踪和定位进行统一度量，MOTA 衡量跟踪器在检测物体和保持轨迹跟踪的性能。IDF1 和 IDs 为体现跟踪目标标号 ID 的变换值指标，其大小反映轨迹跟踪能力。MOTA 数值越高，检测与跟踪效果越好；IDF1、HOTA 和 IDs 更侧重反映跟踪器性能，IDF1 和 HOTA 数值越高，性能越好，IDs 值越小，性能越好。

3　试验结果与分析

3.1　改进 TransTrack 和其他 MOT 方法对比

在公开和私有数据集中，对比 Trackformer、JDE 和 TransTrack，改进 TransTrack 试验结果如表 2 所示。在公开数据集中，改进 TransTrack 的 MOTA 为 92.4%，HOTA 为 72.1%，IDs 为 147。对比 Trackformer、JDE 和 TransTrack 的 MOTA 分别高 3.9，9.0 和 13.1 个百分点，HOTA 分别高 1.3，9.5 和 8.3 个百分点，IDs 分别降低 136，326 和 376。在私有数据集中，改进 TransTrack 的 MOTA 为 91.5%，HOTA 为 62.5%，IDs 为 63。对比 Trackformer 和 TransTrack 的 MOTA 分别高 14.4 和 15.8 个百分点，HOTA 分别高 1.8 和 9.5 个百分点。

试验结果表明，改进 TransTrack 的跟踪性能优于 Trackformer、JDE 和 TransTrack。原因在于改进 TransTrack 在去除重叠框后，采用 CIOU 进行 2 次数据关联，能有效地避免由于光线变化或猪只遮挡导致的猪只漏检情况，并利用 ID 修正处理算法，提高 MOT 性能。因此，改进 TransTrack 算法的性能指标优于其他算法，表明本研究所提改进算法的有效性，适用于复杂场景下群养生猪的多目标跟踪。

表 2 改进 TransTrack 和其他多目标跟踪方法结果对比

Table 2 Comparison of experimental results between improved TransTrack and other multiple object tracking（MOT） approachs

测试视频段 Test video segment	算法 Algorithm	MOTA/%	IDs	HOTA/%	IDF1 score/%	FP	FN	MOTP
公共数据集 Public dataset	Trackformer	88.5	283	70.8	79.5	1 048	3 719	0.157
	JDE	83.4	473	62.6	71.2	3 323	3 455	0.221
	TransTrack	79.3	523	63.8	71.2	3 627	4 910	0.173
	Improved TransTrack	92.4	147	72.1	86.4	1 000	2 335	0.175
私有数据集 Private dataset	Trackformer	77.1	41	60.7	76.0	1 415	879	0.238
	TransTrack	75.7	223	53.0	65.7	1 258	994	0.221
	Improved TransTrack	91.5	63	62.5	77.3	515	364	0.214

注：MOT 为多目标跟踪，MOTA 为多目标跟踪准确率，IDs 为身份变换，HOTA 为高阶跟踪准确率，IDF1 为识别 F1 值，FP 为假阳性，FN 为假阴性，MOTP 为多目标跟踪精度。下同。

Note: MOT is the multiple object tracking, MOTA is the multiple object tracking accuracy, IDs is identity switches, HOTA is the higher order tracking accuracy, IDF1 is the identification F1 score, FP is the False positive, FN is the False negative, MOTP is the multiple object tracking precision. Same below.

改进 TransTrack 算法与其他 MOT 算法在公开数据集和私有数据集的部分可视化结果如图 5 和图 6 所示。在图 5 中，JDE、Trackformer 和 TransTrack 在视频 0602 第 201 帧中最大的跟踪 ID 分别为 93、39 和 31，与猪栏猪只数量 16 头不同，出现猪只 ID 错误的频繁切换；改进 TransTrack 的最大 ID 稳定在 16，没有出现错误的 ID 变换。

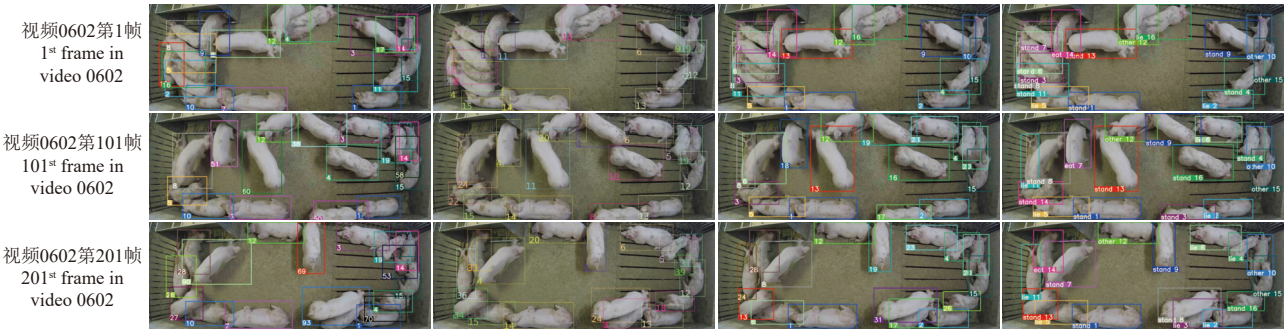


图 5 改进 TransTrack 与其他 MOT 在公开数据集上的跟踪结果对比

Fig.5 Comparison of tracking results between improved TransTrack and the other MOT algorithms in public dataset

在图 6 中，Trackformer 在视频 1 703 第 278 帧中最大的跟踪 ID 达到 12，TransTrack 在视频 1 703 第 199 帧和第 278 帧中最大的跟踪 ID 分别为 12 和 11，出现频繁的错误 IDs。改进 TransTrack 的最大 ID 稳定在 6，和真实的目标数 (ground truth, GT) 一致，保持跟踪 ID 稳定。因此，结合表 2、图 5 和图 6 综合分析，相比于其他 3 种算法，改进 TransTrack 的多目标跟踪指标和跟踪效果均最优，在复杂场景下可以实现准确的生猪行为跟踪。

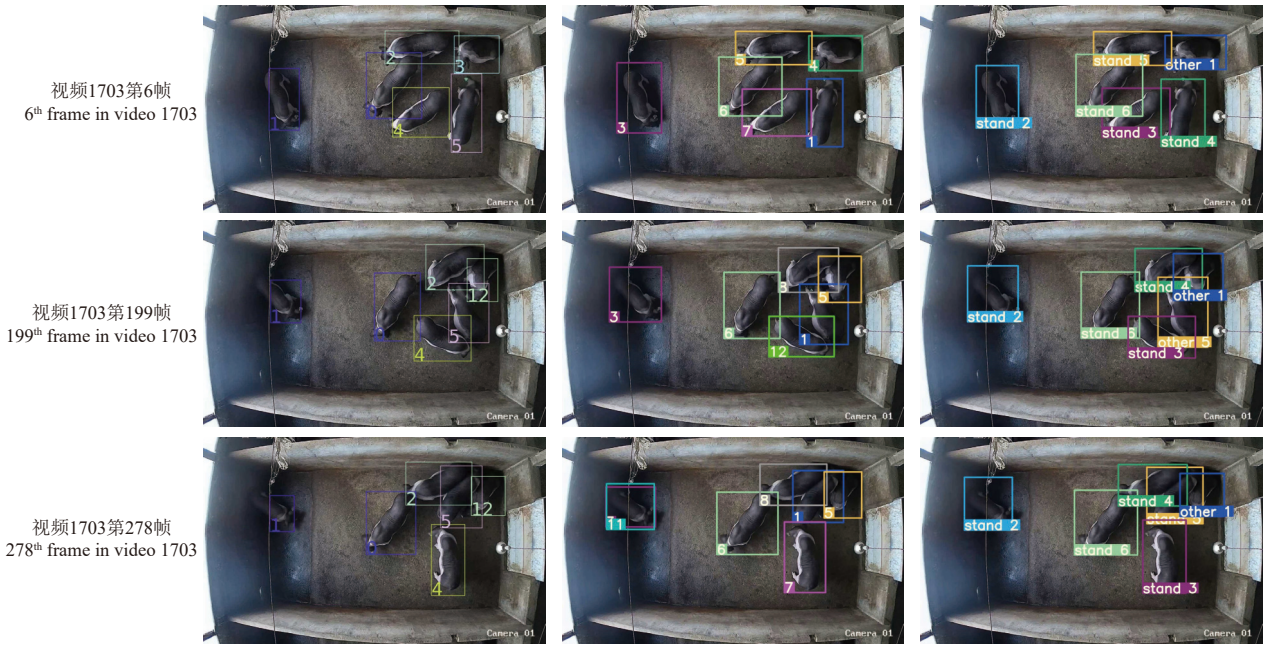


图 6 改进 TransTrack 与其他 MOT 在私有数据集上的跟踪结果对比

Fig.6 Comparison of tracking results between improved TransTrack and the other MOT algorithms in private dataset

3.2　改进 TransTrack 和 TransTrack 在每个测试视频的 MOT 结果与分析

基于 TransTrack 和改进 TransTrack 在每个测试视频的 MOT 结果如表 3 和表 4 所示。TransTrack 在 0502 和 1 701 视频段上 MOTA，HOTA 和 IDF1 分别为 67.9% 和 69.9%，60.0% 和 50.1%，64.4% 和 57.3%，说明跟踪器在这 2 个视频段中，其行为跟踪没有取得很好的效果。TransTrack 只有在 0802 视频中，跟踪性能不错，其 MOTA 取值为 91.0%。在 IDs 上，TransTrack 中每个视频段的 IDs 取值都大于 20 以上，说明 TransTrack 在 ID 具有频繁错误变换，没有实现稳定跟踪。

表 3　TransTrack 的群养猪多目标跟踪结果

Table 3　Pig MOT results based on TransTrack

视频段 Video sequence	MOTA/%	IDs	HOTA/%	IDF1/%	FP	FN	MOTP
0102	74.1	38	43.4	59.0	331	175	0.142
0402	88.7	27	70.3	81.9	239	242	0.147
0502	67.9	38	60.0	64.4	293	439	0.174
0602	73.6	63	56.1	61.3	504	702	0.207
0702	64.8	52	73.8	69.4	777	439	0.162
0802	91.0	40	77.7	82.4	110	202	0.131
0902	83.5	52	71.7	76.0	348	294	0.151
1 002	78.0	70	46.3	62.7	317	535	0.249
1 102	85.6	49	65.4	72.8	221	418	0.137
1 202	74.2	29	67.7	78.7	273	860	0.163
1 502	81.6	65	54.5	66.5	216	604	0.225
1 602	78.7	54	56.8	75.0	342	306	0.243
1 604	74.3	88	47.5	58.2	443	316	0.206
1 701	69.9	59	50.1	57.3	332	150	0.183
1 703	78.6	22	58.2	71.3	141	222	0.190
平均值 Average	77.6	746	60.0	69.1	325.8	393.6	0.181

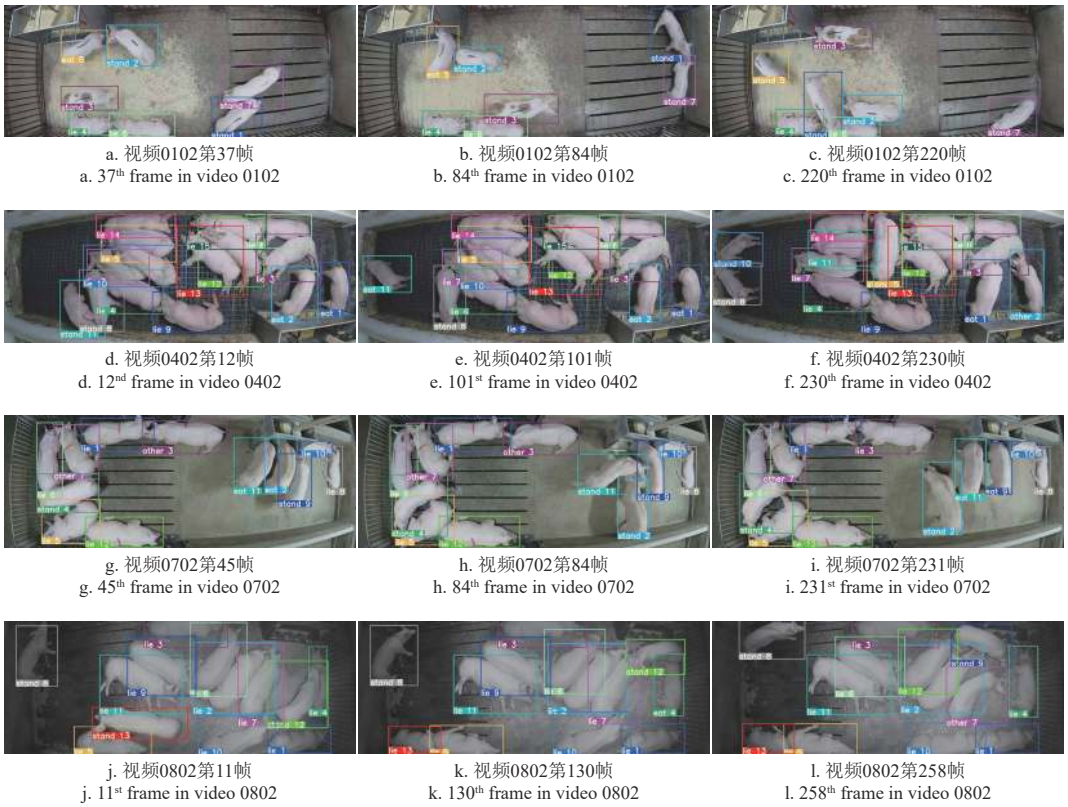
改进 TransTrack 在 0502 和 1 701 视频段上 MOTA，HOTA 和 IDF1 分别为 95.6% 和 91.6%，78.8% 和 69.7%，89.5% 和 82.7%，对比 TransTrack 结果，具有显著提升。同时，改进 TransTrack 在多个测试视频的 IDs、FP 和 FN 显著减少，在体现检测效果的指标 MOTA 中提升 14.6%，在体现跟踪效果的指标 HOTA 提升 9.8%，在跟踪轨迹连续性的指标 IDF1 提升 15.9%，在测试视频 0702、0802 和 0902 的 IDs 分别只有 1、2 和 2。结果表明改进 TransTrack 能更加稳定地对目标进行跟踪。分析原因在于，改进 TransTrack 由于去除重叠目标检测框，提高检测的 MOTA 指标；根据高低匹配阈值进行 2 次数据关联，提高 HOTA 和 IDF1 跟踪性能指标；限制猪只 ID 错误增长，降低了 IDs。

表 4　改进 TransTrack 的群养猪多目标跟踪结果

Table 4　Pig MOT results based on improved TransTrack

视频段 Video sequence	MOTA/%	IDs	HOTA/%	IDF1/%	FP	FN	MOTP
0102	91.1	5	65.4	86.4	34	148	0.145
0402	97.0	6	78.0	89.5	36	91	0.156
0502	95.6	8	78.8	96.5	51	47	0.176
0602	87.4	25	63.4	83.3	298	284	0.209
0702	98.5	1	84.7	98.6	22	32	0.146
0802	99.5	2	83.3	96.3	4	12	0.140
0902	95.6	2	80.6	94.5	29	155	0.152
1 002	85.5	11	56.0	75.5	157	442	0.251
1 102	93.0	22	78.0	87.4	150	161	0.138
1 202	83.8	14	71.2	84.9	85	632	0.170
1 502	89.2	51	51.0	66.4	134	331	0.224
1 602	87.6	34	55.8	74.5	208	168	0.247
1 604	90.4	18	59.7	72.3	156	143	0.208
1 701	91.6	8	69.7	82.7	110	33	0.184
1 703	96.4	3	71.1	86.0	41	20	0.192
平均值 Average	92.0	210	69.8	85.0	101.0	179.9	0.183

改进 TransTrack 的群养生猪 MOT 部分效果如图 7 所示。



a. 视频0102第37帧
a. 37th frame in video 0102

b. 视频0102第84帧
b. 84th frame in video 0102

c. 视频0102第220帧
c. 220th frame in video 0102

d. 视频0402第12帧
d. 12nd frame in video 0402

e. 视频0402第101帧
e. 101st frame in video 0402

f. 视频0402第230帧
f. 230th frame in video 0402

g. 视频0702第45帧
g. 45th frame in video 0702

h. 视频0702第84帧
h. 84th frame in video 0702

i. 视频0702第231帧
i. 231st frame in video 0702

j. 视频0802第11帧
j. 11st frame in video 0802

k. 视频0802第130帧
k. 130th frame in video 0802

l. 视频0802第258帧
l. 258th frame in video 0802

图 7　改进 TransTrack 群养猪多目标跟踪效果

Fig.7　Pig MOT effect based on improved TransTrack

其中，视频 0102、0402、0702 和 0802 的猪只最大 ID 号分别维持在 7、15、12 和 13，与真实 GT 值相同，说明改进 TransTrack 能维持跟踪轨迹持续和稳定，具有较好的跟踪效果。

图 8 给出改进 TransTrack 和 TransTrack 在 3 个视频段的跟踪效果对比图。白色的虚线框表示错误的 ID 号，红色的虚线框表示漏检，从图 8 中发现，TransTrack 存在错误的检测和漏检，导致出现错误的 ID 变换，改进 TransTrack 能避免这些问题，提升群养生猪的多目标跟踪性能。

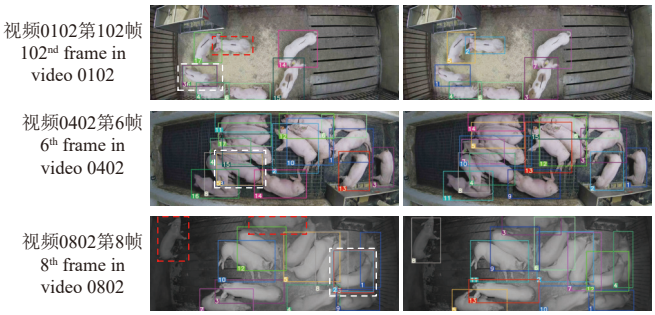


图 8 改进 TransTrack 与 TransTrack 的跟踪结果对比

Fig.8 Comparison of tracking results between improved TransTrack and TransTrack

3.3 改进 TransTrack 和 TransTrack 在不同光照和遮挡下的结果对比

对不同光照强度和遮挡程度的测试视频，给出改进 TransTrack 和 TransTrack 的对比结果。对不同光照下，其结果对比图如图 9 所示。1 701 和 1 703 视频段分别拍摄于同一个猪舍的上午和下午时间段。采用 TransTrack 算法，图 9a 中 1 701 视频第 288 帧和视频 1 703 第 73 帧中存在一头猪只发生误检，是左侧窗户的不同光照导致，而采用基于 CIOU 匹配算法的改进 TransTrack 算法，能避免目标丢失并且去除重叠目标检测框，如图 9b 中结果所示。

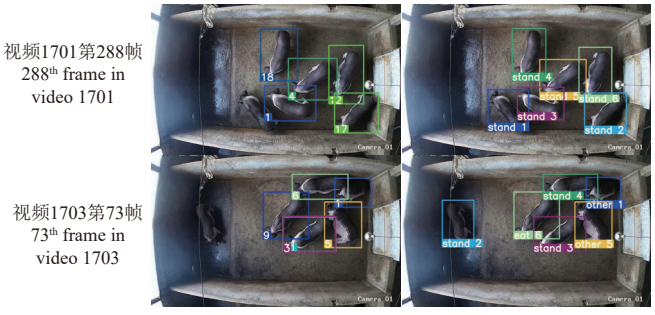


图 9 改进 TransTrack 与 TransTrack 在不同光照条件下的跟踪结果对比

Fig.9 Comparison of tracking results between improved TransTrack and TransTrack under different light conditions

在不同遮挡情况下，改进 TransTrack 和 TransTrack 结果对比图如图 10 所示。图 10a 视频 0702 第 147 帧和视频 1 703 第 280 帧中，分别存在一只猪只由于被遮挡，导致该猪只漏检，同时，图 10a 视频 0702 和视频 1 703 中存在频繁的错误 IDs，其最大 ID 分别为 18 和 13；改进 TransTrack 中则没有漏检，也没有错误的 IDs，结果如图 10b 所示，视频 0702 和视频 1 703 中最大的 ID 分别为 12 和 6，和 GT 相同。结果表明，改进 TransTrack 提高了检测性能，根据高低匹配阈值进行两次数据关联，提高跟踪性能，同时，限制猪只 ID 错误增长，降低了 IDs。

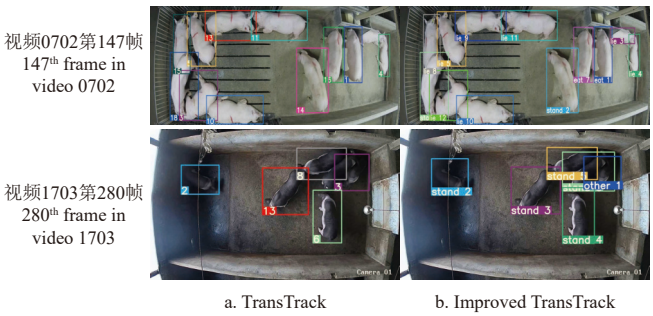


图 10 改进 TransTrack 与 TransTrack 在不同遮挡条件下的跟踪结果对比

Fig.10 Comparison of tracking results between improved TransTrack and TransTrack under different occlusion conditions

4 结 论

本文基于 TransTrack 构建了改进群养生猪行为跟踪模型，实现自然场景下稳定猪只身份的群养生猪行为跟踪。主要结论如下：

1）针对光照变化及猪只密集遮挡导致的目标误检和漏检挑战，提出一种改进 TransTrack 的猪只行为跟踪模型。在 TransTrack 基础上，首先采用基于改进的并集交并比，去除重叠目标检测框；然后利用高低匹配阈值进行 2 次数据关联；最后提出猪只身份修正算法，提高改进 TransTrack 在群养生猪行为跟踪的性能。

2）在所有测试视频中，改进 TransTrack 方法的多目标跟踪准确率（multiple object tracking accuracy，MOTA）为 92.0%，高阶跟踪准确率（higher order tracking accuracy，HOTA）为 69.8%，身份变换（identity switches，IDs）次数为 210。在公开数据集中，改进 TransTrack 的 MOTA 为 92.4%，HOTA 为 72.1%，IDs 为 147。对比 Trackformer，JDE 和 TransTrack 方法，改进 TransTrack 在 MOTA 上分别高 3.9，9.0 和 13.1 个百分点，在 HOTA 上分别高 1.3，9.5 和 8.3 个百分点，在 IDs 上分别降低 136，326 和 376。在私有数据集中，改进 TransTrack 的 MOTA 为 91.5%，HOTA 为 62.5%，IDs 为 63。对比 Trackformer 和 TransTrack 模型，改进 TransTrack 方法在 MOTA 上分别高 14.4 和 15.8 个百分点，HOTA 上分别高 1.8 和 9.5 个百分点。实现更稳定的群养生猪行为跟踪。

本文将生猪基本行为识别研究结果整合到多目标跟踪研究中，后续工作就是在群养环境下统计一段长时间中（例如一整天），每栏每头猪每种行为的时间及行为动态变化跟踪结果，用以分析个体行为异常及群体行为变化，为后续无接触式的生猪监控提供理论和技术支持。

［参　考　文　献］

[1] 程玉兰. 非洲猪瘟背景下推进生猪产业发展的新思考[J]. 中国畜禽种业，2021，17(3)：29-30.
CHENG Yulan. New thoughts on promoting the development of pig industry in the context of African swine fever [J]. The Chinese Livestock and Poultry Breeding, 2021, 17(3): 29-30. (in Chinese with English abstract)

[2] 沈明霞，陈金鑫，丁奇安，等. 生猪自动化养殖装备与技术研究进展与展望[J]. 农业机械学报，2022，53(12)：1-19.
SHEN Mingxia, CHEN Jinxin, DING Qian, et al. Current situation and development trend of pig automated farming equipment application[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12): 1-19. (in Chinese with English abstract)

[3] 华利忠，冯志新，张永强，等. 以史为鉴，浅谈中国非洲猪瘟的防控与净化[J]. 中国动物传染病学报，2019，27(2)：96-104.
HUA Lizhong, FENG Zhixin, ZHANG Yongqiang, et al. Prevention and control of African swine fever in China: Lessons from past outbreaks[J]. Chinese Journal of Animal Infectious Diseases, 2019, 27(2): 96-104. (in Chinese with English abstract)

[4] ARULMOZHI E, BHUJEL A, MOON B. The application of cameras in precision pig farming:An overview for swine-keeping professionals[J]. Animals, 2021, 11(8): 2343.

[5] Sébastien F, Alain N R, Benoit L. Rethinking environment control strategy of confined animal housing systems through precision livestock farming[J]. Biosystems Engineering, 2017, 155: 96-123.

[6] ZHENG C, ZHU X M, YANG X F, et al. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. Computers and Electronics in Agriculture, 2018, 147: 51-63.

[7] CHEN C, ZHU W X, MA C H, et al. Image motion feature extraction for recognition of aggressive behaviours among group-housed pigs[J]. Computers and Electronics in Agriculture, 2017, 142: 380-387.

[8] SHAO H, PU J, MU J. Pig-posture recognition based on computer vision: Dataset and exploration[J]. Animals, 2021, 11, 1295.

[9] RIEKERT M, KLEIN A, ADRION F, et al. Automatically detecting pig position and posture by 2D camera imaging and deep learning[J]. Computers and Electronics in Agriculture, 2020, 174: 105391.

[10] 王鲁，刘晴，曹月，等. 基于改进 Cascade Mask R-CNN 与协同注意力机制的群猪姿态识别[J]. 农业工程学报，2023，39(4)：144-153.
WANG Lu, LIU Qing, CAO Yue, et al. Posture recognition of group-housed pigs using improved Cascade Mask R-CNN and cooperative attention mechanism[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(4): 144-153. (in Chinese with English abstract)

[11] LUO W, XING J, MILAN A, et al. Multiple object tracking: A literature review[J]. Artificial Intelligence, 2021, 293: 103448.

[12] BEWLEY A, GE Z Y, OTT L, et al. Simple online and realtime tracking[C]//2016 IEEE International Conference on Image Processing,Phoenix, USA:IEEE, 2016: 3464-3468.

[13] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]// 2017 IEEE International Conference on Image Processing, Beijing, China:IEEE, 2017: 3645-3649.

[14] WANG Z D, LIANG Z, LIU Y X, et al. Towards real-time multi-object tracking [C]// European Conference on Computer Vision, Glasgow, UK, 2020: 107–122.

[15] ZAHNG Y F, SUN P Z, JIANG Y, et al. ByteTrack: Multi-Object tracking by associating every detection box [C]// European Conference on Computer Vision, Tel Aviv, Israel, 2022: 1-21.

[16] MEINHARDT T, KIRILLOV A, LEAL-TAIXÉ L, et al. TrackFormer: Multi-object tracking with transformers[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, USA: IEEE, 2022: 8844-8854.

[17] SUN P Z, JIANG Y, ZHANG R F, et al. Transtrack: multiple object tracking with transformer[EB/OL]. (2020-12-31)[2021-05-04].https://arxiv.org/pdf/2012.15460v1.pdf.

[18] 张伟，沈明霞，刘龙申，等. 基于 CenterNet 搭配优化 DeepSORT 算法的断奶仔猪目标跟踪方法研究[J]. 南京农业大学学报，2021，44(5)：973-981.
ZHANG Wei, SHEN Mingxia, LIU Longshen, et al. Research on weaned piglet target tracking method based on CenterNet collocation optimized DeepSORT algorithm[J]. Journal of Nanjing Agricultural University, 2021, 44(5): 973-981. (in Chinese with English abstract)

[19] TU S Q, ZENG Q T, LIANG Y, et al. Automated behavior recognition and tracking of group-housed pigs with an improved DeepSORT method[J]. Agriculture-Basel, 2022, 12(11): 1907-1926.

[20] 涂淑琴，刘晓龙，梁云，等. 基于改进 DeepSORT 的群养生猪行为识别与跟踪方法[J]. 农业机械学报，2022，53(8)：345-352.
TU Shuqin, LIU Xiaolong, LIANG Yun, et al. Behavior recognition and tracking method of group-housed pigs based on improved DeepSORT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8): 345-352. (in Chinese with English abstract)

[21] 涂淑琴，黄磊，梁云，等. 基于 JDE 模型的群养生猪多目标跟踪[J]. 农业工程学报，2022，38(17)：186-195.
TU Shuqin, HUANG Lei, LIANG Yun, et al. Multiple object tracking of group-housed pigs based on JDE model[J].

Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(17): 186-195. (in Chinese with English abstract)

[22] 涂淑琴，汤寅杰，李承桀，等. 基于改进 ByteTrack 算法的群养生猪行为识别与跟踪技术[J]. 农业机械学报，2022，53(12)：264-272.
TU Shuqin, TANG Yinjie, LI Chengjie, et al. Behavior recognition and tracking of group-housed pigs based on improved ByteTrack algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12): 264-272. (in Chinese with English abstract)

[23] PSOTA E T, SCHMIDT T, MOTE B, et al. Long-term tracking of group-housed livestock using keypoint detection and MAP estimation for individual animal identification[J]. Sensors , 2020, 20(13): 3670.

[24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(2) : 318-327.

[25] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA:IEEE, 2019: 658-666.

[26] LUITEN J, OSEP A, DENDORFER P, et al. HOTA: A higher order metric for evaluating multi-object tracking[J]. International Journal of Computer Vision, 2021, 192(2): 548–578

[27] 黄成龙，华向东，黄诗豪，等. 基于 Micro-CT 和改进 DeepSORT 的再生稻再生芽追踪计数与再生力评价[J]. 农业工程学报，2023，39(11)：165-174.
HUANG Chenglong, HUA Xiangdong, HUANG Shihao, et al. Regenerated buds tracking and regenerative ability evaluation of ratooning rice using Micro-CT imaging and improved DeepSORT[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(11): 165-174. (in Chinese with English abstract)

Methods for multi-target tracking of pig action using improved TransTrack

TU Shuqin, HUANG Zhengxin, LIANG Yun※, HUANG Lei, LIU Xiaolong

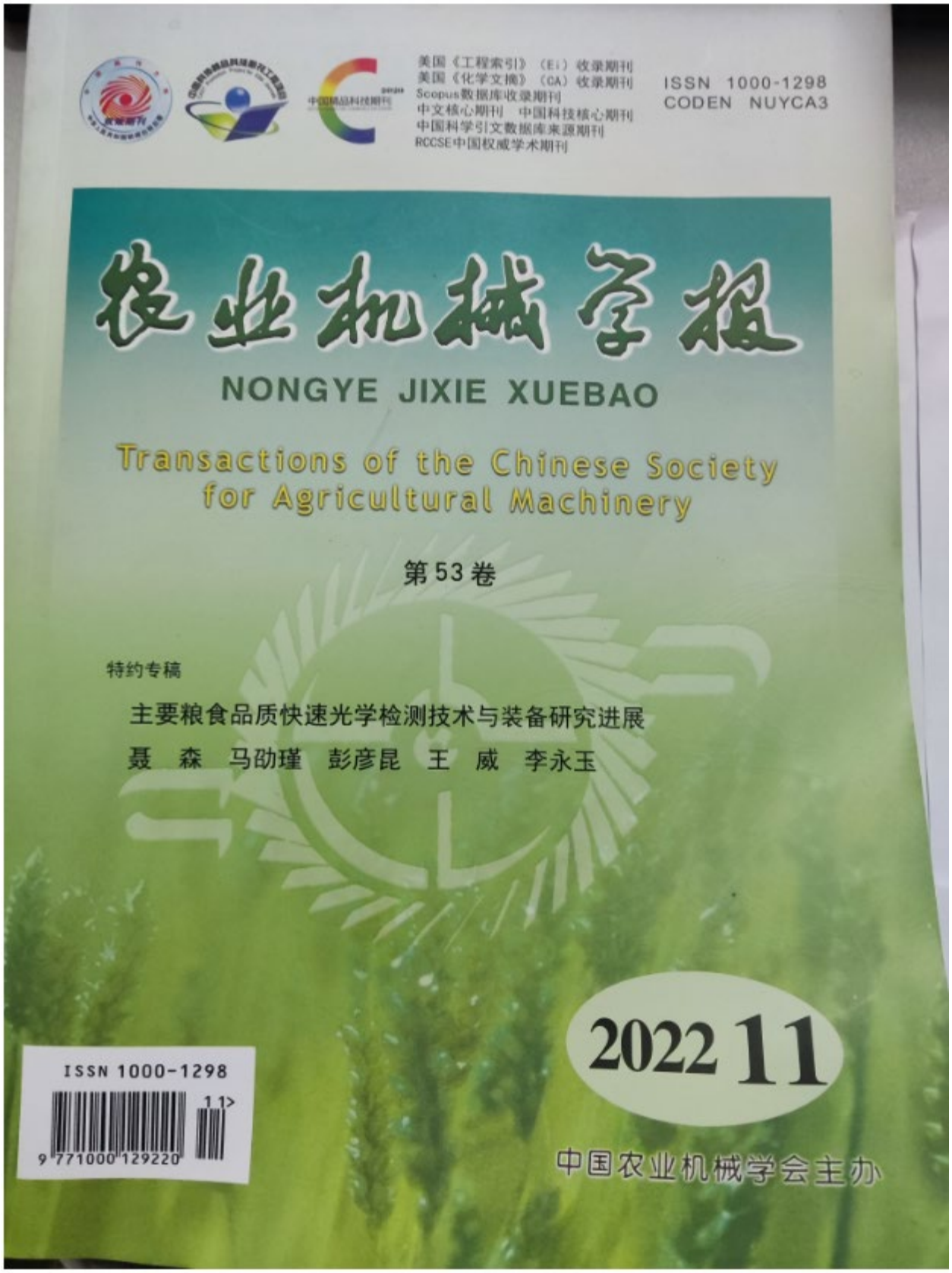
(*College of Mathematics and Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: Pig production is dominated in the livestock industry, particularly for the food safety, social stability, and the coordinated development of the national economy. An accurate tracking of pig behavior can greatly contribute to the health and well-being of pigs, in terms of the detection of abnormal conditions, such as diseases and dangerous movements. However, manual monitoring cannot fully meet the large-scale production in recent years, due to the time consuming, subjective, and labor intensity. Fortunately, the video surveillance technology has been carried out to detect and track pigs. But it is still lacking on accurate tracking and detection of pig behavior in various complex scenes (day or night, sparse or dense condition). In this study, an improved TransTrak multiple object tracking (MOT) was proposed to automatically detect the pigs, and then track the behavior of each detected pig, with considering the motion information of the behavior. Three improvements included an improved CIOU matching to remove the overlapping detections, the behavior category learning with MOT, and the data association. Therefore, the improved TransTrack approach performed better to identify the pig objects during behavior tracking in complex scenarios. The improved TransTrack was validated on the special dataset under a variety of settings. The specific dataset included the public dataset with 23 video sequences and 6 900 images, and the private dataset with 8 video segments and 2 400 images. The experimental results show that the MOT accuracy (MOTA), the higher order tracking accuracy (HOTA), and ID switches (IDs) on all test videos were 92.0%, 69.8%, and 210, respectively, using the improved TransTrack. On the public dataset, the improved TransTrack achieved the MOTA with 92.4%, HOTA with 72.1%, and IDs with 147. And the improved TransTrack improved the MOTA by 3.9, 9.0 and 13.1 percentage points, HOTA by 1.3, 9.5, and 8.3 percentage points, and decreased in the IDs by 136, 326, and 376, respectively, compared with the Trackformer, JDE, and TransTrack. On the private dataset, the improved TransTrack obtained the MOTA with 91.5%, HOTA with 62.5%, and IDs with 63. The MOTA was increased by 14.4, and 15.8 percentage points, and the HOTA was improved by 1.8, and 9.5 percentage points, respectively, compared with the Trackformer and TransTrack. The improved TransTrack can be expected to obtain the best performance in the tracking evaluation metrics (MOTA, HOTA, and IDs) among the four methods. Therefore, the improved TransTrack can also provide the scalable technical support for the automatic monitoring pigs.

Keywords: recognition; multiple object tracking; pigs; TransTrack; data association

3.以通讯作者发表本专业论文情况

3.1. 基于YOLOv4和双重回归的复杂环境檀香树缺苗定位方法



农业机械学报
NONGYE JIXIE XUEBAO
2022 年　第 11 期(第 53 卷)

目　次

农业生物环境与能源工程

农产品加工工程

车辆与动力工程

机械设计制造及其自动化

农业机械学报

NONGYE JIXIE XUEBAO

2022 年　第 53 卷　第 11 期

（月刊，1957 年创刊）

2022 年 11 月 25 日出版

Transactions of the Chinese Society for Agricultural Machinery

No. 11　Vol. 53　2022

（Monthly，started in 1957）

Published on 25，November 2022

主　管	中国科学技术协会
主　办	中国农业机械学会 中国农业机械化科学研究院
编辑出版	《农业机械学报》编辑部 （地址：北京德外北沙滩 1 号 6 信箱） 邮政编码　100083
主　编	任露泉
印　刷	北京富泰印刷有限责任公司
国内发行	中国邮政集团公司北京报刊发行局
订 购 处	全国各地邮局
国外发行	中国国际图书贸易集团有限公司 （北京 399 信箱）

Responsible Department:	China Association for Science and Technology
Sponsored by:	Chinese Society for Agricultural Machinery Chinese Academy of Agricultural Mechanization Sciences
Published by:	Editorial Office of Transactions of the Chinese Society for Agricultural Machinery
Editor in Chief:	Ren Luquan
Editorial Office:	No. 1 Beishatan, Deshengmen Wai, Beijing 100083, China
Tel:	86-10-64882610/64882231
Fax:	86-10-64867367
http:	//www.j-csam.org
E-mail:	njxb@caams.org.cn
Overseas Distributor:	China International Book Trading Corporation (P.O. Box 399, Beijing, China)

ISSN　1000－1298
CN　11－1964/S

国内邮发代号　2－363　　国外发行代号　M289　　国内定价　100.00 元

doi:10.6041/j.issn.1000-1298.2022.11.030

基于 YOLOv4 和双重回归的复杂环境檀香树缺苗定位方法

张　宇[1]　徐浩然[1]　牛家俊[1]　涂淑琴[2]　赵文锋[1]

(1. 华南农业大学电子工程学院(人工智能学院), 广州 510642; 2. 华南农业大学数学与信息学院, 广州 510642)

摘要: 在檀香树大面积种植过程中,存在人工排查缺苗效率低、成本高和难以监管等问题,而且檀香树必备的伴生植物和树间穿插的其它作物,更加大了查补难度。针对这些问题,本文提出一种基于 YOLOv4 和双重回归的复杂环境檀香树缺苗检测和精准定位方法。首先,采用 YOLOv4 目标检测算法,处理无人机采集的遥感图像,实现檀香树植株的智能检测。然后,以双重线性回归结合延长列线补漏策略为核心,构建缺苗定位算法(Missing seedling localization algorithm,MSL):选任意檀香树作基准,根据像素坐标划分列区域,对各列区域中檀香树用线性回归拟合列线;对拟合后仍未归入列的遗漏檀香树,用延长回归线策略重新判断归属,并再次线性回归优化列线。最后,根据种植间距规划,实现缺苗检测和定位。试验结果表明,檀香树缺苗检测精确率 86.82%、召回率 82.25%、F1 值 84.47%、运行时间 8.19 s。该方法融合了大疆无人机遥感图像采集系统的快速性、YOLOv4 算法和双重回归策略的精准性,可实现对复杂生长状况下檀香树的实时智能缺苗检测和精准定位。

关键词: 檀香树; 缺苗定位; 目标检测; YOLOv4; 双重线性回归

中图分类号: TP753　　**文献标识码**: A　　**文章编号**: 1000-1298(2022)11-0299-07　　**OSID**:

Missing Seedling Localization Method for Sandalwood Trees in Complex Environment Based on YOLOv4 and Double Regression Strategy

ZHANG Yu[1]　XU Haoran[1]　NIU Jiajun[1]　TU Shuqin[2]　ZHAO Wenfeng[1]

(1. *College of Electronic Engineering* (*College of Artificial Intelligence*), *South China Agricultural University*, *Guangzhou* 510642, *China*
2. *College of Mathematics and Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: In the process of planting sandalwood trees on a large scale, there are problems such as low efficiency, high cost, and difficulty in the supervision of manual ranking of missing seedlings, and the necessary companion plants for each sandalwood tree and other crops interspersed between the trees, further deepening the difficulty of checking and replenishing. For these problems, a seedling deficiency detection and precise localization method in complex environment was proposed based on YOLOv4 algorithm and double regression strategy. Firstly, the YOLOv4 target detection model was used to achieve sandalwood plant detection from remote sensing images collected by UAV. Then the missing seedling localization algorithm (MSL) was constructed based on the double linear regression and extended column line fixing strategy: arbitrary sandalwood trees were selected as the benchmark, column regions were divided according to the pixel coordinates, and column lines were fitted to the sandalwood trees in each column region by using linear regression; for the omitted sandalwood trees that were not classified into columns after fitting, the attribution was judged again with the extended regression line strategy, and the column lines were optimized by linear regression again. Finally, the missing seedlings were calculated and localized according to the spacing at the time of planting. The results showed that the precision was 86.82%, the recall was 82.25%, the F1-score was 84.47%, and the running time was 8.19 s, respectively. In summary, this method combined the rapidity of DJI UAV remote sensing image acquisition system, the accuracy of YOLOv4 algorithm and double regression strategy, which can be used to achieve real-time intelligent seedling deficiency detection and accurate localization of sandalwood trees under complex growth conditions.

Key words: sandalwood trees; missing seedling localization; object detection; YOLOv4; double linear regression

收稿日期: 2022-08-06　修回日期: 2022-09-25

基金项目: 广东省企业科技特派员项目(GDKTP20210557700)

作者简介: 张宇(1976—),女,讲师,主要从事计算机视觉研究,E-mail: zhangyu@scau.edu.cn

通信作者: 涂淑琴(1978—),女,讲师,主要从事计算机视觉研究,E-mail: tushuqin@163.com

0 引言

檀香树是一种在亚热带地区种植的农作物，其生长缓慢，经济价值极高[1]。如果在其生长过程中，缺苗补种不及时，会造成檀香树生长时差长、植株差异大，导致更为严重的遮挡死亡，因此及时检补缺苗，才能保证檀香园产量。目前，主要依赖人工进行缺苗排查，该方法成本高、效率低且效果难以监管，这些问题也同样发生在其它大规模作物种植园。因此，亟需一种智能缺苗检测技术取代人工，进行智能实时缺苗检测和定位，该技术对智慧种植、作物生长过程智能化监控和提高作物产量具有重要意义。

近年来，智能缺苗检测逐步成为智慧农业的一个研究热点，国内外研究团队从穴盘到自然环境种植场景均取得了研究成果。在标准穴盘种植中，王永维等[2]采用灰度化和 Otsu 阈值分割等数字图像处理技术，依次获取幼苗和穴盘二值图、穴孔位置、穴孔内苗的生长情况，实现缺苗判断；张国栋等[3]采用类似方法识别好苗、坏苗和空穴，并开发了穴盘苗检测系统；文永双等[4]利用光纤传感器信号进行缺苗判断，同时设计了穴盘种植移栽缺苗补偿系统。

在自然环境下，对作物检测[5-8]、行识别[9-15]是实现缺苗检测的关键。尽管这些方法对相对规则、种植单一的自然场景缺苗检测效果良好，然而却很难适应檀香树种植，原因在于，檀香树人工种植行列不齐或受台风、暴雨等影响，行列有时会偏离直线排列很大；必备的伴生植物和插种的木薯等作物，使檀香树植株处于非常杂乱的环境；另外，檀香园通常位于不规则的湿热地带，地块边缘的植株容易出现被漏检或错检等，这种情况需要算法进一步优化。

基于处于复杂生长环境且需要及时补苗的檀香树种植需求，本文拟充分运用无人机遥感技术的快速采集、YOLOv4 的精准目标检测，并构建以列线双重回归及延长补漏策略为核心的缺苗定位算法，以期为其提供快速、精确、有效的技术支持。

1 数据获取与处理

1.1 数据采集

于 2020 年 9 月 29 日在广东省台山市白沙镇研究区(22.14°N，112.83°E，如图 1 所示)，使用大疆 Phantom4 型无人机进行了檀香树遥感图像样本采集。设置云台拍摄角度为垂直地面，航向重叠率 80%，拍摄高度分别为 10、20、30 m，每幅图像尺寸为 4 000 像素 ×3 000 像素，航拍区域覆盖略大于试验区域。共采集 1 051 幅遥感图像，其中拍摄高度 10、20、30 m 的图像分别有 696、155、200 幅。所有拍摄高度的遥感图像中，随机取出 557 幅用于檀香树目标检测模型训练、验证和测试；从拍摄高度为 30 m 的遥感图像中取 40 幅用于檀香树缺苗定位任务。

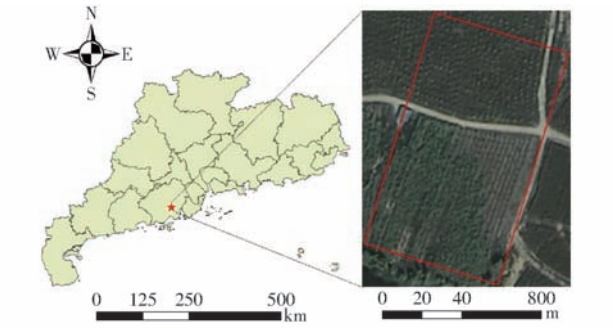


图 1 研究区示意图

Fig. 1 Schematic of study area

1.2 数据集构建

将采集的遥感图像用 LabelImg 工具进行标注，根据检测和缺苗定位任务，进行 2 种不同的标注。

首先，按照 Pascal VOC[16]数据集格式，对检测任务的 557 幅遥感图像中的檀香树植株标注，标签为“sandalwood”，标注文件包含每幅图像目标檀香树的像素坐标及标签，并划分成训练集、验证集和测试集，分别占比 80%、10% 和 10%，用于 YOLOv4 模型的训练、验证和测试任务。

其次，按照 YOLO 数据集格式，对进行缺苗定位任务的 40 幅遥感图像，用 150 像素 ×150 像素的框标注 1 746 个真实缺苗区域，标签为“empty”，同时对其中的檀香树植株标注与检测任务集一致。评测檀香树缺苗定位算法(missing seeding localization algorithm，MSL)前，需要根据种植列方向对图像进行旋转。两部分数据集的标注示意如 图 2 所示。

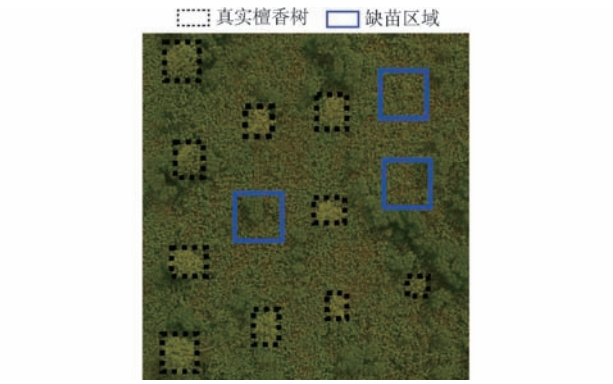


图 2 檀香树及缺苗位置

Fig. 2 Sandalwood trees and location of missing seedlings

2　研究方法

首先用YOLOv4目标检测模型检测檀香树目标。然后构建缺苗定位算法：基于檀香树目标的像素坐标关系划分列区域；利用双重列线性回归和延长回归线策略拟合生成、优化檀香树的列回归线排列；在各条列回归线上，根据檀香树植株间距和规划距离 30 cm 的关系，实现缺苗定位。算法缺苗定位方法流程如图 3 所示。图中，[,] 表示缺苗位置信息，如[2,3]表示在第 2 列、第 3 个位置缺苗。

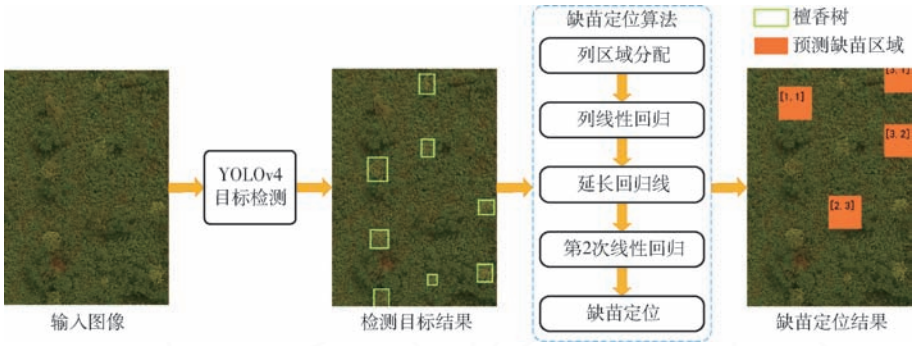


图 3　缺苗定位方法流程图

Fig. 3　Flow chart for missing seedling localization method

2.1　YOLOv4 算法

檀香树植株检测是实现缺苗定位的前提，需要充分利用深层语义特征的深度学习方法进行。基于深度学习的目标检测算法分为两类算法：基于候选区的两阶段算法，主要代表为 R－CNN、Faster R－CNN、Mask R－CNN[17-20]算法；基于回归的单阶段算法，主要代表为 SSD[21]、CenterNet[22]、YOLO[23-27]系列算法。其中单阶段算法相比两阶段算法，更具速度优势，因此适用于无人机遥感影像的快速检测应用。考虑到运行效率与精度的需要，选择 YOLOv4 作为檀香树检测器。

YOLOv4 网络结构分为输入端、Backbone、Neck 和 Head。在输入端，YOLOv4 算法创新了 Mosaic 数据增强和 Self-adversarial-training(SAT)，提高了模型的泛化能力；在 Backbone 阶段采取了 CSPNet[28]的方法，将基础层的特征映射为两部分，并使用跨阶段层次结构将两部分合并，减少了计算量，进而从图像中提取丰富的特征信息；在 Neck 阶段采取 PANet[29]的方法，PANet 采用了一种自下向上路径的特征金字塔网络(Feature pyramid networks，FPN)结构，避免信息丢失问题，提高了检测不同尺寸和尺度物体的能力。YOLOv4 网络结构如图 4 所示。

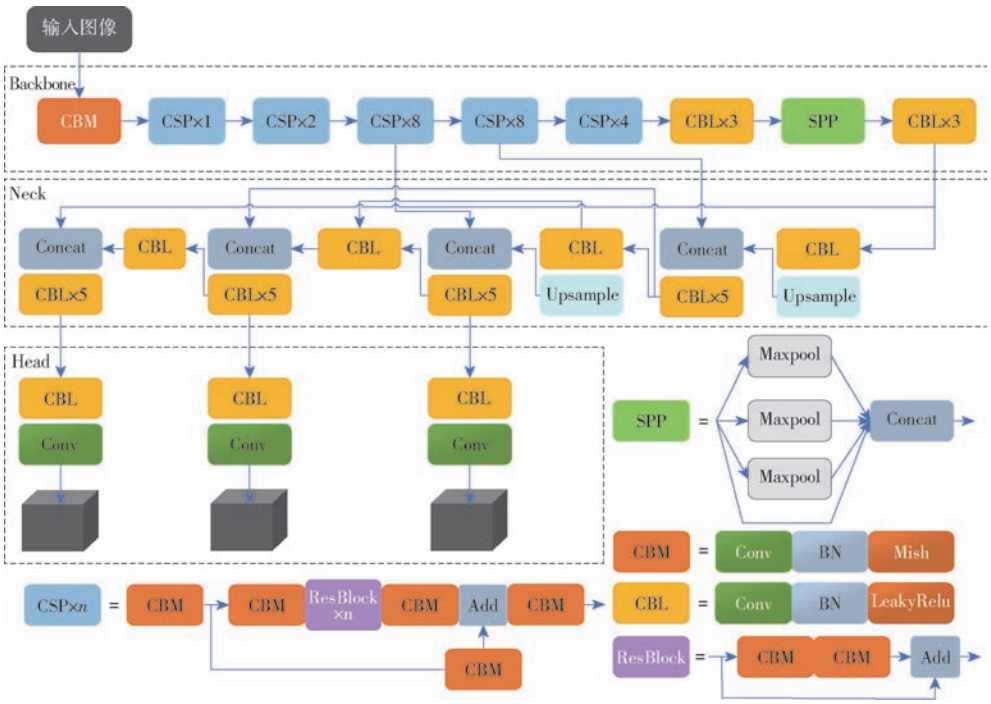


图 4　YOLOv4 网络结构示意图

Fig. 4　YOLOv4 network structure diagram

2.2 缺苗定位算法

遥感图像上可观察到檀香园中植株间距大、种植列线偏离理想直线、行线混乱问题。为应对这些问题,本文充分利用双重线性回归结合延长列回归线补漏的策略,并在此基础上构建缺苗定位算法(MSL)。完整算法如图 5 所示,主要分为以下 5 个步骤:

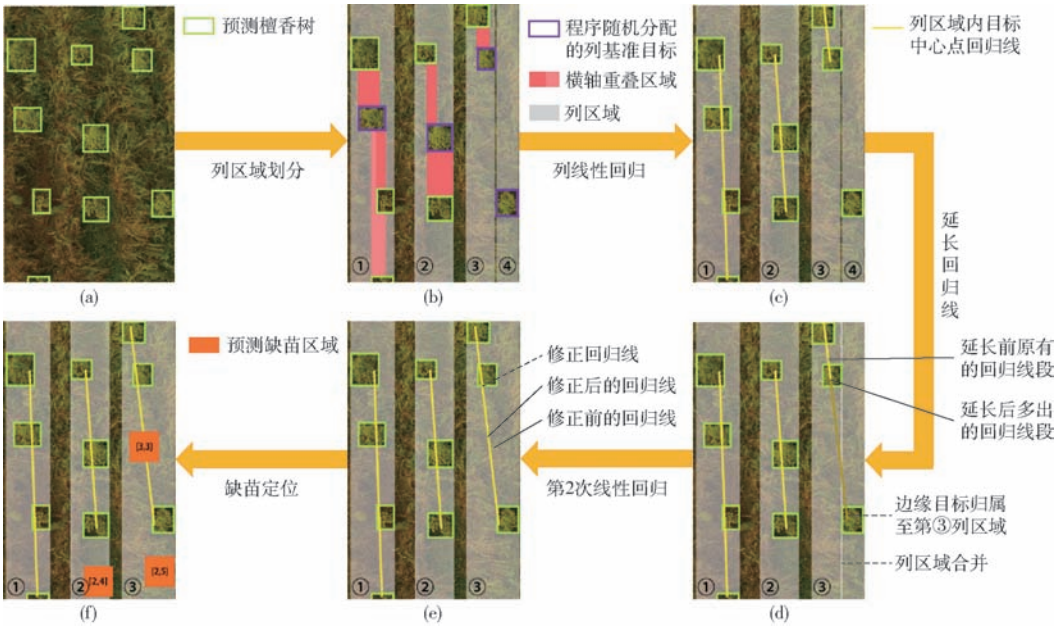


图 5 缺苗定位算法流程图

Fig. 5 Flow chart of missing seedling localization algorithm

(1)基于 YOLOv4 预测的檀香树像素坐标确定列基准位置,并进行列区域分配。

用 $p^i(x_{lt}^i, y_{lt}^i, x_{rb}^i, y_{rb}^i)$ 表示 YOLOv4 预测的檀香树目标 $i(i=1,2,\cdots,n)$ 的定位信息,其中 x_{lt}^i、y_{lt}^i 表示目标 p^i 左上角点的横、纵坐标;x_{rb}^i、y_{rb}^i 表示目标 p^i 右下角点的横、纵坐标;所有檀香树目标定位信息记为 $P=\{p^1,p^2,\cdots,p^n\}$,n 为预测的檀香树目标的个数,檀香树目标 p^i 的横坐标范围记为 $X^i=(x_{lt}^i, x_{rb}^i)$。

在 P 中随机取出一个檀香树目标 p^i,将 p^i 作为一个列区域的基准,并从 P 中移除 p^i。在 P 中进行遍历:若存在一个檀香树目标 $p^j(p^j \in P)$,且 p^j 的横坐标范围 X^j 与 p^i 的横坐标范围 X^i 有交集,表示为 $X^i \cap X^j \neq \varnothing$,则将该檀香树目标 p^j 归属到 p^i 作为基准的第 k 个列区域 R^k,并从 P 中移除 p^j;将所有横坐标范围与 X^i 有交集的檀香树目标分配至同一个列区域 R^k,则该列区域 R^k 由 m 个檀香树目标共同组成,将该列区域记为 $R^k=\{p^{k1},p^{k2},\cdots,p^{ki},\cdots,p^{kj},\cdots,p^{km}\}$;当 P 中除 p^i 以外的所有檀香树目标均遍历完成,则随机选取下一个檀香树目标并作为一个新的列区域的基准,重复上述遍历操作,直至 P 中所有的檀香树目标分配至列区域。

(2)对所有列区域进行线性回归,得到相应的列回归线。

给定任一列区域 $R^i=\{p^{i1},p^{i2},\cdots,p^{im}\}$,$R^i$ 中任一檀香树目标 $p^{ik}(x_{lt}^{ik}, y_{lt}^{ik}, x_{rb}^{ik}, y_{rb}^{ik})$ 的目标中心点记为 $c^{ik}(x_c^{ik}, y_c^{ik})$,其中 $x_c^{ik}=\dfrac{x_{lt}^{ik}+x_{rb}^{ik}}{2}$,$y_c^{ik}=\dfrac{y_{lt}^{ik}+y_{rb}^{ik}}{2}$。区域 R^i 的檀香树目标列向分布线性回归表达式为

$$\hat{x}_c^{ik}=f^i(y_c^{ik})=w^i y_c^{ik}+b^i$$

式中 $\hat{x}_c^{ik}$——对目标中心 c^{ik} 利用 y_c^{ik} 预测的横坐标

$f^i(y_c^{ik})$——列区域 R^i 的线性回归表达式

w^i——线性回归表达式权重

b^i——线性回归表达式偏置

该线性表达式通过最小二乘法训练得出。

将 w^i 初始值设为 0,b^i 初始值设为该列区域 R^i 的 x 坐标平均值。若列区域目标数大于 1,则进行线性回归;否则,将该列区域线性回归表达式 f^i 的 w^i、b^i 固定为初始值。

(3)延长列回归线,将位于延长段上的边缘目标归属至该直线所属的列区域中。

如图 5c 所示,列区域④仅有一个檀香树目标,无法与相同列区域中的其他目标共同组成种植列,而该目标本应属于列区域③,却被列区域③漏拟合。则将该目标称为边缘目标 a^k(k 表示边缘目标序号),坐标定位信息记为 $(x_{lt}^{a(k)}, y_{lt}^{a(k)}, x_{rb}^{a(k)}, y_{rb}^{a(k)})$,且该边缘目标中心点为 $c^{a(k)}(x_c^{a(k)}, y_c^{a(k)})$。

此时,若存在一个其他列区域 R^i 相对应的线性回归表达式 f^i,其代表的列回归线触碰到边缘目标 a^k,表示为 $x_{lt}^{a(k)}<f^i(y_c^{a(k)})<x_{rb}^{a(k)}$,则将该边缘目标

a^k 重新归属至列区域 R^i。

实际种植中的檀香树分布存在偏离理想直线的问题。在此情况下，步骤(1)中列区域分配并未很好地找到偏离种植直线的目标，而步骤(2)线性回归拟合出大致的列向分布规律，在此基础上步骤(3)进行延长回归直线，将其他边缘目标重新归属至正确的列区域，一定程度上可以补回偏离种植直线的目标。

(4)对扩充后列区域中的目标进行第 2 次线性回归，修正线性回归直线，使之更准确表达每列植株排列分布规律。同时，未扩充的列区域不参与第 2 次线性回归，该设定减小了性能浪费，同时保持拟合精度。

(5)将真实规定种植距离 30 cm 转换为遥感图像中像素距离(本场景为 300 像素)，并在每个列区域的线性回归直线上进行缺苗定位。

在每一个线性回归函数 $f^i(q)$ 中，q 表示像素数且 $q \in [0,H]$，H 表示遥感图像像素高度。若存在一个 q，使得位置 $(f^i(q),q)$ 与其同一回归线上相邻的檀香树目标或缺苗目标中心像素欧氏距离大于设定距离，将其设定为缺苗位置，并在遥感图像中标识。重复上述操作，直至所有列回归线遍历完成，最终输出带有缺苗位置标识的图像。

3　试验

3.1　试验环境

训练 YOLOv4 模型的 Ubuntu 服务器环境配置有 Xeon E5－2678 v3 CPU，NVIDIA Tesla K80 显卡。

缺苗定位试验所用的计算机硬件环境配置 CPU 为 AMD 5800H，系统环境配置操作系统为 Windows 10、科学计算的 Python 发行版 anaconda(Python 3.8)、Pytorch 1.8.2 深度学习框架和 Cuda 10.2 加速程序。

3.2　评测指标

采用精确率(Precision)、召回率(Recall)、F1 值(F1-score)、运行时间作为评测指标，对本文缺苗定位方法进行性能评测。

3.3　结果与分析

3.3.1　完整算法与残缺算法结果对比与分析

对 1.2 节中缺苗定位任务 40 幅遥感图像进行评测，MSL 完整算法共判断 1 654 个缺苗位置，准确判断 1 436 个，错误判断 218 个，漏判断 310 个。测试平均运行时间 8.19 s、精确率 86.82%、召回率 82.25%、F1 值 84.47%。从图 6 不同图像色调、场景的缺苗定位结果和表 1 评测结果可以看出，本文方法可以针对不同条件的复杂自然环境，精确识别檀香树真实缺苗位置，同时向种植园工作人员提供清晰位置(即行、列)信息。

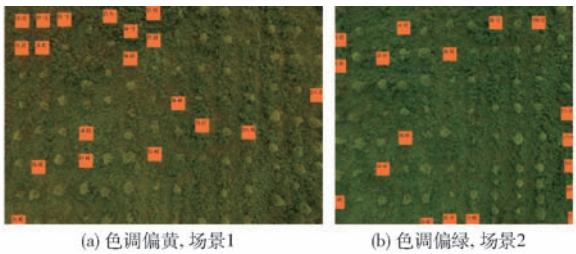


图 6　缺苗定位结果

Fig. 6　Localization results of missing seedlings

表 1　算法评测结果对比

Tab. 1　Comparison of evaluation results

算法步骤	列区域分配	列线性回归	延长回归直线	第 2 次列线性回归	缺苗定位	精确率/%	召回率/%	F1 值/%	运行时间/s
MSL－PL	√				√	80.87	81.10	80.98	7.48
MSL－PRL	√	√			√	85.96	82.42	84.15	7.84
MSL－Full	√	√	√	√	√	86.82	82.25	84.47	8.19

注："√"表示算法包含的步骤。

为了验证 MSL 完整算法的必要性，对完整步骤算法和残缺步骤算法进行比较。完整步骤算法中，经过了列区域分配(Partition)、列线性回归(Regression)、延长回归线(Extension)、第 2 次列线性回归(Regression)和缺苗定位(Localization)，称为 MSL－Full 算法；残缺步骤算法为经过列区域分配、列线性回归、缺苗定位的算法，称为 MSL－PRL 算法；仅经过列区域分配和缺苗定位的残缺步骤算法称为 MSL－PL 算法。

算法评测结果对比如表 1 所示。其中经过完整步骤的 MSL－Full 算法评测性能最好，精确率较 MSL－PRL 提高 0.86 个百分点，F1 值提高 0.32 个百分点；精确率较 MSL－PL 提高 5.95 个百分点，F1 值提高 3.49 个百分点。

图 7 直观地对比了 3 种算法的表现。在图 7a 对比中，如红色框所示，MSL－PL 所预测的植株回归直线并未拟合种植列线，且斜率为 0，因而预测的缺苗区域偏离真实缺苗区域；MSL－PRL 与 MSL－Full 均拟合了真实种植列线，且延长回归直线时并未检测到边缘目标，所以两者预测的回归直线以及

缺苗位置相同。在图 7b 对比中,如红色框所示,MSL – PL、MSL – PRL 在同一真实种植列区域中均预测出双列回归线,该误差由列区域分配中随机选取基准点步骤产生;而 MSL – Full 将 MSL – PRL 产生的回归直线延长,将该列区域附近的边缘目标纳入,删除多余列区域及对应的回归直线,再通过第 2 次线性回归操作,使回归直线更贴合真实种植直线,最后预测的缺苗位置也较 MSL – PL、MSL – PRL 算法更为精确。

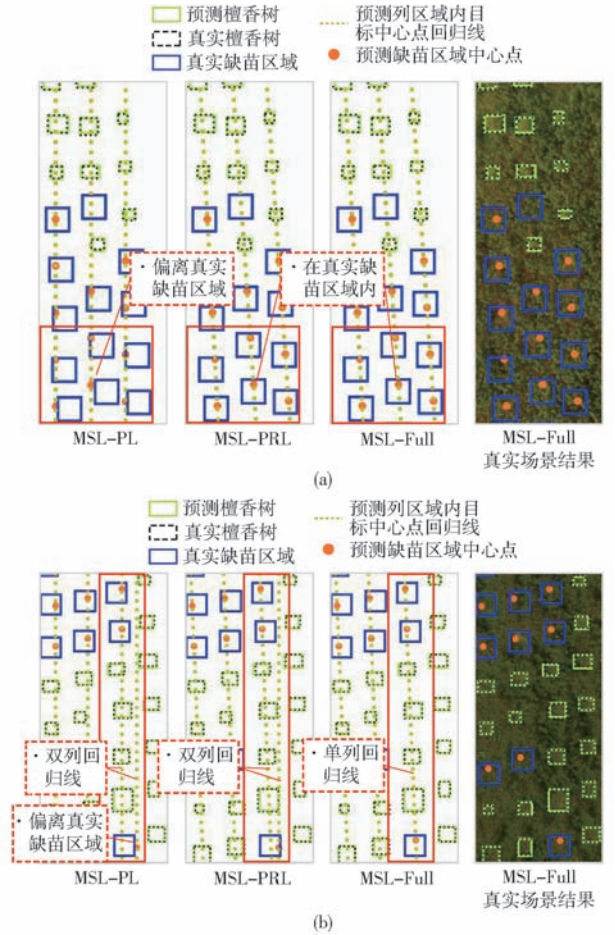


图 7 完整步骤算法与残缺步骤算法比较示例

Fig. 7 Comparison between complete algorithm and incomplete algorithm

综上,完整步骤算法比残缺步骤算法在缺苗定位的过程中更好地拟合了植株分布规律,因此具有更好的缺苗定位性能。

3.3.2 缺苗定位中檀香树检测训练结果与分析

YOLOv4 训练设置如下:

(1)参数设置:设置网络输入图像尺寸为 640 像素 ×640 像素;每次迭代训练的样本数为 2;开启冻结训练 150 轮,学习率为 0.001;关闭冻结训练后训练 100 轮,学习率为 0.000 1;两阶段训练均设置余弦退火学习率调整策略。

(2)训练策略:训练前,载入预训练权重;在训练过程中,通过采用 Mosaic 增强、旋转、翻转、调整色调等方式在线生成更多训练样本,提高网络鲁棒性和准确性。

如图 8 评测结果所示,在檀香树目标检测任务测试集中,YOLOv4 模型平均精度(Average precision,重叠率 0.5)为 96.66%;在缺苗定位任务数据集中平均精度为 98.52%。表明该目标检测模型能够应对复杂的自然环境,精确检测出檀香树目标,为进一步缺苗定位做好准备。

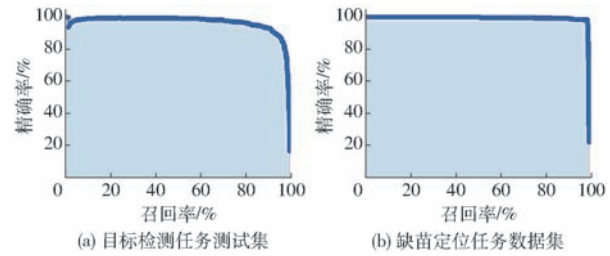


图 8 YOLOv4 评测结果

Fig. 8 YOLOv4 evaluation results

3.3.3 缺苗定位误差分析

本文缺苗定位方法还存在一定误差,误差主要来源于目标检测算法对檀香树多检和漏检。

如图 9a 红框内容所示,目标检测算法无法识别该位置的檀香树,根据植株距离设定,缺苗定位算法将该位置判断为缺苗位置,从而造成目标检测漏检误差。如图 9b 红框内容所示,对非檀香树误判定为檀香树,并将该目标所在列判断为种植列,最终对此列错误地缺苗预测,造成目标检测多检误差。这两类误差是因为目标检测模型对小目标识别精度低,无法区分部分檀香树小苗与周围杂草植株而导致的。

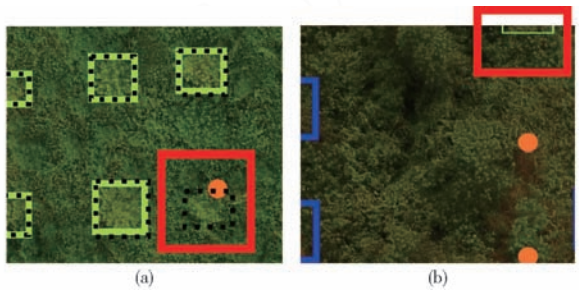


图 9 目标检测漏检、多检误差

Fig. 9 Error results of target detection

4 结束语

针对复杂的生长环境,基于 YOLOv4 算法和双重回归策略,提出了一种檀香树实时智能缺苗定位方法。通过将遥感技术、YOLOv4 目标检测算法与 MSL 算法相结合,先定位檀香树目标,再依次使用列区域分配、列线性回归、延长回归直线、第 2

次线性回归方法，最后实现了在回归直线上定位缺苗位置。同时，提出了一种根据真实缺苗区域与预测目标点位置关系验证的评测方法，评测结果表明，该缺苗定位方法的精确率 86.82%、召回率 82.25%、F1 值 84.47%、运行时间 8.19 s。相比人工巡查的传统方案，该缺苗定位方法结合了遥感技术，利用了快速飞行的无人机与高效的缺苗定位算法，更体现高效实时优势；与蔬菜穴盘苗缺苗定位方法对比，无需过多结构化要求，便能解决复杂的自然环境下檀香树植株缺苗定位问题。

参 考 文 献

[1] 武丽琼. 檀香种植栽培技术综述[J]. 中国园艺文摘, 2018, 34(3): 164 - 166.

[2] 王永维, 肖玺泽, 梁喜凤, 等. 蔬菜穴盘苗自动补苗试验台穴孔定位与缺苗检测系统[J]. 农业工程学报, 2018, 34(12): 35 - 41.
WANG Yongwei, XIAO Xize, LIANG Xifeng, et al. Plug hole positioning and seedling shortage detecting system on automatic seedling supplementing test-bed for vegetable plug seedlings[J]. Transactions of the CSAE, 2018, 34(12): 35 - 41. (in Chinese)

[3] 张国栋, 范开钧, 王海, 等. 基于机器视觉的穴盘苗检测试验研究[J]. 农机化研究, 2020, 42(4): 175 - 179.
ZHANG Guodong, FAN Kaijun, WANG Hai, et al. Experimental study on detection of plug tray seedlings based on machine vision[J]. Journal of Agricultural Mechanization Research, 2020, 42(4): 175 - 179. (in Chinese)

[4] 文永双,张宇,田金元,等.蔬菜移栽钵苗检测与缺苗补偿系统设计与试验[J].农业机械学报,2020,51(增刊1):123 - 129.
WEN Yongshuang,ZHANG Yu,TIAN Jinyuan, et al. Design and experiment of detection and supply system of vegetable plug seedlings for transplanting[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(Supp. 1):123 - 129. (in Chinese)

[5] 梁胤豪, 陈全, 董彩霞, 等. 基于深度学习和无人机遥感技术的玉米雄穗检测研究[J]. 福建农业学报, 2020, 35(4): 456 - 464.
LIANG Yinhao, CHEN Quan, DONG Caixia, et al. Application of deep-learning and UAV for field surveying corn tassel [J]. Fujian Journal of Agricultural Sciences, 2020, 35(4): 456 - 464. (in Chinese)

[6] LIU Y, CEN C, CHE Y, et al. Detection of maize tassels from UAV RGB imagery with Faster R - CNN[J]. Remote Sensing, 2020, 12(2):338.

[7] 杨蜀秦, 刘江川, 徐可可, 等. 基于改进 CenterNet 的玉米雄蕊无人机遥感图像识别[J]. 农业机械学报, 2021, 52(9): 206 - 212.
YANG Shuqin, LIU Jiangchuan, XU Keke, et al. Improved CenterNet based maize tassel recognition for UAV remote sensing image[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(9): 206 - 212. (in Chinese)

[8] 朱学岩, 张新伟, 顾梦梦, 等. 基于无人机可见光图像的云杉计数方法[J]. 林业工程学报, 2021, 6(4): 140 - 146.
ZHU Xueyan, ZHANG Xinwei, GU Mengmeng, et al. Spruce counting method based on UAV visible images[J]. Journal of Forestry Engineering, 2021, 6(4): 140 - 146. (in Chinese)

[9] 赵瑞娇, 李民赞, 张漫, 等. 基于改进 Hough 变换的农田作物行快速检测算法[J]. 农业机械学报,2009,40(7): 163 - 165.
ZHAO Ruijiao, LI Minzan, ZHANG Man, et al. Rapid crop-row detection based on improved Hough transformation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(7): 163 - 165. (in Chinese)

[10] 吴刚, 谭彧, 郑永军, 等. 基于改进 Hough 变换的收获机器人行走目标直线检测[J]. 农业机械学报,2010,41(2):176 - 179.
WU Gang, TAN Yu, ZHENG Yongjun, et al. Walking goal line detection based on improved Hough transform on harvesting robot [J]. Transactions of the Chinese Society for Agricultural Machinery, 2010, 41(2):176 - 179. (in Chinese)

[11] JI R, QI L. Crop-row detection algorithm based on random Hough transformation[J]. Mathematical and Computer Modelling, 2011, 54(3 - 4): 1016 - 1020.

[12] BAH M, HAFIANE A, CANALS R. CRowNet: deep network for crop row detection in UAV images[J]. IEEE Access, 2019, 8:5189 - 5200.

[13] 姜国权, 杨小亚, 王志衡, 等. 基于图像特征点粒子群聚类算法的麦田作物行检测[J]. 农业工程学报, 2017, 33(11): 165 - 170.
JIANG Guoquan, YANG Xiaoya, WANG Zhiheng, et al. Crop rows detection based on image characteristic point and particle swarm optimization-clustering algorithm[J]. Transactions of the CSAE, 2017, 33(11): 165 - 170. (in Chinese)

[14] 王侨, 孟志军, 付卫强, 等. 基于机器视觉的玉米苗期多条作物行线检测算法[J]. 农业机械学报, 2021, 52(4): 208 - 220.
WANG Qiao, MENG Zhijun, FU Weiqiang, et al. Detection algorithm of multiple crop row lines based on machine vision in maize seedling stage[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 208 - 220. (in Chinese)

[15] 王爱臣, 张敏, 刘青山, 等. 基于区域生长和均值漂移聚类的苗期作物行提取方法[J]. 农业工程学报, 2021, 37(19): 202 - 210.
WANG Aichen, ZHANG Min, LIU Qingshan, et al. Seedling crop row extraction method based on regional growth and mean shift clustering[J]. Transactions of the CSAE, 2021, 37(19): 202 - 210. (in Chinese)

[16] EVERINGHAM M, ESLAMI A, VAN L, et al. The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98 - 136.

(下转第 340 页)

13(10):715 - 723.
[25] JAYME A. Angiografia digital[J]. Arquivos Brasileiros De Oftalmologia,2018,58(5):381 - 383.
[26] ZHANG T Y,SUEN C Y. A fast parallel algorithm for thinning digital patterns[J]. Communications of the ACM,1984,27(3):236 - 239.
[27] 常庆贺,吴敏华,骆力明. 基于改进 ZS 细化算法的手写体汉字骨架提取[J]. 计算机应用与软件,2020,37(7):107 - 113,164.
CHANG Qinghe,WU Minhua,LUO Liming. Framework extraction of hand written Chinese characters based on improved ZS thinning algorithm[J]. Computer Applications and Software,2020,37(7):107 - 113,164. (in Chinese)
[28] 樊仲藜,张力,王庆栋,等. SAR 影像和光学影像梯度方向加权的快速匹配方法[J]. 测绘学报,2021,50(10):1390 - 1403.
FAN Zhongli,ZHANG Li,WANG Qingdong,et al. Fast matching method based on gradient direction weighting for SAR image and optical image[J]. Acta Geodaetica et Cartographica Sinica,2021,50(10):1390 - 1403. (in Chinese)
[29] 周小成,王锋克,黄洪宇,等. 基于无人机遥感的伐区造林坑穴数量与参数提取[J]. 农业机械学报,2021, 52(12):201 - 206.
ZHOU Xiaocheng,WANG Fengke,HUANG Hongyu,et al. Number and parameter extraction of afforestation pits in logging area based on UAV remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(12):201 - 206. (in Chinese)
[30] 周润,张新,王仁浩. 大口径光子筛的小孔环带衍射模型设计[J]. 光学学报,2019,39(10):75 - 81.
ZHOU Run,ZHANG Xin,WANG Renhao. Design of small-hole loop diffraction model for large-aperture photonic screen[J]. Acta Photonica Sinica,2019,39(10):75 - 81. (in Chinese)
[31] 曾升,耿国华,邹林波,等. 第一人称视角地形轮廓草图的真实空间重建[J]. 光学精密工程,2020,28(8):1861 - 1871.
ZENG Sheng, GENG Guohua, ZOU Linbo, et al. Real spatial reconstruction of terrain contour sketches from first person perspective[J]. Optics and Precision Engineering,2020,28(8):1861 - 1871. (in Chinese)
[32] 马志艳,朱熠,张徐康,等. 基于视觉的玉米苗期作物识别与定位方法研究[J]. 中国农机化学报,2020,41(9):131 - 137.
MA Zhiyan,ZHU Yi,ZHANG Xukang,et al. Research on recognition and location method of maize seedling based on vision [J]. Journal of Chinese Agricultural Mechanization,2020,41(9):131 - 137. (in Chinese)

(上接第 305 页)

[17] GIRSHICK R, DONAHUA J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.
[18] GIRSHICK R. Fast R - CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440 - 1448.
[19] REN S, HE K, GIRSHICK R, et al. Faster R - CNN: towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 91 - 99.
[20] HE K, GKIOXARI G, DOLLAR P, et al. Mask R - CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 - 2969.
[21] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham, 2016: 21 - 37.
[22] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[J]. arXiv preprint arXiv: 1904.07850, 2019.
[23] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 - 788.
[24] REDMON J, FARHADIS A. YOLO9000: better, faster, stronger[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263 - 7271.
[25] REDMON J, FARHADI A. YOLO v3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
[26] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934, 2020.
[27] GE Z, LIU S, WANG F, et al. YOLOX: exceeding yolo series in 2021[J]. arXiv preprint arXiv: 2107.08430, 2021.
[28] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390 - 391.
[29] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759 - 8768.

四、科研成果

1.

（1）可溯源农产品安全监控关键技术推广应用

广东省农业技术推广奖

获奖证书

为表彰在农业技术推广工作中做出贡献的单位和个人，特颁发此证书，以资鼓励。

获奖项目：可溯源农产品安全监控关键技术推广应用

奖励等级：二等奖

奖励单位：涂淑琴

奖励日期：2017 年 10 月 31 日

证 书 号：2016-2-N02-R13

广东省农业技术推广奖

评审委员会

（2）农产品安全溯源智能化关键技术研究及示范

广州市科学技术奖励

证　书

为表彰广州市科学技术奖励获得者，
特颁发此证书。

奖励类别：科学技术进步奖

奖励等级：二等奖

获奖名称：农产品安全溯源智能化关键技术研究及示范

获 奖 者：涂淑琴

广州市人民政府

2017年3月23日

通报文号：穗府〔2017〕8号
证 书 号：2016B215R04

2.

2.1.　　MaskR-CNN　　Soft-NMS

证书号第5863097号

发明专利证书

发 明 名 称：基于Mask R-CNN和Soft-NMS融合的群养粘连猪实例分割方法

发 明 人：涂淑琴;梁云;刘浩锋;黄健;庞婧;刘姝慧;庄楠

专 利 号：ZL 2019 1 1320756.3

专利申请日：2019年12月19日

专 利 权 人：华南农业大学

地 址：510642 广东省广州市天河区五山路483号

授权公告日：2023年04月07日　　授权公告号：CN 111178197 B

国家知识产权局依照中华人民共和国专利法进行审查，决定授予专利权，颁发发明专利证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。专利权期限为二十年，自申请日起算。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

国家知识产权局
2023年04月07日

其他事项参见续页

证书号第5863097号

专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年12月19日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

申请日时本专利记载的申请人、发明人信息如下：

申请人：

华南农业大学

发明人：

涂淑琴;梁云;刘浩锋;黄健;庞婧;刘姝慧;庄楠

2.2

证书号第1921554号

发明专利证书

发明名称：一种基于压缩感知的选择集成人脸识别方法

发明人：涂淑琴;薛月菊;黄晓琳

专利号：ZL 2013 1 0053703.6

专利申请日：2013年02月19日

专利权人：华南农业大学

授权公告日：2016年01月13日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年02月19日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

中华人民共和国国家知识产权局
2016年01月13日

第 1 页 (共 1 页)

2.3 RGB-D

证书号第3681106号

发明专利证书

发明名称：一种RGB-D图像分类方法及系统

发　明　人：涂淑琴;薛月菊;胡月明;梁云

专　　利　　号：ZL 2015 1 0402298.3

专利申请日：2015年07月09日

专利权人：华南农业大学

地　　址：510642 广东省广州市天河区五山路483号

授权公告日：2020年02月04日　　授权公告号：CN 105224942 B

国家知识产权局依照中华人民共和国专利法进行审查，决定授予专利权，颁发发明专利证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。专利权期限为二十年，自申请日起算。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

国家知识产权局
2020年02月04日

其他事项参见背面

证书号第3681106号

国家知识产权局 国家税务总局北京市海淀区税务局 印花税代扣专用章

专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年07月09日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

申请日时本专利记载的申请人、发明人信息如下：

申请人：

华南农业大学

发明人：

涂淑琴；薛月菊；胡月明；梁云

2.4 MaskR-CNN

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第5499324号

软件名称：基于Mask R-CNN的群养猪关键点检测系统
[简称：KP Detection of Pig BS Mask R-CNN]
V1.0

著作权人：华南农业大学

开发完成日期：2020年04月14日

首次发表日期：未发表

权利取得方式：原始取得

权利范围：全部权利

登记号：2020SR0620628

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

No. 05863511

中华人民共和国国家版权局
计算机软件著作权
登记专用章

2020年06月15日

2.5 MSR-CNN

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第5506038号

软件名称：基于MS R-CNN算法融合对抗网络的群养猪实例检测分割系统
[简称：Detection of Pig BS MS R-CNN fusion GAN]
V1.0

著作权人：华南农业大学

开发完成日期：2020年04月11日

首次发表日期：未发表

权利取得方式：原始取得

权利范围：全部权利

登记号：2020SR0627342

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局
计算机软件著作权登记专用章

No. 05866646

2020年06月15日

2.6 V1.0

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第9445417号

软件名称：百香果视频跟踪系统 V1.0

著作权人：华南农业大学

开发完成日期：2022年03月09日

首次发表日期：2022年03月09日

权利取得方式：原始取得

权利范围：全部权利

登记号：2022SR0491218

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局 计算机软件著作权登记专用章

No. 10531724

2022年04月20日

2.7 V1.0

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第9561421号

软件名称：群养猪行为识别检测系统 V1.0

著作权人：华南农业大学

开发完成日期：2022年03月20日

首次发表日期：未发表

权利取得方式：原始取得

权利范围：全部权利

登记号：2022SR0607222

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局 计算机软件著作权登记专用章

No. 10732781

2022年05月19日

2.8 RepVgg V1.1

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第9561420号

软件名称：RepVgg病虫害检测分类软件
V1.1

著作权人：华南农业大学

开发完成日期：2022年03月20日

首次发表日期：未发表

权利取得方式：原始取得

权利范围：全部权利

登记号：2022SR0607221

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。

中华人民共和国国家版权局
计算机软件著作权
登记专用章

No. 10732780

2022年05月19日

五、其他业绩

1.

1.1. C/C++ A T2

1.2. C/C++ B T2

蓝桥杯大赛

获奖证书

华南农业大学涂淑琴：

指导赖伟鑫荣获第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛C/C++程序设计大学B组二等奖，被评为优秀指导教师。

特发此证，以资鼓励。

证书编号：021454843

证件号码：360122197805063947

工业和信息化部人才交流中心

蓝桥杯大赛组委会

2023年6月16日

1.3. C/C++ B T2

蓝桥杯大赛

获奖证书

华南农业大学涂淑琴：

指导郑楚炳荣获第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛C/C++程序设计大学B组三等奖，被评为优秀指导教师。

特发此证，以资鼓励。

证书编号：021454876

证件号码：360122197805063947

工业和信息化部人才交流中心

蓝桥杯大赛组委会

2023年6月16日

1.4. C/C++ B T2

蓝桥杯大赛

获奖证书

华南农业大学涂淑琴：

指导陈浩锾荣获第十四届蓝桥杯全国软件和信息技术专业人才大赛全国总决赛C/C++程序设计大学B组三等奖，被评为优秀指导教师。

特发此证，以资鼓励。

证书编号：021454857

证件号码：360122197805063947

工业和信息化部人才交流中心

蓝桥杯大赛组委会

2023年6月16日

1.5 B

蓝桥杯大赛

获奖证书

华南农业大学涂淑琴：

指导董跃飞荣获第十四届蓝桥杯全国软件和信息技术专业人才大赛广东赛区C/C++程序设计大学A组一等奖，被评为优秀指导教师。

特发此证，以资鼓励。

证书编号：011412577

证件号码：360122197805063947

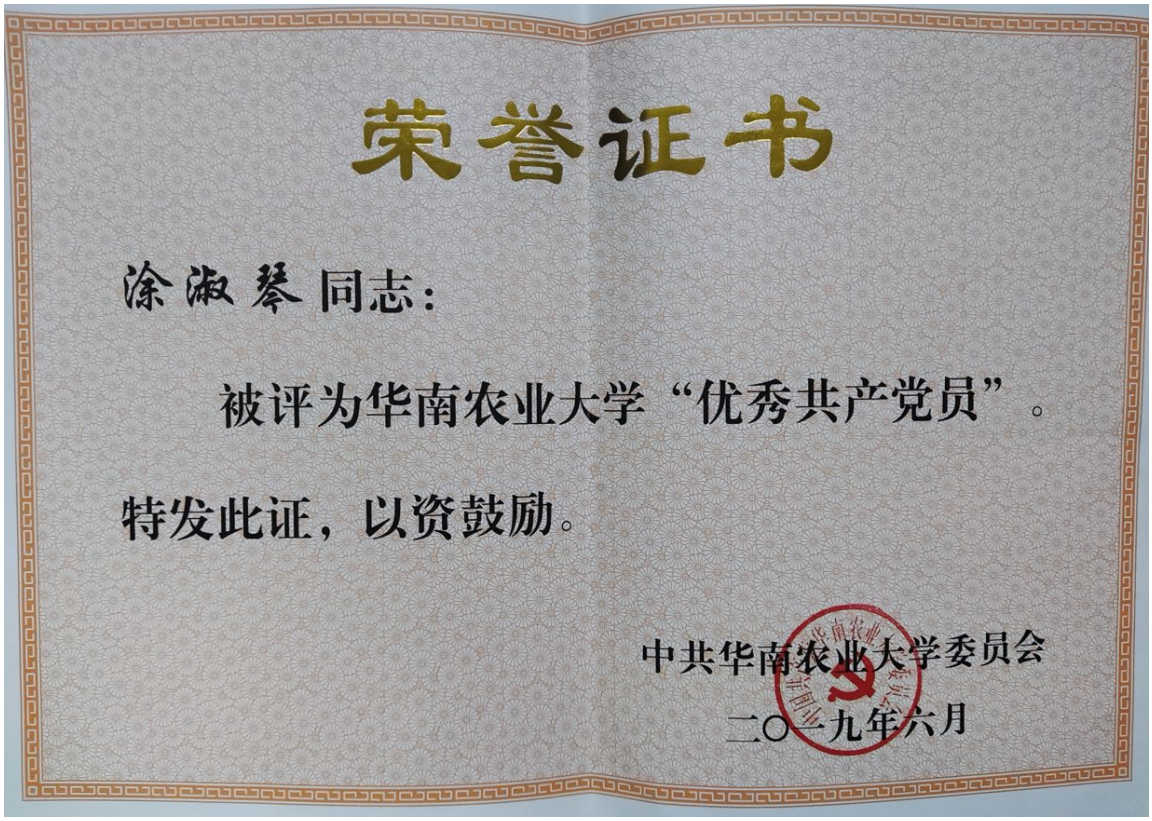
工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2023年4月23日

2.

2.1. “ ”


荣誉证书

涂淑琴 同志：

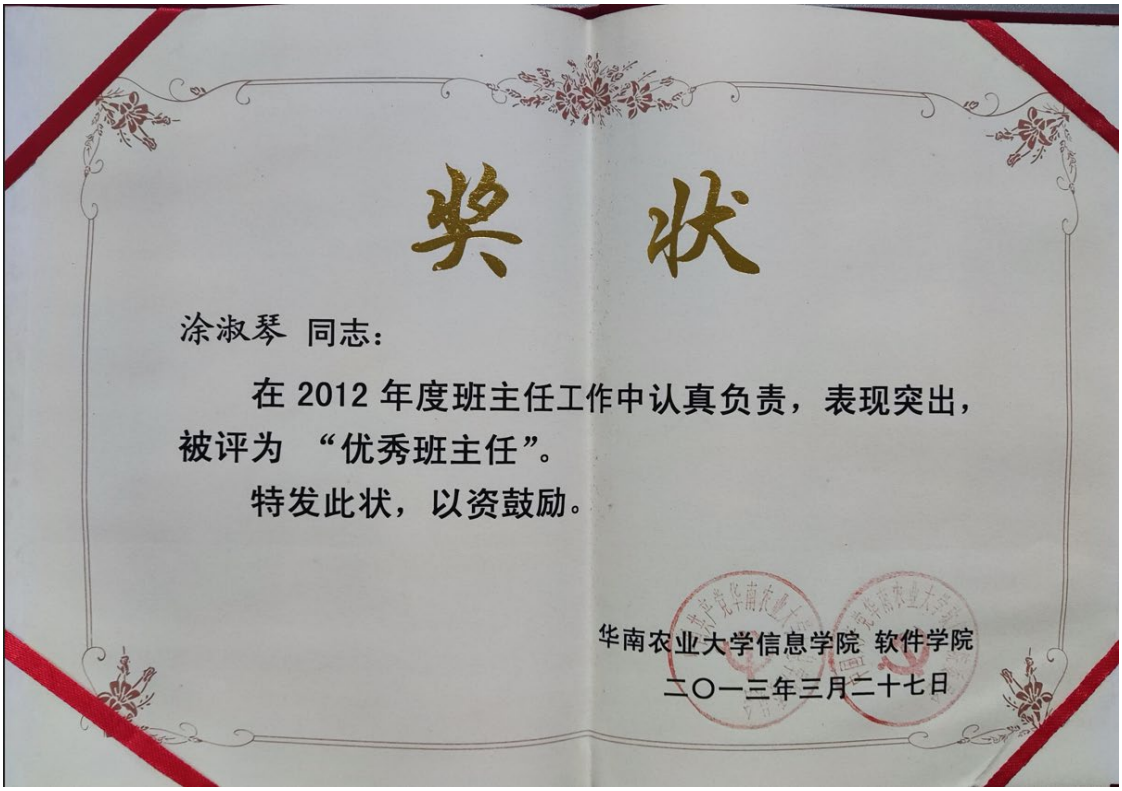
被评为华南农业大学“优秀共产党员”。

特发此证，以资鼓励。

中共华南农业大学委员会

二〇一九年六月

2.2 “ ”


奖 状

涂淑琴 同志：

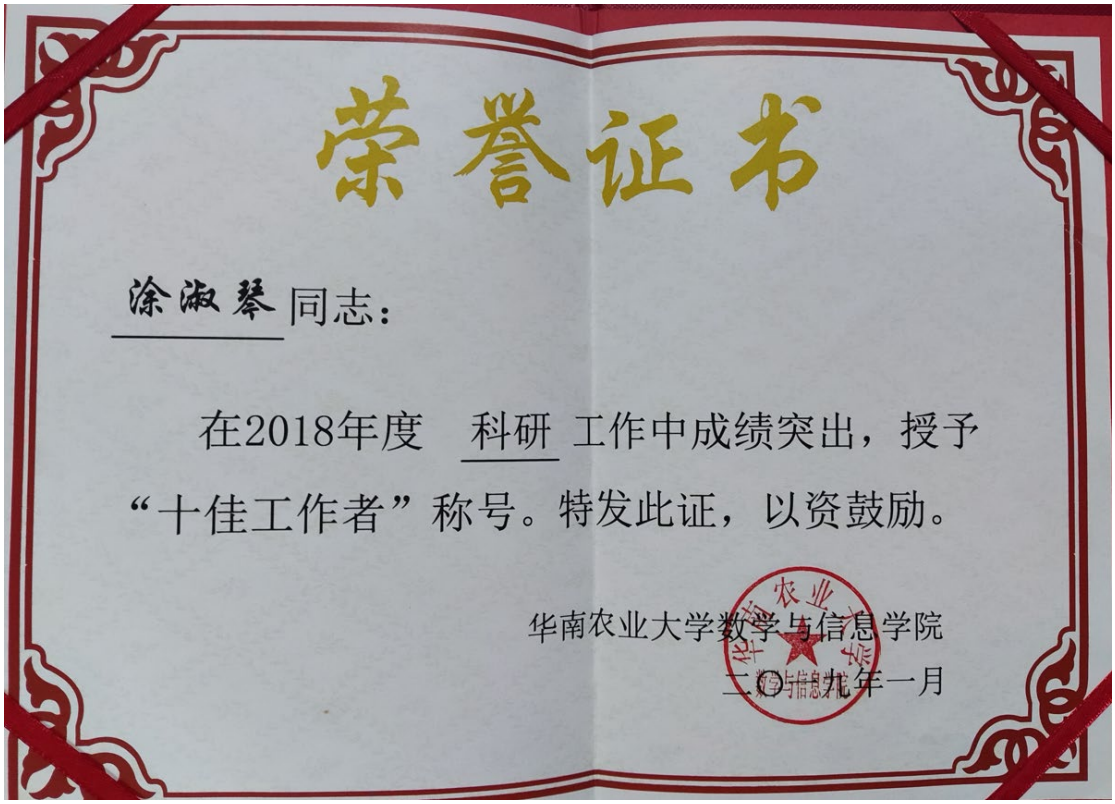
在 2012 年度班主任工作中认真负责，表现突出，被评为 “优秀班主任”。

特发此状，以资鼓励。

华南农业大学信息学院 软件学院

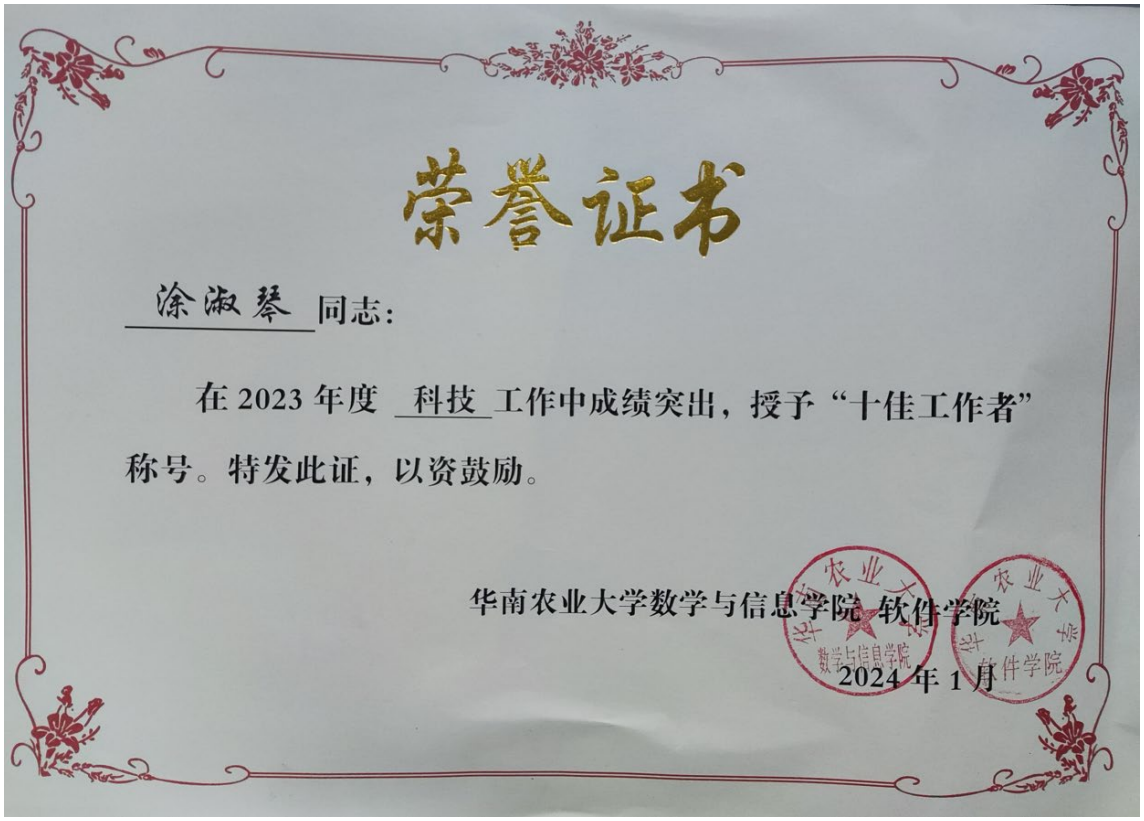
二〇一三年三月二十七日

2.3 2018 “ ”



2.4 2021

2.5 2023 “ ”



荣誉证书

涂淑琴 同志：

在 2023 年度 科技 工作中成绩突出，授予“十佳工作者”称号。特发此证，以资鼓励。

华南农业大学数学与信息学院、软件学院

2024 年 1 月