申报	系列：教师系列 教学科研并重型
	专业：计算机科学与技术
	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）数学与信息学院、软件学院

姓　名　周子涵

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

《人工智能理论及应用》

研究生教材证明函

兹证明，由梁云、林琛、涂淑琴、杨磊担任主编，周婧、周子涵、钟灿琨、蒋荣金担任副主编，周凡担任主审的《人工智能理论及应用》（ISBN：978-7-5615-9475-9，厦门大学出版社 2024 年 8 月出版），系我社正式出版的研究生层次教材。

本教材面向计算机科学与技术、人工智能、控制科学与工程等相关学科的硕士研究生，内容涵盖人工智能核心理论、前沿技术及应用场景，符合我国研究生教育的教学大纲要求与人才培养目标，已被华南农业大学数学与信息学院选作研究生课程教学用书。

特此证明。

厦门大学出版社

2026 年 3 月 4 日

人工智能
理论及应用

主 编 梁 云 林 琛 涂淑琴 杨 磊

厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

人工智能
理论及应用

主　审　周　凡

主　编　梁　云　林　琛　涂淑琴　杨　磊

副主编　周　婧　周子涵　钟灿琨　蒋荣金

厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

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

国家自然科学基金资助项目批准通知

（包干制项目）

周子涵　先生/女士：

根据《国家自然科学基金条例》、相关项目管理办法规定和专家评审意见，国家自然科学基金委员会（以下简称自然科学基金委）决定资助您申请的项目。项目批准号：62401210，项目名称：视觉-语言模型引导下的内容-退化解耦及其在图像质量评价中的应用，资助经费：30.00万元，项目起止年月：2025年01月至 2027年 12月，有关项目的评审意见及修改意见附后。

请您尽快登录科学基金网络信息系统（https://grants.nsfc.gov.cn），**认真阅读《国家自然科学基金资助项目计划书填报说明》并按要求填写《国家自然科学基金资助项目计划书》（以下简称计划书）**。对于有修改意见的项目，请您按修改意见及时调整计划书相关内容；如您对修改意见有异议，须在电子版计划书报送截止日期前向相关科学处提出。

请您将电子版计划书通过科学基金网络信息系统（https://grants.nsfc.gov.cn）提交，由依托单位审核后提交至自然科学基金委。自然科学基金委审核未通过者，将退回的电子版计划书修改后再行提交；审核通过者，打印纸质版计划书（一式两份，双面打印）并在项目负责人承诺栏签字，由依托单位在承诺栏加盖依托单位公章，且将申请书纸质签字盖章页订在其中一份计划书之后，一并报送至自然科学基金委项目材料接收工作组。纸质版计划书应当保证与审核通过的电子版计划书内容一致。**自然科学基金委将对申请书纸质签字盖章页进行审核，对存在问题的，允许依托单位进行一次修改或补齐。**

向自然科学基金委提交电子版计划书、报送纸质版计划书并补交申请书纸质签字盖章页截止时间节点如下：

1. **2024年9月9日16点：**提交电子版计划书的截止时间；

2. **2024年9月16日16点：**提交修改后电子版计划书的截止时间；

3. **2024年9月23日：**报送纸质版计划书（一式两份，其中一份包含申请书纸质签字盖章页）的截止时间。

4. **2024年10月8日：**报送修改后的申请书纸质签字盖章页的截止时间。

请按照以上规定及时提交电子版计划书，并报送纸质版计划书和申请书纸质签字盖章页，逾期不报计划书或申请书纸质签字盖章页且未说明理由的，视为自动放弃接受资助；未按要求修改或逾期提交申请书纸质签字盖章页者，将视情况给予暂缓拨付经费等处理。

附件：项目评审意见及修改意见表

国家自然科学基金委员会

2024年8月23日

项目批准号	62401210
申请代码	F0116
归口管理部门	
依托单位代码	51064208A0499-0932

国家自然科学基金

资助项目计划书

（包干制项目）

资助类别：青年科学基金项目

亚类说明：

附注说明：

项目名称：视觉-语言模型引导下的内容-退化解耦及其在图像质量评价中的应用

资助经费：30万元　　执行年限：2025.01-2027.12

负 责 人：周子涵　　BRID：05802.00.67908

通讯地址：广东省广州市天河区华南农业大学数学与信息楼514

邮政编码：510642　　电　　话：18819472687

电子邮件：zhouzihan@scau.edu.cn

依托单位：华南农业大学

联 系 人：唐家林　　电　　话：020-85280070

填表日期：2024年08月27日

国家自然科学基金委员会制

Version：1.002.341

国家自然科学基金资助项目计划书填报说明

（包干制项目）

一、项目负责人收到《国家自然科学基金资助项目批准通知》（以下简称《批准通知》）后，请认真阅读本填报说明，参照国家自然科学基金相关项目管理办法和新修订的《国家自然科学基金资助项目资金管理办法》（以下简称《资金管理办法》，请查阅国家自然科学基金委员会官方网站首页“政策法规”栏目），按《批准通知》的要求认真填写和提交《国家自然科学基金资助项目计划书》（以下简称《计划书》）。

二、填写《计划书》时要科学严谨、实事求是、表述清晰、准确。《计划书》经国家自然科学基金委员会相关项目管理部门审核批准后，将作为项目研究计划执行、检查和验收的依据。

三、《计划书》各部分填写要求如下：

（一）简表：由系统自动生成。

（二）摘要及关键词：各类获资助项目都应当填写中、英文摘要及关键词。

（三）正文：

1. 青年科学基金项目、青年学生基础研究项目：如果《批准通知》所附“项目评审意见及修改意见表”中“修改意见”栏目没有修改要求的，只需选择“研究内容和研究目标按照申请书执行”即可；如果《批准通知》中上述栏目明确要求调整研究期限或研究内容等的，须选择“根据研究方案修改意见更改”并填报相关修改内容。

2. 国家杰出青年科学基金项目和优秀青年科学基金项目按下列提纲撰写：

（1）研究方向；

（2）结合国内外研究现状，说明研究工作的学术思想和科学意义（限两个页面）；

（3）研究内容、研究方案及预期目标（限两个页面）；

（4）年度研究计划；

四、资助经费相关要求：

1. 资助经费批准时不再区分直接费用和间接费用。

2. 项目负责人在提交计划书时需签署承诺书，承诺尊重科研规律，弘扬科学家精神，遵守科研伦理道德和作风学风诚信要求，认真开展科学研究工作；承诺项目经费全部用于与本项目研究工作相关的支出，不得用于与本项目研究无关的支出。

3. 项目负责人提交计划书时，无需编制项目预算。项目资金由项目负责人自主决定使用，按照《资金管理办法》第九条规定的开支范围列支。有关管理费用的补助支出，由依托单位根据实际管理需要，在充分征求项目负责人意见基础上合理确定。绩效支出由项目负责人根据实际科研需要和相关薪酬标准自主确定，依托单位按照工资制度进行管理。对于青年学生基础研究项目，支付给项目负责人本人的劳务费用，应符合相关比例要求。其余用途经费无额度限制，由项目负责人根据实际需要自主决定使用。

4. 项目结题时，项目负责人根据实际使用情况编制项目经费决算，经依托单位财务、科研管理部门审核后，报自然科学基金委。依托单位应当在单位内部公开非涉密项目立项、主要研究人员、资金使用（重点是间接费用、外拨资金、结余资金使用等）、决算、大型仪器设备购置以及项目研究成果等情况，接受内部监督。
5. 自然科学基金委结合项目管理，对经费使用情况和依托单位管理情况定期开展抽查。

简表

项目负责人信息	姓　　名	周子涵	性 别	女	出生年月	1996年02月	民 族	汉族
	学　　位	博士			职称	讲师		
	是否在站博士后	否			电子邮件	zhouzihan@scau.edu.cn		
	电　　话	18819472687			个人网页	https://zzihan.mysxl.cn/		
	工 作 单 位	华南农业大学						
	所 在 院 系 所	数学与信息(软件)学院						
依托单位信息	名　　称	华南农业大学					代码	51064208A0499
	联 系 人	唐家林			电子邮件	kycjhk@scau.edu.cn		
	电　　话	020-85280070			网站地址	http://kjc.scau.edu.cn/		
合作单位信息	单 位 名 称							
项目基本信息	项 目 名 称	视觉-语言模型引导下的内容-退化解耦及其在图像质量评价中的应用						
	资 助 类 别	青年科学基金项目			亚 类 说 明			
	附 注 说 明							
	申 请 代 码	F0116:图像信息处理						
	基 地 类 别							
	执 行 年 限	2025.01-2027.12						
	资 助 经 费	30万元						

项目摘要

中文摘要：

随着数字图像的广泛应用,自动化评估图像质量以指导后续应用变得日益迫切。然而,准确评价复杂真实场景中图像的视觉质量仍是一个极具挑战性的问题。图像质量由内容和退化共同决定。但由于两者高度耦合,现有方法难以有效分离和分析其对质量的影响,限制了评价模型的可解释性和准确性。为此,本项目拟结合语言模型,引入文本-图像跨模态监督,并设计合适的信息表征约束,以推进内容-退化解耦驱动的无参考图像质量评价研究。研究拟从以下方面展开:⑴研究面向图像质量的内容-退化描述的自动生成策略,建立退化图像的质量描述数据集；⑵研究视觉-语言模型引导的内容-退化解耦,设计信息约束策略,以进行准确的特征解耦；⑶研究多模态内容-退化特征的动态交互聚合策略,设计自适应机制,以提高评价模型在真实场景的泛化能力。此项目不仅推动图像质量评价技术的发展及其在计算机视觉领域的广泛应用,还可丰富图像表达,提升图像恢复等任务性能。

Abstract：

With the widespread application of digital images, automatic assessment of image quality to guide subsequent applications has become increasingly urgent. However, accurately evaluating the visual quality of images in complex real-world scenarios remains a highly challenging problem. Image quality is jointly determined by content and degradation. Yet, due to their high coupling, existing methods struggle to effectively separate and analyze their influence on quality, limiting the interpretability and accuracy of assessment models. To address this, this project aims to integrate language models, introduce text-image cross-modal supervision, and design appropriate information representation constraints to advance research on content-degradation decoupling-driven no-reference image quality assessment. The research will focus on the following aspects: (1) Investigating automatic generation strategies for content-degradation descriptions oriented towards image quality, and establishing a quality description dataset for degraded images; (2) Exploring content-degradation decoupling guided by vision-language models, designing information constraint strategies for accurate feature disentanglement; (3) Studying dynamic interaction and aggregation strategies for multimodal content-degradation features, and designing adaptive mechanisms to improve the generalization ability of assessment models in real-world scenarios. This project not only promotes the development of image quality assessment techniques and their widespread application in the field of computer vision but also enriches image representation and enhances the performance of tasks such as image restoration.

关键词(用分号分开)：图像质量评价；视觉-语言模型；特征解耦；动态神经网络；无参考评价

Keywords(用分号分开)： Image quality assessment; Visual-language model; Feature decoupling; Dynamic neural network; No-reference assessment

报告正文

研究内容和研究目标按照申请书执行。

国家自然科学基金项目负责人、依托单位承诺书

国家自然科学基金项目负责人承诺书

本人郑重承诺：我接受国家自然科学基金的资助，严格遵守中共中央办公厅、国务院办公厅《关于进一步加强科研诚信建设的若干意见》《关于进一步弘扬科学家精神加强作风和学风建设的意见》《关于加强科技伦理治理的意见》《科技伦理审查办法（试行）》等规定，和国家自然科学基金委员会关于资助项目管理、项目资金管理等各项规章，在《计划书》填写及项目执行过程中：

（一）按照《批准通知》《国家自然科学基金资助项目计划书填报说明》的要求填写《计划书》，未自行降低、更改目标任务或约定要求，或缩减研究（研制）内容；

（二）树立“红线”意识，严格履行科研合同义务，按照《计划书》负责实施本项目（批准号：62401210），切实保证研究工作时间，按时报送有关材料，及时报告重大情况变动，不违规将科研任务转包、分包他人，不以项目实施周期外或不相关成果充抵交差；

（三）遵守科研诚信、科技伦理规范和学术道德，认真开展研究工作，对资助项目发表的论著和取得的研究成果按规定进行标注，不在非本项目资助的成果或其他无关成果上标注本项目批准号，反对无实质学术贡献者“挂名”，不在成果署名、知识产权归属等方面侵占他人合法权益，并如实报告本人及项目组成员发生的违背科研诚信要求的任何行为；

（四）尊重科研规律，弘扬科学家精神，严谨求实，追求卓越，反对浮夸浮躁、投机取巧，不人为夸大学术或技术价值，不传播未经科学验证的现象和观点；

（五）将项目资金全部用于与本项目研究工作相关的支出，并结合科研活动需要，科学合理安排项目资金支出进度；

（六）做好项目组成员的教育和管理，确保遵守以上相关要求。

如违背上述承诺，本人愿接受国家自然科学基金委员会和相关部门做出的各项处理决定。

项目负责人（签字）：

年　　月　　日

国家自然科学基金项目依托单位承诺书

我单位同意承担上述国家自然科学基金项目，将保证项目负责人及其研究队伍的稳定和研究项目实施所需的条件，严格遵守中共中央办公厅、国务院办公厅《关于进一步加强科研诚信建设的若干意见》《关于进一步弘扬科学家精神加强作风和学风建设的意见》《关于加强科技伦理治理的意见》《科技伦理审查办法（试行）》等规定，和国家自然科学基金委员会有关资助项目管理、项目资金管理、科研诚信管理和科技伦理管理等各项规定，并督促实施。

依托单位（公章）

年　　月　　日

国家自然科学基金资助项目签批审核表

本栏目由自然科学基金委填写	科学处审查意见： 负责人（签章）： 年　　月　　日
	科学部审查意见： 负责人（签章）： 年　　月　　日

受理编号：c25140500000666

项目编号：2025A1515011539

文件编号：粤基金字〔2025〕10号

广东省基础与应用基础研究基金项目

任务书

项目名称：基于语言协同式特征解耦的无参考图像质量评价方法研究

项目类别：广东省自然科学基金-面上项目

项目起止时间：2025-01-01　至　2027-12-31

管理单位（甲方）：广东省基础与应用基础研究基金委员会

依托单位（乙方）：华南农业大学

通讯地址：广东省广州市天河区五山路483号

邮政编码：510642　　单位电话：020-85283435

项目负责人：周子涵　　联系电话：18819472687

（广东科技微信公众号）　　（查看任务书信息）　　（受理纸质材料二维码）

广东省基础与应用基础研究
基金委员会
二〇二〇年制

填写说明

一、项目任务书内容原则上要求与申报书相关内容保持一致，不得无故修改。

二、项目承担单位通过广东省科技业务管理阳光政务平台下载项目任务书，按要求完成签名盖章后扫描上传到广东省科技业务管理阳光政务平台。

三、签名盖章说明。请分别在单位工作分工及经费分配情况页、人员信息页、签约各方页等地方按要求签字或盖章，签章不合规或错漏将不予受理。其中，人员信息页要求所有参与人员本人亲笔签名，代签或印章无效，漏签将不予受理。

四、本任务书自签字并加盖公章之日起生效，各方均应负本任务书的法律责任，不应受机构、人事变动影响。

五、根据《广东省科学技术厅广东省财政厅关于深入推进省基础与应用基础研究基金项目经费使用“负面清单+包干制”改革试点工作的通知》(粤科规范字〔2022〕2号)，2022年度及以后立项资助的全部省基金项目(包括省自然科学基金、省市联合基金、省企联合基金项目等)均适用“负面清单+包干制”，项目提交申请书和任务书时无需编制费用明细科目预算。

一、主要研究内容和要达到的目标

主要研究内容：

本项目从人类视觉感知特性和无参考质量评价需求出发，通过挖掘面向图像质量的语言先验，从跨模态质量评价数据的分析和生成、解耦模型的构建及其在图像质量评价中的应用三个层面进系统性研究，以提高无参考评价模型在复杂环境下的准确性和泛化性，为图像质量评价领域开辟新思路。本项目拟从以下三个方面展开研究：

（1）针对面向图像质量的内容-退化的文本描述，研究自动生成策略，构建质量描述文本生成新范式。通过设计多维度图像质量描述文本的生成模型，为内容-退化解耦和图像质量评价提供更为丰富和准确的数据支持。

（2）针对内容-退化特征分解，研究语言协同引导的图像内容-退化解耦模型。通过设计语言信息协同约束的对比重构解耦框架，并结合视觉-语言模型，实现有效的内容-退化特征解耦。

（3）针对基于内容-退化分解的图像质量评价，研究多模态内容-退化特征的动态交互聚合策略，通过建立内容-退化驱动的动态交互机制,以显式建模两者对质量的交互影响。通过建立跨模态多任务学习框架，并设计自适应机制，以增强评价模型在复杂真实场景下的准确性和泛化能力。

研究目标：

本项目针对真实复杂场景下的无参考图像质量评价问题，对语言协同式内容-退化解耦机制及其在图像质量评价中的应用展开深入研究，旨在推动图像质量评价技术的发展及其在计算机视觉领域的广泛应用，并丰富图像表达，从而提升图像恢复等任务性能。具体目标包括：

（1）建立质量描述文本生成新范式，并提出一个多维度图像质量描述文本的生成模型，为内容-退化解耦和图像质量评价任务提供数据支持。

（2）提出视觉语言模型引导的图像内容-退化解耦模型，设计语言信息协同约束的对比重构解耦框架，以实现内容和退化的有效解耦。

（3）结合迁移学习技术，提出基于内容-退化特征动态交互的图像质量评价方法。通过设计自适应策略提高图像质量评价模型在真实场景的准确性和泛化能力。

（4） 整合以上成果，提出一套完整有效的基于特征解耦的无参考质量评价方法或系统。

基于上述研究成果，本项目旨在形成科技报告一篇，申请专利1项，发表 JCR 二区及以上学术期刊和 CCF-A 类会议论文共4篇。协助培养硕士3名。

二、项目预期获得的研究成果及形式

<table>
<tr><td rowspan="3">论文及专著情况</td><td colspan="2">国家统计源刊物以上刊物发表论文（篇）</td><td colspan="2">3</td><td colspan="2">科技报告（篇）</td><td colspan="2">1</td></tr>
<tr><td colspan="2">其中被SCI/EI/ISTP收录论文数（篇）</td><td colspan="2">0</td><td colspan="2">培养人才（人）</td><td colspan="2">3</td></tr>
<tr><td colspan="2">专著（册）</td><td colspan="2">0</td><td colspan="2">引进人才（人）</td><td colspan="2">0</td></tr>
<tr><td rowspan="3">专利情况(项)</td><td colspan="2">发明专利</td><td colspan="2">实用新型专利</td><td colspan="2">外观设计专利</td><td colspan="2">国外专利</td></tr>
<tr><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请</td><td>授权</td></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
</table>

三、项目进度和阶段目标

（一）项目起止时间： 2025-01-01 至 2027-12-31		
（二）项目实施进度及阶段主要目标:		
开始日期	结束日期	主要工作内容
2025-01-01	2025-12-31	收集和整理相关资料，调研最新研究进展，完善实验方案。完善复合退化函数模型，收集和准备无主观标记的合成图像数据集和有主观分数的图像质量评价数据集。针对上述视觉图像数据，细化结构化思考链和提示文本内容，微调大语言模型以构建多维度质量描述语料库，重点包括内容和退化描述。基于该方法开发质量描述文本生成软件原型系统，发表高水平学术论文1-2篇，并申请专利或软件著作权。
2026-01-01	2026-12-31	完善自监督学习框架的设计，包括细化内容编/解码器和退化编/解码器的结构。细化面向语言模型和图像质量的对比损失函数和训练方案，进行视觉内容-退化解耦网络的训练和验证。汇总本阶段和上阶段成果，发表高水平学术论文1-2篇，并申请专利。
2027-01-01	2027-12-31	进一步对相关领域进行调研，完善视觉-语言多模态特征融合机制、内容和退化特征动态交互机制和动态质量分数回归机制，并验证有效性。结合之前成果，将语言先验驱动的内容-退化解耦模型应用在无参考图像质量评价中，完善针对多模态的多任务框架，细化损失函数及多阶段训练策略，并验证有效性。汇总本阶段研究成果，发表高水平学术论文2-3篇，并申请专利。总结该项目在理论方法研究，模型设计和实际应用效果等方面所取得的成果，并提出与此项目相关的后续研究内容和发展方向。

四、项目总经费及省基金委经费预算

1.省基金委经费下达总额：　（大写）壹拾万圆整；（小写　）10万元；					
2.省基金委经费年度下达计划：					
年度	2025 年	年	年	年	年
经费(万元)	10.00				

六、工作分工及财政经费分配

承担/参与单位名称（盖章）	工作分工	省级财政科技资金分配（万元）
华南农业大学	本项目团队将全面负责项目的开展和进行，对各个研究内容进行统筹规划，对各项任务和其科研成果进展进行全面的把控。本团队还将负责科研任务的相关数据收集和处理、模型设计、算法设计以及程序设计。	10.00
	合计	10.00

五、人员信息

项目负责人

姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
周子涵	342601199602060028	29	女	讲师	博士研究生	项目负责人	华南农业大学	

项目组主要成员

姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
黄立峰	431202199002010815	35	男	讲师	博士研究生	模型设计	华南农业大学	
钟灿琨	445202199502277730	30	男	讲师	博士研究生	算法设计	华南农业大学	
肖新杰	460006200005106817	25	男	未取得	本科	实验测试	华南农业大学	
李亮辉	440902199711303230	28	男	未取得	本科	实验测试	华南农业大学	
肖雨婷	420112200011282740	25	女	未取得	本科	数据管理和清洗	华南农业大学	

七、任务书条款

第一条 甲方与乙方根据《中华人民共和国民法典》及国家有关法规和规定，按照《广东省自然科学基金及联合基金项目管理实施细则》（粤科规范字〔2024〕5号）《省级科技计划项目任务书管理细则》（粤科规范字〔2022〕8号）等规定，为顺利完成（2025）年基于语言协同式特征解耦的无参考图像质量评价方法研究专项项目（项目编号：2025A1515011539）经协商一致，特订立本任务书，作为甲乙双方在项目实施管理过程中共同遵守的依据。

第二条 甲方的权利义务：1.按任务书规定进行经费核拨的有关工作协调。2.根据甲方需要，在不影响乙方工作的前提下，定期或不定期对乙方项目的实施情况和经费使用情况进行检查或抽查。3.根据《广东省科学技术厅科技计划项目科研诚信管理办法》（粤科规范字〔2024〕2号）《广东省基础与应用基础研究基金项目科研不端行为调查处理实施细则（试行）》（粤科规范字〔2023〕1号）等规定对乙方进行科技计划信用管理。

第三条 乙方的权利义务：1.确保落实自筹经费及有关保障条件。2.按任务书规定，对甲方核拨的经费实行专款专用，单独列账，并随时配合甲方进行监督检查。3.经费使用按照广东省级财政科研项目经费使用及省基金项目经费使用“负面清单+包干制”等有关规定进行管理。4.项目依托单位应制定经费使用“负面清单+包干制”内部管理制度并报甲方备案。5.使用财政资金采购设备、原材料等，按照《广东省实施〈中华人民共和国招标投标法〉办法》有关规定，符合招标条件的须进行招标。6.项目任务书任务完成后，或任务书规定的任务、指标及经费投入等提前完成的，乙方可提出验收结题申请，并按甲方要求做好项目验收结题工作。7.若项目发生需要终止结题的情况，乙方须提出终止结题申请，并按甲方要求做好项目终止结题工作。8.在每年规定时间内向甲方如实提交上年度工作情况报告，报告内容包含上年度项目进展情况、经费决算和取得的成果等。9.按照国家和省有关规定，提交科技报告及其他材料。10.利用甲方的经费获得的研究成果，项目负责人和参与者应当注明获得“广东省基础与应用基础研究基金（英文：Guangdong Basic and Applied Basic Research Foundation）（项目编号）”资助或作有关说明。11.乙方要恪守科学道德准则，遵守科研活动规范，践行科研诚信要求，不得抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；不得购买、代写、代投论文，虚构同行评议专家及评议意见；不得违反论文署名规范，擅自标注或虚假标注获得科技计划（专项、基金等）等资助；不得弄虚作假，骗取科技计划（专项、基金等）项目、科研经费以及奖励、荣誉等；不得有其他违背科研诚信要求的行为。12.确保本项目开展的研究工作符合我国科技伦理管理相关规定。

第四条 在履行本任务书的过程中，如出现广东省相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。

第五条 在履行本任务书的过程中，当事人一方发现可能导致项目整体或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第六条 本项目技术成果的归属、转让和实施技术成果所产生的经济利益的分享，除双方另有约定外，按国家和广东省有关法规执行。

第七条 根据项目具体情况，经双方另行协商订立的附加条款，作为本任务书正式内容的一部分，与本任务书具有同等效力。

第八条 本任务书一式三份，各份具有同等效力。甲、乙方及项目负责人各执一份，三方签字、盖章后即生效，有效期至项目结题后一年内。各方均应负任务书的法律责任，不应受机构、人事变动的影响。

第九条 乙方必须接受甲方聘请的本项目任务书监理单位的监督和管理。监理单位按照甲方赋予的权利对本项目任务书的履行进行审核、进度调查，对项目任务书变更、经费使用情况进行监督管理及组织项目验收。

说明：1.本任务书中，凡是当事人约定无需填写的内容，应在空白处划（/）。

2.委托代理人签订本任务书的，应出具合法、有效的委托书。

八、本任务书签约各方

管理单位（甲方）：　　广东省基础与应用基础研究基金委员会 （盖章）

法定代表人（或法人代理）：　曾路　（签章）

2025 年　03 月　21 日

依托单位（乙方）：　华南农业大学　（盖章）

法定代表人（或法人代理）：　薛红卫　（签章）

联系人（项目主管）姓名：　夏杰　（签章）

Email：　kjcgxk@scau.edu.cn

电话：　020-85283435　/　13711345768

开户单位名称：　华南农业大学

开户银行名称：　广东广州工行五山支行

开户银行账号：　3602002609000310520

2025 年 4 月 3 日

联系人（项目负责人）姓名：　周子涵　（签名）

Email：　joannezzh@outlook.com

电话：　18819472687

25 年 4 月 3 日

受理编号：c232019102400000948

项目编号：2023A1515110646

文件编号：粤基金字〔2024〕4号

广东省基础与应用基础研究基金项目

任务书

项目名称：基于深度自适应机制的无参考图像质量评价方法研究

项目类别：区域联合基金-青年基金项目

项目起止时间：2023-11-01　至　2026-10-31

管理单位（甲方）：广东省基础与应用基础研究基金委员会

依托单位（乙方）：华南农业大学

通讯地址：广东省广州市天河区五山路483号

邮政编码：510642　　单位电话：020-85283435

项目负责人：周子涵　　联系电话：18819472687

（广东科技微信公众号）　　（查看任务书信息）　　（受理纸质材料二维码）

广东省基础与应用基础研究
基金委员会
二〇二〇年制

填写说明

一、项目任务书内容原则上要求与申报书相关内容保持一致，不得无故修改。

二、项目承担单位通过广东省科技业务管理阳光政务平台下载项目任务书，按要求完成签名盖章后扫描上传到广东省科技业务管理阳光政务平台。

三、签名盖章说明。请分别在单位工作分工及经费分配情况页、人员信息页、签约各方页等地方按要求签字或盖章，签章不合规或错漏将不予受理。其中，人员信息页要求所有参与人员本人亲笔签名，代签或印章无效，漏签将不予受理。

四、本任务书自签字并加盖公章之日起生效，各方均应负本任务书的法律责任，不应受机构、人事变动影响。

五、根据《广东省科学技术厅广东省财政厅关于深入推进省基础与应用基础研究基金项目经费使用“负面清单+包干制”改革试点工作的通知》(粤科规范字〔2022〕2号)，2022年度及以后立项资助的全部省基金项目(包括省自然科学基金、省市联合基金、省企联合基金项目等)均适用“负面清单+包干制”，项目提交申请书和任务书时无需编制费用明细科目预算。

一、主要研究内容和要达到的目标

1. 主要研究内容：

从“预训练-训练-测试”流程范式入手，分析人类视觉感知特性和无参考质量评价需求，结合迁移学习的思想和技术，针对范式中的三个阶段分别进行自适应基础、模型自适应及域自适应的研究，以在不同阶段以输入为条件生成/更新更符合人类视觉感知的特征提取规则和质量预测规则，从而显著提升面向真实场景的质量预测泛化性能和稳定性。研究内容包括：

1）基于内容-退化分解的预训练模型

2）内容感知的动态质量评价网络

3）质量感知的自监督测试时间自适应方案

2. 研究目标：

针对真实复杂场景下的无参考图像质量评价问题，对深度自适应机制及模型展开深入研究。具体目标包括：

1）提出有效的基于内容-退化分解的预训练模型

通过设计有效的自监督学习框架，以有效引导模型学习自适应的内容和退化表示，从而提高预训练模型的区分性内容和退化表征能力，为后续自适应机制下模型微调提供更好的初始化模型架构和参数。

2）提出内容感知的动态卷积网络

通过模型自适应以提高模型在训练和测试阶段适应图像内容的特征表征与分数回归能力。通过显式地建模根据图像内容更新滤波器和回归器参数的机制，建立动态滤波模块和动态分数回归规则模块，以实现动态的特征提取和分数回归。

3）提出质量感知的自监督测试时间自适应方法

通过域自适应以提高模型对真实测试图像的泛化能力。通过设计质量感知的自监督辅助任务和损失函数，以有效利用无标签测试图像更新源模型，减轻源训练数据和测试数据分布之间的偏移，从而实现适应测试图像的特征提取。

4） 整合以上成果，提出一套完整有效的基于深度自适应机制的无参考质量评价方法或系统。

二、项目预期获得的研究成果及形式

<table>
<tr><td rowspan="3">论文及专著情况</td><td colspan="2">国家统计源刊物以上刊物发表论文（篇）</td><td colspan="2">3</td><td colspan="2">科技报告（篇）</td><td colspan="2">1</td></tr>
<tr><td colspan="2">其中被SCI/EI/ISTP收录论文数（篇）</td><td colspan="2">3</td><td colspan="2">培养人才（人）</td><td colspan="2">1</td></tr>
<tr><td colspan="2">专著（册）</td><td colspan="2">0</td><td colspan="2">引进人才（人）</td><td colspan="2">0</td></tr>
<tr><td rowspan="3">专利情况(项)</td><td colspan="2">发明专利</td><td colspan="2">实用新型专利</td><td colspan="2">外观设计专利</td><td colspan="2">国外专利</td></tr>
<tr><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请</td><td>授权</td></tr>
<tr><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>其他</td><td colspan="8">无</td></tr>
</table>

三、项目进度和阶段目标

（一）项目起止时间： 2023-11-01 至 2026-10-31		
（二）项目实施进度及阶段主要目标:		
开始日期	结束日期	主要工作内容
2023-11-01	2024-06-30	对学术界在无参考图像质量评估方面的最新动态和研究进展进行调研。搭建数据采集环境,严格控制实验条件,模拟真实退化,生成大量具有丰富内容和多变/复杂退化的无主观标记的退化图像。进一步完善自适应图像内容和退化的预训练方案的设计,并进行简单论证。接着针对确定好的方案,进行实验方案的细化和实施,并进行可靠的图像质量评估实验验证。整理本阶段成果,申请专利1项。
2024-07-01	2025-04-30	完善内容感知的动态滤波器生成模块和动态质量回归模块的设计，并验证其有效性。细化内容自适应动态网络的设计方案，完善前面的研究结果并进行汇总整理，提出完整的基于动态网络无参考图像质量评价方法。整理本阶段成果，发表高水平学术论文1篇。
2025-05-01	2026-02-28	完善质量感知的自辅助任务和目标函数的设计，细化面向图像质量评估的测试时间自适应方案，并进行简单验证。结合多组质量评估源模型和不同目标域数据，进行测试时间自适应方案的调整，并最终提出一个有效的质量感知驱动的自监督质量评价方法。汇总本阶段成果，发表高水平学术论文1 篇。
2026-03-01	2026-10-31	汇总在预训练-训练-测试三个阶段的成果，提出一套完整系统的基于深度自适应机制的去参考图像质量评估方法。有机结合各个阶段，并分别验证各种组合在图像质量评估中的有效性,。汇总项目研究成果，申请专利1项；总结该项目在理论方法研究、模型设计和实际应用效果等方面所取得的成果，并提出与此项目相关的后续研究内容和发展方向。

四、项目总经费及省基金委经费预算

（一）省基金委经费下达总额：　（大写）壹拾万圆整；（小写 ）10万元；					
（二）省基金委经费年度下达计划：					
年度	2023 年	年	年	年	年
经费(万元)	10.00				

五、人员信息

项目负责人								
姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
周子涵	342601199602060028	28	女	未取得	博士研究生	项目负责人	华南农业大学	周子涵

六、工作分工及财政经费分配

承担/参与单位名称 （盖章）	工作分工	省级财政科技资金分配 （万元）
华南农业大学	本人作为项目负责人，全面负责项目的开展和进行，对各项任务进行统筹规划，对论文进展进行全面的把控。本人还将负责科研任务的相关数据收集和处理、模型设计、算法设计以及程序设计。	10
	合计	10

七、任务书条款

第一条 甲方与乙方根据《中华人民共和国民法典》及国家有关法规和规定，按照《广东省科学技术厅关于广东省基础与应用基础研究基金（省自然科学基金、联合基金等）项目管理的实施细则（试行）》《省级科技计划项目任务书管理细则》《广东省省级科技计划项目验收结题工作规程（试行）》等规定，为顺利完成（2023）年基于深度自适应机制的无参考图像质量评价方法研究专项项目（项目编号：2023A1515110646）经协商一致，特订立本任务书，作为甲乙双方在项目实施管理过程中共同遵守的依据。

第二条 甲方的权利义务：

1.按任务书规定进行经费核拨的有关工作协调。

2.根据甲方需要，在不影响乙方工作的前提下，定期或不定期对乙方项目的实施情况和经费使用情况进行检查或抽查。

3.根据《广东省科研诚信管理办法(试行)》等规定对乙方进行科技计划信用管理。

第三条 乙方的权利义务：

1.确保落实自筹经费及有关保障条件。

2.按任务书规定，对甲方核拨的经费实行专款专用，单独列账，并随时配合甲方进行监督检查。

3.经费使用按照广东省级财政科研项目经费使用等有关规定进行管理。

4.项目依托单位应制定经费使用“负面清单+包干制”内部管理制度并报甲方备案。

5.使用财政资金采购设备、原材料等，按照《广东省实施〈中华人民共和国招标投标法〉办法》有关规定，符合招标条件的须进行招标。

6.项目任务书任务完成后，或任务书规定的任务、指标及经费投入等提前完成的，乙方可提出验收结题申请，并按甲方要求做好项目验收结题工作。

7.若项目发生需要终止结题的情况，乙方须提出终止结题申请，并按甲方要求做好项目终止结题工作。

8.在每年规定时间内向甲方如实提交上年度工作情况报告，报告内容包含上年度项目进展情况、经费决算和取得的成果等。

9.按照国家和省有关规定，提交科技报告及其他材料。

10.利用甲方的经费获得的研究成果，项目负责人和参与者应当注明获得“广东省基础与应用基础研究基金（英文：Guangdong Basic and Applied Basic Research Foundation）（项目编号）”资助或作有关说明。

11.乙方要恪守科学道德准则，遵守科研活动规范，践行科研诚信要求，不得抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；不得购买、代写、代投论文，虚构同行评议专家及评议意见；不得违反论文署名规范，擅自标注或虚假标注获得科技计划（专项、基金等）等资助；不得弄虚作假，骗取科技计划（专项、基金等）项目、科研经费以及奖励、荣誉等；不得有其他违背科研诚信要求的行为。

12.确保本项目开展的研究工作符合我国科研伦理管理相关规定。

第四条 在履行本任务书的过程中，如出现广东省相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。

第五条 在履行本任务书的过程中，当事人一方发现可能导致项目整体或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第六条 本项目技术成果的归属、转让和实施技术成果所产生的经济利益的分享，除双方另有约定外，按国家和广东省有关法规执行。

第七条 根据项目具体情况，经双方另行协商订立的附加条款，作为本任务书正式内容的一部分，与本任务书具有同等效力。

第八条 本任务书一式三份，各份具有同等效力。甲、乙方及项目负责人各执一份，三方签字、盖章后即生效，有效期至项目结题后一年内。各方均应负任务书的法律责任，不应受机构、人事变动的影响。

第九条 乙方必须接受甲方聘请的本项目任务书监理单位的监督和管理。监理单位按照甲方赋予的权利对本项目任务书的履行进行审核、进度调查，对项目任务书变更、经费使用情况进行监督管理及组织项目验收。

说明：1.本任务书中，凡是当事人约定无需填写的内容，应在空白处划（/）。

2.委托代理人签订本任务书的，应出具合法、有效的委托书。

八、本任务书签约各方

管理单位（甲方）：　　广东省基础与应用基础研究基金委员会　（盖章）

法定代表人（或法人代理）：　曾路　（签章）

2024 年　04 月　08 日

依托单位（乙方）：　华南农业大学　（盖章）

法定代表人（或法人代理）：　薛红卫　（签章）

联系人（项目主管）姓名：　倪慧群　（签章）

Email：　kjcgxk@scau.edu.cn

电话：　020-85283435　/　15920301530

开户单位名称：　华南农业大学

开户银行名称：　广东广州工行五山支行

开户银行账号：　3602002609000310520

2024 年 4 月 10 日

联系人（项目负责人）姓名：　周子涵　（签名）

Email：　joannezzh@outlook.com

电话：　18819472687

2024 年 4 月 10 日

任务书编号：2025A04J4502

广州市科技计划项目

任务书

项目名称：	基于自适应特征分解与聚合的图像质量评价研究
承担单位：	华南农业大学
项目负责人：	周子涵
计划类别：	基础研究计划
专题名称：	2025年度基础与应用基础研究专题
支持方向：	青年博士“启航”项目
组织单位：	华南农业大学
起止时间：	2025-01-01 至 2026-12-31
主管处室：	引进智力管理处（科技人才处）

广州市科学技术局制

二〇二五年

填写说明

1.任务书甲方为广州市科学技术局；乙方为项目承担单位；丙方为项目组织单位。

2.任务书基于项目申报书转换而成，请按照“广州科技GI（广州科技大脑）”提示在线填写核实，若存在不填写内容的栏目，请用“无”表示；任务书中的单位名称应为规范全称，并与单位公章一致。

3.乙方与合作单位的合作协议自动从项目申报书中读取，如需变化调整，须待任务书签订后，按要求及时办理重大变更。

4.乙方完成项目任务书在线填写，依次提交丙方和甲方审核确认后，按要求完成签章。具备电子签章条件的单位，在“广州科技GI（广州科技大脑）”完成任务书签署；不具备电子签章条件的单位，经与业务主管处室沟通对接后，可下载电子版项目任务书用A4纸双面打印装订签章，一式六份报甲方和丙方签章，其中甲方两份丙方两份，项目承担单位和项目负责人各一份。

5.涉密项目请在“广州科技GI（广州科技大脑）”下载项目任务书模板，按保密要求离线填写报送。

6.项目申报书是项目任务书填报的重要依据，未经甲方许可，乙方不得修改考核指标，调整主要研究内容。项目任务书将作为项目实施管理、验收结题和监督评估的重要依据。

7.项目任务书中的“备注”，包括重要的必须补充的内容。

8.“广州科技GI（广州科技大脑）”是项目管理过程中重要通知和文书的电子送达平台，电子送达与书面送达具有同等法律效力。为确保电子送达渠道畅通，乙方和项目负责人应及时更新维护“广州科技GI（广州科技大脑）”的单位和个人信息。

9.项目涉及科技伦理、科技安全（如临床研究、生物安全、信息安全等）和科技保密相关问题的，申报单位须严格执行国家有关法律法规和伦理原则，完成相关审查工作；项目负责人在项目任务书签订环节，须提供符合国家有关法律法规和伦理准则要求的审查

批准文件，项目承担单位负责审核批准文件的真实性和有效性。

一、项目基本信息

项目基本信息	项目名称	基于自适应特征分解与聚合的图像质量评价研究
	申请市财政科技经费	5(万元)
	研究期限	2（年）
项目摘要	准确评价复杂真实场景中图像的视觉质量是一个重要且极具挑战性的问题。图像质量由内容和退化共同决定，但两者高度耦合,现有方法难以有效分离和分析其对质量的影响。为此,本项目拟结合语言模型,引入文本-图像跨模态监督,并设计合适的信息表征约束，实现有效的自适应特征分解和聚合，以提高图像质量评价模型的可解释性和泛化性。此项目不仅推动图像质量评价技术的发展,还可丰富图像表达,提升图像恢复等任务性能。	

二、项目单位情况

项目承担单位	单位名称	华南农业大学	统一社会信用代码	12440000455416563 4
	注册时间	1952-01-01	单位类型	高等院校
	注册地址	广东省广州市天河区五山路483号		
	办公地址（联系地址）	广东省广州市天河区五山路483号		
	联系人	姓名	夏杰	
		手机号码	13711345768	
		电子邮箱	kjcgxk@scau.edu.cn	
	开户银行	广东广州工行五山支行		
	开户户名	华南农业大学		
	银行账号	3602002609000310520		

注：如果办公地址（联系地址）等相关信息有变更，项目单位应当在变更之日起三个工作日内告知我单位变更后的地址。

三、项目负责人信息

姓名	周子涵	证件类型	身份证
证件号码	342601199602060028	性别	女
出生日期	1996-02-06	民族	汉族
国籍	中国	学历	博士研究生
学位	博士	学位授予国家（或地区）	中国
职务	无	职称	无
所学专业	计算机科学与技术	手机号码	18819472687
办公电话	020-85285383	电子邮箱	joannezzh@outlook.com

四、项目经费信息

本项目总投入：¥（5）万元，其中，市财政科技经费：¥（5）万元，自筹经费：¥（0）万元。

经费下达计划			
资金来源	小计	市财政科技经费	自筹经费
2025	5	5	0
总计	5	5	0

（单位：万元）

注：本专题纳入“包干制”，市财政科技经费按市科技计划项目经费“包干制”相关规定执行。

五、预期代表性成果

项目负责人在项目实施期内，以该项目作为资助项目获得以下5种情形之一且经费使用符合规定的，由组织单位审核后通过验收。

（一）项目实施期内，以第一作者/通讯作者发表论文1篇或以上（须标注资助项目编号）；

（二）项目实施期内，以第一完成人申请或授权专利、软件著作1项或以上；

（三）项目实施期内，获省级以上科技计划项目或人才项目支持1项或以上；

（四）项目实施期内，获省级以上科技奖励（含列入获奖团队成员名单）1项或以上；

（五）项目实施期内，获得职称晋升。

六、备注

专题补充约定条款：

甲方对未履行勤勉尽责义务的相关责任主体，自作出处理结论之日起，依照法律法规规定或任务书约定实施惩戒5年，取消相关责任主体申报市科技计划项目、申领市科技计划项目经费的资格。

预期代表性成果需在实施期内获得。

项目承担单位（乙方）及项目负责人承诺书

承诺书

本单位/本人作为广州市科技计划项目承担单位/项目负责人，将严格遵守广州市科技计划管理相关规定，严格履行自身责任，加强对项目组人员及合作单位的管理，在此郑重承诺：

（一）确保与本项目有关的全部材料真实、合法、有效，未侵犯其他方知识产权等权利，不存在多头申报、重复申报行为；

（二）严格遵守《广州市科技创新条例》《广州市科技计划项目管理办法》《广州市科技计划项目经费管理办法》《广州市科技计划科技报告管理办法》等相关规定，实施项目和经费管理；

（三）严格遵守国家、省、市关于科研诚信、科技伦理、科技安全（如临床研究、生物安全、信息安全等）和科技保密的有关法律、法规，相关政策以及各项规定，加强项目实施过程中的科研诚信、科技伦理、科技安全（如临床研究、生物安全、信息安全等）和科技保密管理，恪守科研道德准则。

如有违反，本单位/本人愿意接受相关部门做出的各项处理决定，包括但不限于终止项目、停拨经费、核减经费、追回经费，取消一定期限广州市科技计划项目申报资格，记入科研失信行为数据库，将不良行为向社会公开等。

项目承担单位：华南农业大学

日期：2024年12月26日

项目负责人：周子涵

日期：2024年12月26日

电子送达确认书

告知事项	1.为便于本项目承担单位（受送达人）及时收到相关通知和文书，请受送达人知悉项目管理过程中重要通知和文书（如：项目验收结果通知书、配合结余资金审计通知书、项目终止通知书等）的送达方式是通过“广州科技GI（广州科技大脑）”平台（以下简称平台）电子送达。 2．确认的送达方式适用于行政执法全过程程序。 3．请受送达人在项目管理过程中及时、主动查看平台相应模块，以免错过相关通知和文书。 4．此电子送达方式，以发送方设备显示发送成功视为送达。但接收方证明其到达平台的日期与发送方对应系统显示发送日期不一致的，以受送达人证明到达平台的日期为准。
电子送达	受送达人同意：“广州科技GI（广州科技大脑）”是项目管理过程中重要通知和文书的电子送达平台，电子送达与书面送达具有同等法律效力。为确保电子送达渠道畅通，受送达人同意“广州科技GI（广州科技大脑）”作为电子送达平台。 “广州科技GI（广州科技大脑）”网站地址：https://gzsti.gzsi.gov.cn。
受送达人确认	我单位已阅读本确认书的告知事项，接受并确认了上栏送达方式。 受送达人：华南农业大学 2024年12月26日
备注	

任务书签署

甲乙丙三方根据《广州市科技计划项目管理办法》《广州市科技计划项目经费管理办法》《广州市科技计划科技报告管理办法》等有关文件规定，以及有关法律、政策和管理要求，签署本任务书。

签订地点：广州市越秀区

广州市科学技术局（甲方）：广州市科学技术局
局项目经办人：陈良　　联系电话：83124036
责任处室负责人：洪雪妍

2024年12月31日

项目承担单位（乙方）：华南农业大学
二级部门：华南农业大学数学与信息学院
项目负责人：周子涵
项目经费汇入账号
账户名：华南农业大学　　账号：3602002609000310520
开户银行：广东广州工行五山支行

2024年12月26日

组织单位（丙方）：华南农业大学
项目经办人：夏杰

2024年12月27日

受理编号：c25140500001891

项目编号：2025A1515010030

文件编号：粤基金字〔2025〕10号

广东省基础与应用基础研究基金项目

任务书

项目名称： 黑箱对抗场景下自动驾驶视觉模型的鲁棒优化与集成研究

项目类别： 广东省自然科学基金-面上项目

项目起止时间： 2025-01-01　至　2027-12-31

管理单位（甲方）： 广东省基础与应用基础研究基金委员会

依托单位（乙方）： 华南农业大学

通讯地址： 广东省广州市天河区五山路483号

邮政编码： 510642　　**单位电话：** 020-85283435

项目负责人： 黄立峰　　**联系电话：** 13929500478

（广东科技微信公众号）　　（查看任务书信息）　　（受理纸质材料二维码）

广东省基础与应用基础研究
基金委员会
二〇二〇年制

填写说明

一、项目任务书内容原则上要求与申报书相关内容保持一致，不得无故修改。

二、项目承担单位通过广东省科技业务管理阳光政务平台下载项目任务书，按要求完成签名盖章后扫描上传到广东省科技业务管理阳光政务平台。

三、签名盖章说明。请分别在单位工作分工及经费分配情况页、人员信息页、签约各方页等地方按要求签字或盖章，签章不合规或错漏将不予受理。其中，人员信息页要求所有参与人员本人亲笔签名，代签或印章无效，漏签将不予受理。

四、本任务书自签字并加盖公章之日起生效，各方均应负本任务书的法律责任，不应受机构、人事变动影响。

五、根据《广东省科学技术厅广东省财政厅关于深入推进省基础与应用基础研究基金项目经费使用“负面清单+包干制”改革试点工作的通知》(粤科规范字〔2022〕2号)，2022年度及以后立项资助的全部省基金项目(包括省自然科学基金、省市联合基金、省企联合基金项目等)均适用“负面清单+包干制”，项目提交申请书和任务书时无需编制费用明细科目预算。

一、主要研究内容和要达到的目标

项目围绕当前自动驾驶视觉模型在黑箱对抗场景面临的安全隐患问题展开研究。针对现有方案面临的对抗噪音数据质量低、模型优化对抗偏差大和模型集成方案选择难等问题，项目拟从对抗数据生成、鲁棒模型优化和集成方案选择三方面展开研究，采用数据驱动和模型驱动的方式提升自动驾驶视觉模型在黑箱对抗场景下的鲁棒性，以防范恶意对抗噪音干扰。具体而言，项目拟深入研究点云通用对抗噪音生成、鲁棒模型去偏差优化和模型集成方案安全评估与选择等关键技术：

（1）研究黑箱场景可迁移的点云通用对抗噪音生成方法，从点云数据自身特性出发，在不依赖替代模型的前提下，设计并度量在黑箱场景下更加通用和泛化的空间重要性测度，旨在构造更高质量的点云对抗数据，为模型鲁棒优化（关键技术二）和模型集成方案评估（关键技术三）提供数据基础；

（2）研究基于归因理解的鲁棒模型去偏差优化方法，以高质量的通用对抗噪音数据（关键技术一）为基础，深入理解鲁棒模型的推理偏好并将其纳入优化过程，旨在降低对抗偏差，并进一步提升鲁棒模型在黑箱对抗场景的鲁棒性，为模型集成方案评估与选择（关键技术三）提供模型支撑。

（3）研究模型集成方案多样化的安全评估与选择方法，通过缓解现有方法在评估视觉模型安全性时面临的差异敏感性问题，结合黑箱场景点云对抗噪音数据（关键技术一），为不同的多模态鲁棒模型组合（关键技术二）建立黑箱场景安全性基准，进而在有限资源约束下选择安全性高的模型集成部署方案，为自动驾驶系统提供安全支撑。

本项目整体研究目标是实现黑箱场景下对抗鲁棒的自动驾驶视觉模型，为实现安全可信的自动驾驶提供理论和实践支撑。具体研究目标包括：

（1）设计黑箱场景可迁移的点云通用对抗噪音生成方法，构造高质量的点云对抗噪音数据，作为模型优化的学习数据和模型安全评估的测试数据；

（2）研究基于归因理解的鲁棒模型去偏差优化方法，提高视觉模型在黑箱场景下的鲁棒性，为模型集成方案提供鲁棒模型集合。

（3）提出模型集成方案多样化的安全评估与选择方案，在资源约束下从多种鲁棒模型组合中选择黑箱场景下安全性高的集成部署方案，为自动驾驶系统提供安全支撑。

二、项目预期获得的研究成果及形式

<table>
<tr><td rowspan="3">论文及专著情况</td><td colspan="2">国家统计源刊物以上刊物发表论文（篇）</td><td colspan="2">3</td><td colspan="2">科技报告（篇）</td><td colspan="2">1</td></tr>
<tr><td colspan="2">其中被SCI/EI/ISTP收录论文数（篇）</td><td colspan="2">2</td><td colspan="2">培养人才（人）</td><td colspan="2">3</td></tr>
<tr><td colspan="2">专著（册）</td><td colspan="2"></td><td colspan="2">引进人才（人）</td><td colspan="2"></td></tr>
<tr><td rowspan="3">专利情况(项)</td><td colspan="2">发明专利</td><td colspan="2">实用新型专利</td><td colspan="2">外观设计专利</td><td colspan="2">国外专利</td></tr>
<tr><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请</td><td>授权</td></tr>
<tr><td>2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

三、项目进度和阶段目标

（一）项目起止时间： 2025-01-01 至 2027-12-31		
（二）项目实施进度及阶段主要目标：		
开始日期	结束日期	主要工作内容
2025-01-01	2025-12-31	1、研究黑箱场景可迁移的点云通用对抗噪音生成方法； 2、构建高质量的点云对抗数据集； 3、搭建视觉模型对抗攻防仿真平台，并完成验证测试； 4、发表1篇学术论文，申请发明专利1件； 5、参加1-2次国内外知名学术会议或者相关学术活动。
2026-01-01	2026-12-31	1、研究基于归因理解的鲁棒模型去偏差优化方法； 2、构建黑箱场景下高鲁棒性的视觉模型集合，完成验证测试； 3、发表1篇学术论文，申请发明专利1件； 4、参加1-2次国内外知名学术会议或者相关学术活动。
2027-01-01	2027-12-31	1、研究集成部署方案多样化的安全评估与选择方案； 2、建立鲁棒模型组合在黑箱场景下的安全性基准，选择安全性高的集成部署方案，完成验证测试； 3、发表1-2篇学术论文，申请发明专利1件； 4、参加1-2次国内外知名学术会议或者相关学术活动； 5、准备结题工作，进行技术和研究成果汇总，完成结题报告。

四、项目总经费及省基金委经费预算

1.省基金委经费下达总额：　（大写）壹拾万圆整；（小写　）10万元；					
2.省基金委经费年度下达计划：					
年度	2025 年	年	年	年	年
经费(万元)	10.00				

五、人员信息

项目负责人								
姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
黄立峰	431202199002010815	35	男	讲师	博士研究生	项目负责人	华南农业大学	黄立峰

项目组主要成员								
姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
周子涵	342601199602060028	29	女	讲师	博士研究生	算法设计、架构设计	华南农业大学	周子涵
罗浩宇	360302198907132536	36	男	副教授	博士研究生	算法设计、架构设计	华南农业大学	罗浩宇
王涵	440105200008055432	25	男	未取得	本科	系统设计、软件开发	华南农业大学	王涵
刘名	370921200009074235	25	男	未取得	本科	系统设计、软件开发	华南农业大学	刘名

六、工作分工及财政经费分配

承担/参与单位名称（盖章）	工作分工	省级财政科技资金分配（万元）
华南农业大学	本项目由华南农业大学研究团队负责项目的整体组织、制定总体方案和研发规划，具体包括：1. 项目整体规划、研究方案制定和项目实施规划，确保项目按时按质地有序推进；2. 针对自动驾驶系统视觉模型，开展黑箱场景下的点云通用对抗噪音生成关键技术公关，从数据生成方面为模型安全提供基础。 聚焦于黑箱场景下的视觉模型鲁棒优化关键技术公关，研究基于归因理解的鲁棒模型去偏差优化方法，从模型构建方面提供支撑。研究自动驾驶系统中高鲁棒性视觉模型集成部署的关键技术公关，研究模型集成方案多样化的安全评估与选择方法。公开相关技术和模型。3. 实现自动驾驶视觉攻防原型系统，主要包含架构扩展模块、数据生成模块和攻防评估模块，涵盖本申请中研究方法以及已有的主流攻防算法，将重点面向互联网科技公司、创新创业企业和研究机构等，为自动驾驶系统相关开发者、使用者和研究者提供自动驾驶汽车视觉模型鲁棒性评估服务。4. 针对项目的研究成果进行总结，充分准备并配合项目的验收工作。	10.00
	合计	10.00

七、任务书条款

第一条 甲方与乙方根据《中华人民共和国民法典》及国家有关法规和规定，按照《广东省自然科学基金及联合基金项目管理实施细则》（粤科规范字〔2024〕5号）《省级科技计划项目任务书管理细则》（粤科规范字〔2022〕8号）等规定，为顺利完成（2025）年黑箱对抗场景下自动驾驶视觉模型的鲁棒优化与集成研究专项项目（项目编号：2025A1515010030）经协商一致，特订立本任务书，作为甲乙双方在项目实施管理过程中共同遵守的依据。

第二条 甲方的权利义务：1.按任务书规定进行经费核拨的有关工作协调。2.根据甲方需要，在不影响乙方工作的前提下，定期或不定期对乙方项目的实施情况和经费使用情况进行检查或抽查。3.根据《广东省科学技术厅科技计划项目科研诚信管理办法》（粤科规范字〔2024〕2号）《广东省基础与应用基础研究基金项目科研不端行为调查处理实施细则（试行）》（粤科规范字〔2023〕1号）等规定对乙方进行科技计划信用管理。

第三条 乙方的权利义务：1.确保落实自筹经费及有关保障条件。2.按任务书规定，对甲方核拨的经费实行专款专用，单独列账，并随时配合甲方进行监督检查。3.经费使用按照广东省级财政科研项目经费使用及省基金项目经费使用“负面清单+包干制”等有关规定进行管理。4.项目依托单位应制定经费使用“负面清单+包干制”内部管理制度并报甲方备案。5.使用财政资金采购设备、原材料等，按照《广东省实施〈中华人民共和国招标投标法〉办法》有关规定，符合招标条件的须进行招标。6.项目任务书任务完成后，或任务书规定的任务、指标及经费投入等提前完成的，乙方可提出验收结题申请，并按甲方要求做好项目验收结题工作。7.若项目发生需要终止结题的情况，乙方须提出终止结题申请，并按甲方要求做好项目终止结题工作。8.在每年规定时间内向甲方如实提交上年度工作情况报告，报告内容包含上年度项目进展情况、经费决算和取得的成果等。9.按照国家和省有关规定，提交科技报告及其他材料。10.利用甲方的经费获得的研究成果，项目负责人和参与者应当注明获得“广东省基础与应用基础研究基金（英文：Guangdong Basic and Applied Basic Research Foundation）（项目编号）”资助或作有关说明。11.乙方要恪守科学道德准则，遵守科研活动规范，践行科研诚信要求，不得抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；不得购买、代写、代投论文，虚构同行评议专家及评议意见；不得违反论文署名规范，擅自标注或虚假标注获得科技计划（专项、基金等）等资助；不得弄虚作假，骗取科技计划（专项、基金等）项目、科研经费以及奖励、荣誉等；不得有其他违背科研诚信要求的行为。12.确保本项目开展的研究工作符合我国科技伦理管理相关规定。

第四条 在履行本任务书的过程中，如出现广东省相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。

第五条 在履行本任务书的过程中，当事人一方发现可能导致项目整体或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第六条 本项目技术成果的归属、转让和实施技术成果所产生的经济利益的分享，除双方另有约定外，按国家和广东省有关法规执行。

第七条 根据项目具体情况，经双方另行协商订立的附加条款，作为本任务书正式内容的一部分，与本任务书具有同等效力。

第八条 本任务书一式三份，各份具有同等效力。甲、乙方及项目负责人各执一份，三方签字、盖章后即生效，有效期至项目结题后一年内。各方均应负任务书的法律责任，不应受机构、人事变动的影响。

第九条 乙方必须接受甲方聘请的本项目任务书监理单位的监督和管理。监理单位按照甲方赋予的权利对本项目任务书的履行进行审核、进度调查，对项目任务书变更、经费使用情况进行监督管理及组织项目验收。

说明：1.本任务书中，凡是当事人约定无需填写的内容，应在空白处划（/）。

2.委托代理人签订本任务书的，应出具合法、有效的委托书。

八、本任务书签约各方

管理单位（甲方）：　　广东省基础与应用基础研究基金委员会（盖章）

法定代表人（或法人代理）：　曾路　（签章）

2025 年　03 月　21 日

依托单位（乙方）：　华南农业大学　（盖章）

法定代表人（或法人代理）：　薛红卫　（签章）

联系人（项目主管）姓名：　夏杰　（签章）

Email：　kjcgxk@scau.edu.cn

电话：　020-85283435　/　13711345768

开户单位名称：　华南农业大学

开户银行名称：　广东广州工行五山支行

开户银行账号：　3602002609000310520

2025 年 4 月 9 日

联系人（项目负责人）姓名：　黄立峰　（签名）

Email：　huanglf6@scau.edu.cn

电话：　13929500478

2025 年 4 月 5 日

SCAULIB202519676

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学院 周子涵 3 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	作者支持单位	收录情况	影响因子	中科院大类分区
1	对应章节三-3文章 Underwater Image Enhancement via Adaptive Bi-Level Color-Based Adjustment	IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 出版年：2025 卷期：74 页码：- 文献号：5018916 文献类型：Article	唯一通讯作者	A 类	华南农业大学数学与信息学院	SCI	IF2-year=5.9 IF5-year=6.0 （2024）	工程技术 2 区 Top 期刊：否 （2025）
2	对应章节三-2.2文章 Deep Blind Image Quality Assessment Using Dynamic Neural Model With Dual-Order Statistics	IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY 出版年：2024 出版日期：JUL 卷期：34 7 页码：6279-6290 文献类型：Article	第一作者	T2 类	华南农业大学数学与信息学院	SCI	IF2-year=11.1 IF5-year=9.8 （2024）	计算机科学 1 区 Top 期刊：是 （2025）
3	Nighttime image dehazing via a physics-aware dynamic neural model with progressive contrastive	PATTERN RECOGNITION 出版年：2026 出版日期：JAN	唯一通讯作者	T2 类	华南农业大学数学与信息学院	SCI	IF2-year=7.6 IF5-year=7.9 （2024）	计算机科学 1 区 Top 期刊：是

	regularization	卷期：169 页码：- 文献号：111863 文献类型：Article						（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：卢炳卫

华南农业大学图书馆

2025-08-01

SCAULIB202625126

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学院、软件学院 周子涵(学科类型:自然科学) 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	Dual perception-aware blind image quality assessment with semantic-distortion integration and dynamic global-local refinement	DISPLAYS 出版年：2026 出版日期：JUL 卷期：93 页码：- 文献号：103357 文献类型：Article	通讯作者	A 类	华南农业大学数学与信息学院	SCI	IF2-year=3.4 IF5-year=3.5 (2024)	计算机科学 2 区 Top 期刊：否 OA 期刊：否 (2025)
2	Enhancing CNN-Based Blind Image Quality Assessment via Deep Cross-Layer Pattern Encoding 对应章节三-2.1文章	IEEE TRANSACTIONS ON MULTIMEDIA 出版年：2025 卷期：27 页码：9712-9722 文献类型：Article	第一作者	T2 类	华南农业大学数学与信息学院	SCI	IF2-year=9.7 IF5-year=9.0 (2024)	计算机科学 1 区 Top 期刊：是 OA 期刊：否 (2025)
3	CDINet: Content Distortion Interaction Network for Blind Image Quality Assessment	IEEE TRANSACTIONS ON MULTIMEDIA 出版年：2024 卷期：26 页码：7089-7100	第三作者	T2 类	华南农业大学数学与信息学院	SCI	IF2-year=9.7 IF5-year=9.0 (2024)	计算机科学 1 区 Top 期刊：是 OA 期刊：否 (2025)

		文献类型：Article						
4	Hierarchical hashing-based multi-source image retrieval method for image denoising	APPLIED SOFT COMPUTING 出版年：2021 出版日期：DEC 卷期：113 页码：- 文献号：108028 文献类型：Article	第二作者	A 类	华南理工大学计算机科学与工程学院	SCI	IF2-year=8.263 IF5-year=7.595 (2021)	计算机科学 2 区 Top 期刊：是 OA 期刊：否 (2021)
5	Spatial Adaptive Filter Network With Scale-Sharing Convolution for Image Demoireing	IEEE SIGNAL PROCESSING LETTERS 出版年：2024 卷期：31 页码：2495-2499 文献类型：Article	第四作者	B 类	华南农业大学数学与信息学院	SCI	IF2-year=3.9 IF5-year=3.9 (2024)	计算机科学 3 区 Top 期刊：否 OA 期刊：否 (2025)
6	Deep Underwater Image Quality Assessment via Progressive Physics-aware Multi-Prior Collaboration	IEEE Transactions on Circuits and Systems for Video Technology 出版年：2025 卷期： 页码：- 文献号： 10.1109/TCSVT.2025.3639385 文献类型：	第 一作者	T2 类	华南农业大学数学与信息学院	已发表，暂未被 SCI 收录	IF2-year=11.1 IF5-year=9.8 (2024)	计算机科学 1 区 Top 期刊：是 OA 期刊：否 (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：尹银怀

华南农业大学图书馆

信息检索专用章

2026-03-04

检索证明

根据委托方提供的论文目录（2024 年），经《工程索引》EI-VILLAGE 数据库的检索，**周子涵**（Zhou, Zihan）发表的论文被《工程索引》EI Compendex 收录了 1 篇，其中第一作者 1 篇。题录如下：

1. Title: No-Reference Image Quality Assessment Using Local Binary Patterns: A Comprehensive Performance Evaluation

Authors: Zhou, Zihan[1]; Xu, Yong[1,5]; Wan, Xi[1]; Quan, Yuhui[1,6]; Xu, Ruotao[2]; Li, Jing[3]; Le Callet, Patrick[4]

Source: QoEVMA 2024 - Proceedings of the 3rd Workshop on Quality of Experience in Visual Multimedia Applications, p 2-11, October 28, 2024, QoEVMA 2024 - Proceedings of the 3rd Workshop on Quality of Experience in Visual Multimedia Applications

Language: English

Document type: Conference article (CA)

对应章节三-2.3文章

Author affiliation:

[1] South China Agricultural University, Guangzhou, China;

[2] Institute for Super Robotics, Guangzhou, China;

[3] Moku Lab, Alibaba Group, Beijing, China;

[4] Nantes Universite, Nantes, France;

[5] PaZhou Laboratory of Guangzhou, Guangdong Provincial Key Laboratory of Multimodal Big Data Intelligent Analysis, China;

[6] Pazhou Laboratory of Guangzhou, China

E.I. COMPENDEX No: 20244917465054

查证员（签字）：袁嘉仪

华南理工大学图书馆

2026 年 3 月 5 日

20260306-36815906242576384-1

查看地址:https://eseal.scut.edu.cn/esign/i/s/ZBFn2q　　提取码:gRtyMm

Enhancing CNN-Based Blind Image Quality Assessment via Deep Cross-Layer Pattern Encoding

Zihan Zhou, Yong Xu, Yuhui Quan, Yun Liang, Jing Li, and Patrick Le Callet

***Abstract*—Evaluating image quality without reference images, known as blind image quality assessment (BIQA), is crucial for image communication. Recently, convolutional neural networks (CNNs) have emerged as a prominent BIQA approach due to their feature learning power. Usually, both high-level semantic information and low-level details significantly impact perceived visual quality. However, most existing CNN-based methods focus on high-level semantic information via aggregating features on top of the last convolutional layer into a global descriptor, neglecting the importance of shallow, low-level cues. To address this limitation, this paper proposes a novel approach that exploits local encoding and histogram-based pyramid pooling on cross-layer features produced by a CNN, achieving a joint local and global analysis. Specifically, we introduce a cross-layer pattern encoding model that characterizes features generated along convolutional layers via a soft histogram of local 3D binary patterns. This leads to a highly informative yet compact descriptor for score regression. By building this module into a ResNet backbone, we present an effective BIQA model demonstrating state-of-the-art performance in extensive experiments on synthetic and authentic datasets.**

Received 18 August 2024; revised 20 January 2025; accepted 23 March 2025. Date of publication 7 October 2025; date of current version 17 December 2025. The work of Zihan Zhou was supported in part by the National Natural Science Foundation of China under Grant 62401210, in part by the Natural Science Foundation of Guangdong Province under Grant 2025A1515011539, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515110646, and in part by Guangzhou Basic and Applied Basic Research Project under Grant 2025A04J4502. The work of Yong Xu was supported in part by the National Key Research and Development Program of China under Grant 2024YFE0105400, in part by the National Natural Science Foundation of China under Grant 62472179, and in part by Guangzhou Science and Technology Plan Project - Key R and D Plan under Grant 2024B01W0007. The work of Yun Liang was supported in part by the Key Research and Development Project of Guangzhou under Grant 202206010091, in part by the Fund of Southern Marine Science and Engineering Guangdong Laboratory at Zhanjiang under Grant ZJW-2023-04, and in part by the Special Fund for Marine Economic Development (Six Marine Industries) of Guangdong Province, China, under Grant GDNRC[2024] No.18. The associate editor coordinating the review of this article and approving it for publication was Yuming Fang. *(Corresponding author: Yuhui Quan.)*

Zihan Zhou and Yun Liang are with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China (e-mail: zhouzihan@scau.edu.cn; yliang@scau.edu.cn).

Yong Xu and Yuhui Quan are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510642, China (e-mail: yxu@scut.edu.cn; csyhquan@scut.edu.cn).

Jing Li is with the Moku Lab, Alibaba Group, Beijing 100085, China (e-mail: lj225205@alibaba-inc.com).

Patrick Le Callet is with the Équipe Image, Perception et Interaction, Laboratoire des Sciences du Numérique de Nantes, Université de Nantes, 44000 Nantes, France (e-mail: patrick.lecallet@univ-nantes.fr).

Digital Object Identifier 10.1109/TMM.2025.3618566

***Index Terms*—Image quality assessment, convolutional networks, feature encoding, feature aggregation.**

I. Introduction

IMAGE quality is fundamental to multimedia systems, directly impacting the efficacy of visual communication across diverse applications. In multimedia-driven fields such as telemedicine, video surveillance, and entertainment, high-quality visual content is crucial for accurate analysis, decision-making, and user engagement. However, the multimedia pipeline is vulnerable to quality degradation from factors like noise, compression artifacts, and transmission errors. Consequently, robust multimedia systems require sophisticated image quality evaluation capabilities for real-time monitoring, adaptive streaming, and prompt issue resolution. Image Quality Assessment (IQA) addresses this need by developing computational models and objective metrics that align with human perception, obviating the need for subjective evaluations in large-scale multimedia operations. The applications IQA are diverse, encompassing areas like providing immediate feedback for image acquisition, serving as objective functions for image processing models, and serving as criteria for compression, watermarking, and camera settings [1], [2].

In scenarios where no information about a reference pristine image is available, the task of IQA is termed non-reference IQA or blind IQA (BIQA). BIQA holds significant practical importance but presents great challenges. Over the recent decades, significant efforts have been devoted to the development of BIQA methods; see e.g. [3], [4], [5], [6]. Traditional approaches generally follow a two-stage process: quality-aware feature extraction and quality score regression. Early studies usually adopted hand-crafted designs for feature extraction, such as natural scene statistics [3], [7] and local pattern encoding [8], [9]. Regression is then done using well-established learning-based models, such as support vector regression [10] and random forest [11]. In recent years, inspired by the power of deep neural networks in capturing semantic features for image classification, many convolutional neural network (CNN) models have been proposed for BIQA, where feature extraction and quality score regression are jointly optimized in an end-to-end manner, as seen in [12], [13], [14].

Many current CNN-based BIQA models (*e.g.* [12], [15], [16], [17]) follow the spirit of the common framework used in image

classification [18], [19]. Spatial features with high-level semantics are initially derived through a series of convolutional blocks. Subsequently, quality scores are regressed using a multi-layer perceptron (MLP). Between these two stages, global aggregation achieved through global pooling like global average pooling (GAP) [4], [12], [17], [20], is commonly employed. More specifically, a global aggregation module is added on top of convolutional blocks, generating a fixed-size global descriptor from the final convolutional feature tensor. This descriptor is then utilized for the subsequent quality score regression. By aggregating features across spatial coordinates, this global descriptor diminishes the sensitivity of MLP-based quality score regression to spatial transformations in images. By converting spatially-distributed features into a single position-invariant representation, it ensures consistent quality predictions regardless of where distortions appear in the image. While this architectural framework has demonstrated efficacy in image classification, it is sub-optimal for BIQA since BIQA necessitates the consideration of low-level or middle-level features, local patterns, and distortion awareness [21], [22], [23].

Local features are critical in NR-IQA, as image quality often depends on preserving specific details, such as local structures and texture patterns. Distortions like blocking artifacts or localized blur, though less noticeable globally, remain prominent at regional or pixel levels. However, several existing CNN-based methods primarily rely on global feature aggregation from the final convolutional layer, overlooking quality cues tied to local details or global information from earlier layers. This limitation constrains the ability to capture structural and textural changes caused by local distortions, and the issue is exacerbated when pre-trained models from image classification tasks are transferred to BIQA, as commonly seen in prior work (*e.g.* [12], [15], [16], [17]). To address these challenges, we propose a CNN-based approach that jointly leverages local and global features. Our method repurposes features from multiple CNN layers to construct a comprehensive quality descriptor for regression. Lower layers capture fine-grained, localized details, while higher layers focus on global, abstract features, reflecting the hierarchical organization of features within CNNs. By integrating hierarchical features, the quality descriptor can combine feature patterns from various complementary scales, seamlessly incorporating quality cues ranging from localized to holistic views and evaluating distortions from micro to macro levels. This paradigm also can mimic the multi-resolution processing of human visual systems [24].

Furthermore, unlike current multi-scale IQA methods, often rely on down-sampling to align feature shapes across different scales or layers, which can discard critical local structural details. To address this, our approach integrates features across CNN depths and incorporates up-sampling of multi-scale features. This design preserves intricate local information while seamlessly combining low- and mid-level features essential for accurate distortion perception. Besides, many NR-IQA methods treat multi-scale feature extraction and fusion separately, overlooking interactions between low- and high-level information critical for IQA. Our method reorganizes multi-scale features channel-wise for local encoding on cross-layer features, allowing specific channels to detect distortions and patterns at different spatial scales. This enables simultaneous processing of low-level textures, mid-level structures, and high-level semantics. This design mimics human visual perception effectively without added computational complexity.

To efficiently extract quality-aware 3D local patterns from cross-layer feature tensors, we propose a Cross-Layer Pattern Encoding (CLPE) module. This module collects features from shallow to deep layers, resizes them, and organizes them into cross-layer tensors. Within these tensors, local 3D patterns are extracted and aggregated using a soft spatial pyramid histogram, resulting in a compact and robust cross-layer descriptor. Refer to Fig. 1 for an illustration of the main idea of proposed approach.

Inspired by Local Binary Patterns (LBP) [26], CLPE encodes 3D features into binary codes by comparing each feature value with its neighbors, converting them into decimal labels. This identifies texture patterns and local structures across scales while preserving spatial information. As shown in Fig. 2, these binary codes detect subtle distortions and capture degradation in structural and textural features, highlighting CLPE's ability to extract quality-indicative details, especially in CNN-based features. CLPE also integrates a histogram-based pooling approach that preserves local feature distributions and fine-grained details. Unlike GAP, which emphasizes global representations and smooths out subtle distortions, our histogram-based pooling captures diverse distortion patterns by encoding feature statistics comprehensively. This enhances distortion sensitivity, parameter efficiency, and generalization, making pyramid pooling distortion-aware.

By integrating the CLPE module into a ResNet backbone, we present CLPENet. This effective deep BIQA model achieves noticeable performance gains by fully exploiting multi-level image features by cross-layer feature encoding. Extensive experiments are conducted on both synthetic and authentic datasets, and the experimental results have demonstrated the excellent performance of CLPENet.

In summary, our contributions to this work are as follows:

- We propose leveraging local encoding and histogram-based pyramid pooling on cross-layer features to enhance BIQA, enabling the joint utilization of fine-grained local details and global semantic information.
- We introduce the CLPE module, designed to explicitly and efficiently encode both local and global information from multi-layer features into a comprehensive descriptor for score regression.
- We develop an effective deep end-to-end NN for BIQA with state-of-the-art performance.

The remainder of this paper is organized as follows. Section II gives a literature review on BIQA, particularly focusing on CNN-based BIQA methods. Section III is denoted to the detailed description of our proposed CLPE module and CLPENet. Section IV is for the experimental evaluation. Section V concludes the paper and discusses the future work.

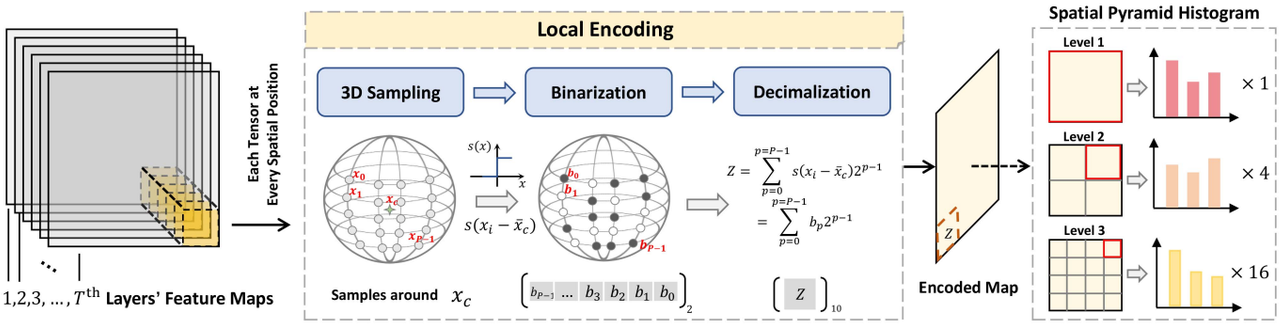
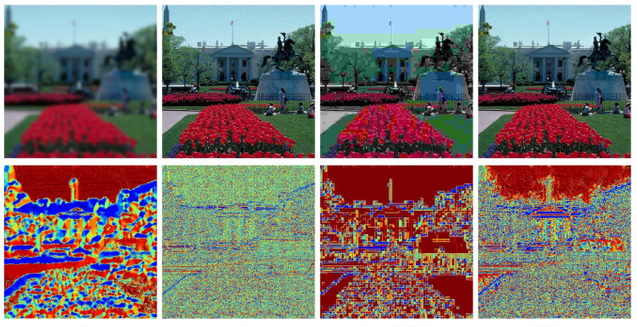


Fig. 1. Illustration of proposed CLPE module. CLPE stacks the resized and grouped feature maps from a series of convolutional layers into a cross-layer feature tensor, encodes local 3D patches on the feature tensor into binary codes, and finally aggregates the codes into a histogram-based global descriptor.



(a) Blur (b) Noise (c) JPEG (d) JPEG2000

Fig. 2. LBP maps for images with different degradation types. Viewing RGB channels of an image as a 3D feature tensor, the 3D LBP codes [25] are calculated. The binary codes, represented in decimal values, are visualized as color maps, with different colors representing different patterns captured. We can see that these maps differ for different degradation types, and they are sensitive to subtle, visually imperceptible distortions.

II. Related Work

A. Handcrafted Models for BIQA

Built upon expert knowledge, traditional BIQA methods utilized handcrafted features to capture specific statistical or perceptual image properties for quality modeling. A variety of handcrafted features have been derived from natural image statistics and human visual system principles. BRISQUE [27] utilizes asymmetric Gaussian fitting on features of luminescence, contrast and correlations. BLIINDS-II [28] uses generalized Gaussian modeling of DCT coefficients. CORNIA [29] constructs a codebook via k-means clustering of image patches. GM-LOG [30] extracts gradient and Laplacian of Gaussian statistics to train a support vector regression model. NIQE [31] utilizes a learning-free multivariate Gaussian model. ILNIQE [32] applies multivariate Gaussian modeling and average pooling for opinion-unaware BIQA. HOSA [33] performs high-order statistical aggregation, extracting local statistics and constructing a codebook via k-means clustering. LBPs are adopted for BIQA in [8], [9]. Handcrafted features have subjective bias and may fail in capturing complex image content and high level features adequately, yielding limited descriptive power. In addition, they lack the adaptivity to diverse image content.

B. Deep Learning-Based Methods for BIQA

Recently, inspired by the success of CNN in image classification, a growing number of BIQA methods utilize CNNs as powerful feature extractors. However, BIQA remains a small sample problem due to the limited availability of subjective scores. This poses challenges for CNN-based feature extractors in handling diverse distortions and variable image contents encountered in real-world scenarios, requiring new network design for improving the effectiveness of quality-informed feature extraction.

Many existing CNN-based BIQA methods focus on content-aware distortion-sensitive feature representation for improved generalization, such as content-adaptive filtering [22], content-adaptive regression [17], and attention-weighting mechanisms [34], [35], [36]. Zhou et al. [37] proposes to combine content-adaptive feature extraction and quality regression for better generalization ability. Besides, Zhou et al. [23] proposed a quality predictor that dynamically balances the representation of image content and distortions. Chen et al. [38] utilized reinforcement learning to capture attention information better.

Global pooling mechanisms are another key in CNN-based methods. Often-used global pooling mechanisms include GAP [4], [12], [17], [20], [35], [39], max pooling [40], [41], spatial pyramid pooling [42], [43]. These methods focus primarily on improving the discriminative power of global feature aggregation, while local feature encoding has received less attention. In contrast, our work centers on the exploitation of cross-layer hierarchical features from various scales, achieving performance gain.

Another direction of CNN-based methods to address the small sample problem in BIQA is directly addressing the insufficiency of labeled training data. This can be usually achieved by four different strategies: exploring patch-wise pseudo labels [12], [44], utilizing transfer learning from image classification data [12], [17], [20], self-supervised pre-training with large-scale human-unlabelled data [45], [46], [47], or introducing auxiliary tasks on unlabelled data, *e.g.* distortion type classification [20], [39], [41], image quality ranking [4], [48], and quality map estimation [49]. Unlike ours, these methods focus on training schemes, not CNN architecture.

While CNN-based methods have become one dominant approach in BIQA due to their efficiency, there is another line of

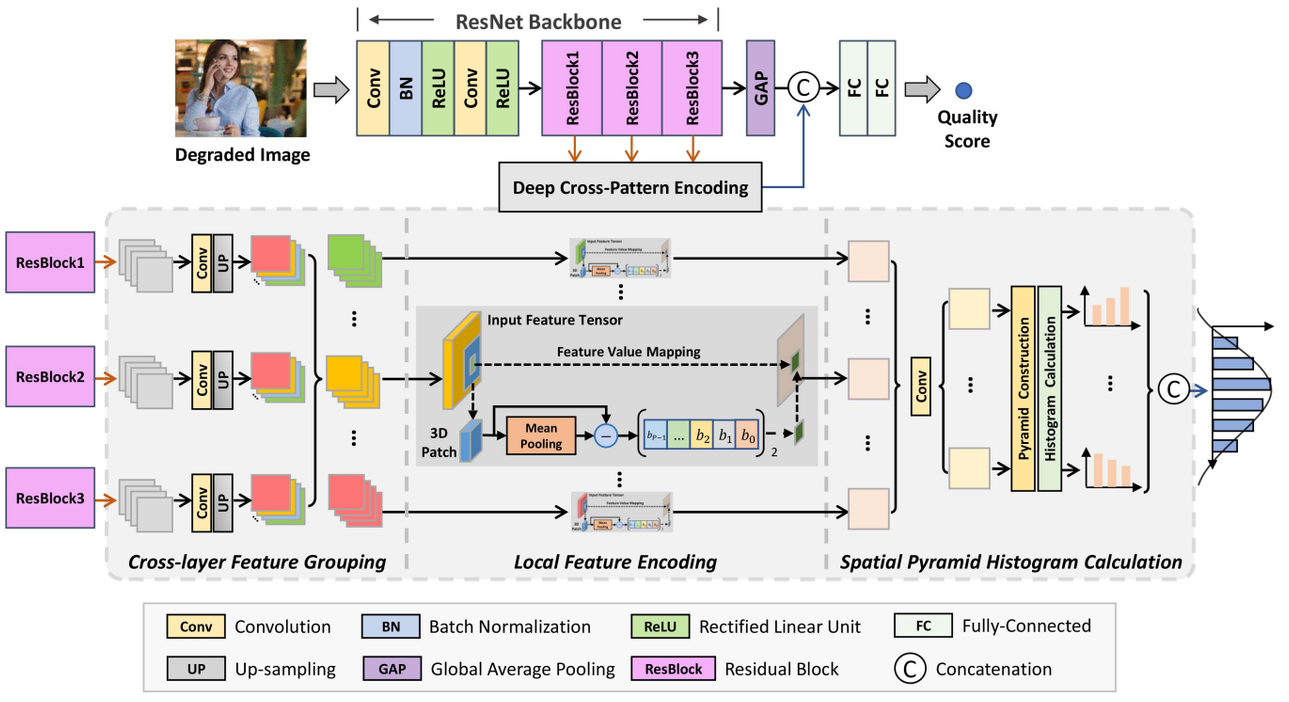


Fig. 3. Architecture of proposed CLPENet. A CLPE module (shown in gray rectangular) is integrated into a ResNet backbone to extract cross-layer feature patterns for BIQA. It groups feature maps into cross-layer tensors (left), encodes local 3D patches into feature descriptors (middle), and aggregates descriptors into a cross-layer histogram via soft spatial pyramid pooling (right). This cross-layer descriptor is concatenated with a global descriptor from GAP and passed through FC layers for quality score regression.

research using transformer models (*e.g.* [50], [51], [52], [53], [54]) which have also achieved state-of-the-art results. These transformer models exploit CNNs to extract local features from image patches and utilize self-attention to capture global information for BIQA. Notable approaches include TRIQ [50], TranSLA [51], TFIQA [55], and TReS [52], each innovating in areas like quality embedding, saliency-guided attention, and model robustness. Despite their performance, transformers typically have larger sizes and higher computational costs than CNNs. This work focuses on enhancing CNN-based methods for BIQA through our proposed CLPE module, aiming to improve performance while maintaining efficiency.

III. Proposed Method

A. Overall Framework

Our proposed CLPENet for BIQA is constructed by integrating a CLPE module into a ResNet backbone. See Fig. 3 for an illustration of its architecture. The CLPENet extracts a cross-layer descriptor by applying the CLPE module to features from various convolutional layers. This cross-layer descriptor, capturing cross-layer feature patterns, is combined with the global semantic descriptor generated via GAP on the features produced at the final convolutional layer. The features follow two parallel paths: through the ResNet backbone which applies global pooling to produce 1024-dimensional features, and through the CLPE module which outputs $M \times N$ dimensional features, where M represents the number of regions separated by spherical grid for histogram calculation and N is the number of histogram bins. The combination of these complementary descriptors enhances the model's descriptive power. Finally, an MLP with two fully connected (FC) layers is employed for quality score regression.

The CLPE module is input with the features from convolutional layers of selected residual blocks (RBs) of the ResNet backbone and then outputs a cross-layer feature descriptor. It consists of three components: (i) cross-layer feature grouping, which organizes the input feature maps into multiple cross-layer feature tensors; (ii) local feature encoding, which extracts multi-layer feature patterns from the cross-layer feature tensor; and (iii) spatial pyramid histogram pooling, which aggregates the extracted cross-layer feature patterns into a single descriptor via computing a soft spatial pyramid histogram. This descriptor is used as the module's output.

Formally, the CLPE module can be written as

$$\boldsymbol{f}_{\text{CLPE}} = \text{CLPE}(\mathcal{F}_1, \ldots, \mathcal{F}_T), \tag{1}$$

where $\mathcal{F}_i \in \mathbb{R}^{H_i \times W_i \times D_i}$ denotes the feature tensor output by the i-th of T selected RBs, with a spatial size of $H_i \times W_i$ and a channel number of D_i, and $\boldsymbol{f}_{\text{CLPE}}$ is output descriptor.

B. Cross-Layer Feature Grouping

The spatial size $H_i \times W_i$ and channel number D_i of the RB-produced feature tensor $\mathcal{F}_i$ vary across different RBs. In the first phase, the CLPE module firstly normalizes the inputs $\mathcal{F}_i \in \mathbb{R}^{H_i \times W_i \times D_i}$ into the same shape. Toward this end, we apply a 1×1 convolution to transform each $\mathcal{F}_i$ into a D-channel feature tensor and then upsample this newly generated tensor into a

unified spatial resolution $H \times W$, with $H = \max_i H_i$ and $W = \max_i W_i$. In implementation, we use bi-linear interpolation for the upsampling.

Let $\mathcal{N}_1, \ldots, \mathcal{N}_T \in \mathbb{R}^{H\times W\times D}$ denote the shape-normalized feature tensors. To capture patterns along layers, we reorganize $\mathcal{N}_i$s into D feature tensors as

$$\mathcal{G}_d = [\boldsymbol{N}_1^d; \boldsymbol{N}_2^d; \cdots; \boldsymbol{N}_T^d] \in \mathbb{R}^{H\times W\times T},\ \forall d \in \{1, \ldots, D\}, \tag{2}$$

where $\boldsymbol{N}_i^d$ denotes the d-th channel of $\mathcal{N}_i$ and $[\cdot;\cdot]$ denotes the concatenation along the channel dimension. In other words, the feature maps from different layers but sharing the same sequential channel number are grouped into a tensor. These reorganized feature tensors $\mathcal{G}_d$s contain aligned cross-layer feature maps, on which cross-layer feature patterns are extracted.

C. Local Feature Encoding

The re-organized feature tensor $\mathcal{G}_d$ contains rich structures that encode complex spatial relationship and hierarchical characteristics along layers. The second phase of CLPE aims at extracting discriminative and local features in $\mathcal{G}_d$. Towards this end, we use a spatially-sliding window to sample 3D local patches at each spatial location of $\mathcal{G}_d$, and extract a binary code for each 3D local patch. Specifically, for a 3D local patch denoted by $\mathcal{X} \in \mathbb{R}^{k\times k\times T}$, a set of P equally-spaced pixels distributed on a ball around the patch center with radius r is sampled; see Fig. 1. Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_P]$ denote the values of sampled pixels and M the mean value of $\mathcal{X}$. We define a binary code for $\mathcal{X}$ as follows:

$$\mathcal{B}(\boldsymbol{x}, M) = \sum_{p=1}^{P} \mathcal{S}(x_p - M) 2^{p-1}, \tag{3}$$

where $\mathcal{S}(x) = 1$ if $x \geq 0$ and 0 otherwise. Image rotation, an often-seen image transformation, will cause the permutation within $\boldsymbol{x}$, changing the binary code. To achieve the robustness against rotation, a minimized code among different permutations is further calculated as

$$\mathcal{C}(\mathcal{X}) = \min_{\boldsymbol{T}\in\mathbb{T}} \mathcal{B}(\boldsymbol{T}\boldsymbol{x}, M), \tag{4}$$

where $\mathbb{T}$ is a set of permutation matrices corresponding to rotations on the sampling ball. The minimization over these permutations leads to rotation-invariance on the code $\mathcal{C}(\mathcal{X})$.

Let $\mathcal{X}_{\boldsymbol{p},d}$ denote the 3D local patch locating at the position $\boldsymbol{p}$ of the d-th grouped tensor $\mathcal{G}_d$, where

$$\boldsymbol{p} \in \mathbb{I} = \{1, \ldots, H\} \times \{1, \ldots, W\} \in \mathbb{Z}_+^2. \tag{5}$$

For local feature encoding, we calculate the binary codes $c_{\boldsymbol{p},d} = \mathcal{C}(\mathcal{X}_{\boldsymbol{p},d})$ for all $\boldsymbol{p}$ and all d. Then, the binary codes for the same position $\boldsymbol{p}$ but different tensors $\mathcal{G}_d s$ are concatenated into a feature vector:

$$\boldsymbol{c}_{\boldsymbol{p}} = [c_{\boldsymbol{p},1}, \ldots, c_{\boldsymbol{p},D}] \in \mathbb{R}^D. \tag{6}$$

Afterwards, a linear projection, denoted by $\boldsymbol{W} \in \mathbb{R}^{N\times D} (N < D)$ and implemented by 1×1 convolutions, is further applied to obtaining refined lower-dimensional features $\boldsymbol{f}_{\boldsymbol{p},d}$:

$$\boldsymbol{f}_{\boldsymbol{p}} = \boldsymbol{W}\boldsymbol{c}_{\boldsymbol{p}} \in \mathbb{R}^N,\ \boldsymbol{p} \in \mathbb{I}, \tag{7}$$

where the linear projection is implemented by 1×1 convolutions, and N is the projected feature length.

D. Spatial Pyramid Histogram Pooling

In this phase, the generated set of local features $\{\boldsymbol{f}_{\boldsymbol{p}} | \boldsymbol{p} \in \mathbb{I}\}$ aggregated into a single cross-layer descriptor. Instead of simply using GAP, we consider calculating a soft histogram for each dimension $n \in \{1, \ldots, N\}$ in $\boldsymbol{f}_{\boldsymbol{p}} \in \mathbb{R}^N$ over all spatial position $\boldsymbol{p}$. Specifically, the soft histogram, denoted by $\boldsymbol{h}_n = \text{Hist}(\{\boldsymbol{f}_{\boldsymbol{p}}(n) | \boldsymbol{p} \in \mathbb{I}\})$, is calculated as

$$\boldsymbol{h}_n(b) = \sum_{\boldsymbol{p}\in\mathbb{I}} \frac{\exp(-\gamma_{n,b}^2 \cdot (\boldsymbol{f}_{\boldsymbol{p}}(n) - \mu_{n,b})^2)}{\sum_{k'=1}^{B} \exp(-\gamma_{n,b'}^2 \cdot (\boldsymbol{f}_{\boldsymbol{p}}(n) - \mu_{n,b'})^2)}, \tag{8}$$

for $k = 1, \ldots, K$, where $\mu_{d,b'}$s and $\gamma_{d,b'}$s are learnable bin centers and scaling factors, respectively, and K is the number of bins.

As the global soft histogram defined by (8) completely discards spatial orders, potentially limiting the descriptive capacity, we apply the idea of the spatial pyramid [56] in the histogram calculation. Specifically, we partition the image into $2^\ell \times 2^\ell$ regions for the ℓ-th scale, with $\ell = 0, 1, \ldots, L$, resulting in $M = (4^{L+1} - 1)/3$ regions. The soft histograms are then calculated for all the regions and finally concatenated into a descriptor vector for the whole image. Let $\mathbb{I}^\ell$ denote the spatial indices of the ℓ segment and $\boldsymbol{h}^{(\ell)} = \text{Hist}(\{\boldsymbol{f}_p | p \in \mathbb{I}^\ell\})$ denote the corresponding histogram. The global feature for the n-th dimension is then constructed via concatenation:

$$\boldsymbol{f}_n = [\boldsymbol{h}_n^{(0)}, \ldots, \boldsymbol{h}_n^{(M-1)}].$$

The bin centers and scaling factors in $\boldsymbol{h}_n^{(\ell)}$ are shared among different ℓ but varied across different n. The final CLPE feature vector is then defined as

$$\boldsymbol{f}_{\text{CLPE}} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_N]. \tag{9}$$

E. Training Loss

Given a set of images and their subjective scores $\{s_i^*\}_{i=1}^M$ measured by human. Let $\{s_i\}_{i=1}^M$ denote the objective scores predicted by our CLPENet model. The training loss is then defined as

$$\mathcal{L} = \sum_{i=1}^{M} \mathcal{L}_\delta(s_i, s_i^*), \tag{10}$$

where $\mathcal{L}_\delta$ is the parametrized Huber loss defined by

$$\mathcal{L}_\delta(s, s^*) = \begin{cases} \frac{1}{2}(s - s^*)^2, & \text{for } |s - s^*| \leq \delta, \\ \delta\left(|s - s^*| - \frac{1}{2}\delta\right), & \text{otherwise.} \end{cases} \tag{11}$$

Here δ is a threshold to select the way to penalty outliers, which is set to $1/9$, as suggested in [16]. We use the Huber loss for training due to its stronger robustness to outliers over the commonly-used mean square error loss, as demonstrated in existing literature.

TABLE I
PERFORMANCE COMPARISON ON SYNTHETICALLY-DISTORTED DATASETS. THE BEST RESULTS UNDER EACH METRIC ARE **BOLDFACED**

Method	CSIQ		TID2013		Kadid-10K	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NIQE [31]	0.632	0.726	0.343	0.378	0.374	0.428
ILNIQE [32]	0.832	0.873	0.570	0.598	0.531	0.563
PQR [66]	0.872	0.901	0.745	0.798	-	-
DeepIQA [44]	0.955	0.973	0.761	0.787	0.628	0.647
DBCNN [20]	0.946	0.959	0.816	0.865	-	-
SGDNet [13]	0.883	0.903	0.843	0.861	-	-
NSSADNN [67]	0.893	0.927	0.844	0.910	-	-
MetaIQA [15]	-	-	-	-	0.767	0.774
CaHDC [68]	0.874	0.915	0.862	0.878	-	-
HyperNet [17]	0.933	0.947	0.857	0.898	0.896	0.907
SiamIQA [69]	0.962	-	0.855	-	0.913	-
AIGQA [70]	0.927	0.952	0.871	0.893	0.864	0.863
UNIQUE [48]	0.902	0.927	0.855	0.879	0.876	0.878
OLNet [45]	0.966	0.975	0.863	0.889	-	-
VCRNet [71]	0.943	0.955	0.856	0.875	-	-
MSBIQA [72]	0.969	0.978	0.835	0.859	-	-
CLRNet [73]	0.915	0.938	0.837	0.863	-	-
GraphIQA [74]	0.947	0.959	-	-	-	-
VIPNet [75]	0.963	0.966	0.891	0.902	0.954	0.955
SAWAR [23]	0.952	0.960	0.884	0.896	0.928	0.932
KGANet [77]	0.954	0.963	0.927	0.933	0.940	0.943
DDNet [37]	0.971	0.977	0.956	0.961	0.925	0.934
HITS [76]	-	-	0.884	0.857	0.891	0.887
CLPENet [Ours]	**0.978**	**0.977**	**0.961**	**0.968**	**0.955**	**0.961**

TABLE II
PERFORMANCE COMPARISON ON AUTHENTICALLY-DISTORTED DATASETS. THE BEST RESULTS UNDER EACH METRIC ARE **BOLDFACED**

Method	LIVE-C		KonIQ-10K		SPAQ	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NIQE [31]	0.464	0.515	0.601	0.597	0.703	0.712
ILNIQE [32]	0.469	0.536	0.552	0.573	0.714	0.721
PQR [66]	0.857	0.882	0.881	0.884	-	-
DeepIQA [44]	0.671	0.686	0.797	0.805	-	-
DBCNN [20]	0.851	0.869	0.875	0.884	0.911	0.915
SGDNet [13]	0.851	0.872	0.897	0.917	-	-
NSSADNN [67]	0.745	0.813	-	-	-	-
MetaIQA [15]	0.802	0.835	0.851	0.887	0.875	0.877
CaHDC [68]	0.738	0.744	-	-	0.827	0.834
HyperNet [17]	0.859	0.882	0.906	0.917	0.916	0.918
SiamIQA [69]	0.851	-	0.894	-	-	-
AIGQA [70]	0.751	0.761	-	-	-	-
UNIQUE* [48]	0.854	**0.890**	0.896	0.901	-	-
OLNet [45]	0.849	0.858	0.877	0.882	-	-
VCRNet [71]	0.856	0.865	-	-	-	-
CLRNet [73]	0.832	0.866	-	-	-	-
GraphIQA [74]	0.845	0.862	0.911	0.915	-	-
VIPNet [75]	0.827	0.848	0.915	**0.928**	-	-
SAWAR [23]	0.853	0.871	0.898	0.906	-	-
DDNet [37]	0.853	0.876	0.916	0.926	0.917	0.920
HITS [76]	0.718	0.697	0.821	0.798	-	-
CLPENet [Ours]	**0.861**	0.880	**0.921**	**0.928**	**0.918**	**0.921**

IV. EXPERIMENTS

A. Experimental Setup

1) Implementation Details of CLPENet: Our proposed method is implemented in PyTorch on an NVIDIA RTX 3090 GPU. In the CLPE module, (i) the parameters for cross-layer feature grouping are set as: $T = 5$ and $D = 16$; (ii) the parameters for cross-layer feature encoding are set as: $k = 5$ for 3D local patches, $P = 14$ and $r = 2$ for the sampling ball, and $N = 16$ for the linear projection; and (iii) the parameters for spatial pyramid histograms are set as: $B = 8$ and $L = 2$. Following [16], [57], [58], we use ResNet-101 [59] as the backbone. Specifically, the first three blocks (conv1, conv2_x, conv3_x) are used as ResBlock1, ResBlock2 and ResBlock3. The ResNet backbone is initialized with pre-trained models and its other parameters are initialized using Kaiming. For training, the AdamW optimizer is called for 100 epochs, with an initial learning rate of 0.2 and a weight decay factor of $1e{-}2$. The weights of ResBlock1 are frozen in training. Our code will be released via GitHub upon paper's acceptance.

2) Datasets: Totally six publicly available BIQA datasets of natural images are used for experimental evaluation, including (i) three synthetic image datasets with artificial distortions: CSIQ [60], TID2013 [61] and Kadid-10K [62]; and (ii) three authentic datasets with realistic distortion: LIVE-C [63], KonIQ-10K [64] and SPAQ [65]. See followings for the details of these datasets:

- CSIQ: 30 pristine images and 866 distorted images with 6 distortion types at 4 to 5 distortion levels.
- TID2013: 25 pristine images and a total of 3000 distorted images with 17 distortion types at 4 degradation levels.
- Kadid-10 K: 81 pristine images individually degraded by 25 distortions in 5 levels, resulting in 10125 distorted images.
- LIVE-C: 1,162 realistically natural pictures with a resolution of 500×500 pixels.
- KonIQ-10 K: 10073 realistically and complexly distorted images with resolution of 1024×768 pixels.
- SPAQ: 11125 images captured by 66 smartphones with diverse resolutions.

Following [16], [37], images are used in their original resolutions in the datasets to test generalization to different sizes. Following [12], [16], 80% of images are randomly sampled for training and the rest for testing. For synthetic datasets, the split ensures pristine image content does not intersect between sets. Same as [16], data augmentation of horizontal/vertical flips and $\pm 1^{\circ}$ rotation is randomly applied during training, improving performance slightly in most cases. Extra borders from rotation are removed by cropping. For performance comparison, median values of evaluation metrics over 10 test set sessions are reported.

3) Evaluation Metrics and Compared Methods: Two commonly-used performance metrics in BIQA, namely the Spearman Rank Order Correlation Coefficient (SROCC) and the Pearson Linear Correlation Coefficient (PLCC), are adopted for performance comparison. The SROCC measures prediction monotonicity, while the PLCC measures linear correlation. An effective IQA metric should yield high values for both PLCC and SROCC.

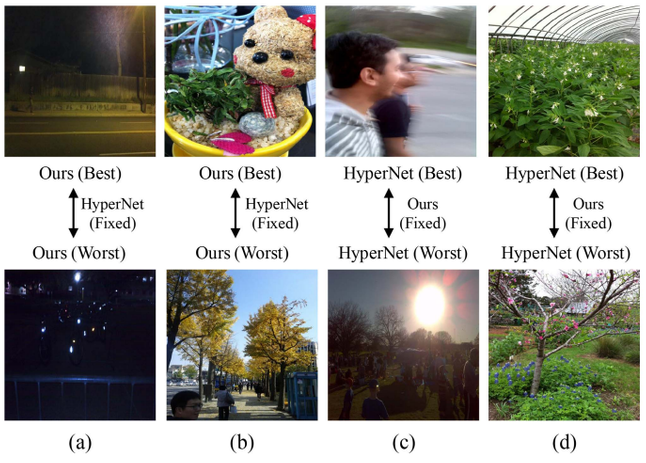


Fig. 4. Representative gMAD pairs between CLPENet and HyperNet on the Live-C dataset. (a) Fixing HyperNet at a low quality level. (b) Fixing HyperNet at a high quality level. (c) Fixing CLPENet at a low quality level. (d) Fixing CLPENet at a high quality level.

Two well-known traditional handcrafted BIQA models, namely NIQE [31] and ILNIQE [32], are selected for performance comparison. In addition, several recent deep learning-based methods are selected as baselines: PQR [66], DeepIQA [44], DBCNN [20], SGDNet [13], NSSADNN [67], CaHDC [68], HyperNet [17], MetaIQA [15], SiamIQA [69], AIGQA [70], UNIQUE [48], OLNet [45], VCRNet [71], MSBIQA [72], CLRNet [73], GraphIQA [74], VIPNet [75], SAWAR [23], DDNet [37], HITS [76] and KGANet [77], where last method are just desinged for synthetic data. Among the deep learning-based BIQA methods, ten of them are the most advanced ones developed in the past three years. Their experimental results are directly quoted from the original papers whenever available and fair, or otherwise obtained by running the codes provided by the authors.

B. Comparison Against State-of-The-Arts

1) Evaluation on Individual Datasets: We first conduct evaluations on individual datasets. The results on synthetically distorted images from the CSIQ, TID2013 and Kadid-10 K datasets are reported in Table I. Generally, all deep models outperform the two hand-crafted methods, indicating the superiority of deep learning. Among the deep learning-based methods, our CLPENet performs best in terms of both SROCC and PLCC. Noticeable performance gains are also observed on the more challenging datasets including TID2013 and Kadid-10 K, which contains more diverse content types and more complex distortion types. It is worth noting that while the compared methods performed well on CSIQ, they showed a larger decrease in performance on TID2013 and Kadid-10 K.

Table II presents the results on authentically-distorted images from the LIVE-C, Koniq-10 K and SPAQ datasets. Again, deep models outperform hand-crafted methods by a wide margin. Among the deep models, our CLPENet significantly outperforms others on KonIQ-10 K, a large-scale dataset, while achieving comparative performance to UNIQUE on LIVE-C. Note that UNIQUE$*$ is marked because it uses all datasets for training, while ours uses individual ones respectively.

TABLE III
PERFORMANCE COMPARISON ON MODELS TRAINED ON KONIQ-10 K AND TESTED ON LIVE-C/SPAQ WITHOUT FINE-TUNING. THE BEST RESULT UNDER EACH SETTING IS **BOLDFACED**

LIVE-C	VIPNet	DBCNN	HyperNet	UNIQUE	CLPENet
SROCC	0.800	0.755	0.785	0.786	**0.801**
PLCC	-	-	0.818	-	**0.821**

SPAQ	DBCNN	CaHDC	HyperNet	CLPENet
SROCC	0.783	0.730	0.807	**0.809**
PLCC	0.792	0.778	0.818	**0.819**

TABLE IV
RESULTS OF D-TEST, L-TEST AND P-TEST

Method	D-test	L-test	P-test
DeepIQA [44]	0.907	0.947	0.963
DBCNN [45]	0.937	0.950	**0.999**
HyperNet [17]	0.941	0.977	**0.999**
VIPNet [76]	0.941	0.986	**0.999**
DDNet [37]	0.939	0.979	0.997
Ours	**0.942**	**0.988**	**0.999**

All above experiments have demonstrated The superior performance of CLPENet over the compared methods. This superiority is mainly attributed to its capability of learning more accurate quality-aware features by the combination of global semantic and cross-layer features, enabling better generalization of complex distortions.

2) Cross-Dataset Evaluation: To further assess the generalization capability of CLPENet, cross-dataset experiments are conducted using KonIQ-10 K as the training set while using LIVE-C and SPAQ for test. See Table III for the results, where only methods with available results are included for comparison. The results show that in the cross-dataset case, CLPENet can still closely aligns with the subjective perception of quality by the human visual system, outperforming other methods. It clearly indicates that CLPENet effectively generalizes in predicting image quality for images with diverse resolutions, varying contents, and real-world complex distortions.

To further validate our method's ability to discriminate image quality differences, we conduct three complementary tests on the Waterloo Exploration Database using our model trained on KonIQ. These include (1) the pristine/distorted image discriminability test (D-test), (2) the listwise ranking consistency test (L-test), and (3) the pairwise preference consistency test (P-test). The D-test evaluates the model's ability to distinguish between pristine and distorted images. The L-test assesses whether the predicted quality scores maintain consistent ranking across different distortion levels of the same content. The P-test measures prediction accuracy on quality-discriminable image pairs. As shown in Table IV, our method achieves top-tier performance across all three tests, demonstrating its strong capability in quality discrimination across different evaluation scenarios.

TABLE V
SROCC RESULTS ON DIVERSE DISTORTION TYPES OF TID2013. THE BEST RESULT FOR EACH DISTORTION TYPE IS **BOLDFACED**

Distortion	VIPNet	MetaIQA	HyperNet	AIGQA	VCRNet	CLPENet
AGN	0.859	0.947	0.769	0.932	0.844	**0.933**
ANC	0.895	0.924	0.613	0.916	0.785	**0.923**
SCN	0.932	**0.955**	0.918	0.944	0.787	0.943
MN	0.655	0.728	0.448	0.662	**0.795**	0.721
HFN	0.933	0.952	0.839	0.953	0.942	**0.968**
IN	0.824	0.866	0.758	0.911	0.876	**0.921**
QN	0.924	0.745	0.828	**0.908**	0.847	0.903
GB	0.921	**0.977**	0.873	0.917	0.906	0.938
DEN	0.943	0.938	0.804	0.914	**0.937**	**0.933**
JPEG	0.948	0.934	0.860	0.945	0.934	**0.951**
JP2K	**0.950**	0.957	0.888	0.932	0.906	0.949
JGTE	0.819	0.931	0.723	0.858	0.762	**0.934**
J2TE	0.832	0.903	0.846	0.898	0.865	**0.918**
NEPN	**0.777**	0.729	0.369	0.130	0.457	0.675
Block	**0.759**	0.391	0.428	0.723	0.601	0.609
MS	0.461	0.402	0.424	0.554	0.509	**0.543**
CTC	0.866	0.764	0.740	0.830	0.595	**0.867**
CCS	0.613	0.829	0.710	0.689	**0.855**	0.831
MGN	0.947	0.939	0.767	0.948	0.845	**0.948**
CN	0.865	**0.952**	0.786	0.886	0.804	0.938
LCNI	0.902	**0.978**	0.879	0.897	0.816	0.974
ICQD	0.898	0.859	0.785	0.908	**0.945**	**0.916**
CHA	0.814	0.927	0.739	0.889	**0.932**	0.917
SSR	0.950	0.974	0.910	0.908	0.948	**0.977**

Following [22], a gMAD (Group MAximum Differentiation) competition [78] is performed on the Live-C dataset to visually compare our CLPENet with the HyperNet, a top competing CNN model in most experiments, especially on authentic data and a code-available model. The gMAD competition efficiently selects challenging cross-quality image pairs to reveal differences in two models' robustness. In gMAD, one model attacks by identifying pairs it ranks further apart in quality, while the other defends by grouping images at the same level. Human observers are shown with selected pairs and asked to judge which model better matches the perceived quality. Fig. 4 shows representative image pairs using models trained on KonIQ-10 K. In Fig. 4(a) and (b), the upper images have slightly better perceptual quality and clearer structures compared to the images below, indicating CLPENet successfully attacked HyperNet. As the defender, CLPENet survived HyperNet's attacks at low and good quality levels, as Fig. 4(c) and (d) show. Both models successfully recognize obvious low-quality images.

3) Comparison on Diverse Distortion Types: We further analyze the effectiveness across different distortion types, using the TID2013 and CSIQ datasets. The results are collected in Tables V and VI. Our CLPENet achieves competitive results on individual distortion types compared to other CNN-based methods. On the TID2013 dataset, CLPENet attains the best performance on 13 out of 24 distortion types. On the CSIQ dataset, CLPENet surpasses all other methods on five distortion types. This performance advantage of CLPENet across different degradation types stems from CLPENet's effective exploitation of cross-layer hierarchical features, taking account into both local-scale and global-scale characteristics during the BIQA process.

TABLE VI
SROCC RESULTS ON DIVERSE DISTORTION TYPES OF CSIQ. THE BEST RESULT FOR EACH DISTORTION TYPE IS **BOLDFACED**

Distortion	GraphIQA	DBCNN	HyperNet	OLNet	VCRNet	CLPENet
GB	0.947	0.947	0.915	0.965	0.950	**0.967**
AWGN	0.948	0.948	0.927	0.945	0.939	**0.972**
JPEG	0.947	0.940	0.934	0.968	0.956	**0.970**
JP2K	0.947	0.953	0.960	0.945	0.962	**0.969**
APN	0.948	0.941	0.931	0.953	0.899	**0.963**
CTD	**0.947**	0.872	0.874	0.925	0.919	0.933

TABLE VII
SROCC RESULTS IN ABLATION STUDY. THE BEST RESULTS ARE **BOLDFACED**

Dataset	w/o CLPE	Backbone$^+$	CLPENet
CSIQ	0.931	0.939	**0.978**
TID2013	0.921	0.934	**0.961**
Live-C	0.832	0.821	**0.861**
KonIQ-10K	0.899	0.911	**0.921**

TABLE VIII
SROCC COMPARISON OF DIFFERENT CROSS-LAYER PATTERN ENCODING SCHEMES. THE BEST RESULTS ARE **BOLDFACED**

Cross-layer Encoding Scheme		Kadid-10K	Live-C	KonIQ-10K
Baseline-1	None	0.929	0.825	0.902
Baseline-2	None	0.935	0.841	0.904
3D Statics	Mean	0.935	0.838	0.904
	Variance	0.939	0.842	0.906
CLPENet	CLPE	**0.961**	**0.861**	**0.921**

C. Ablation Study and More Analysis

1) Ablation Study on CLPE: We conduct ablation study to verify the effectiveness of our proposed CLPE module, using the CSIQ, TID2013, LIVE-C and Koniq-10 K datasets in terms of SROCCs under several settings. Towards this end, we construct two variants of CLPENet. (i) 'w/o CLPE': a baseline model built by removing the CLPE module in CLPENet; (ii) 'Backbone$^+$': the ResNet101 model with more channels in each layer to maintain a similar number of parameters with CLPENet, which can be viewed as 'w/o CLPEE' with more parameters. See Table VII for the results. Notably, removing CLPE significantly decreases performance across the datasets. In addition, CLPENet also outperforms 'Backbone$^+$', an enlarged version of 'w/o CLPEE'. This implies the gains stem from the specific CLPE module rather than simply increasing model size.

2) Further Verification on Cross-Layer Feature Encoding: We compare CLPENet against (i) Baseline-1, which is the ResNet101 backbone using GAP at the last convolutional layer; (ii) Baseline-2, which separately performs GAP on features produced at each convolutional layer and concatenates the resulting descriptors; and (iii) two common statistical descriptors on 3D multi-scale cross-layer features. FC layers are used in these compared models to map image features to quality scores. As shown in Table VIII, Baseline-2 is better than Baseline-1, indicating the

TABLE IX
SROCC RESULTS OF ABLATION STUDY ON KADID-10 K IN TERMS OF DIFFERENT DISTORTION TYPES. THE BEST RESULTS ARE **BOLDFACED**

Distortion Type		Baseline-2	CLPENet
Blur	Gaussian blur	0.958	**0.979**
	Motion blur	0.937	**0.955**
Noise	White noise	0.937	**0.965**
	Impulse noise	0.944	**0.969**
Spatial distortion	Pixelate	0.935	**0.975**
	Color block	0.944	**0.972**

TABLE X
PERFORMANCE OF CLPENET USING VARIED CNN BACKBONES

Backbone	TID2013		KonIQ-10K	
	SROCC	PLCC	SROCC	PLCC
ResNet152	0.966	0.969	0.922	0.926
ResNet101	0.961	0.968	0.921	0.927
ResNet50	0.960	0.963	0.917	0.926
VGG16	0.945	0.947	0.786	0.795
MobileNetv3_small	0.885	0.889	0.793	0.794
ShuffleNetv2	0.822	0.827	0.758	0.757

benefit of using multi-layer features. Moreover, all cross-layer statistical encoding improves IQA performance over Baseline-1, validating our motivation to capture cross-layer statistics for BIQA. Further, our CLPENet achieved the largest performance gains on most datasets, demonstrating the effectiveness of its design in utilizing cross-layer features.

We also evaluate the performance of several common distortions on the Kadid-10 K dataset. The results, summarized in Table IX, are compared against Baseline-2 to demonstrate the performance gain of our CLPE scheme over the commonly-used GAP that is less efficient in capturing distortions that minimally alter average image features. Indeed, global-preserving distortions are prevalent, such as zero-mean noise corruption and blurring that maintains average intensity due to the sum-to-one property of blur kernels. Our CLPENet significantly outperformed Baseline-2 on these average-insensitive distortions, including blur, noise, and spatial distortions. This has validated the capability of our proposed CLPE to effectively identify and assess diverse degradation types.

3) Performance Using Varied CNN Backbones: Our CLPE module can be incorporated with other CNN backbones. We replace the ResNet101 backbone in CLPENet by ResNet152, ResNet101, ResNet50, and VGG16, respectively. The results are listed in Table X, showing that varied ResNet backbones maintain state-of-the-art performance compared to previous benchmarks in Tables I and II. The deeper ResNet152 backbone provides additional gains on some datasets versus ResNet101. However, the computational cost increases with larger backbones. The consistently strong performance across ResNet models demonstrates the effectiveness of our method in releasing the potential of CNN-based BIQA approaches.

To further compress CLPENet for efficiency, especially for real-time applications on mobile devices. We also evaluated CLPENet performance with famous lightweight backbones, including MobileNetv3 small and ShuffleNetv2. The results are shown in Table X, indicating that using very lightweight backbones of smaller model complexity in quality assessment models tends to lead to decreased but acceptable performance.

TABLE XI
COMPARISON OF MODEL COMPLEXITY BY #FLOPS

Koncept512	SGDNet	CaHDC	HyperNet	CLPENet
2.58×10^{11}	1.23×10^{11}	0.37×10^{11}	1.98×10^{11}	1.31×10^{11}

TABLE XII
INFLUENCE OF PARAMETER IN LOSS FUNCTION

Kadid-10k	SROCC	PLCC
δ = 0	0.949	0.951
δ = 1/9	0.955	0.961
δ = 1	0.953	0.959
δ = 100	0.951	0.956

TABLE XIII
INFLUENCE OF PARAMETERS IN CLPE MODULE

KonIQ-10k	SROCC	PLCC
$k = 3$	0.919	0.921
$k = 5$	0.921	0.928
$k = 7$	0.915	0.916
$T = 3$	0.920	0.924
$T = 5$	0.921	0.928
$T = 7$	0.915	0.915
$P = 10$	0.909	0.911
$P = 14$	0.921	0.928
$P = 18$	0.913	0.919

4) Model Complexity: The computational complexity of CLPENet is measured by the number floating point operations per second (#FLOPS) in handling a 1024×768 color image. We include four recent methods for comparison: Koncept512 [64], SGDNet, CaHDC, and HyperNet. The evaluation is based on their released Python implementations. The results are listed in Table XI, showing that the complexity of CLPENet is smaller than that of HyperNet, one of the top-performing competitors in previous experiments. In addition, CLEPNet has a comparable #FLOPS to SGDNet. This indicates that CLPENet achieves state-of-the-art performance with an efficient model and this complexity is suitable for various multimedia systems and applications.

5) Influence of Parameter in Loss Function: Our empirical results show that $\delta = 1/9$ yields optimal performance, which aligns with the findings in other works, as shown in Table XII. When δ is too small, the loss function behaves similarly to L1 loss, potentially under-penalizing small errors. Conversely, when δ is too large, it approximates MSE loss, becoming more sensitive to outliers. The chosen threshold of 1/9 balances robustness to outliers and penalization of prediction errors.

6) Influence of Parameters in CLPE Module: The performance of the CLPE module is significantly influenced by three key parameters: the spatial dimension k, channel dimension T, and the number of sampling points P. Through extensive experiments in Table XIII, we found that $k = 5$ provides the optimal spatial window size for capturing local distortion patterns - smaller values miss crucial quality-relevant features, while larger values introduce unnecessary computational overhead

without meaningful performance gains. Similarly, the temporal dimension $T = 5$ was determined to be the sweet spot for cross-layer pattern extraction, as it enables sufficient depth for feature correlation while avoiding overfitting. The number of sampling points P directly impacts the module's ability to characterize local pattern variations, with our experiments showing that $P = 14$ sampling points achieves the best trade-off between feature representation capability and computational efficiency.

V. Conclusion and Future Work

This paper presented a novel CNN-based BIQA approach that exploits cross-layer features inside a CNN for image quality prediction. The cross-layer features are extracted by the CLPE module, which extracts local 3D binary patterns from grouped shape-normalized cross-layer feature tensors and calculates a soft histogram on these patterns to achieve an informative descriptor. Building the CLPE module into a ResNet101 backbone, we developed the CLPENet, with state-of-the-art performance demonstrated by extensive experiments across multiple datasets.

Our research advances BIQA through cross-layer feature integration and provides valuable insights for broader image communication and processing applications. Looking forward, the current work can be extended in three key directions. First, developing advanced local pattern descriptors through multi-valued pattern extractors, adaptive sampling strategies, and learnable pattern templates to discover more meaningful quality-related patterns. Second, aligning channel-wise features with specific artifact types to improve quality assessment accuracy by strengthening correlations between features and distortion characteristics. Third, the framework should be extended to incorporate perceptual priors, cross-modal learning, and domain-specific adaptations, broadening its applicability to tasks such as video quality assessment and other image processing challenges.

Acknowledgment

The authors express their deepest gratitude to Professor Yuhui Quan for his invaluable contributions to this work. His expertise, dedication, and insightful guidance were instrumental in shaping this research. Tragically, Professor Quan passed away in January 2025 due to malignant melanoma. The authors are profoundly saddened by his loss and dedicate this paper to his memory, honoring his lasting impact on our field and this study.

References

[1] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Ssim-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.

[2] P. Kancharla and S. S. Channappayya, "Completely blind quality assessment of user generated video content," *IEEE Trans. Image Process.*, vol. 31, pp. 263–274, 2022.

[3] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *J. Vis.*, vol. 17, no. 1, pp. 32–38, 2017.

[4] X. Liu, J. Van De Weijer, and A. D. Bagdanov, "Rankiqa: Learning from rankings for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1040–1049.

[5] Q. Jiang et al., "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 20, pp. 2035–2048, 2018.

[6] D. Li, T. Jiang, W. Lin, and M. Jiang, "Which has better visual quality: The clear blue sky or a blurry animal?," *IEEE Trans. Multimedia*, vol. 21, pp. 1221–1234, 2019.

[7] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, pp. 2457–2469, 2016.

[8] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 207–210, Feb. 2015.

[9] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 541–545, Apr. 2016.

[10] Y. Ding, Y. Zhao, and X. Zhao, "Image quality assessment based on multi-feature extraction and synthesis with support vector regression," *Signal Process. Image Commun.*, vol. 54, pp. 81–92, 2017.

[11] S.-C. Pei and L.-H. Chen, "Image quality assessment using human visual dog model fused with random forest," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3282–3292, Nov. 2015.

[12] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.

[13] S. Yang, Q. Jiang, W. Lin, and Y. Wang, "SGDNet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment," in *Proc. ACM Int. Conf. Multimedia*, 2019, pp. 1383–1391.

[14] L. Zheng, Y. Luo, Z. Zhou, J. Ling, and G. Yue, "Cdinet: Content distortion interaction network for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 26, pp. 7089–7100, 2024.

[15] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetaIQA: Deep meta-learning for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 14143–14152.

[16] Y. Su and J. Korhonen, "Blind natural image quality prediction using convolutional neural networks and weighted spatial pooling," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 191–195.

[17] S. Su et al., "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3664–3673.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[20] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 1, pp. 36–47, Jan. 2020 .

[21] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 12116–12128.

[22] Z. Zhou, Y. Xu, R. Xu, and Y. Quan, "No-reference image quality assessment using dynamic complex-valued neural model," in *Proc. ACM Int. Conf. Multimedia*, 2022, pp. 1006–1015.

[23] Z. Zhou, F. Zhou, and G. Qiu, "Blind image quality assessment based on separate representations and adaptive interaction of content and distortion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2484–2497, Apr. 2024.

[24] D. Temel and G. AlRegib, "PerSIM: Multi-resolution image quality assessment in the perceptually uniform color domain," in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 1682–1686.

[25] P. G. Freitas, W. Y. L. Akamine, and M. C. Q. Farias, "No-reference image quality assessment using orthogonal color planes patterns," *IEEE Trans. Multimedia*, vol. 20, pp. 3353–3360, 2018.

[26] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[27] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[28] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[29] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[30] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4850–4862, Nov. 2014.
[31] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[32] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[33] J. Xu et al., "Blind image quality assessment based on high order statistics aggregation," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4444–4457, Sep. 2016.
[34] J. Gu, G. Meng, S. Xiang, and C. Pan, "Blind image quality assessment via learnable attention-based pooling," *Pattern Recognit.*, vol. 91, pp. 332–344, 2019.
[35] Z. Wang, L. Yuan, and G. Zhai, "Channel attention for no-reference image quality assessment in DCT domain," *IEEE Signal Process. Lett.*, vol. 31, pp. 1274–1278, 2024.
[36] J. Guan, S. Yi, X. Zeng, W.-K. Cham, and X. Wang, "Visual importance and distortion guided deep image quality assessment framework," *IEEE Trans. Multimedia*, vol. 19, pp. 2505–2520, 2017.
[37] Z. Zhou, J. Li, D. Zhong, Y. Xu, and P. Le Callet, "Deep blind image quality assessment using dynamic neural model with dual-order statistics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 7, pp. 6279–6290, Jul. 2024.
[38] D. Chen, Y. Wang, and W. Gao, "No-reference image quality assessment: An attention driven approach," *IEEE Trans. Image Process.*, vol. 29, pp. 6496–6506, 2020.
[39] A. Li, J. Wu, Y. Liu, and L. Li, "Bridging the synthetic-to-authentic gap: Distortion-guided unsupervised domain adaptation for blind image quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 28422–28431.
[40] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
[41] K. Ma et al., "End-to-end blind image quality assessment using deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1202–1213, Mar. 2018.
[42] D. Varga, D. Saupe, and T. Szirányi, "Deeprn: A content preserving deep architecture for blind image quality assessment," in *Proc. IEEE Conf. Multimedia Expo*, 2018, pp. 1–6.
[43] K. Ma, X. Liu, Y. Fang, and E. P. Simoncelli, "Blind image quality assessment by learning from multiple annotators," in *Proc. Int. Conf. Image Process.*, 2019, pp. 2344–2348.
[44] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2018.
[45] X. Yao, Q. Cao, X. Feng, G. Cheng, and J. Han, "Learning to assess image quality like an observer," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8324–8336, Nov. 2023.
[46] P. C. Madhusudana, N. Birkbeck, Y. Wang, B. Adsumilli, and A. C. Bovik, "Image quality assessment using contrastive learning," *IEEE Trans. Image Process.*, vol. 31, pp. 4149–4161, 2022.
[47] N. C. Babu, V. Kannan, and R. Soundararajan, "No reference opinion unaware quality assessment of authentically distorted images," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2023, pp. 2459–2468.
[48] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.*, vol. 30, pp. 3474–3486, 2021.
[49] D. Pan et al., "Blind predicting similar quality map for image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6373–6382.
[50] J. You and J. Korhonen, "Transformer for image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, 2021, pp. 1389–1393.
[51] M. Zhu et al., "Saliency-guided transformer network combined with local embedding for no-reference image quality assessment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1953–1962.
[52] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2022, pp. 1220–1230.
[53] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "MUSIQ: Multi-scale image quality transformer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 5148–5157.
[54] K. Xu et al., "Boosting image quality assessment through efficient transformer adaptation with local feature enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 2662–2672.
[55] C. Zeng and S. Kwong, "Learning transformer features for image quality assessment," 2021, *arXiv:2112.00485*.
[56] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[57] A. Chetouani, "Image quality assessment without reference by mixing deep learning-based features," in *Proc. IEEE Conf. Multimedia Expo*, 2020, pp. 1–6.
[58] D. Li, T. Jiang, and M. Jiang, "Norm-in-norm loss with faster convergence and better performance for image quality assessment," in *Proc. ACM Int. Conf. Multimedia*, 2020, pp. 789–797.
[59] S.-H. Gao et al., "Res2net: A new multi-scale backbone architecture," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.
[60] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, 2010, Art. no. 11006.
[61] N. Ponomarenko et al., "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, 2015.
[62] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted IQA database," in *Proc. IEEE Conf. Qual. Multimedia Experience*, 2019, pp. 1–3.
[63] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 372–387, Jan. 2016.
[64] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 4041–4056, 2020.
[65] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone photography," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3677–3686.
[66] H. Zeng, L. Zhang, and A. C. Bovik, "A probabilistic quality representation approach to deep blind image quality prediction," 2017, *arXiv:1708.08190*.
[67] B. Yan, B. Bare, and W. Tan, "Naturalness-aware deep no-reference image quality assessment," *IEEE Trans. Multimedia*, vol. 21, pp. 2603–2615, 2019.
[68] J. Wu et al., "End-to-end blind image quality prediction with cascaded deep neural network," *IEEE Trans. Image Process.*, vol. 29, pp. 7414–7426, 2020.
[69] W. Zhang, K. Zhai, G. Zhai, and X. Yang, "Learning to blindly assess image quality in the laboratory and wild," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 111–115.
[70] J. Ma et al., "Blind image quality assessment with active inference," *IEEE Trans. Image Process.*, vol. 30, pp. 3650–3663, 2021.
[71] Z. Pan et al., "Vcrnet: Visual compensation restoration network for no-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 31, pp. 1613–1627, 2022.
[72] M. Liu, J. Huang, D. Zeng, X. Ding, and J. Paisley, "A multiscale approach to deep blind image quality assessment," *IEEE Trans. Image Process.*, vol. 32, pp. 1656–1667, 2023.
[73] F.-Z. Ou, Y.-G. Wang, J. Li, G. Zhu, and S. Kwong, "A novel rank learning based no-reference image quality assessment method," *IEEE Trans. Multimedia*, vol. 24, pp. 4197–4211, 2022.
[74] S. Sun, T. Yu, J. Xu, W. Zhou, and Z. Chen, "GraphIQA: Learning distortion graph representations for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 25, pp. 2912–2925, 2023.
[75] X. Wang, J. Xiong, and W. Lin, "Visual interaction perceptual network for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 25, pp. 8958–8971, 2023.
[76] C. Nandhini and M. Brindha, "Hierarchical patch selection: An improved patch sampling for no reference image quality assessment," *IEEE Trans. Artif. Intell.*, vol. 5, no. 2, pp. 541–555, Feb. 2024.
[77] T. Zhou, S. Tan, B. Zhao, and G. Yue, "Multitask deep neural network with knowledge-guided attention for blind image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 7577–7588, Aug. 2024.
[78] K. Ma et al., "Group mad competition-a new methodology to compare objective image quality models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1664–1673.

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3359667

Deep Blind Image Quality Assessment Using Dynamic Neural Model with Dual-order Statistics

Zihan Zhou, Jing Li, Dexiang Zhong*, Yong Xu, Patrick Le Callet

***Abstract*—Deep convolutional neural networks (CNNs) have increasingly become a prominent method for blind image quality assessment (BIQA). The process of quality assessment typically involves feature extraction, average-based pooling, and quality regression. Based on this process, as well as the consensus that the visual quality of an image mainly relies on its content and distortions, this work improves CNNs for BIQA in two ways. First, considering the content-awareness of visual quality perception, we incorporate content-awareness via a dynamic filtering module to extract content-adaptive features and a dynamic regression module to learn content-adaptive perception rules based on local content and global semantics. Second, considering distortion-sensitivity in visual quality perception, we introduce second-order global variance pooling and combine it with global average pooling (GAP). First-order pooling methods like GAP are limited in distinguishing complex distortions that cause local degradation while preserving global features. Thus, pooling with dual-order statistics enables a more distortion-sensitive and discriminative global representation. These two improvements result in a content-adaptive BIQA model with a dual-order global pooling mechanism, improving generalization on diverse images with varying contents and distortion types. Extensive experiments on synthetic and authentic distortion datasets demonstrate state-of-the-art performance of the proposed approach.**

***Index Terms*—Blind Image Quality Assessment, Convolutional Networks, Dynamic Filtering, Feature Pooling, Dynamic Regression**

I. Introduction

DIGITAL images are ubiquitous in modern life and work for communication, entertainment and data analysis. However, the quality of images can be compromised by various distortions that occur during acquisition, processing, transmission, and display. This has prompted interest in objective image quality assessment (IQA) algorithms for automated quality evaluation. Accurately quantifying the human perception of image quality has a plenty of potential applications, such as instant feedback generation for image collection systems, automatic optimization of camera settings or post-processing parameters, guidance of designing image restoration models [1], [2], automatic adjustment of objective parameters based on given image quality in image compression [3] and image watermarking [4], and quality labeling of user-generated-content visual data.

IQA can be divided into three categories: full-reference (FR), reduced-reference (RR), and no-reference (NR). FR-IQA approaches (*e.g.* [5], [6], [7], [8], [9]) require a pristine reference of very high quality. Therefore, FR-IQA is suitable for scenarios such as image compression and image watermarking, where a reference image is available. RR-IQA methods (*e.g.* [10], [11], [12]) enable quality evaluation using partial information of reference images, in the form of features, rather than the complete image. RR-IQA is particularly suitable for image transmission and communication, where the reference images are not accessible to the receivers. In these cases, the features of reference images, rather than the reference images themselves, are transmitted to the receivers and compared to those of the transmitted images, as demonstrated in [10].

NR (*i.e.* Blind)-IQA aims to estimate the quality scores of distorted images without access to any information about reference versions, making it more challenging but applicable to more scenarios. Due to the relaxed requirements on the references, NR-IQA approaches are more flexible than the FR and RR ones. Over the past few decades, significant efforts have been dedicated to the development of blind-IQA (BIQA) methods (*e.g.* [13], [14], [15], [16]). This topic has garnered considerable attention from both industry and academia, owing to its immense potential and value in many practical applications [17], [18].

BIQA approaches broadly follow a two-stage process: quality-aware feature extraction and quality score regression. Early approaches usually adopt hand-crafted designs for feature extraction, such as natural scene statistics [19], [20] and variants of local binary patterns [21], [22]. Regression then utilizes well-established learning-based models, such as support vector regression, random forest, and Gaussian process. In recent years, inspired by the success of deep learning in image recognition, a series of CNN-based and transformer-based BIQA approaches have been proposed, where the two phases are jointly optimized in an end-to-end manner, as seen in [23], [24], [25], [26]. These data-driven models, consisting of convolutional-layer-based feature extractors, average-based pooling mechanisms, and fully connected-layer-based score regressors, have shown impressive performance. Following this line of research, this paper aims to improve the power of CNN-based models for NR-IQA in the following two aspects.

Introducing dynamic filtering and dynamic regression for improving content-awareness. An effective IQA model should be capable of handling images with diverse types and degrees of distortion, as well as varying semantic content. In

Zihan Zhou, Dexiang Zhong are with the College of Mathematics and Informatics at South China Agricultural University, China. Jing Li is with the Moku Lab, Alibaba Group, Beijing, China. Yong Xu is with the School of Computer Science and Engineering at South China University of Technology, China; PengCheng Lab, Shenzhen, China; Pazhou Lab, GuangZhou, China. Patrick Le Callet is with Nantes Universite, Ecole Centrale Nantes, CAPACITES SAS, CNRS, LS2N, UMR 6004, F-44000 Nantes, France. Email: zhouzihan@scau.edu.cn, dxzhong@scau.edu.cn, lj225205@alibaba-inc.com, yxu@scut.edu.cn, patrick.lecallet@univ-nantes.fr. Asterisk indicates the corresponding author.

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3359667

general, different image contents may be sensitive to different types of distortions. For example, flattened or uniform areas are often more sensitive to noise distortions, while textured areas are more sensitive to blurring. Consider a large flattened area - it may be perceived as high quality if it is part of a clear blue sky image, but could be seen as a serious blurring artifact if it occurs in a highly textured image. As demonstrated in [27], human judgments of visual quality for identical distortion patterns even vary based on the surrounding content and overall semantic meaning of the image.

A traditional convolutional neural network (CNN) model is typically content-agnostic. This is because the spatially-invariant convolutional filters are shared across all spatial locations and input images for a trained model. However, optimal IQA models often require content-aware feature extraction. Additionally, quality-related feature extraction should be adaptive to image content, with filters tailored to the specific patterns being analyzed, as depicted in Fig. 1. Here, red spots in different semantic regions demand different filters for quality-aware perception. For most IQA methods relying on standard convolution, an image of clear blue sky may be erroneously judged as low quality due to the large uniform areas, which could be mistaken as blurring without content-aware processing [28]. Due to the immense diversity of possible distortions and image contents, it is impractical to design a fixed filter bank covering all potential patterns, especially for authentically distorted images. Therefore, dynamic filters adapted to local image patterns and conditioned on the input image content are desirable for IQA, as introduced in this work. The capability to adapt filter parameters based on input content has shown promise for modeling complex visual patterns [29], [30], which is advantageous for the complex nature of human visual quality perception.

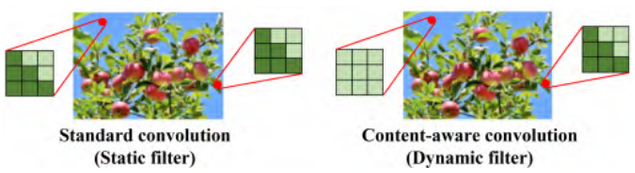
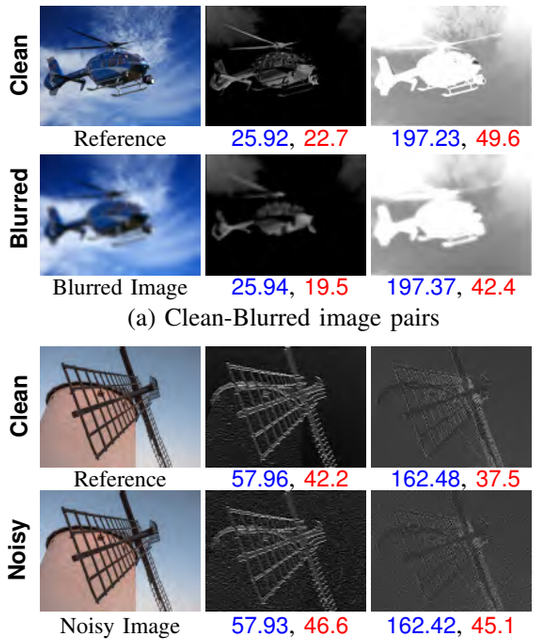


Fig. 1: Two ways of feature extraction for a blurred image. Left: Convolution shares a static filter among pixels, conflicting with content-aware feature extraction. Right: The dynamic filter generates one filter for each pixel adapted to distorted image content.

In addition to dynamic feature extraction, dynamic quality regression can also benefit IQA models in handling wide content variation. The process of quality estimation can be viewed as a perceptual rule for humans to judge image quality, corresponding to the parameters of the quality regression module [28]. Using a fixed rule for predicting quality of varying images is insufficient to cover different structures. To generate rules dynamically for diverse contents and distortions, dynamic quality regression involves predicting network weights conditioned on image content. Our proposed parameter generation modules can learn adaptive perception rules as parameters based on recognized image content. These



(a) Clean-Blurred image pairs

(b) Clean-Noisy image pairs

Fig. 2: Examples of two pairs of images (the first column) and their corresponding paired feature maps (the last two columns) outputted by the conv_3 block of pre-trained ResNet-101. The blue and red digits denote the mean and standard variance values of feature maps respectively.

rules (i.e. parameters) are then applied in the quality regression network to produce the final quality score.

To improve content-awareness, we introduce dynamic filtering and dynamic regression techniques simultaneously into the CNN framework for NR-IQA. By extracting features using the filters conditioned on both local contents and global semantics, and making different rules of judging quality based on image content, the proposed network aims to provide predictions more consistent with human perceptual judgments.

Introducing the second-order statistics for improving distortion-sensitivity in feature pooling. Typical CNNs in computer vision often connect convolutional layers to fully-connected (FC) layers. The FC layers encode feature maps into a single quality-aware representation for regression. However, such architectures can only process images of a fixed size. Rescaling, a common solution in image recognition, is unsuitable for IQA as it alters visual quality. Therefore, global pooling, which aggregates score maps or feature maps of varying sizes into a single score or fixed-length vector, is critical for CNN-based IQA models. Some methods use score pooling strategies to combine predicted quality scores from local regions or patches [31], [32], [33]. However, without real supervision on local patch quality, the accuracy of these local scores is uncertain, limiting performance. Another approach is pooling quality-aware feature maps to generate a fixed-size image representation for regression. Various pooling mechanisms are employed, including GAP [24], [28] and other average-based pooling [34], [35], [36].

The GAP and other average-based pooling methods primar-

ily capture first-order statistics of feature distributions. While these summary statistics can indicate holistic changes from distortions, they are insensitive to distortions that minimally alter image features on average. These global-preserving distortions are common real-world scenarios. For example, noise degradation oscillates pixel values around a zero mean, thus approximately preserving average intensity. Similarly, image blurring tends to maintain average intensity levels due to the sum-to-one property of blur kernels. Since a blind IQA CNN aims to preserve and propagate distortion effects for accurate quality prediction, relying on GAP can reduce effectiveness in identifying degradation. As illustrated in Fig. 2, feature maps of noisy and blurred images exhibit minimal change in global averages compared to clean versions. Furthermore, for some local distortions, differences in feature maps may only occur in localized regions. Although varying within these small areas, such changes can be diminished by averaging operations, leading to similar mean values. This analysis motivates the exploration of alternative statistics beyond first-order means for global pooling in CNN-based BIQA, in order to better identify degradation.

In this work, we investigate the use of second-order statistics (*e.g.* variance) for global pooling mechanisms in CNN-based BIQA. Incorporating such second-order information can provide improved descriptions of distorted image feature distributions. Consider average-insensitive distortions like white noise and Gaussian blur - although typically preserving the mean, they alter variances, with noise increasing and blur decreasing variance due to their oscillating and smoothing properties, respectively. Furthermore, second-order statistics are often more sensitive to outliers. Quadratic calculations magnify large local changes during pooling. While unexplored for BIQA, second-order pooling has proven effective in recognition by capturing self and cross-channel feature map similarities and describing complex decision boundaries [37]. This success in recognition motivates the exploration of second-order statistics in BIQA for the complex nature of human visual processing.

To capture second-order statistics, we introduce global variance pooling (GVP) for feature aggregation. By combining GVP with first-order global average pooling, we propose an end-to-end framework incorporating dual-order global pooling for BIQA.

Based on the aforementioned two improvements, we have developed a dynamic CNN model with dual-order statistics for NR-IQA. Our contributions in this work are three-fold:

- We introduce simultaneously dynamic filtering and dynamic quality regression into the CNN framework for content-awareness. Dynamic filtering enables locally and globally content-aware feature extraction. Dynamic quality regression is combined with dynamic filtering to better model human perceptual quality judgments across diverse image contents and distortions. To the best of our knowledge, this is the first work to simultaneously apply dynamic feature extraction and dynamic quality estimation in CNN-based models for BIQA.
- We investigate the use of second-order statistics for global pooling in CNN-based BIQA, for distortion-sensitivity. The proposed dual-order pooling strategies, which combine GVP and GAP, enable discrimination between average-insensitive and average-sensitive distortions.
- We propose an end-to-end dynamic model with a dual-order global pooling mechanism for BIQA. Benefiting from its effectiveness, this model can be efficiently trained without backbone fine-tuning. Extensive experiments conducted on benchmark datasets, containing both synthetic and authentic distortions, demonstrate the effectiveness of our model.

II. Related Work

In this section, we provide a brief review of major works on BIQA and recent advances in pooling strategies and dynamic filtering.

A. NR-IQA Models

Traditional NR-IQA involves handcrafted feature extraction using natural scene statistics [19], [20] or local binary patterns [21], [22] and so on, followed by learning-based quality regression, *e.g.*, support vector regressor, random forest, and Gaussian process regression. Recently, many CNN-based NR-IQA methods have shown significant improvements. However, IQA remains a small sample problem due to the limited availability of subjective scores. This poses challenges in handling diverse distortions and variable image contents encountered in real-world scenarios.

To address these challenges, two research directions have been explored in recent years. The first direction focuses on directly addressing the insufficiency of labeled data for training NR-IQA CNNs, aiming to obtain better generalization. This can be achieved by four different stratigies: exploring patch-wise pseudo labels [31], [24], utilizing transfer learning from image classification data [24], [38], [28], self-supervised pre-training with large-scale human-unlabelled data [39], [40], [41] or utilizing multi-task framework with quality-related tasks and unlabelled distorted images, *e.g.* distortion type classification [42], [38], image quality ranking [43], [44], and quality map estimation [33]. However, these often lack explicit content-adaptive or distortion-aware feature extraction, which limits their ability to handle images with varying contents and distortion types.

The second research direction focuses on content/distortion-aware strategies (*e.g.*, content-adaptive regression and attention weighting) for improved generalization. In detail, Su *et al.* [28] proposed a content-aware regressor whose weights are predicted with image features. Gu *et al.* [32] introduced attention mechanisms to weight different image areas based on their predicted perceptual importance. Pan *et al.* [45] proposed two distortion feature extractors for learning synthetic and authentic distortions and a weight-adaptive fusion module to explore distortion-awareness. In [46], an intermediary enhancement-based bilateral network with an iterative training strategy has been proposed to address the challenges of distribution shift and long-tailed distribution in evaluating low-quality images for authentic distortions. To explore distortion-awareness and content-awareness simultaneously, Zhou *et al.* [47] separately

represents the content and distortion information using collaborative autoencoders and adaptively balances their contributions to image quality prediction. For further improvement, Chen *et al.* [48] used reinforcement learning to better capture attention information in the input image. Ke [49] introduced the transformer architecture to NR-IQA, where self-attention builds a similarity-based transform for image-adaptive feature extraction. These content-aware strategies generate content-adaptive weights on extracted features, incorporating the spatial information of the image content. However, the feature extraction process itself remains content-agnostic, which may limit the capability of the model to capture fine-grained content-adaptive information.

B. Pooling Strategies in NR-IQA

This section gives a brief review on global pooling techniques in deep learning-based no-reference methods. Global pooling can generate new lower-resolution feature maps to greatly reduce input spatial size. Pooling also provides quality-aware statistics for IQA, which not only decreases the number of parameters, reducing computational cost, but also helps control network overfitting. Moreover, pooling allows arbitrary input sizes, avoiding training noise and ensuring stability. Additionally, pooling strategies improve IQA metrics by pooling quality-related features and discarding irrelevant information.

Based on whether fully-connected layers serve as the score generator between the last convolutional layer and final quality output, pooling strategies can be divided into two types. The first type involves pooling before fully-connected layers, where arbitrary-sized feature maps are aggregated into fixed-length vectors as inputs for regression [35], [50]. Unlike outputting global features, the second type operates on local maps or scores, converting them into quality scores using weighted statistics [32], [48].

Earlier IQA networks such as NIMA [24] mainly focused on various backbone frameworks. These models primarily adopted global average pooling [43], [24], [50], [28] and max pooling [51], [42] techniques, calculating spatial means or maxima. Recently, newer methods have been proposed to capture better statistics and spatial relationships on convolutional outputs. One such method is spatial pyramid pooling [52], [53], which considers the discriminability of representations across different resolutions. By capturing features at multiple scales, spatial pyramid pooling improves the representation capability of the model. Another approach is attention-based pooling [35], where attention mechanisms are used to highlight degraded regions in the image. Additionally, region-of-interest pooling [34] has been introduced, which takes quality perception into account. These effective pooling strategies that align with human perception are crucial for NR-IQA [54].

C. Dynamic Filtering

Dynamic filtering has emerged as a powerful technique for building adaptive CNNs. It overcomes the limitations of standard convolutional layers and improve the adaptivity of CNNs for spatially-varying image structures or effects. There are two types of works, *i.e.*, applying convolutions in a spatially-shared way or in a spatially-varying way.

For spatially-shared dynamic filtering, techniques such as conditionally-parameterized convolutions [29], [55] and dynamic convolutions [30], [56] predict coefficients that combine multiple expert filters in a spatially-shared manner. By dynamically adjusting the coefficients, these methods enable the network to learn to adapt to different image structures effectively.

On the other hand, the spatially-varying dynamic filtering approach generates individual filters for each pixel location. For instance, the kernel prediction network [57] generates a unique filter for every pixel to capture spatially-varying image features. However, this approach can be computationally expensive. To address this, decoupled dynamic filtering [58] uses a composite strategy with decoupled spatial and channel dynamic filters, reducing the computational cost while maintaining the adaptive filtering capability. To further enhance the efficiency of dynamic filtering, Quan *et al.* [59] propose even kernel mixture learning for the deblurring problem by decomposing the predicted kernels into a group of spatially-shared bases and some spatially-varying coefficients. This decomposition significantly reduces the model size while still capturing the spatial variations of image structures.

While dynamic filtering has shown success in various tasks such as recognition, recovery, and generation [60], [61], [62], its exploration for IQA remains limited. The adaptability of filter parameters based on input content shows promise for modeling the complexity of human visual quality perception. Further research into dynamic filtering strategies could provide flexible frameworks to incorporate content-aware localized processing into CNN architectures for IQA.

III. Proposed Method

The NR-IQA model proposed in this study is called DDNet (Dynamic Network with Dual-order Statistics). It consists of four main components: 1) a static NN for initial feature extraction; 2) a dynamic filtering module for content-aware feature extraction; 3) a dual-order feature pooling mechanism for feature aggregation; 4) a dynamic regressor for predicting the quality score. DDNet can handle images of arbitrary sizes due to the employment of global pooling. To enhance its performance, DDNet is combined with a pre-trained backbone, similar to many existing works. The architecture of the proposed model is depicted in Fig. 3, and its specific details are elaborated below.

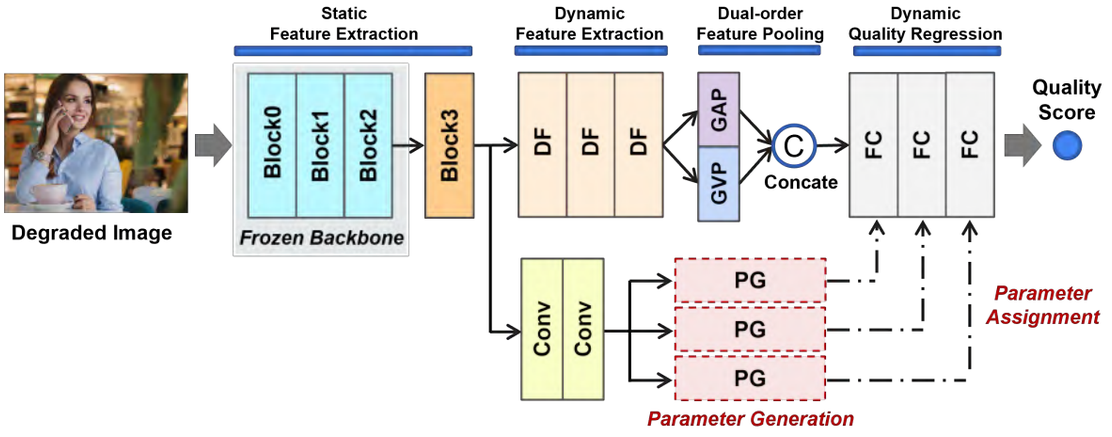
A. Pre-Stage Static Feature Extraction

Given the input image $\mathcal{I}$, the pre-stage feature extraction procedure can be formulated as:

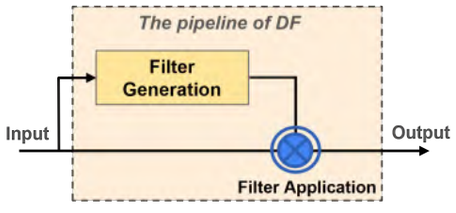
$$\boldsymbol{A} = f_0 \circ f_1 \circ \mathcal{I}, \tag{1}$$

where $\circ$ denotes function composition, f_0 denotes the frozen pre-trained backbone consisting of Block0-Block2, and f_1 denotes the trainable Block3. Specifically, we employ the first four layers, conv2_x and conv3_x, of a pretrained ResNet backbone as Block0-Block2 in $f_0(\cdot)$, respectively. Building

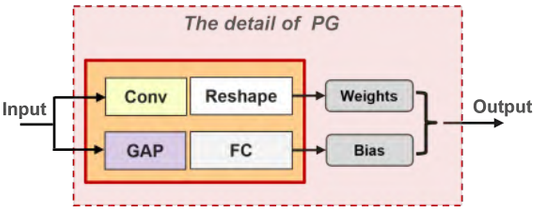
This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3359667



(a) Framework of DDNet



(b) Diagram of DF



(c) Diagram of PG

Fig. 3: Proposed DDNet for NR-IQA. Block_x (x=0,1,2,3) represents the first four layes, conv2_x, conv3_x and conv4_x in ResNet Backbone. Conv for Convolution, DF for Dynamic Filtering, GAP for Global Average Pooling, GVP for Global Variance Pooling, PG for Parameter Generation and FC for Fully-Connected.

block conv4_x is employed for Block3, i.e., $f_1(\cdot)$. Note that Block0-Block2 are frozen during training.

B. Dynamic Feature Extraction

To refine the pre-stage features in a content-aware manner, we stack three dynamic filtering modules to construct a dynamic feature extraction process, denoted as $f_2(\cdot)$. This module is illustrated in Figures 3(b) and 4.

For one dynamic filtering module and for each channel, we calculate

$$\boldsymbol{Y}_j(p) = \sum_{p' \in \Omega(p)} \boldsymbol{X}_j(p') \boldsymbol{D}_j(p', p). \tag{2}$$

where $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{h \times w \times c}$ denote the input and output tensors, $\Omega(\cdot)$ denotes convolution window around the p-th pixel, and $\boldsymbol{D}$ denotes the dynamic kernel tensor. Compared to the static convolution, the value $\boldsymbol{D}_j(p', p)$ in the dynamic kernel not only depends on the relative position $(p - p')$, but also involves the absolute position p, enabling spatially-varying processing. Furthermore, unlike the standard convolution where kernels are shared across different inputs, the kernels in Eq.(2) are predicted conditioned on the input.

For the j-th channel, the desired kernel tensor $\boldsymbol{D}$ in the size of $h \times w \times c \times k \times k$ is first calculated by

$$\mathcal{W} = g(\boldsymbol{X}), \tag{3}$$

where $g(\cdot)$ is a learnable function to predict dynamic filters. Each pixel-wise filter $\boldsymbol{D}_j(\cdot, p) \in \mathbb{R}^{k \times k}$ is then extracted from the kernel $\mathcal{W}_j(p) \in \mathbb{R}^{k \times k}$ located at p-th pixel and j-th channel of $\mathcal{W}$. The large output size of $g(\cdot)$ introduces many parameters. Even with a 1×1 convolution, k^2c^2 parameters are still needed. To reduce the number of parameters, following [58], $g(\cdot)$ is separated into a spatially-varying channel-shared component $\phi(\cdot)$, and a spatially-shared channel-varying component $\psi(\cdot)$:

$$\boldsymbol{D}_j(\cdot, p) = \mathcal{W}_j(p) = [\phi(\boldsymbol{X})](p) \odot [\psi(\boldsymbol{X})]_j = \boldsymbol{S}(p) \odot \boldsymbol{C}_j, \tag{4}$$

for each pixel index p and channel index j, where $\boldsymbol{S}(p) \in \mathbb{R}^{k \times k}$, $\boldsymbol{C}_j \in \mathbb{R}^{k \times k}$ are extracted and reshaped from $\boldsymbol{S} = \phi(\boldsymbol{X}) \in \mathbb{R}^{h \times w \times k^2}$, $\boldsymbol{C} = \psi(\boldsymbol{X}) \in \mathbb{R}^{c \times k \times k}$, and $\odot$ denotes the element-wise product, see Fig. 4 for intuitive understanding.

The spatial filter branch $\phi(\cdot)$ takes a local receptive field and predicts local-content-aware filters, while the channel filter branch $\psi(\cdot)$ takes the whole image as receptive field and predicts global-semantic-aware filters. The composition strategy, shown in Fig. 4, enables a locally and globally content-aware process with far fewer parameters. Our model uses a 1×1 convolution for $\phi(\cdot)$; GAP, two fully-connected layers and an in-between ReLU, for $\psi(\cdot)$, reducing parameters to $k^2c + 2c^2$.

C. Dual-order Global Pooling

The dual-order global pooling module consists of two submodules: first-order global pooling and second-order global pooling. The first-order global pooling module is designed to primarily process average-sensitive distortions, while

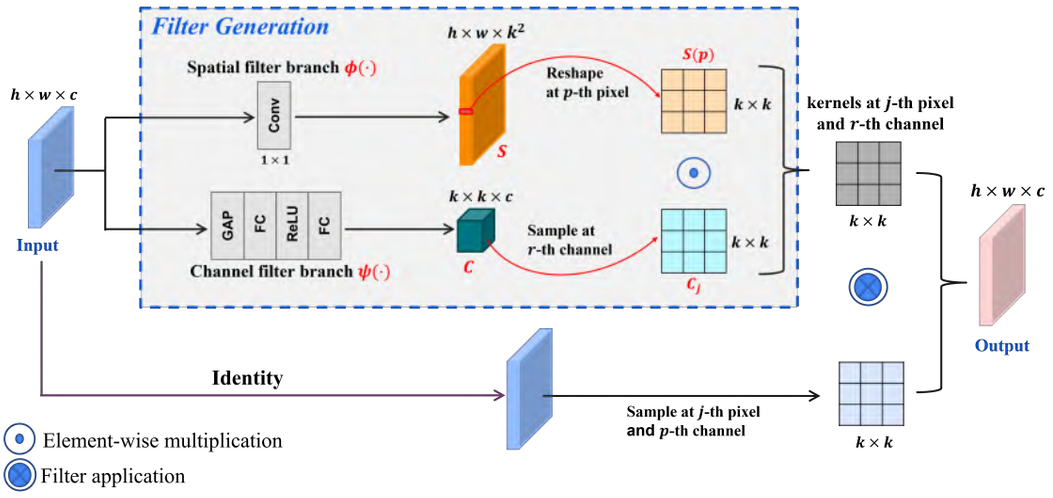


Fig. 4: DF module. The filter application means applying convolution at a single position.

the second-order pooling module is intended for average-insensitive distortions. The details of these two pooling operators are illustrated below.

First-order Global Pooling. Let $\mathcal{I} \in \mathbb{R}^{H_0 \times W_0 \times C_0}$ denote the input image for assessment, and $\mathcal{X} = f(\mathcal{I}) \in \mathbb{R}^{H \times W \times C_1}$ denote the extracted features. Given the extracted feature $\mathcal{X}$ with an arbitrary spatial size $H \times W$, the global pooling mechanism is designed to aggregate $\mathcal{X}$ into a quality-aware global feature $\boldsymbol{z}$, which is then fed into the regressor for quality prediction. In the pooling scheme, the feature tensor $\mathcal{X}$ is treated as a set of samples $\{\boldsymbol{x}_i\}_{i=1}^N$ with $N = HW$ and the sample $\boldsymbol{x}_i \in \mathbb{R}^{C_1}$ is the feature vector located at i-th position of $\mathcal{X}$.

First-order statistics are employed to characterize feature distributions. In BIQA, first-order statistics estimated by averaging are discriminative for average-sensitive distortions such as color and brightness changes. To measure first-order statistics, we introduce GAP which computes the average:

$$\mu(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x}_i. \tag{5}$$

Second-order Global Pooling. Although GAP handles average-sensitive distortions effectively, it often fails to capture feature changes caused by average-insensitive distortions like blur and zero-mean noise. Moreover, for spatial distortions such as color blocks and non-eccentricity patches, quality only drops locally. GAP tends to weaken or eliminate such local changes during global averaging. To address these issues, second-order statistics come into consideration. We propose GVP for feature aggregation, which captures second-order statistics of the feature distribution.

Treat the feature tensor $\mathcal{X}$ as a set of samples $\{\boldsymbol{x}_i\}_{i=1}^N$ with $N = HW$, GVP estimates the variance of the samples as

$$\Sigma(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^2, \tag{6}$$

where $\bar{\boldsymbol{x}}$ denotes the mean of samples, *i.e.* $\bar{\boldsymbol{x}} = \sum_{i=1}^N \boldsymbol{x}_i / N$.

Dual-order Pooling. After content-aware feature extraction, we get $\boldsymbol{B} = f_2(\boldsymbol{X}) = f_2 \circ f_1 \circ f_0 \circ \mathcal{I}$. Then the first-order pooling-GAP module $\mu(\cdot)$ and the second-order pooling-GVP module $\Sigma(\cdot)$ are employed to aggregate the feature maps $\boldsymbol{B}$ into global feature vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, respectively, so as to process average-sensitive and average-insensitive distortions. The output of the dual-order global pooling module concatenates $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ into $\boldsymbol{z} = [\boldsymbol{z}_1, \boldsymbol{z}_2] = [\mu(\boldsymbol{B}), \Sigma(\boldsymbol{B})]$, which is also the input to the FC layers for score prediction.

D. Dynamic Quality Regression

The dynamic quality regression (DQR) module maps learned features to a quality score using adaptive weights conditioned on image content. The module, denoted as $f_3(\cdot)$, consists of three fully connected layers. It takes the feature vector $\boldsymbol{z}$ as input and propagates through the parameter-determined layers to output the final quality score. During regression, the sigmoid activation function is adopted.

Inspired by [28], our parameter determined layers $\varphi(\cdot)$ consist of two 3×3 convolutional layers for dimension reduction and three parameter generations (PG) layers for weights and bias generation of FC layers. The input to $\varphi(\cdot)$ is $f_1 \circ f_0 \circ \mathcal{I}$, which are the semantic features extracted from the input image $\mathcal{I}$ by the static feature extraction. The three outputs of $\varphi(\cdot)$ are parameters $\theta_{\mathcal{I}}^1, \theta_{\mathcal{I}}^2, \theta_{\mathcal{I}}^3$ to be applied to three FC layers in $f_3(\cdot)$ respectively. For easier understanding, parameters $\theta_{\boldsymbol{x}}$ can be regarded as quality perceiving rules. The role of the parameter layers is to learn the mapping from image content to the rule of how to judge image quality.

Since FC layers are used for quality estimation in our network, the parameter layers $\varphi(\cdot)$ must generate two parameter types - FC weights and biases. Each PG layer generates a group of FC weights and biases corresponding to an FC layer in $f_3(\cdot)$. Different generation approaches are used for each type of parameter. FC weights for each FC layer are generated by operating one 3×3 convolution layer and reshaping the extracted features, while biases use an average pooling and a FC, as they have far fewer parameters, see Fig. 3(c) for

illustration. The output channels of the convolution and FC layers match the dimensions of the corresponding FC layers in $f_3(\cdot)$. The generated weights represent quality perception rules that instruct the dynamic regression module to predict quality in a content-aware manner. By constructing adaptive filters and quality indicators based on content, our model becomes more content-aware.

Given a set of images $\{\mathcal{I}_i\}_{i=1}^M$ and their subjective scores measured by human $\{s_i^*\}_{i=1}^M$. Let $\{s_i\}_{i=1}^M$ denote the scores predicted by the proposed model. We use the Huber loss for training due to its stronger robustness to outliers over the commonly-used mean square error loss, which is defined as

$$\ell = \sum_{i=1}^{M} \ell_\delta(s_i, s_i^*), \tag{7}$$

where ℓ_δ is the paramterized Huber loss defined by

$$\ell_\delta(s, s^*) = \begin{cases} \frac{1}{2}(s-s^*)^2, & \text{for } |s-s^*| \le \delta \\ \delta\left(|s-s^*| - \frac{1}{2}\delta\right), & \text{otherwise} \end{cases}. \tag{8}$$

And δ is a parameter to choose the way to penalty outliers. In implementation, we set $\delta = 1/9$ as suggested in [35].

IV. Experiments

A. Experimental Setups

1) Implementation details: For our network backbone, following [35], [63], [64], we choose ResNet-101 [65], as it has been found to be effective in feature extraction. Specifically, the first 3 blocks (conv1, conv2_x, conv3_x) are used as the frozen $f_0(\cdot)$ to extract low/mid-level semantics. Block 4 (conv4_x) is used for the trainable $f_1(\cdot)$. These blocks are initialized with ImageNet-trained ResNet-101 weights. For the dynamic filtering, the output kernel size is 3×3. All convolutions *i.e.*, yellow blocks in Fig. 3, use 3×3 kernels. The output sizes of three FC layers in $f_3(\cdot)$ are set to 192, 64 and 1 respectively. Refer to more details in the code repository DDNet.

For training, we utilize the AdamW optimizer with betas=(0.9, 0.999), weight decay=$1e{-}2$ for 100 epochs. The initial learning rate is 0.2 with decays on plateaus. The model and loss function are implemented in PyTorch on an NVIDIA RTX 3090 GPU.

2) Datasets: Six publicly available natural image quality databases are used for experimental evaluation: (i) Three artificially-distorted sets: CSIQ [66], TID2013 [67] and Kadid-10k [68]. (ii) Three realistically-distorted sets: LIVE-C [69], KonIQ-10k [70] and SPAQ [71].

- CSIQ: 30 pristine images and 866 distorted images with 6 distortion types at 4 to 5 distortion levels.
- TID2013: 25 pristine images and a total of 3,000 distorted images with 17 distortion types at 4 degradation levels.
- Kadid-10k: 81 pristine images each of which is degraded by 25 distortions in 5 levels, resulting in 10,125 distorted images.
- LIVE-C: 1,162 realistically natural pictures with resized resolution of 500×500 pixels.
- KonIQ-10k: 10,073 realistically and complexly distorted images with resolution of 1024×768 pixels.
- SPAQ: 11,125 images captured by 66 smartphones with diverse resolutions.

Images are used in their original resolutions as input to test generalization to different sizes. Following [24], [35], 80% of images are randomly sampled for training and the rest for testing. For synthetic datasets, the split ensures pristine image content does not intersect between sets. As ing [35], data augmentation of horizontal/vertical flips and $\pm 1^\circ$ rotation is randomly applied during training, which improves performance slightly in most cases. Extra borders from rotation are removed by cropping. For performance comparison, median values of evaluation metrics over 10 test set sessions are reported.

3) Evaluation metric and compared methods: Two commonly-used evaluation metrics, namely the Spearman Rank Order Correlation Coefficient (SROCC) and the Pearson Linear Correlation Coefficient (PLCC), are adopted for performance comparison. The SROCC measures prediction monotonicity, while the PLCC measures linear correlation. An effective IQA metric should yield high values for both PLCC and SROCC.

Two well-known traditional handcrafted NR-IQA models, namely NIQE [72] and ILNIQE [73], are selected for performance comparison. In addition, several recent CNN-based methods are selected: PQR [23], deepIQA [31], DBCNN [38], SGDNet [26], MetaIQA [25], CaHDC [36], HyperNet [28], SiamIQA [74], AIGQA [75], UNIQUE [44], OLNet [39], and SAWAR [47]. Their experimental results are quoted from the original papers whenever available, or otherwise obtained by running the codes provided by the authors.

B. Performance Evaluation

1) Evaluation on individual databases: We first conducted evaluations on individual databases containing both synthetic and authentic distortions. The results on synthetically distorted datasets (i.e. CSIQ, TID2013, and Kadid-10k) are reported in Table I. As shown, our DDNet performs best and outperforms all state-of-the-art methods except for SAWAR [47] by a wide

TABLE I: Performance comparison on synthetically-distorted databases. The best result for each metric is **boldfaced**.

Method	CSIQ		TID2013		Kadid-10k	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NIQE [72]	0.632	0.726	0.343	0.378	0.374	0.428
ILNIQE [73]	0.832	0.873	0.570	0.598	0.531	0.563
PQR [23]	0.872	0.901	0.745	0.798	-	-
deepIQA [31]	0.955	0.973	0.761	0.787	0.628	0.647
DBCNN [38]	0.946	0.959	0.816	0.865	-	-
SGDNet [26]	0.883	0.903	0.843	0.861	-	-
MetaIQA [25]	-	-	0.854	0.887	0.767	0.774
CaHDC [36]	0.874	0.915	0.862	0.878	-	-
HyperNet [28]	0.933	0.947	0.857	0.898	0.896	0.907
SiamIQA [74]	0.962	-	0.855	-	0.913	-
AIGQA [75]	0.927	0.952	0.871	0.893	0.864	0.863
UNIQUE [44]	0.902	0.927	0.855	0.879	0.876	0.878
OLNet [39]	0.966	0.975	0.863	0.889	-	-
SAWAR [47]	0.952	0.960	0.884	0.896	**0.928**	0.932
DDNet [Ours]	**0.971**	**0.979**	**0.956**	**0.961**	0.925	**0.934**

This article has been accepted for publication in IEEE Transactions on Circuits and Systems for Video Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2024.3359667

margin on all three databases in terms of both SROCC and PLCC metrics. The larger performance gains are observed on the more challenging TID2013 and Kadid-10k datasets, which feature diverse content types and complex distortion types. It is worth noting that while the compared methods performed well on CSIQ, they showed a larger decrease in performance on TID2013 and Kadid-10k. The enhanced performance of DDNet can be attributed to its ability to learn more accurate content-adaptive filters for extracting quality-related features and content-aware quality regression, as well as more accurate global statistics. These factors contribute to better generalization on complex distortions.

Results on authentically-distorted databases (*i.e.* LIVE-C, Koniq-10k and SPAQ) are reported in Table II. As the results indicate, DDNet significantly outperforms all methods on the large-scale KonIQ-10k, while achieving comparative performance to UNIQUE* on LIVE-C. Images in LIVE-C were resized to uniform square sizes, disrupting objects' native pixel distributions and probably causing generalization issues. Fortunately, content-aware models such as HyperNet [28] and ours still perform well on it. Note that UNIQUE* [44] is marked because it uses all datasets for training, while ours uses individual ones respectively. Even that, DDNet still competes against it. DDNet's strong performance likely stems from content-aware filtering and content-adaptive quality regression, as well as dual-order statistics of feature maps.

TABLE II: Performance comparison on authentically-distorted databases. The best result for each metric is **boldfaced**.

Method	LIVE-C		KonIQ-10k		SPAQ	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
NIQE [72]	0.464	0.515	0.601	0.597	0.703	0.712
ILNIQE [73]	0.469	0.536	0.552	0.573	0.714	0.721
PQR [23]	0.857	0.882	0.881	0.884	-	-
deepIQA [31]	0.671	0.686	0.797	0.805	-	-
DBCNN [38]	0.851	0.869	0.875	0.884	0.911	0.915
SGDNet [26]	0.851	0.872	0.897	0.917	-	-
MetaIQA [25]	0.802	0.835	0.851	0.887	0.875	0.877
CaHDC [36]	0.738	0.744	-	-	0.827	0.834
HyperNet [28]	**0.859**	0.882	0.906	0.917	0.916	0.918
SiamIQA [74]	0.851	-	0.894	-	-	-
AIGQA [75]	0.751	0.761	-	-	-	-
UNIQUE* [44]	0.854	**0.890**	0.896	0.901	-	-
OLNet [39]	0.849	0.858	0.877	0.882	-	-
SAWAR [47]	0.853	0.871	0.898	0.906	-	-
DDNet [Ours]	0.853	0.876	**0.916**	**0.926**	**0.917**	**0.920**

2) Cross-dataset evaluation: To further evaluate the generalization capability of DDNet, we conduct cross-dataset experiments using KonIQ-10k for training and LIVE-C/SPAQ for testing. Only methods with available results are compared (Table III). The results demonstrate that DDNet generalizes well in predicting quality for images with arbitrary resolutions, varying contents, and real-world complex distortions.

Following [76], a gMAD competition [77] is conducted on the SPAQ [71] database for direct visualization. gMAD efficiently aims to select image pairs with maximum quality difference by an attacking IQA model to challenge a defending model which partitions images to the same level of quality. Selected pairs are shown to observers to determine if the attacker or defender is more robust. In a gMAD competition, DDNet competes with the top CNN competitor in most experiments, especially for authentic data, HyperNet [71], in cross-dataset evaluation. Figures 5 show representative SPAQ pairs with models trained on KonIQ-10k. In Figs 5(a)-(b), the first row has slightly better perceptual quality on clear structures, indicating DDNet successfully attacked HyperNet. As defender, DDNet survived HyperNet's attacks at low and good quality levels (Figs 5(c)-(d)). Both models successfully recognized obvious low-quality images.

TABLE III: Performance comparison on models trained on KonIQ-10k and tested on LIVE-C/SPAQ without fine-tuning. The best result under each setting is **boldfaced.**

LIVE-C	PQR	DBCNN	HyperNet	UNIQUE	**DDNet**
SROCC	0.772	0.755	0.785	0.786	**0.794**
PLCC	0.817	-	0.818	-	**0.822**
SPAQ	DBCNN	CaHDC	HyperNet	**DDNet**	
SROCC	0.783	0.730	0.807	**0.809**	
PLCC	0.792	0.778	0.818	**0.819**	

3) Comparison on diverse distortion types: To evaluate performance on diverse distortion types, results of CNN models on TID2013 and CSIQ categories are collected in Tables IV and V. DDNet shows competitive performance on individual types compared to other methods. It achieved the best results for 19 out of 24 types on TID2013 and all types on CSIQ. Specifically, it is observed that DDNet achieves better evaluation accuracy compared to other CNN-based methods on common distortion categories such as AWGN and JPEG.

TABLE IV: SROCC results on diverse distortion types of TID2013. The best result for each distortion type is **boldfaced.**

Distortion	DBCNN	MetaIQA	HyperNet	AIGQA	**DDNet**
AGN	0.790	0.947	0.769	0.932	**0.954**
ANC	0.700	0.924	0.613	0.916	**0.925**
SCN	0.826	0.955	0.918	0.944	**0.963**
MN	0.646	0.728	0.448	0.662	**0.744**
HFN	0.879	0.952	0.839	0.953	**0.967**
IN	0.708	0.866	0.758	0.911	**0.916**
QN	0.825	0.745	0.828	0.908	**0.910**
GB	0.859	0.977	0.873	0.917	**0.978**
DEN	0.865	0.938	0.804	0.914	**0.953**
JPEG	0.894	0.934	0.860	0.945	**0.950**
JP2K	0.916	0.957	0.888	0.932	**0.942**
JGTE	0.772	0.931	0.723	0.858	**0.931**
J2TE	0.822	0.903	0.846	0.898	**0.916**
NEPN	0.270	0.729	0.369	0.130	**0.731**
Block	0.444	0.391	0.428	**0.723**	0.715
MS	-0.009	0.402	0.424	0.554	**0.625**
CTC	0.548	0.764	0.740	0.830	**0.849**
CCS	0.631	**0.829**	0.710	0.689	0.736
MGN	0.711	0.939	0.767	0.948	**0.958**
CN	0.752	**0.952**	0.786	0.886	0.947
LCNI	0.860	**0.978**	0.879	0.897	0.977
ICQD	0.833	0.859	0.785	0.908	**0.929**
CHA	0.732	**0.927**	0.739	0.889	0.907
SSR	0.902	0.974	0.910	0.908	**0.984**

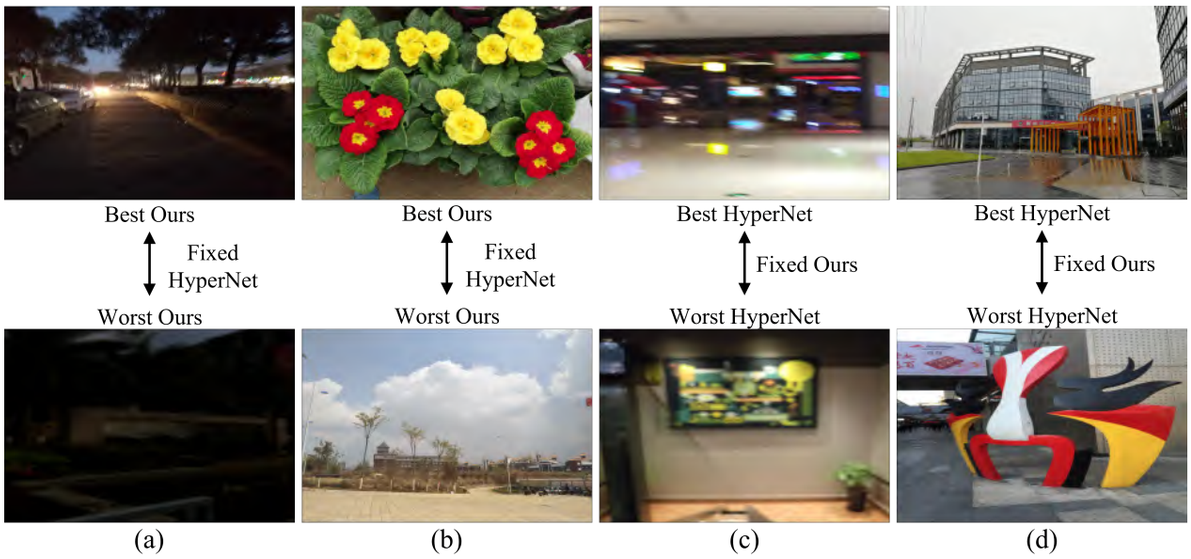


Fig. 5: Representative gMAD pairs between DDNet and HyperNet on the SPAQ database. (a) Fixing HyperNet at low quality level. (b) Fixing HyperNet at high quality level. (c) Fixing DDNet at low quality level. (d) Fixing DDNet at high quality level.

TABLE V: SROCC results on diverse distortion types of CSIQ. **Bold** on digits denote the best result for each distortion type.

Distortion	PQR	DBCNN	HyperNet	OLNet	**DDNet**
GB	0.921	0.947	0.915	0.965	**0.967**
AWGN	0.915	0.948	0.927	0.945	**0.965**
JPEG	0.934	0.940	0.934	0.968	**0.974**
JP2K	0.955	0.953	0.960	0.945	**0.977**
APN	0.926	0.941	0.931	0.953	**0.974**
CTD	0.837	0.872	0.874	0.925	**0.940**

C. Ablation Studies

To verify the effectiveness of the dynamic mechanism for featuere extraction and score regression, and deual-order statistics of DDNet, we conduct ablation studies on the TID2013, LIVE-C and Koniq-10k databases in terms of SROCCs and PLCCs under several settings, using the following DDNet variants. (i) Baseline: using only Block0-Block3 for static feature extraction, GAP for feature aggregation and three FC layers for static score regression. (ii) A: Removing the dynamic filtering module (*i.e.*, 3 DF layers) and parameter determined layers (2 convolutional layers and three PG layers) for dynamic quality regression in DDNet. (iii) B: Removing the dynamic quality regression in DDNet. (iv) C: Removing the parameter determined layers in DDNet. (v) D: Removing global variance pooling in DDNet. (vi) E: Removing global average pooling in DDNet. (vii) DDNet: Full proposed model with all modules.

Ablation results are in Table VI. Three FC layers remained for static score regression without DQR. As can be seen,each component noticeably contributes: (i) Model A and Model D improve over Baseline, demonstrating effectiveness of dual-order pooling and dynamic mechanisms for content-aware perception. (ii) Model B and Model C improve over Model A, showing dynamic feature extraction and rule learning benefits. (iii) Model D and Model E decrease compared to DDNet, indicating dual-order pooling improves dynamics. (iv) DDNet outperforms all variants on all databases, further demonstrating overall effectiveness.

In summary, the ablation studies validate the contributions of the proposed dynamic mechanisms, dual-order pooling, and their synergistic integration in DDNet.

D. More Analysis

1) Model complexity: Model complexity is evaluated by floating-point operations per second (FLOPS) on 1024×768 color images. Recent methods SGDNet, CaHDC, and HyperNet are compared using released Python codes. See Table VII for the results. DDNet has comparable FLOPS to HyperNet, one of the top competitors in most experiments experimentally.

2) Performance with different backbones: The backbone model can influence DDNet performance. Table VIII shows results applying orresponding layers of ResNet152, ResNet101, ResNet50, and VGG16 backbones to $f_0(\cdot)$ and $f_1(\cdot)$. Compared to previous experiments in Tables I and II, DDNet with different ResNet backbones maintains state-of-the-art performance. Larger backbones like ResNet152 can provide further gains in some cases.

To further compress DDNet for efficiency, especially for real-time applications on mobile devices. We also evaluated DDNet performance with famous lightweight backbones including MobileNetv3_small and ShuffleNetv2 without fine-tunning. The results are shown in Table VIII, indicating that using very lightweight backbones of smaller model complexity in quality assessment models tends to lead to decreased performance, primarily due to under-fitting with only a few learnable parameters. This trade-off can be considered in real-world applications on mobile terminals.

TABLE VI: Median SROCC and PLCC results of ablation study on the test sets of three IQA databases. **Bold** on digits denote the best result for each criteria.

Model	Module				TID2013		LIVE-C		KonIQ-10k	
	DF	GAP	GVP	DQR	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
Baseline	✘	✔	✘	✘	0.898	0.892	0.812	0.803	0.894	0.872
A	✘	✔	✔	✘	0.935	0.929	0.836	0.810	0.908	0.897
B	✔	✔	✔	✘	0.953	0.951	0.846	0.831	0.915	0.911
C	✘	✔	✔	✔	0.946	0.943	0.843	0.824	0.911	0.903
D	✔	✔	✘	✔	0.949	0.947	0.856	0.844	0.912	0.910
E	✔	✘	✔	✔	0.951	0.949	0.849	0.842	0.911	0.907
DDNet	✔	✔	✔	✔	**0.961**	**0.956**	**0.876**	**0.853**	**0.926**	**0.916**

TABLE VII: Comparison of model complexity by FLOPS

SGDNet	CaHDC	HyperNet	**DDNet**
1.23×10^{11}	0.37×10^{11}	1.98×10^{11}	1.71×10^{11}

TABLE VIII: Performance on different backbones

Backbones	TID2013		KonIQ-10k	
	SROCC	PLCC	SROCC	PLCC
ResNet152	0.957	0.958	0.910	0.917
ResNet101	0.956	0.961	0.916	0.926
ResNet50	0.953	0.957	0.915	0.926
VGG16	0.941	0.943	0.686	0.745
MobileNetv3_small	0.783	0.819	0.733	0.744
ShuffleNetv2	0.792	0.807	0.750	0.750

V. Conclusion

This paper improves the effectiveness of CNNs in NR-IQA by incorporating dynamic feature extraction, dynamic quality regression, and dual-order global statistics. The resulting dynamic model and modules increase content-awareness by employing dynamic filtering and dynamic regressed perception rule learning to emulate human perception. The introduced GVP enables the characterization of second-order feature statistics and produces a quality-aware global feature encoding for distortion-sensitivity. By combining GVP with GAP, the dual-order pooling generates more distortion-sensitive and discriminative representations. Experiments conducted on six public datasets demonstrate the superior performance of our approach compared to state-of-the-art methods for both synthetic and authentic distortions. Future work involves extending the approach to video quality assessment tasks.

References

[1] X. Liu, D. Zhai, J. Zhou, S. Wang, D. Zhao, and H. Gao, "Sparsity-based image error concealment via adaptive dual dictionary learning and regularization," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 782–796, 2017.

[2] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE Transactions on Neural Network and Learning Systems*, vol. 29, no. 4, pp. 1301–1313, 2018.

[3] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Ssim-motivated rate-distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 4, pp. 516–529, 2012.

[4] A. Mishra, A. Jain, M. Narwaria, and C. Agarwal, "An experimental study into objective quality assessment of watermarked images," *International Journal of Image Processing*, vol. 5, no. 2, p. 199, 2011.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions Image Processing*, vol. 13, no. 4, pp. 600–12, 2004.

[6] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

[7] H. W. Chang, H. Yang, Y. Gan, and M. H. Wang, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Transactions Image Processing*, vol. 22, no. 10, pp. 4007–18, 2013.

[8] Z. Zhou, J. Li, Y. Xu, and Y. Quan, "Full-reference image quality metric for blurry images and compressed images using hybrid dictionary learning," *Neural Computing and Applications*, vol. 32, pp. 12 403–12 415, 2020.

[9] Z. Zhou, J. Li, Y. Quan, and R. Xu, "Image quality assessment using kernel sparse coding," *IEEE Transactions on Multimedia*, vol. 23, pp. 1592–1604, 2021.

[10] L. Ma, S. Li, and K. N. Ngan, "Reduced-reference video quality assessment of compressed video sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 10, pp. 1441–1456, 2012.

[11] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Sciences*, vol. 351, pp. 18–29, 2016.

[12] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 379–391, 2018.

[13] Z. Wang, A. C. Bovik, and B. L. Evan, "Blind measurement of blocking artifacts in images," in *Proccedings of International Conference on Image Processing*, 2000, pp. 981–984.

[14] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000." *IEEE Transactions Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.

[15] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.

[16] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 12, 2013.

[17] F. A. G. Peña, P. D. M. Fernández, T. I. Ren, J. d. J. G. Leandro, and R. M. Nishihara, "Burst ranking for blind multi-image deblurring," *IEEE Transactions on Image Processing*, vol. 29, pp. 947–958, 2019.

[18] P. Kancharla and S. S. Channappayya, "Completely blind quality assessment of user generated video content," *IEEE Transactions on Image Processing*, vol. 31, pp. 263–274, 2021.

[19] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[20] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *Journal of Vision*, vol. 17, no. 1, pp. 32–38, 2017.

[21] M. Zhang, C. Muramatsu, X. Zhou, T. Hara, and H. Fujita, "Blind image quality assessment using the joint statistics of generalized local binary pattern," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 207–210, 2014.

[22] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541–545, 2016.

[23] H. Zeng, L. Zhang, and A. C. Bovik, "A probabilistic quality representation approach to deep blind image quality prediction," *CoRR*, vol. abs/1708.08190, 2017.

[24] H. Talebi and P. Milanfar, "Nima: Neural image assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3998–4011, 2018.

[25] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "Metaiqa: Deep meta-learning for no-reference image quality assessment," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 143–14 152.

[26] S. Yang, Q. Jiang, W. Lin, and Y. Wang, "Sgdnet: An end-to-end saliency-guided deep neural network for no-reference image quality assessment," in *Proceedings of ACM International Conference on Multimedia*, 2019, pp. 1383–1391.

[27] D. Li, T. Jiang, W. Lin, and M. Jiang, "Which has better visual quality: The clear blue sky or a blurry animal?" *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1221–1234, 2019.

[28] S. Su, Q. Yan, Y. Zhu, C. Zhang, X. Ge, J. Sun, and Y. Zhang, "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3664–3673.

[29] B. Yang, G. Bender, Q. V. Le, and J. Ngiam, "Condconv: Conditionally parameterized convolutions for efficient inference," in *Proceedings of Advances in Neural Information Processing Systems*, vol. 32, 2019.

[30] Y. Zhang, J. Zhang, Q. Wang, and Z. Zhong, "Dynet: Dynamic convolution for accelerating convolutional neural networks," *arXiv preprint arXiv:2004.10694*, 2020.

[31] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, 2018.

[32] J. Gu, G. Meng, S. Xiang, and C. Pan, "Blind image quality assessment via learnable attention-based pooling," *Pattern Recognition*, vol. 91, pp. 332–344, 2019.

[33] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu, and Y. Zhang, "Blind predicting similar quality map for image quality assessment," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6373–6382.

[34] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, and A. Bovik, "From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3575–3585.

[35] Y. Su and J. Korhonen, "Blind natural image quality prediction using convolutional neural networks and weighted spatial pooling," in *Proceedings of IEEE International Conference on Image Processing*, 2020, pp. 191–195.

[36] J. Wu, J. Ma, F. Liang, W. Dong, G. Shi, and W. Lin, "End-to-end blind image quality prediction with cascaded deep neural network," *IEEE Transactions on Image Processing*, vol. 29, pp. 7414–7426, 2020.

[37] P. Li, J. Xie, Q. Wang, and W. Zuo, "Is second-order information helpful for large-scale visual recognition?" in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2070–2078.

[38] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits System and Video Technology*, vol. 30, no. 1, pp. 36–47, 2018.

[39] X. Yao, Q. Cao, X. Feng, G. Cheng, and J. Han, "Learning to assess image quality like an observer," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2022.

[40] P. C. Madhusudana, N. Birkbeck, Y. Wang, B. Adsumilli, and A. C. Bovik, "Image quality assessment using contrastive learning," *IEEE Transactions on Image Processing*, vol. 31, pp. 4149–4161, 2022.

[41] N. C. Babu, V. Kannan, and R. Soundararajan, "No reference opinion unaware quality assessment of authentically distorted images," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, January 2023, pp. 2459–2468.

[42] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, and W. Zuo, "End-to-end blind image quality assessment using deep neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1202–1213, 2017.

[43] X. Liu, J. Van De Weijer, and A. D. Bagdanov, "Rankiqa: Learning from rankings for no-reference image quality assessment," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1040–1049.

[44] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Transactions on Image Processing*, vol. 30, pp. 3474–3486, 2021.

[45] Z. Pan, H. Zhang, J. Lei, Y. Fang, X. Shao, N. Ling, and S. Kwong, "Dacnn: Blind image quality assessment via a distortion-aware convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7518–7531, 2022.

[46] T. Song, L. Li, P. Chen, H. Liu, and J. Qian, "Blind image quality assessment for authentic distortions by intermediary enhancement and iterative training," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7592–7604, 2022.

[47] Z. Zhou, F. Zhou, and G. Qiu, "Blind image quality assessment based on separate representations and adaptive interaction of content and distortion," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[48] D. Chen, Y. Wang, and W. Gao, "No-reference image quality assessment: An attention driven approach," *IEEE Transactions on Image Processing*, vol. 29, pp. 6496–6506, 2020.

[49] J. Ke, Q. Wang, Y. Wang, P. Milanfar, and F. Yang, "Musiq: Multi-scale image quality transformer," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 5148–5157.

[50] W. Zhang, K. Ma, J. Yan, D. Deng, and Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 36–47, Jan. 2020.

[51] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 206–220, 2016.

[52] D. Varga, D. Saupe, and T. Szirányi, "Deeprn: A content preserving deep architecture for blind image quality assessment," in *Proceedings of IEEE Conference on Multimedia Expo*, 2018, pp. 1–6.

[53] K. Ma, X. Liu, Y. Fang, and E. P. Simoncelli, "Blind image quality assessment by learning from multiple annotators," in *Proceedings of International Conference on Image Processing*, 2019, pp. 2344–2348.

[54] J. Gui, X. Cong, Y. Cao, W. Ren, J. Zhang, J. Zhang, J. Cao, and D. Tao, "A comprehensive survey and taxonomy on single image dehazing based on deep learning," *ACM Computing Surveys*, vol. 55, no. 13s, pp. 1–37, 2023.

[55] Y. Pu, Y. Wang, Z. Xia, Y. Han, Y. Wang, W. Gan, Z. Wang, S. Song, and G. Huang, "Adaptive rotated convolution for rotated object detection," *arXiv preprint arXiv:2303.07820*, 2023.

[56] T. Verelst and T. Tuytelaars, "Dynamic convolutions: Exploiting spatial sparsity for faster inference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2020.

[57] B. Mildenhall, J. T. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll, "Burst denoising with kernel prediction networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2502–2510.

[58] J. Zhou, V. Jampani, Z. Pi, Q. Liu, and M.-H. Yang, "Decoupled dynamic filter networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6647–6656.

[59] Y. Quan, Z. Wu, and H. Ji, "Gaussian kernel mixture network for single image defocus deblurring," *Advances in Neural Information Processing Systems*, vol. 34, pp. 20 812–20 824, 2021.

[60] Y. Jo, S. W. Oh, J. Kang, and S. J. Kim, "Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3224–3232.

[61] J. He, Z. Deng, and Y. Qiao, "Dynamic multi-scale filters for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3562–3572.

[62] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 030–11 039.

[63] A. Chetouani, "Image quality assessment without reference by mixing deep learning-based features," in *Proceedings of IEEE Conference on Multimedia Expo*, Jul. 2020, pp. 1–6.

[64] D. Li, T. Jiang, and M. Jiang, "Norm-in-norm loss with faster convergence and better performance for image quality assessment," in *Proceedings of ACM International Conference on Multimedia*, 2020, pp. 789–797.

[65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[66] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, p. 011006, 2010.

[67] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Image database

TID2013: Peculiarities, results and perspectives," *Signal Process: Image Communication*, vol. 30, pp. 57–77, 2015.

[68] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqa database," in *Proceedings of IEEE Conference on Quality Multimedia Experience*, 2019, pp. 1–3.

[69] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2015.

[70] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.

[71] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, "Perceptual quality assessment of smartphone photography," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3677–3686.

[72] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[73] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, 2015.

[74] W. Zhang, K. Zhai, G. Zhai, and X. Yang, "Learning to blindly assess image quality in the laboratory and wild," in *Proceedings of IEEE International Conference on Image Processing*, 2020, pp. 111–115.

[75] J. Ma, J. Wu, L. Li, W. Dong, X. Xie, G. Shi, and W. Lin, "Blind image quality assessment with active inference," *IEEE Transactions on Image Processing*, vol. 30, pp. 3650–3663, 2021.

[76] Z. Wang, Z.-R. Tang, J. Zhang, and Y. Fang, "Learning from synthetic data for opinion-free blind image quality assessment in the wild," *arXiv preprint arXiv:2106.14076*, 2021.

[77] K. Ma, Q. Wu, Z. Wang, Z. Duanmu, H. Yong, H. Li, and L. Zhang, "Group mad competition-a new methodology to compare objective image quality models," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1664–1673.

No-Reference Image Quality Assessment Using Local Binary Patterns: A Comprehensive Performance Evaluation

Zihan Zhou
South China Agricultural University
Guangzhou, China
zhouzihan@scau.edu.cn

Yong Xu*
South China University of Technology
Guangzhou, China
yxu@scut.edu.cn

Xi Wan
South China University of Technology
Guangzhou, China
csxwan@mail.scut.edu.cn

Yuhui Quan†‡
South China University of Technology
Guangzhou, China
csyhquan@scut.edu.cn

Ruotao Xu
Institute for Super Robotics
Guangzhou, China
rtxu@superobots.com

Jing Li
Moku Lab, Alibaba Group
Beijing, China
lj225205@alibaba-inc.com

Patrick Le Callet
Nantes Universite
Nantes, France
patrick.lecallet@univ-nantes.fr

Abstract

One key in image quality assessment (IQA) is the design of image representations that can capture the changes in image structures caused by distortions. In the last decades, local binary patterns (LBP) have been proven to be a powerful tool as a statistical model for texture and local structure representation. LBP and its variants, as the texture descriptor with low computational complexity, have been applied widely and successfully in several specific applications, such as texture classification, face recognition and IQA. Generally, visible impairments alter the statistics of LBP descriptors, making it possible to measure degradation and then estimate image quality. However, only a few variants of LBP are applied in no-reference (NR) IQA and a few characteristics of LBP descriptors are explored for quality prediction, while some characteristics useful in the rest variants are ignored for IQA. Two previous studies give a review of 16/4 LBP operators and compare the performance with them separately in IQA application only on synthetic distorted data. To extend this work, we provide a review of LBP methodologies to assist the scientific community and new researchers, as well as explore more LBP descriptors in NR-IQA methods under various distortion conditions, particularly real-world cases. Specifically, we comprehensively review 30 widely-used or effective LBP-based operators, including recent variations. We then utilize a common framework for applying LBP descriptors in NR-IQA. The practicality of the reviewed descriptors is demonstrated and analyzed using experimental results under synthetic and authentic cases, indicating suitable LBP descriptors and characteristics for NR-IQA. Codes implementing the reviewed LBP operators is available at https://github.com/csxwan/Local-Binary-Patterns-for-IQA.

CCS Concepts

• **Computing methodologies** → **Image representations**.

Keywords

No-Reference Image Quality Assessment, Local Binary Patterns, Variants, Texture Descriptors

ACM Reference Format:
Zihan Zhou, Yong Xu, Xi Wan, Yuhui Quan, Ruotao Xu, Jing Li, and Patrick Le Callet. 2024. No-Reference Image Quality Assessment Using Local Binary Patterns: A Comprehensive Performance Evaluation. In *Proceedings of the 3rd Workshop on Quality of Experience in Visual Multimedia Applications (QoEVMA '24), October 28-November 1 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3689093.3689180

1 Introduction

Digital images are ubiquitous in modern life and work for communication, entertainment and data analysis. However, the quality of images can be compromised by various distortions that occur during acquisition, processing, transmission, and display. These introduced distortions negatively impact image quality, which further limits the performance of applications relying on distorted images, such as object detection [28, 38, 39, 70], face recognition [7, 10, 36, 46] and video stream recognition systems. Consequently, there is growing interest in developing objective image quality assessment (IQA) algorithms to automatically evaluate perceptual quality. IQA has numerous potential applications, including instant feedback generation for image collection systems, automatic optimization of camera settings or post-processing parameters, the guidance of designing image restoration models [22, 42].

*Yong Xu is also with PaZhou Laboratory of Guangzhou, and Guangdong Provincial Key Laboratory of Multimodal Big Data Intelligent Analysis.
†Yuhui Quan is also with Pazhou Laboratory of Guangzhou, China.
‡Corresponding Author: Yuhui Quan.

QoEVMA '24, October 28-November 1 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-1204-3/24/10
https://doi.org/10.1145/3689093.3689180

There are three categories in objective IQA: full-reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA. The FR approaches (*e.g.* [6, 73, 80, 84]) are designed for the scenarios where the reference image of very high quality is given. Such scenarios can be found in image compression, image watermarking, as well as the training stages of learning-based image processing methods. The RR approaches (*e.g.* [43, 44, 74]) assume that only partial information (in the form of features) of the reference image can be accessed. RR-IQA is particularly suitable for image transmission, where receivers lack the reference images. Instead, reference features are sent and compared to transmitted image features (see [44] for an example). In contrast, NR-IQA (*i.e.*, Blind) is more challenging since blind approaches (e.g. [20, 62, 72, 85–87]) estimate quality without reference versions. Due to their relaxed requirements, blind IQA methods are more flexible for real-world use than FR and RR, better handling cases like image recovery and identifying low-quality images, especially for wild user-generated content.

The key in NR-IQA is the design of image representations that can capture the changes of image structures caused by distortions-that is, being discriminative to diverse distortions and various image contents. Several non-deep-learning techniques have been applied successfully to the feature extraction of degraded images, such as natural scene statistics [50, 54, 61, 71], independent component analysis [32, 78], sparse coding [76, 84] and local binary pattern description [4, 25, 58]. Among these, Local Binary Pattern (LBP) has emerged as a powerful statistical model for texture/structure representation [52]. With only fast binary operations, LBP has flexible real-world NR-IQA applicability on low- power-consumtion devices and in real-time scenarios. Furthermore, distortions often damage image textures and statistics, enabling LBP descriptors to quantify quality and proving useful for IQA.

The original local binary pattern (LBP) is a non-parametric descriptor that efficiently summarizes local image structures by comparing each pixel value to its surrounding neighbor pixel values. LBP converts an image into an array or map, where each pixel is labeled with a decimal number representing the LBP value. LBP values are derived from binary codes of pixel comparisons. The LBP maps are further converted into histograms used for specific applications. The LBP family has been developed until now. Many LBP variants have been developed, overcoming limitations and improving performance for specific applications like statistic/dynamic texture classification [45, 55, 65], face recognition [2, 35, 37], object detection [33, 77], image quality assessment [14, 83], etc. Table 1 summarizes the applications of LBP operators reviewed in this paper. For NR-IQA, although some variants help characterize structural degradation, many operators and characteristics are unexplored, especially for real-world images with a huge diversity of distortion and a large variation of image contents.

Although successful in various computer vision applications, basic LBP operators have some limitations [31]. For example, using the central pixel as a threshold makes them sensitive to noise. To address the limitations of conventional LBP, several variants have been proposed in the literature. Some variants are designed to increase description capability, *e.g.* Complete LBP [23], Averaged LBP [26]. Some aim to increase robustness to noise, *e.g.* Local Ternary Pattern (LTP) [69], Adjacent evaluation LBP [64], Improved Local Quinary Patterns [3]. Some aim to robust to blur, *e.g.*, Local Phase Quantization [53]. Some determine other types of neighboring pixels to utilize more information, *e.g.* Pyramid LBP [57], and Multi-scale LBP [11]. Some calculate LBP codes in other domain with specific transforms such as wavelet transform, Gabor transform, see [60, 81] for example.

Obviously, it would be impossible to consider all (*i.e.*, more than 100) variants of LBP have been proposed in the literature. Therefore, we selected a representative set of popular, widely-used, and recent LBP variants from typical applications like static/dynamic texture analysis, face recognition, and image quality assessment. Since our focus is on image quality assessment rather than video, we limited the scope to 2D image variants. While not an exhaustive list, the chosen variants provide a broad overview of key LBP developments. Table 2 briefly introduces the basic LBPs and reviewed variants, including common abbreviations, with details provided in the following sections and the released docs.

Table 1: The applications of the reviewed LBP operators.

Application	Method
Texture Classification	LBP[52], Rotation Invariant LBP[52], Uniform LBP[52], Averaged LBP[26], Median LBP[24], Complete LBP[23], Dominant LBP[40], Multi-Scale LBP[51], Pyramid LBP[57], Opponent Color LBP[48], Adjacent Evaluation LBP[64], Path Integration-based LBP[41], Decorrelated LBP[29], Adjacent Evaluation LTP[64], Local Phase Quantization[53], Local Binary Count[82], Completed Local Binary Count[82], Completed Local Derivative Patterns[30], Improved Local Quinary Patterns[3]
Face Recognition	LTP[69], Local Gabor Binary Patterns[81], Local Phase Quantization[1], Local Gradient Patterns[34], Eight Local Directional Patterns[9]
Image Quality Assessment	LBP[49], LTP[12], Multi-Scale LBP[11], Wavelet Domain LBP[60], Salient LBP[15], Multi-Scale Salient LBP[16], Local Phase Quantization[18], Local Variance Patterns[14], Orthogonal Color Planes Patterns[13], Opponent Color[19], Complete LBP[18]
Image Retrieval	Multi-Scale LBP[66]
Writer Identification	Wavelet Domain LBP[60]
Fingerprint Liveness	Weber Local Binary Descriptor[75]

Table 2: Brief descriptions of LBP variants

LBP Variants	Abbreviation	Description	Ref.
Local Binary Patterns	LBP	Encode binary codes via intensity comparison in a local neighborhood	[52]
Rotation Invariant LBP	RILBP	Clockwise shift the LBP code bit-wisely to get the smallest decimal value	[52]
Uniform LBP	ULBP	Conunt the number of bitwise 0/1 changes of LBP codes	[52]
Rotated Uniform LBP	RIULBP	Combine rotation-invariant LBP and uniform LBP	[52]
Averaged LBP	ALBP	Use the mean of local neighbourhood as the threshold	[26]
Median LBP	MLBP	Use the median value over a local neighborhood as the threshold	[24]
Complete LBP	CLBP	Consider the sign and magnitude of the local differences, and the global intensity	[23]
Dominant LBP	DLBP	Make use of the most frequently occurred patterns of LBP	[40]
Multi-Scale LBP	MSLBP	Encode LBP codes across multiple circular regions at different scales with different sampling points	[11]
Pyramid LBP	PLBP	Cascade the LBP information of the hierarchical spatial pyramid	[57]
Opponent Color LBP	OCLBP	Calulate intra-channel and inter-channel LBP codes for color images	[48]
Wavelet Domain LBP	WDLBP	Calculate LBP maps on wavelet coefficients of different sub-bands	[60]
Local Gabor Binary Patterns	LGBP	Calculate LBP maps on multi-scale and multi-orientation Gabor magnitude maps	[81]
Adjacent Evaluation LBP	AELBP	Set the averages in adjacent evaluation windows as the neighbors of the neighborhood center for binary code encoding	[64]
Path Integration-based LBP	pi-LBP	Filter and encode pixels to binary values at different scales along a particular path	[41]
Decorrelated LBP	dLBP	Operate discrete cosine transform on local differences of different scales	[29]
Salient LBP	SLBP	Weight the LBP map with the saliency map	[15]
Multi-Scale Salient LBP	MS-SLBP	Utilize the saliency map as weights for the LBP maps at multiple scales	[16]
Local Ternary Patterns	LTP	Endode three-digit codes by intensity comparison with a user-specific threshold	[69]
Adjacent Evaluation LTP	AELTP	Consider the means in adjacent evaluation windows as the neighbors of the neighborhood center for three-value code encoding	[64]
Local Phase Quantization	LPQ	Quantize phase of the discrete Fourier transform on images computed in local neighborhoods	[53]
Local Gradient Patterns	LGP	Encode binary codes by comparing absolute values of the intensity difference with their average	[34]
Local Binary Count	LBC	Count the number of value 1's in the binary neighborhood	[82]
Completed Local Binary Count	CLBC	Combine the LBC with the measures of local intensity difference and central gray level	[82]
Eight Local Directional Patterns	ELDP	Encode the local neighborhood filtered by eight Kirsch compass masks	[9]
Local Variance Patterns	LVP	Compute the spread of the texture local energy based on the square of intensities	[14]
Completed Local Derivative Patterns	CLDP	Complement the CLBP operator with the local directional derivative information	[30]
Orthogonal Color Planes Patterns	OCPP	Consider three two-dimensional planes decomposed from the tri-dimensional (XYZ) space for color images	[13]
Improved Local Quinary Patterns	ILQP	Endode five-digit codes by intensity comparison with two definative thresholds	[3]
Weber Local Binary Descriptor	WLBD	Combine the LBP under Weber's law and compute gradients from center-symmetric pixel pairs	[75]

This work expands upon previous studies by Freitas et al. [17, 18], which compared 16 and 4 LBP-based operators separately for IQA using synthetically distorted data. We extend them by going beyond synthetic distortions to focus more on real-world degradation scenarios. Additionally, we review more LBP variants used in broader computer vision tasks. This paper aims to familiarize the scientific community and support new researchers with past and current local binary pattern methodologies used in different fields, as well as explore more LBP descriptors for NR-IQA. In this paper, we first give a review of four basic LBP operators and 26 variants. We then

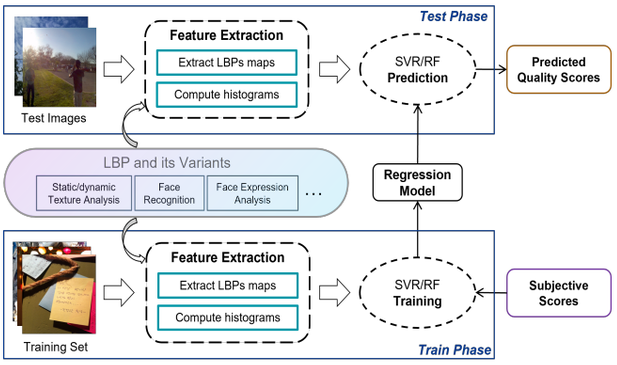


Figure 1: The framework of NR-IQA used in this paper. Two phases are included: 1) training the quality estimation model and 2) predicting quality scores.

apply these LBP variants to an NR-IQA framework, analyzing their pros and cons. Due to page limitations, we focus on key ideas of LBPs, with references for further details.

To apply LBP and its variants to NR-IQA, we utilize a two-step process: 1) Extracting visual image features using LBPs, then 2) applying learning-based regression models, *i.e.*, support machine regression (SVR) or random forest (RF), to derive quality scores from extracted features. See Figs. 1 for the whole framework which includes training and testing stages of LBPs-based NR-IQA methods.

Our contributions in this work are two-fold:

- This work provides a comprehensive review of 30 LBP descriptors to familiarize the scientific community and support new researchers with past and current local binary pattern methodologies.
- We implement/reuse and apply these LBP variants to an NR-IQA framework. Experiments on synthetic and authentic data reveal suitable LBP descriptors and characteristics for NR-IQA.

The remainder of this paper is organized as follows. Section 2 presents a brief review of basic LBP operators and variants. The details of remaining LBP variants will be presented in the released projects including docs and codes. Section 3 describes the experimental setup, experimental results on three synthetic and three real-world IQA databases, and a discussion of these results. Section 4 concludes the paper.

2 LBP Descriptors

LBP is a statistical method that summarizes the local structure of the image [59]. It is regarded as an effective local texture descriptor. Texture is a fundamental attribute of images, but there is no consensus on its definition. In this paper, local structures such as textures refer to regional characteristics perceived as combinations of basic image patterns, following the definition in [17]. These basic image patterns exhibit a certain regularity that can be obtained through statistical measurement, which are then utilized for IQA.

This section only describes the basic LBP descriptor due to page limitations. Table 3 and Table 4 show the classification of the reviewed LBP variants in terms of the behavior for improving standard LBP and design objectives, respectively.

Throughout the paper, unless specified, for calculating various LBPs, R is the sampling radius of the neighborhood of the central pixel, P is the total number of neighbors sampled with a distance R. And g_c is gray (*i.e.*, intensity) value of the center pixel and g_p is gray value of the neighbors. $s(x)$ is the function what equals to 0 when x is less than 0 and equals to 1 otherwise. $U(\cdot)$ is a measure to count the number of spatial transitions (*i.e.*, 0-1, 1-0 jumps) in the binary code.

Table 3: Classification for LBP variants based on behavior

No.	Behavior	Method
1	Refining the encoding strategy	Rotation Invariant LBP, Uniform LBP, Dominant LBP, Local Ternary Patterns, Local Binary Count, Local Variance Patterns, Improved Local Quinary Patterns, Weber Local Binary Descriptor
2	Changing the threshold to be compared	Averaged LBP, Median LBP
3	Refining the neighbors of the neighborhood center	Adjacent Evaluation LBP, Path Integration-based LBP, Adjacent Evaluation LTP, Local Gradient Patterns, Eight Local Directional Patterns
4	Considering more information in the spatial domain	Complete LBP, Multi-Scale LBP, Pyramid LBP, Salient LBP, Multi-Scale Salient LBP, Completed Local Binary Count, Completed Local Derivative Patterns
5	Encoding in the frequency domain	Wavelet Domain LBP, Local Gabor Binary Patterns, Decorrelated LBP, Local Phase Quantization
6	Extension to color images	Opponent Color LBP, Orthogonal Color Planes Patterns

Generally, the LBP operator marks the pixels of an image with decimal numbers, called Local Binary Patterns, which encode the local structure around each pixel. In this section, 4 basic LBP operators (*i.e.*, original LBP, rotated LBP, uniform LBP and rotated-uniform LBP) are introduced. The original LBP was first proposed by [52]. It records a 0/1 sequence by comparing each pixel value with all surrounding neighboring pixel values and calculates a decimal number of binary strings. The traditional LBP operator is defined as follows:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad \text{where } s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases} \tag{1}$$

Table 4: Classification for LBP variants based on objective

Objective	Method
Improving the robustness to noise	Uniform LBP, Averaged LBP, Median LBP, Pyramid LBP, Adjacent Evaluation LBP, Local Ternary Patterns, Adjacent Evaluation LTP, Improved Local Quinary Patterns
Improving the robustness to local intensity variation	Local Gradient Patterns, Eight Local Directional Patterns, Weber Local Binary Descriptor
Improving the sensitivity to distortions	Salient LBP, Multi-Scale Salient LBP, Local Variance Patterns
Increasing the description capability	Complete LBP, Dominant LBP, Multi-Scale LBP, Opponent Color LBP, Wavelet Domain LBP, Local Gabor Binary Patterns, Path Integration-based LBP, Decorrelated LBP, Completed Local Binary Count, Orthogonal Color Planes Patterns

where R is the radius of neighborhood of the central pixel, P is the total number of neighbors sampled with a distance R, g_c is gray (*i.e.*, intensity) value of the center pixel and g_p is gray value of neighboring pixel. An example to calculate LBP binary code and LBP decimal value for one center pixel value with $P = 8, R = 3$ is shown in Fig. 2. The LBP feature of images is combining all LBP values (*i.e.*, decimal numbers) of all pixels. To reduce feature dimension, histograms can be calculated from LBP feature vectors/maps as feature vectors of images.

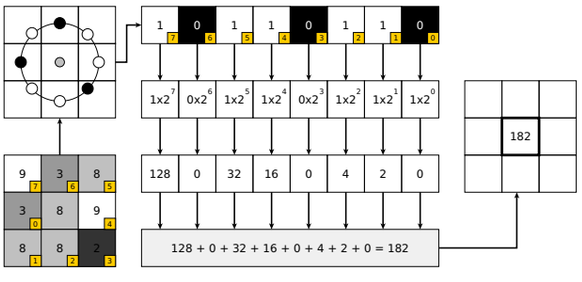


Figure 2: An illustration of the original LBP [52].

Such LBP is sensitive to image rotation. To address this issue, rotation invariant LBP ("ri" strategy) is made in which the LBP binary code is rotated one turn bit by bit to take the smallest decimal value among all candidate rotated binary codes. It is usually defined as:

$$LBP^{ri}_{P,R} = \min\{ROR(LBP_{P,R}, i),\ i = 0, 1, \ldots, P-1\}, \tag{2}$$

where $ROR(x, i)$ is the clockwise bit-wise shift operator that shifts the binary code x by i times.

The number of spatial transitions (bitwise 0/1 changes) in the pattern is defined as follows:

$$\begin{aligned} U\left(LBP_{P,R}\right) &= \left|s\left(g_{P-1} - g_c\right) - s\left(g_0 - g_c\right)\right| \\ &+ \sum_{p=1}^{P-1} \left|s\left(g_P - g_c\right) - s\left(g_{p-1} - g_c\right)\right|. \end{aligned} \tag{3}$$

The uniform pattern means that bitwise 0/1 changes (*i.e.*, 0 to 1 and 1 to 0 jumps) are not more than 2. Uniformity is important because it characterizes the patches that contain primitive structural information such as edges and corners. Based on the observation that more than 80 percent of binary patterns are uniform and many of the remaining ones contain essentially noise, Uniform LBP [52], are defined to reduce the number of local binary patterns. The non-uniform patterns are grouped into one mode (*i.e.*, single bin), typically without losing too much information. Combining rotation-invariant LBP and Uniform LBP can retain the minimum binary code with no more than two spatial transitions is called rotation-invariant uniform LBP ("riu2" strategy).

$$LBP^{riu2}_{P,R} = \begin{cases} LBP_{P,R}, & \text{if } U\left(LBP^{ri}_{P,R}\right) \leq 2 \\ P+1, & \text{else.} \end{cases} \tag{4}$$

3 Experiment

In this section, we investigate how each reviewed LBP descriptor affects no-reference image quality evaluation. We present performance comparisons to determine which descriptors are better suited for NR-IQA under synthetic and real-world cases. Additionally, we analyze the relationship between descriptor type and IQA method accuracy.

3.1 Experimental Setups

3.1.1 Databases. Six publicly available natural image quality databases are used for experimental evaluation, including (i) three artificially-distorted sets: LIVE [63], CSIQ [5] and TID2013 [56]; and (ii) three realistically-distorted sets: LIVE-C [21], KonIQ-10k [27] and SPAQ [8]. See below for their details.

- LIVE: 29 pristine images each of which is degraded by 5 distortions at four to five different levels of distortion, resulting in 982 distorted images.
- CSIQ: 30 pristine images and 866 distorted images with 6 distortion types at 4 to 5 distortion levels.
- TID2013: 25 pristine images and a total of 3,000 distorted images with 17 distortion types at 4 degradation levels.
- LIVE-C: 1,162 realistically natural pictures with resized resolution of 500×500 pixels.
- KonIQ-10k: 10,073 realistically and complexly distorted images with resolution of 1024×768 pixels.
- SPAQ: 11,125 images captured by 66 smartphones with diverse resolutions.

3.1.2 Evaluation Metrics. Three commonly-used evaluation metrics are adopted for performance comparison: Spearman Rank Order Correlation Coefficient (SROCC) measuring prediction monotonicity; Pearson Linear Correlation Coefficient (PLCC) measuring linear

correlation; and Root Mean Square Error (RMSE) measuring prediction accuracy. An effective IQA metric should yield high PLCC and SROCC values along with low RMSE values.

3.1.3 Implementation Details. The overall framework including training and testing stages of LBPs-based NR-IQA methods are described in Fig. 1. Following [17], the random forest is used as the regressor. It is implemented by TreeBagger of MATLAB with 50 trees and the optimization search method is not used. The experimental results are generated from the laptop with Intel (R) Xeon (R) e5-2690 V4 processor at 2.60 GHz with MATLAB. LBP codes are implemented in this paper except for LBP, CLBP, LPQ, AECLBP and AELTP whose codes are provided by the authors. For LBP variants designed for the gray images, we expand it to three-channel version for color images by concatenating LBP features for taking three images of one channel as input separately.

Following [67, 68], we randomly sample 80% of the images in each database for training and leave the rest for testing. Specifically, for synthetically distorted datasets, we split the training and test sets according to the pristine images such that the content is not intersected between the two sets. For performance comparison, the average values of evaluation metrics across one hundred sessions on five-fold division are reported for reducing the randomness of data partition.

For parameter settings for LBP variants used in IQA methods, the individual default values in the original papers are adopted if they are provided, otherwise the radius and the number of neighbor points are taken as 1 and 8 experimentally. Concretely, for example, the radius of LBP is 1 and the number of neighbors is 8. The threshold of LTP is 5. The window size of LPQ is set to 3×3. MSLBP sets the radius to 1 and the number of neighbors to 4,8,8, the radius to 2 and the number of neighbors to 4,8,16, the radius to 3 and the number of neighbors to 4,8,16,24, and 9 histograms using the "riu2" strategy are concatenated. MS-SLBP combines 9 SLBP histograms with the "riu2" strategy into a feature vector by concatenating the radius 1 with 4,8,8 neighbors, radius 2 with 4,8,16 neighbors, and radius 3 with 4,8,16,24 neighbors. Pi-LBP chooses the number of samples on the path to be 3. The standard deviation σ of the PLBP is set to 0.5 for the low-pass filter, and a downsampling rate of 2, and a 4-stage pyramid were used. In LGBP, five scales and eight orientations Gabor filters are used. In the experimental tables, LBP means basic operator using "riu2" strategy.

3.2 Individual-database Evaluation

Two well-known traditional handcrafted NR-IQA models NIQE [47] and ILNIQE [79] are selected for performance comparison. We first conduct evaluation on individual databases including both synthetic and authentic distortions.

The results on synthetically-distorted datasets (*i.e.* LIVE, CSIQ and TID2013) are reported in Table 5. The results show almost all LBP variants can be consistently and successfully applied to IQA on the LIVE database, except dLBP which performed poorer than other operators with mean SROCC below 0.6. LVP based on local energy achieved the best performance on LIVE. Most operators also demonstrated acceptable performance on CSIQ compared to NIQE, with WLBD using Weber's law performing best. However, on the more challenging TID2013 dataset, with complex image contents and distortion types, almost all operators struggled to perform consistently well. pi-LBP with strong descriptive capacity performed best on TID2013. dLBP performed worst on CSIQ and TID2013 since the decorrelation reduces distortion sensitivity. Notably, multiscale strategies and gradient computation improved performance on CSIQ and TID2013.

The results on authentically-distorted databases (*i.e.* LIVE-C, Koniq-10k and SPAQ) are reported in Table 5. The results indicate all variants struggled with the authentic data, likely due to the extremely complex distortions and diverse image contents. However, all variants still outperformed standard LBP on real-world data. In this case, saliency-based and adjacent evaluation strategies proved helpful for assessment, while gradient computation did not improve performance as much. Additionally, CLBP demonstrated superior performance over all other operators on the three datasets.

3.3 Cross-database Evaluation

We performed the cross evaluation to investigate the generality of LBP variants on IQA tasks. This evaluation uses all images from one database to train the methods and to test all images in other databases. Table 6 presents the SROCC values for synthetic data and authentic data. We only show some LBP variants with good performance when testing on six datasets individually. From the results, in terms of SROCC and PLCC, we can see that WLBD and CLBP outperform other methods on LIVE and LIVE-C, respectively. However, the best results are not satisfied due to the limitation of the representation power of LBP and the regression power of the random forest.

3.4 Comparison on Diverse Distortion Types

It is important to analyze LBP-based IQA metric performance across different distortion types. For this purpose, we evaluated the performance of each distortion type individually using the same protocol applied to the full datasets. Table 7 summarizes the results on the CSIQ database, which contains common degradation types. The per-distortion CSIQ results show WDLBP performed best for AWGN, pi-LBP for BLUR, OCPP for CONTRAST, OCLBP for FNOISE and JPEG, and LVP for JPEG2k. As expected, dLBP struggled on most types due to decorrelation losing distortion sensitivity. Overall, most variants succeeded on AWGN and BLUR, while JPEG and JPEG2k proved less challenging for operators. Gradient-based variants performed poorly on contrast due to descriptor invariance. Considering performance across datasets and CSIQ distortions, we evaluated standard LBP plus top variants (OCLBP, pi-LBP, WLBD) on 24 TID2013 degradations.

3.5 Computational Cost

The computational complexity is evaluated by comparing their running time. The testing time for LBP-based NR-IQA methods is important when facilitating their use in real-time image quality prediction. The time complexity of general NR-IQA methods includes the time consumed by feature extraction and score prediction, where the former costs the most time. In this experiment, we mainly consider the time consumed by extracting local statistical features by LBP operators. We only calculate the time (in seconds) consumed from reading the image to generating the image feature

Table 5: Performance comparison on synthetically and authentically distorted databases.

Database	LIVE		CSIQ		TID2013		LIVE-C		Koniq-10k		SPAQ	
	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC	SROCC
NIQE	0.9072	0.9060	0.7268	0.6232	0.5432	0.5787	0.5147	0.4643	0.5975	0.6014	0.7123	0.7032
ILNIQE	0.9123	0.9071	0.8736	0.8325	0.5982	0.5704	0.5080	0.4320	0.5230	0.5070	0.7214	0.7141
LBP	0.9225	0.9239	0.8235	0.7742	0.6237	0.5277	0.4369	0.3760	0.3941	0.3890	0.6989	0.6984
ALBP	0.8692	0.8828	0.7748	0.6998	0.6321	0.5483	0.6165	0.5710	0.7228	0.7138	0.7633	0.7603
MLBP	0.8332	0.8545	0.8244	0.7732	0.6270	0.5256	0.4900	0.4578	0.5932	0.5937	0.7422	0.7403
CLBP	0.9318	0.9384	0.8546	0.7996	0.6894	0.6244	0.6215	0.6019	**0.7852**	**0.7742**	**0.8216**	**0.8172**
DLBP	0.8165	0.8417	0.7751	0.7073	0.5296	0.3659	0.3482	0.3217	0.3568	0.3540	0.6499	0.6491
MSLBP	0.9196	0.9208	0.8311	0.7814	0.6922	0.6060	0.5656	0.5340	0.6321	0.6239	0.7614	0.7599
PLBP	0.9228	0.9232	0.8398	0.7934	0.6534	0.5514	0.5780	0.5505	0.6726	0.6599	0.7866	0.7841
OCLBP	0.9289	0.9406	0.8760	0.8379	0.6881	0.6133	0.5992	0.5796	0.6983	0.7061	0.7745	0.7740
WDLBP	0.9158	0.9220	0.8207	0.7620	0.6904	0.6015	0.5890	0.5491	0.6787	0.6653	0.7589	0.7571
LGBP	0.8879	0.8999	0.7854	0.7297	0.5971	0.5174	0.5382	0.5183	0.6851	0.6516	0.8065	0.8008
AELBP	0.8743	0.8943	0.8223	0.7421	0.6520	0.5721	**0.6302**	0.5874	0.7810	0.7615	0.8174	0.8129
pi-LBP	0.9280	0.9302	0.8557	0.8142	**0.7154**	**0.6509**	0.5835	0.5593	0.6948	0.6898	0.8026	0.7843
dLBP	0.5630	0.5565	0.4805	0.3755	0.4689	0.3657	0.4775	0.4645	0.7009	0.6715	0.7759	0.7615
SLBP	**0.9386**	0.9425	0.8316	0.7846	0.6851	0.5957	0.6158	0.5969	0.7054	0.7032	0.8105	0.7963
MS-SLBP	0.8946	0.9073	0.8063	0.7472	0.6679	0.5694	0.5691	0.5513	0.6975	0.6832	0.8042	0.7846
LTP	0.8429	0.8591	0.7970	0.7364	0.6266	0.5328	0.5529	0.5324	0.7108	0.7013	0.7497	0.7471
AELTP	0.8064	0.8269	0.7940	0.7190	0.6215	0.5307	0.5628	0.5203	0.7107	0.6890	0.7778	0.7748
LPQ	0.9032	0.9192	0.7935	0.7376	0.6723	0.6052	0.5562	0.5293	0.6918	0.6853	0.7644	0.7631
LGP	0.9258	0.9349	0.8226	0.7757	0.6312	0.5473	0.5127	0.4577	0.6069	0.5977	0.7482	0.7464
CLBC	0.9134	0.9174	0.8444	0.7959	0.6409	0.5419	0.5626	0.5426	0.6876	0.6820	0.7420	0.7416
ELDP	0.9067	0.9180	0.8748	0.8347	0.6829	0.6041	0.5477	0.5398	0.6048	0.6122	0.7047	0.7053
LVP	0.9380	**0.9448**	0.8763	0.8424	0.7119	0.6242	0.5972	0.5883	0.6771	0.6822	0.7586	0.7571
CLDP	0.9215	0.9262	0.8214	0.7621	0.6616	0.5587	0.5837	0.5560	0.7098	0.6973	0.7901	0.7859
OCPP	0.8383	0.8442	0.7908	0.7608	0.6099	0.5191	0.4522	0.4382	0.6377	0.6267	0.8015	0.8004
ILQP	0.9145	0.9205	0.8616	0.8216	0.6841	0.6008	0.6218	**0.6084**	0.7209	0.7224	0.7567	0.7546
WLBD	0.9302	0.9350	**0.8833**	**0.8547**	0.6757	0.6065	0.5811	0.5656	0.6757	0.6799	0.7504	0.7484

Table 6: Cross-database Validation.

	Train	Test	PLCC	SROCC	RMSE
LBP	TID2013	LIVE	0.4561	0.2859	39.5935
CLBP	TID2013	LIVE	0.5610	0.5620	39.5422
MSLBP	TID2013	LIVE	0.2186	0.0183	39.8800
PLBP	TID2013	LIVE	0.4965	0.4428	39.7700
Pi-LBP	TID2013	LIVE	0.5632	0.5990	39.4966
OCLBP	TID2013	LIVE	0.4849	0.4939	39.5074
LVP	TID2013	LIVE	0.5533	0.5757	**39.3279**
WLBD	TID2013	LIVE	**0.6777**	**0.6886**	39.3439
LBP	KonIQ_10k	LIVE_C	0.2741	0.2606	2.4938
CLBP	KonIQ_10k	LIVE_C	**0.5411**	**0.5043**	2.6038
MSLBP	KonIQ_10k	LIVE_C	0.3155	0.2940	2.4682
PLBP	KonIQ_10k	LIVE_C	0.3388	0.3117	**2.3620**
Pi-LBP	KonIQ_10k	LIVE_C	0.3956	0.3486	2.7598
OCLBP	KonIQ_10k	LIVE_C	0.4875	0.4537	2.6769
LVP	KonIQ_10k	LIVE_C	0.4960	0.4588	2.6753
WLBD	KonIQ_10k	LIVE_C	0.5205	0.4876	2.7012

vector. We randomly chose 100 test images from the LIVE-C dataset with size 500×500 and recorded the average time over images. Then the average time over 100 runs is reported in Table 8. It can be seen that the execution speeds of MS-LBP, pi-LBP, dLBP, MS-SLBP and LGBP are slower than those of other texture descriptors. This is mainly because of special time-consuming operations such as the DLBP decorrelation operation, the multi-path calculation of pi-LBP and the filtering of 40 Gabor filters of LGBP. Other than that, the rest of the texture descriptors take little time to execute and are all suitable for real-time applications in terms of time performance.

3.6 Basic Parameters Analysis

We test some basic but key parameters used mostly in LBP variants for assessing the qualities of real-world images. One is the radius length R of sampling. R indicates the texture scale that the texture descriptor can describe, and too large or too small radius may lead to distortion fluctuations that cannot be detected in the image, so it can affect the performance of IQA methods. Fig. 3 shows the SROCC values on different sampling radii for the LBP^{riu2} operator with $P = 8$ on the LIVE-C database. A total of 12 radius values from 1 to 12 are tested. From the results, we can see that the LBP operator

Table 7: SROCC comparison of NR-IQA methods based on different LBP variants on individual distortion type of CSIQ.

Distortion Type	LBP	ALBP	MLBP	CLBP	DLBP	MSLBP	PLBP	OCLBP	WDLBP	LGBP	AELBP	pi-LBP	dLBP
AWGN	0.8473	0.8777	0.8909	0.9208	0.8169	0.8725	0.9209	0.9196	0.9443	0.8196	0.9298	0.9420	0.8452
BLUR	0.9086	0.8865	0.8187	0.9020	0.8849	0.9071	0.8973	0.9184	0.8822	0.8797	0.8970	0.9318	0.7871
CONTRAST	0.5025	0.6789	0.5755	0.5892	0.6600	0.4796	0.6175	0.5851	0.6119	0.5999	0.2945	0.5472	0.5300
FNOISE	0.7373	0.6172	0.7537	0.8594	0.6388	0.7817	0.8537	**0.9067**	0.8763	0.7277	0.7488	0.8380	0.3512
JPEG	0.8828	0.8258	0.8626	0.9253	0.8286	0.8995	0.8894	**0.9458**	0.8916	0.9306	0.8899	0.9168	0.4997
JPEG2k	0.8600	0.8110	0.7678	0.8914	0.7954	0.8455	0.8492	0.9035	0.8467	0.8703	0.8542	0.8850	0.6546
Distortion Type	SLBP	MS-SLBP	LTP	AELTP	LPQ	LGP	CLBC	ELDP	LVP	CLDP	OCPP	ILQP	WLBD
AWGN	0.8931	0.8351	0.7553	**0.9453**	0.7956	0.9006	0.8881	0.7915	0.9053	0.8816	0.8043	0.8805	0.9332
BLUR	0.9117	0.8912	0.7735	0.8747	0.9032	0.9012	0.8926	0.9103	**0.9284**	0.8975	0.6774	0.9086	0.9129
CONTRAST	0.6217	0.5229	0.6957	0.6157	0.5428	0.4254	0.4845	0.4972	0.5874	0.5330	**0.7478**	0.5997	0.6049
FNOISE	0.8044	0.7194	0.6848	0.8271	0.6906	0.6923	0.8200	0.8250	0.8937	0.7919	0.6194	0.8568	0.8932
JPEG	0.8987	0.8867	0.6700	0.7748	0.9397	0.8884	0.8863	0.9355	0.9434	0.9110	0.7398	0.8937	0.9384
JPEG2k	0.8758	0.8503	0.7106	0.7894	0.8704	0.8496	0.8750	**0.9076**	**0.9079**	0.8714	0.7210	0.8676	0.8876

Table 8: Computational time (in seconds) comparison

Operator	Time	Operator	Time	Operator	Time
LBP	0.0330	LGBP	0.3194	LGP	0.2799
ALBP	0.0905	AELBP	0.0552	CLBC	0.0385
MLBP	0.0917	pi-LBP	0.9757	ELDP	0.0351
CLBP	0.0982	dLBP	0.3308	LVP	0.0409
DLBP	0.0291	SLBP	**0.0084**	CLDP	0.1341
MSLBP	0.3841	MS-SLBP	0.2281	OCPP	0.0637
PLBP	0.0371	LTP	0.0574	ILQP	0.2021
OCLBP	0.0337	AELTP	0.0716	WLBD	0.0679
WDLBP	0.0090	LPQ	0.0181		

with a radius of 1 has the highest performance, probably because the distortion of the image is more sensitive to feature variations with more complete information for the local region. When the P is fixed, smaller R means more concentrated and more completed information can be obtained. As a result, the parameter setting of $R = 1$ is used in the performance comparison experiment for LBP variants without default parameter settings.

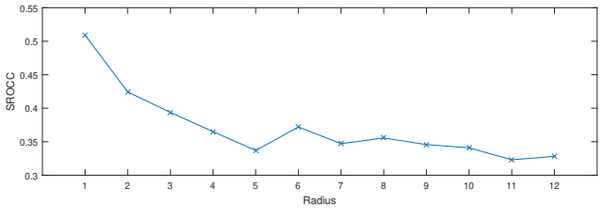


Figure 3: Performance of the LBP operator with the different sampling radius on LIVE-C.

Another is the number of sampled neighbors P. P represents the number of times when the central point of LBP needs to be compared with the surrounding points, and it decides the length of the resulting feature vector. If this value is small, the descriptive power of the LBP operators will be reduced and it is likely that the distortion will not be detected and quantized apparently. Bigger P means utilizing more information in each local region but leading to redundant information and length of features. Fig. 4 shows the SROCC values of the LBP^{riu2} operator with $R = 1$ on the LIVE-C database for different number of sampled neighborhood points. A total of 13 sampling point numbers from 4 to 16 are tested. From the results, we can see that the performance has a tend to increase as P. Considering the trade-off between the description ability of the operator and the feature length, P is set to 8 in the performance comparison experiment for LBP variants without default parameter settings.

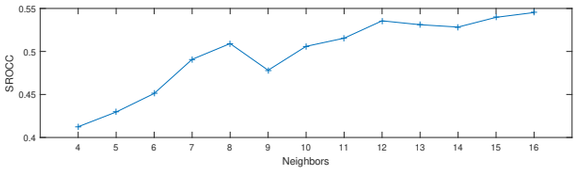


Figure 4: Performance of LBP algorithms with different radius on LIVE-C.

4 Conclusion

In this paper, we first provide a brief review of 30 variants of the basic LBP as texture descriptors. We then investigate their performance when applied to NR-IQA under synthetic and authentic distortion cases. From the results, we verify whether LBP variants can serve as effective feature descriptors for IQA applications and determine which characteristics of LBP operators are most suitable for IQA. Results demonstrate that multiscale approaches and gradient-based methods substantially improve quality prediction performance. This work aims to familiarize the scientific community and support new researchers with past and current local binary pattern methodologies used in different fields. Moving forward, we will integrate the characteristics most beneficial to IQA to design an improved LBP operator for quality prediction in real-world cases.

Acknowledgments

Yuhui Quan would like to acknowledge the partial support from National Natural Science Foundation of China under Grant 62372186, Natural Science Foundation of Guangdong Province under Grants 2022A1515011755 and 2023A1515012841, Fundamental Research Funds for the Central Universities under Grant x2jsD2230220, and National Key Research and Development Program of China under Grant 2024YFE0105400. Zihan Zhou would like to acknowledge the partial support from Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515110646. Yong Xu would like to acknowledge the partial support from National Foreign Expert Project of the Ministry of Science and Technology of China under Grant G2023163015L, National Natural Science Foundation of China under Grant 62072188, Guangzhou Science and Technology Plan Project under Grant 2024B01W0007. Ruotao Xu would like to acknowledge the partial support from National Natural Science Foundation of China under Grant 62106077 and the partial support Natural Science Foundation of Guangdong Province, China under Grant 2022A1515011087.

References

[1] Timo Ahonen, Esa Rahtu, Ville Ojansivu, and Janne Heikkila. 2008. Recognition of blurred faces using local phase quantization. In *International conference on pattern recognition*. IEEE, 1–4.

[2] Bhawna Ahuja and Virendra P. Vishwakarma. 2021. Local Binary Pattern Based ELM for Face Identification. *Advances in Intelligent Systems and Computing* 1164.

[3] Laleh Armi and Shervan Fekri-Ershad. 2019. Texture image Classification based on improved local Quinary patterns. *Multimedia Tools and Applications* 78 (2019), 18995–19018. Issue 14.

[4] Soumendu Chakraborty and Anand Singh Jalal. 2020. A novel local binary pattern based blind feature image steganography. *Multimedia Tools and Applications* 79 (2020). Issue 27-28.

[5] Damon M. Chandler. 2010. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging* 19 (2010). Issue 1.

[6] H. W. Chang, H. Yang, Y. Gan, and M. H. Wang. 2013. Sparse feature fidelity for perceptual image quality assessment. *IEEE Trans. Image Process.* 22, 10 (2013), 4007–18.

[7] Ken Chen, Yichao Wu, Zhenmao Li, Yudong Wu, and Ding Liang. 2020. Face image quality assessment for model and human perception. *Proceedings - International Conference on Pattern Recognition*.

[8] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. 2020. Perceptual quality assessment of smartphone photography. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

[9] Mohammad Reza Faraji and Xiaojun Qi. 2015. Face recognition under illumination variations based on eight local directional patterns. *IET Biometrics* 4 (2015). Issue 1.

[10] Emna Fourati, Wael Elloumi, and Aladine Chetouani. 2020. Anti-spoofing in face recognition-based biometric authentication using Image Quality Assessment. *Multimedia Tools and Applications* 79 (2020). Issue 1-2.

[11] Pedro Garcia Freitas, Welington YL Akamine, and Mylène CQ Farias. 2016. Blind image quality assessment using multiscale local binary patterns. *Journal of Imaging Science and Technology* 60, 6 (2016), 60405–1.

[12] Pedro Garcia Freitas, Welington Y.L. Akamine, and Mylene C.Q. Farias. 2016. No-reference image quality assessment based on statistics of Local Ternary Pattern. *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*.

[13] Pedro Garcia Freitas, Welington Y L Akamine, and Mylène C Q Farias. 2018. No-Reference Image Quality Assessment Using Orthogonal Color Planes Patterns. *IEEE Transactions on Multimedia* 20 (2018), 3353–3360. Issue 12.

[14] Pedro Garcia Freitas, Welington Yorihiko Lima Akamine, and Mylene Christine Queiroz De Farias. 2017. Blind image quality assessment using local variant patterns. *Proceedings of Brazilian Conference on Intelligent Systems* 2018-January.

[15] Pedro Garcia Freitas, Welington Yorihiko Lima Akamine, and Mylène Christine Queiroz De Farias. 2018. No-reference image quality assessment using salient local binary patterns. *IS and T International Symposium on Electronic Imaging Science and Technology*.

[16] Pedro Garcia Freitas, Sana Alamgeer, Welington Y.L. Akamine, and Mylène C.Q. Farias. 2018. Blind image quality assessment based on multiscale salient local binary patterns. *Proceedings of the 9th ACM Multimedia Systems Conference, MMSys 2018*.

[17] Pedro Garcia Freitas, Luísa Peixoto Da Eira, Samuel Soares Santos, and Mylene Christine Queiroz de Farias. 2018. On the Application LBP Texture Descriptors and Its Variants for No-Reference Image Quality Assessment. *Journal of Imaging* 4, 10 (2018).

[18] Pedro Garcia Freitas, Luísa Peixoto da Eira, Samuel Soares Santos, and Mylène C.Q. Farias. 2020. Image quality assessment using BSIF, CLBP, LCP, and LPQ operators. *Theoretical Computer Science* 805 (2020).

[19] Pedro Garcia Freitas and Mylène Christine Queiroz De Farias. 2017. On the Performance of Visual Semantics for Improving Texture-Based Blind Image Quality Assessment. *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 330–337.

[20] Xinbo Gao, Fei Gao, Dacheng Tao, and Xuelong Li. 2013. Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning. *IEEE Trans. Neural Netw. Learn. Syst.* 24, 12 (2013).

[21] Deepti Ghadiyaram and Alan C. Bovik. 2016. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing* 25 (2016). Issue 1.

[22] Ke Gu, Dacheng Tao, Jun-Fei Qiao, and Weisi Lin. 2018. Learning a no-reference quality assessment model of enhanced images with big data. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 4 (2018), 1301–1313.

[23] Zhenhua Guo, Lei Zhang, and David Zhang. 2010. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing* 19 (2010). Issue 6.

[24] Adel Hafiane, Guna Seetharaman, and Bertrand Zavidovique. 2007. Median binary pattern for textures classification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4633 LNCS.

[25] M. Hassaballah, Hammam A. Alshazly, and Abdelmgeid A. Ali. 2019. Ear recognition using local binary patterns: A comparative experimental study. *Expert Systems with Applications* 118 (2019).

[26] M Hassaballah, Hammam A Alshazly, and Abdelmgeid A Ali. 2019. Ear recognition using local binary patterns: A comparative experimental study. *Expert Systems with Applications* 118 (2019), 182–200.

[27] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. 2020. KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment. *IEEE Transactions on Image Processing* 29 (2020).

[28] Han Kai Hsu, Chun Han Yao, Yi Hsuan Tsai, Wei Chih Hung, Hung Yu Tseng, Maneesh Singh, and Ming Hsuan Yang. 2020. Progressive domain adaptation for object detection. *Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020*.

[29] Ran Hu, Xiaolong Li, and Zongming Guo. 2018. Decorrelated local binary patterns for efficient texture classification. *Multimedia Tools and Applications* 77 (2018), 6863–6882. Issue 6.

[30] Yuting Hu, Zhiling Long, and Ghassan AlRegib. 2016. Completed local derivative pattern for rotation invariant texture classification. *IEEE International Conference on Image Processing*, 3548–3552.

[31] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. 2011. Local binary patterns and its application to facial image analysis: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41, 6 (2011), 765–781.

[32] Ying Huang and Kaikai Yao. 2020. Multi-Exposure Image Fusion Method Based on Independent Component Analysis. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*.

[33] R. Rizal Isnanto, Adian Fatchur Rochim, Dania Eridani, and Guntur Dwi Cahyono. 2021. Multi-Object Face Recognition Using Local Binary Pattern Histogram and Haar Cascade Classifier on Low-Resolution Images. *International Journal of Engineering and Technology Innovation* 11 (2021). Issue 1.

[34] Bongjin Jun and Daijin Kim. 2012. Robust face detection using local gradient patterns and evidence accumulation. *Pattern Recognition* 45, 9 (2012), 3304–3316.

[35] Shekhar Karanwal and Manoj Diwakar. 2021. OD-LBP: Orthogonal difference-local binary pattern for Face Recognition. *Digital Signal Processing: A Review Journal* 110 (2021).

[36] Ali Khodabakhsh, Marius Pedersen, and Christoph Busch. 2019. Subjective versus objective face image quality evaluation for face recognition. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*.

[37] Durga Ganga Rao Kola and Srinivas Kumar Samayamantula. 2021. A novel approach for facial expression recognition using local binary pattern with adaptive window. *Multimedia Tools and Applications* 80 (2021). Issue 2.

[38] Lingchao Kong, Ademola Ikusan, Rui Dai, and Jingyi Zhu. 2019. Blind Image Quality Prediction for Object Detection. *Proceedings - 2nd International Conference on Multimedia Information Processing and Retrieval, MIPR 2019*.

[39] Lingchao Kong, Ademola Ikusan, Rui Dai, Jingyi Zhu, and Dara Ros. 2019. A No-Reference Image Quality Model for Object Detection on Embedded Cameras. *International Journal of Multimedia Data Engineering and Management* 10 (2019). Issue 1.

[40] S Liao, Max W K Law, and Albert C S Chung. 2009. Dominant Local Binary Patterns for Texture Classification. *IEEE Transactions on Image Processing* 18 (2009), 1107–1118. Issue 5.

[41] Qiuyan Lin and Wenfa Qi. 2015. Multi-scale local binary patterns based on path integral for texture classification. *IEEE International Conference on Image Processing*, 26–30.

[42] Xianming Liu, Deming Zhai, Jiantao Zhou, Shiqi Wang, Debin Zhao, and Huijun Gao. 2017. Sparsity-based image error concealment via adaptive dual dictionary learning and regularization. *IEEE Trans. Image Process.* 26, 2 (2017), 782–796.

[43] Yutao Liu, Guangtao Zhai, Ke Gu, Xianming Liu, Debin Zhao, and Wen Gao. 2018. Reduced-reference image quality assessment in free-energy principle and sparse representation. *IEEE Trans. Multimedia* 20, 2 (2018), 379–391.

[44] Lin Ma, Songnan Li, and King Ngi Ngan. 2012. Reduced-reference video quality assessment of compressed video sequences. *IEEE Trans. Circuits Syst. Video Technol.* 22, 10 (2012), 1441–1456.

[45] Amandeep Rasool Masrat and Kaur. 2021. A Novel Rotation Invariant Descriptor for Texture Classification with Local Binary Patterns, V Kamakshi, Wang Jiacun, Reddy K T V Reddy V. Sivakumar, and Prasad (Eds.). *Soft Computing and Signal Processing*, 385–396.

[46] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. 2021. MagFace: A Universal Representation for Face Recognition and Quality Assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14225–14234.

[47] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Letters* 20, 3 (2012), 209–212.

[48] T Mäenpää. 2003. The Local Binary Pattern Approach to Texture Analysis: Extensions and Applications. *Oulun Yliopisto: Oulu, Finland* (2003).

[49] I. Nenakhov, V. Khryashchev, and A. Priorov. 2016. No-reference image quality assessment based on local binary patterns. *Proceedings of 2016 IEEE East-West Design and Test Symposium, EWDTS 2016*.

[50] Imran Fareed Nizami, Mobeen ur Rehman, Muhammad Majid, and Syed Muhammad Anwar. 2020. Natural scene statistics model independent no-reference image quality assessment using patch based discrete cosine transform. *Multimedia Tools and Applications* (2020).

[51] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence* 24, 7 (2002), 971–987.

[52] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002). Issue 7.

[53] Ville Ojansivu and Janne Heikkilä. 2008. Blur insensitive texture classification using local phase quantization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5099 LNCS.

[54] Fu Zhao Ou, Yuan Gen Wang, and Guopu Zhu. 2019. A Novel Blind Image Quality Assessment Method Based on Refined Natural Scene Statistics. *Proceedings of International Conference on Image Processing* 2019-September.

[55] Zhibin Pan, Shiqi Hu, Xiuquan Wu, and Ping Wang. 2021. Adaptive center pixel selection strategy in Local Binary Pattern for texture classification. *Expert Systems with Applications* 180 (2021).

[56] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C. C. Jay Kuo. 2015. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication* 30 (2015).

[57] Xueming Qian, Xian-Sheng Hua, Ping Chen, and Liangjun Ke. 2011. PLBP: An effective local binary patterns texture descriptor with pyramid representation. *Pattern Recognition* 44 (2011), 2502–2515. Issue 10. https://www.sciencedirect.com/science/article/pii/S0031320311001336 Semi-Supervised Learning for Visual Content Analysis and Understanding.

[58] Chuan Qin, Yecen Hu, Heng Yao, Xintao Duan, and Liping Gao. 2019. Perceptual Image Hashing Based on Weber Local Binary Pattern and Color Angle Representation. *IEEE Access* 7 (2019).

[59] Ayushman Ramola, Amit Kumar Shakya, and Dai Van Pham. 2020. Study of statistical methods for texture analysis and their modern evolutions. *Engineering Reports* 2, 4 (2020), e12149.

[60] Farshad Rezaie, Mohammad Sadegh Helfroush, and Habibollah Danyali. 2018. No-reference image quality assessment using local binary pattern in the wavelet domain. *Multimedia Tools and Applications* 77 (2018), 2529–2541. Issue 2.

[61] Andleeb Sadiq, Imran Fareed Nizami, Syed Muhammad Anwar, and Muhammad Majid. 2020. Blind image quality assessment using natural scene statistics of stationary wavelet transform. *Optik* 205 (2020).

[62] H. R. Sheikh, A. C. Bovik, and L Cormack. 2005. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Trans. Image Process.* 14, 11 (2005), 1918–1927.

[63] H R Sheikh, Z Wang, L Cormack, and A C Bovik. 2019. LIVE image quality assessment database release 2. (2019).

[64] Kechen Song, Yunhui Yan, Yongjie Zhao, and Changsheng Liu. 2015. Adjacent evaluation of local binary pattern for texture classification. *Journal of Visual Communication and Image Representation* 33 (2015), 323–339.

[65] Tiecheng Song, Yuanjing Han, Jie Feng, Yuanlin Wang, and Chenqiang Gao. 2020. First- And second-order sorted local binary pattern features for grayscale-inversion and rotation invariant texture classification. *Proceedings - International Conference on Pattern Recognition*.

[66] Prashant Srivastava and Ashish Khare. 2018. Utilizing multiscale local binary pattern for content-based image retrieval. *Multimedia Tools and Applications* 77, 10 (2018), 12377–12403.

[67] Yicheng Su and Jari Korhonen. 2020. Blind natural image quality prediction using convolutional neural networks and weighted spatial pooling. In *Proc. Int. Conf. Image Process.* 191–195.

[68] Hossein Talebi and Peyman Milanfar. 2018. NIMA: Neural image assessment. *IEEE Trans. Image Process.* 27, 8 (2018), 3998–4011.

[69] Xiaoyang Tan and Bill Triggs. 2010. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing* 19 (2010). Issue 6.

[70] Peng Tang, Chunyu Wang, Xinggang Wang, Wenyu Liu, Wenjun Zeng, and Jingdong Wang. 2020. Object Detection in Videos by High Quality Object Linking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (2020). Issue 5.

[71] Zhaolin Wan, Ke Gu, and Debin Zhao. 2020. Reduced Reference Stereoscopic Image Quality Assessment Using Sparse Representation and Natural Scene Statistics. *IEEE Transactions on Multimedia* 22 (2020). Issue 8.

[72] Zhou Wang, A. C. Bovik, and B. L. Evan. 2000. Blind measurement of blocking artifacts in images. In *Proc. Int. Conf. Image Process.* 981–984 vol.3.

[73] Zhou Wang, A. C Bovik, H. R Sheikh, and E. P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 4 (2004), 600–12.

[74] Jinjian Wu, Weisi Lin, Guangming Shi, Leida Li, and Yuming Fang. 2016. Orientation selectivity based visual pattern for reduced-reference image quality assessment. *Information Sciences* 351 (2016), 18–29.

[75] Zhihua Xia, Chengsheng Yuan, Rui Lv, Xingming Sun, Neal N Xiong, and Yun-Qing Shi. 2020. A Novel Weber Local Binary Descriptor for Fingerprint Liveness Detection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50 (2020), 1526–1536. Issue 4.

[76] Xichen Yang, Quansen Sun, and Tianshu Wang. 2019. No-reference image quality assessment based on sparse representation. *Neural Computing and Applications* 31 (2019). Issue 10.

[77] Huige Yin, Yuantao Chen, Jie Xiong, Runlong Xia, Jingbo Xie, and Kai Yang. 2021. An improved local binary pattern method for pollen image classification and recognition. *Computers and Electrical Engineering* 90 (2021).

[78] Chuang Zhang, Jiawei Xu, Xiaoyu Huang, and Seop Hyeong Park. 2019. No-Reference Image Quality Assessment Using Independent Component Analysis and Convolutional Neural Network. *Journal of Electrical Engineering and Technology* 14 (2019). Issue 1.

[79] Lin Zhang, Lei Zhang, and Alan C Bovik. 2015. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* 24, 8 (2015), 2579–2591.

[80] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. 2011. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Trans. Image Process.* 20, 8 (2011), 2378–2386.

[81] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang. 2005. Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* 1, 786–791 Vol. 1.

[82] Yang Zhao, De-Shuang Huang, and Wei Jia. 2012. Completed Local Binary Count for Rotation Invariant Texture Classification. *IEEE Transactions on Image Processing* 21 (2012), 4492–4497. Issue 10.

[83] Wujie Zhou, Lu Yu, Weiwei Qiu, Yang Zhou, and Mingwei Wu. 2017. Local gradient patterns (LGP): An effective local-statistical-feature extraction scheme for no-reference image quality assessment. *Information Sciences* 397-398 (2017).

[84] Zihan Zhou, Jing Li, Yuhui Quan, and Ruotao Xu. 2021. Image Quality Assessment Using Kernel Sparse Coding. *IEEE Trans. on Multimedia* 23 (2021), 1592–1604.

[85] Zihan Zhou, Jing Li, Dexiang Zhong, Yong Xu, and Patrick Le Callet. 2024. Deep Blind Image Quality Assessment Using Dynamic Neural Model with Dual-order Statistics. *IEEE Transactions on Circuits and Systems for Video Technology* (2024).

[86] Zihan Zhou, Yong Xu, Yuhui Quan, and Ruotao Xu. 2022. Deep blind image quality assessment using dual-order statistics. In *2022 IEEE International Conference on Multimedia and Expo*. IEEE, 01–06.

[87] Zihan Zhou, Yong Xu, Ruotao Xu, and Yuhui Quan. 2022. No-reference image quality assessment using dynamic complex-valued neural model. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1006–1015.

Underwater Image Enhancement via Adaptive Bi-Level Color-Based Adjustment

Yun Liang, Lianghui Li, Zihan Zhou, Lieyu Tian, Xinjie Xiao, and Huan Zhang

***Abstract*— Underwater images often exhibit severe color distortions and reduced contrast due to light absorption and scattering, presenting substantial challenges for image enhancement techniques. To address these challenges, this article presents BCTA-Net, an adaptive bi-level color-based network specifically engineered to enhance the quality of underwater images by addressing distortions in dynamic and complex environments. The network integrates content-aware global and local restoration strategies. On a local scale, a color-aware attention mechanism is proposed which employs color histograms to adaptively correct nonuniform color distortions and enhance local color fidelity. In addition, a triple attention (TA) module restores spatially varying local details in a content-aware manner, improving clarity and texture precision of enhancement. These elements are combined into a dual-branch architecture aimed at reducing local contrast, color fidelity, and detail precision issues. On a global scale, contrastive learning focused on background lightness corrects color distortions due to uneven illumination. The integration of these components results in a lightweight, dynamic global–local model with robust generalization capabilities across various underwater scenarios, as demonstrated by comprehensive experiments that show significant performance improvements over existing methods.**

***Index Terms*— Color attention, contrastive learning, convolutional network, global and local distortion recovery, underwater image enhancement (UIE).**

I. Introduction

UNDERWATER imaging is essential for applications such as underwater robotics and ocean resource exploration.

Received 16 December 2024; accepted 19 February 2025. Date of publication 17 March 2025; date of current version 28 March 2025. The work of Yun Liang was supported in part by the Key Research and Development Project of Guangzhou under Grant 202206010091; in part by the Fund of Southern Marine Science and Engineering Guangdong Laboratory at Zhanjiang under Grant ZJW-2023-04; and in part by the Special Fund for Marine Economic Development (Six Marine Industries) of Guangdong Province, China, under Grant GDNRC[2024] No.18. The work of Zihan Zhou was supported in part by the National Natural Science Foundation of China under Grant 62401210, in part by the National Key Research and Development Program of China under Grant 2024YFC2814901, in part by the Natural Science Foundation of Guangdong Province under Grant 2025A1515011539, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515110646, and in part by Guangzhou Basic and Applied Basic Research Project under Grant 2025A04J4502. The Associate Editor coordinating the review process was Dr. Xianqiang Yang. *(Corresponding author: Zihan Zhou.)*

Yun Liang, Lianghui Li, Zihan Zhou, and Xinjie Xiao are with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China (e-mail: yliang@scau.edu.cn; 15918411668@stu.scau.edu.cn; zhouzihan@scau.edu.cn; caelum@stu.scau.edu.cn).

Lieyu Tian is with Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou 510075, China (e-mail: tianlieyu23@163.com).

Huan Zhang is with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: huanzhang2021@gdut.edu.cn).

Digital Object Identifier 10.1109/TIM.2025.3551931

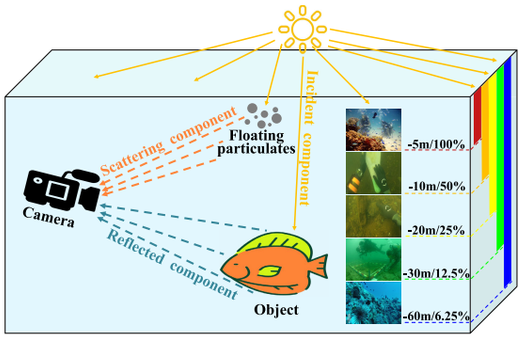


Fig. 1. Underwater imaging model, light attenuation characteristics, and underwater images across different colors.

However, the quality of underwater images is significantly degraded due to the absorption and scattering of light, leading to issues such as color casts, low contrast, haze, and blurriness [1]. These degradations adversely impact the performance and reliability of underwater visual tasks, highlighting the importance of underwater image enhancement (UIE). UIE plays a crucial role in improving the visual quality and fidelity of images, thereby facilitating a more accurate understanding of the underwater environment.

Fig. 1 illustrates the process of underwater optical imaging and the attenuation of light across different wavelengths. The degradation of underwater images can be attributed to two main factors. First, the propagation distance of light varies with wavelength, unlike the uniform attenuation observed in terrestrial light [2]. As the propagation distance increases, longer wavelengths such as red and orange attenuate more dramatically, resulting in a blue–green bias associated with shorter wavelengths [3], as observed in various underwater images at different depths in Fig. 1. This effect becomes more pronounced with increasing distance between the subject and the camera. Second, suspended particles in the water, including organic particles and planktonic microorganisms, cause light scattering, which deflects the light propagation direction and reduces image contrast and clarity [4].

In order to improve the quality of underwater images, numerous methods have emerged, broadly classified into two categories: traditional and deep learning-based approaches. Traditional UIE methods include prior-based and physical model-based approaches. Prior-based methods leverage rich priors and explore the spatial relationships between pixel values in the original underwater image to enhance it by adjusting contrast, brightness, and saturation. See Gray World [5], Max RGB [6], and White Balance [7] for examples. However, these

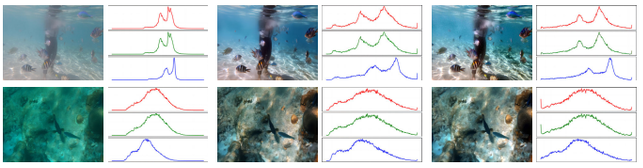


Fig. 2. Three pairs of the image and related histogram from the UIEB dataset [1]. GT means ground truth. (a) Input pairs. (b) Enhanced pairs. (c) GT pairs.

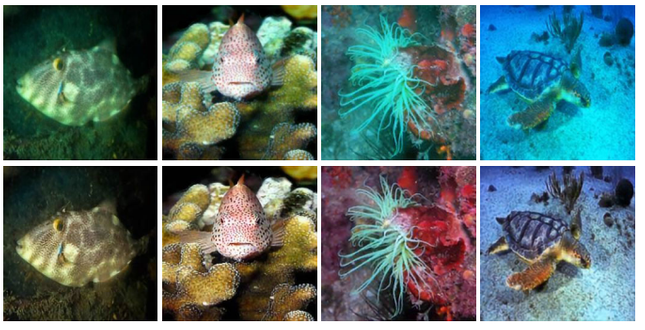


Fig. 3. Examples of underwater images with nonuniform color distortion from the EUVP dataset. The top row displays the original underwater images. The bottom row displays the enhanced results of our method.

methods often overlook the physical degradation processes, limiting their enhancement quality. Physical model-based UIE methods [8], [9], [10] focus on accurately estimating the parameters of the underwater image formation model or medium transmission. These methods aim to obtain a clean image by reversing the physical underwater imaging model. Despite their potential, the performance of traditional UIE methods is often restricted by the complexities and diverse conditions of real-world underwater environments.

Recently, deep learning has driven significant advancements in UIE techniques, with major approaches founded on convolutional neural networks (CNNs) [11], [12], [13], generative adversarial networks (GANs) [14], [15], [16], and Transformers [17], [18], [19]. In this work, we focus on CNN architectures due to their powerful visual feature learning capabilities and flexible training processes. However, despite their potential, CNN-based UIE methods encounter challenges in real-world underwater scenes with complex distortions and diverse image contents, which impede accurate and adaptive restoration. As depicted in underwater imaging models [20], global color distortion caused by background illumination and local complex distortions within the scene radiance pose significant hurdles, substantially degrading image clarity. Thus, this article aims to enhance the capabilities of CNN-based models for UIE in three following key aspects.

1) *Introducing the Color-Aware Attention Mechanism for Adaptive Nonuniform Color Recovery:* UIE faces major challenges due to light attenuation-induced color casts (see Fig. 1), requiring adaptive recovery of degraded color information across channels and spatial locations. Traditional methods like gray world assumption and white balance lack dynamic adaptation capabilities, while existing deep learning approaches are limited in explicitly modeling color distortions. Our analysis reveals significant differences between color histogram distributions of high-quality and raw underwater images (see Fig. 2). Enhanced images show broad, uniform histograms, while raw images exhibit narrow, shifted distributions due to absorption, scattering, and particulates. To address this, we propose a color-aware attention branch that leverages color histograms for adaptive local color restoration. This histogram-based approach enables quantitative color distribution assessment and effective correction of nonuniform color distortions, as demonstrated in Fig. 3.
2) *Designing the Triple Attention (TA) Module for Content-Aware Spatially-Varying Restoration:* An effective UIE model should be capable of handling underwater images with diverse distortion types, as well as varying semantic content. However, a traditional CNN model is typically content-agnostic because the spatially-invariant convolutional filters are shared across all spatial locations and input images for a trained model. Besides, it is impractical to design a fixed filter bank covering all potential patterns for CNN, especially for authentically underwater images. Thus, the explicit adaptive mechanism is necessary for adaptive UIE. In this article, attention mechanisms are exploited and the TA module is proposed, in order to enhance the restoration of detailed and blurred information in underwater images. This module integrates self-attention, channel attention, and spatial attention mechanisms, which helps the model to focus on different aspects of the input data. In detail, spatial attention enhances important features of high-frequency regions and significantly blurred areas by adjusting spatial location weights, crucial for improving clarity and detail in primary content regions [21], [22], [23]. Channel attention prioritizes the most informative color channels [24], [25], [26]. And the self-attention models long-range dependencies across the entire image [27], [28], [29]. This results in a thorough and nuanced enhancement across both local and global scales.
   Despite the effectiveness of existing attention-based methods in enhancing specific aspects of underwater images [30], [31], they often overlook comprehensive improvements across key color channels, crucial regions, and long-range dependencies. The TA module addresses these gaps by integrating spatial, channel, and self-attention mechanisms, ensuring a holistic enhancement across both local and global scales. This approach not only improves visual quality by rectifying detailed and blurred regions but also enhances overall image clarity and contrast.
3) *Introducing Contrastive Learning of the Background Lightness for Global Illumination Restoration:* Underwater background illumination significantly impacts color fidelity and visual perception, with variability across different depths and environments. Contrasting with physical model-based methods that estimate background illumination directly [32], [33], our approach leverages contrastive learning to effectively map diverse

illumination conditions, thereby refining the solution space for global illumination restoration for each training sample. We introduce a strategy that includes the generation and filtering of color-degraded background light, enhancing the negative sample sets to overcome the limitations of small underwater image datasets. Moreover, the integration of contrastive learning in our framework enables precise differentiation and isolation of background light effects from intrinsic image features. This method not only strengthens feature extraction robustness but also significantly improves the visual output quality. As a result, our approach achieves more accurate color restoration and enhanced clarity, substantially improving the generalization ability.

Based on the aforementioned three improvements, we have developed a bi-level color-based TA network (BCTA-Net) for UIE. Our contributions are threefold.

1) *Color-Based Attention Mechanisms for Adaptive Nonuniform Local Distortion Restoration:* We have developed a color-based attention branch that models color information using histograms and integrates a self-attention mechanism, allowing for adaptive correction of nonuniform color distortions in degraded images. In addition, we introduce a TA mechanism that combines spatial, channel, and self-attentions to enhance the analysis of crucial channels, key regions, and global dependencies. The integration of these branches enables content-aware restoration of color biases, blurring, and low-contrast areas in underwater images, effectively addressing both localized and comprehensive distortions based on image content.
2) *Background Illumination-Based Contrastive Learning for Global Color Adjustment:* We introduce a novel approach that integrates global uniform background illumination information with contrastive learning techniques to reduce the solution space of clean images and facilitate global color adjustment in underwater images. By generating pairs of images that simulate different lighting conditions, our model learns to identify and correct deviations from natural color profiles.
3) *BCTA-Net for Content-Aware UIE:* We propose a bi-level color-aided TA network for UIE, the first to address local nonuniform and global uniform distortions with color information simultaneously. This network integrates a local color recovery (LCR) branch for handling nonuniform distortions and a multiscale TA branch coupled with background illumination-based contrastive learning for correcting global uniform distortions. Extensive evaluations on benchmark datasets, encompassing both synthetic and authentic underwater images, demonstrate the model's effectiveness across various underwater conditions.

II. Related Work

A. Traditional Methods

The traditional UIE algorithms can be broadly categorized into prior-based methods and physical model-based methods.

1) Prior-Based Methods: Prior-based methods aim to restore clear underwater images by modifying pixel values [34], primarily through three approaches.

1) *Direct Pixel Manipulation:* This includes techniques such as contrast adjustment [35], histogram equalization [36], and white balance [7] to enhance overall image quality by improving contrast and saturation.
2) *Weighted Fusion:* This approach involves the composite enhancement of images through the weighted fusion of various traditional UIE methods [37], [38], [39].
3) *Retinex-Based Enhancement:* Utilizes Retinex theory for image enhancement, as demonstrated in studies like [40], [41].

Despite their ability to enhance image features, these methods often perform poorly on real underwater images due to complex distortions and varying contents.

2) Physical Model-Based Methods: Physical model-based methods approach UIE as an inverse problem, involving three key steps.

1) *Establishment of Prior Conditions:* Define the assumptions of a hypothetical physical imaging model.
2) *Parameter Estimation:* Estimate essential parameters of the model.
3) *Degradation Reversal:* Reverse the degradation process inherent in underwater imaging to recover clear images.

The foundation of these methods lies in various established priors, including the underwater dark channel prior [42], attenuation curve prior [43], blur prior [44], and minimum information prior [9]. Initial research often adapted the dark channel prior for underwater conditions, with notable revisions by Akkaynak and Treibitz [45], who adjusted the atmospheric image formation model for UIE using RGBD imagery. Zhang et al. [30] proposed a UIE method that minimizes color loss and enhances contrast adaptively. Despite their potential, the efficacy of these methods is limited by their dependency on the accuracy of the physical model relative to the actual underwater scene, resulting in challenges in robustness and visual quality in diverse environments. Recent advancements have begun to combine these physical models with deep learning approaches [46] to improve performance and adaptability.

B. Deep Learning-Based Methods

Recently, deep learning has developed rapidly in UIE, e.g., [47], [48], [49], [50], [51], [52], [53], [54], [55], [56]. In terms of architectures, three main architectures are employed: CNN for hierarchical features, GAN for adversarial training, and vision transformers for modeling global context using self-attention.

1) CNN-Based Methods: Recent advancements in CNNs have significantly advanced UIE. Li et al. [1] established a UIEB dataset and trained it using gated fusion, while their Water-Net enhances images by fusing features from variably processed images based on confidence maps. Further extending these principles, Li et al. [57] trained the UWCNN with synthesized images based on underwater scene knowledge. Innovative approaches include Fu and Cao [58], who developed a global–local network with compressed histogram

equalization for improved results, and Wang et al. [13], who integrated HSV color space adjustments with deep learning in UIEC^2-Net to correct color and optimize brightness. In addition, Li et al. [46] applied transformations across RGB, HSV, and Laboratory spaces, enhancing images through channel attention and medium transfer, fused using residual learning. Sharma et al. [59] introduced WaveNet, utilizing convolutional block attention Modules and shortcut connections for detailed color channel processing. These methods highlight a trend toward complex spatial transformations and sophisticated learning strategies for superior UIE outcomes. Due to the nonuniform attenuation of colors, CNN-based models require both local and global restoration, an area where recent CNN-based methods are limited.

2) GAN-Based Methods: Recent progress in GANs has demonstrated their potential in addressing UIE challenges. GAN-based methods simulate underwater images through adversarial interactions between generators and discriminators. Li et al. [60] pioneered the use of GANs to generate paired data essential for deep learning, applying this to the color correction of underwater images, which is called Water-GAN. In addition, the conditional GAN-based model, FUnIE-GAN [15], was developed to effectively supervise adversarial training. Drawing inspiration from Cycle GAN, both Li et al. [61] and Fabbri et al. [62] adopted weakly supervised models, eliminating the need for paired training data. Xia et al. [63] introduced a hybrid optimization model that combines unsupervised and supervised learning, targeting color correction and channel refinement. To improve the visual quality of underwater images further, Guo et al. [64] developed a multiscale dense GAN, integrating multiscale processing, dense concatenation, and residual learning. Hambarde et al. [14] utilized GAN for a two-step depth estimation process at both coarse and fine levels. Cong et al. [16] introduced the PUGAN model, a physics-guided GAN for UIE that incorporates a degradation quantization module to enhance critical regions and improve the realism and esthetic quality of underwater imagery. Despite their high quantitative performance, these GAN-based approaches often overlook the explicit modeling of color information and the comprehensive integration of attention mechanisms, which are vital for effective image restoration. Modeling of underwater image color information and comprehensive consideration of different attention mechanisms are crucial for effective restoration, which remains a significant limitation.

3) Transformer-Based Methods: Recent advancements in Transformer-based UIE highlight its effectiveness in modeling global information, surpassing traditional CNNs by capturing long-range dependencies [65]. Ren et al. [66] integrated the Swin Transformer [27] with U-Net to improve global dependency detection, while also incorporating CNNs with core attention mechanisms to enhance local attention. Similarly, Wang et al. [23] developed a feature fusion Transformer that merges CNNs for comprehensive global and local feature modeling. Recognizing spatial variability in image degradation, Peng et al. [17] designed a spatially-oriented Transformer focusing on regions with severe color distortion and proposed a U-shaped architecture. To reduce the attention parameter complexity, the adaptive group attention mechanism [19] has been introduced and integrated within the Swin Transformer for dynamically selecting visually complementary channels based on dependencies. Despite its advancements, the method faces challenges in restoring significant features in high-frequency and highly blurred areas, impeding optimal enhancement results. To overcome these deficiencies, UDAformer [67] proposes a novel strategy that modeling local pixel relationship for color channel feature extraction. Despite their strengths, the high computational demands of Transformers, particularly due to self-attention, pose challenges at increased image resolutions. Besides, a few Transformer-based methods explicitly model color information and restore color distortion from both local and global views.

III. Proposed Method

The proposed UIE model, named BCTA-Net, employs a local and global bi-level color adjustment approach with three main branches: 1) LCR for adaptive nonuniform local color correction; 2) TA for content-aware local distortion restoration; and 3) global light recovery (GLR) for alleviating background light color bias. These branches synergistically address local and global distortions in underwater images. The model architecture is depicted in Fig. 4, and its specific details are elaborated below.

A. LCR Branch

The LCR branch focuses on correcting localized color distortions using a color-based attention branch. This branch $f_{\mathrm{LCR}}(\cdot)$ mainly employs histogram analysis and a self-attention mechanism to output a color feature map, ultimately achieving adaptive color correction. In this context, histogram analysis is conducted to extract color information from images and integrate it into the network. The self-attention mechanism is utilized to enhance critical color information and adjust the color distribution, such as adaptively transforming the final color distribution from being narrow and abrupt to broad and uniform, ultimately allowing the network to generate a satisfying color feature map. The following provides a detailed introduction to histogram calculation and color-based attention mechanisms.

1) Histogram Calculation: Given the input image $\mathcal{I}$, the histogram calculation procedure can be formulated as

$$\boldsymbol{h}_c(i) = \sum_{x,y} \delta(\mathcal{I}_c(x, y) - i) \tag{1}$$

where $\boldsymbol{h}_c(i)$ is the histogram for color channel c, and $\mathcal{I}_c(x, y)$ is the pixel value at (x, y) for channel c, δ is indication function. The histogram calculation is performed for each color channel (R, G, B) to understand the distribution of color intensities. Then, the histogram $\boldsymbol{h} = \mathcal{H}(\mathcal{I}) = [\boldsymbol{h}_{\mathrm{R}}, \boldsymbol{h}_{\mathrm{G}}, \boldsymbol{h}_{\mathrm{B}}]$ is transferred to a linear block followed by a convolution layer to obtain the color-related feature $\boldsymbol{C}$. The linear block (L-Block) is designed for dimension alignment and is illustrated in Fig. 5. In the L-Block, we transfer the color histogram from a size of 1×768 to a size of $C \times H \times W$, allowing $\boldsymbol{C} \in \mathbb{R}^{C \times HW}$ can be multiplied with $\boldsymbol{K}$ and the color feature map output by the

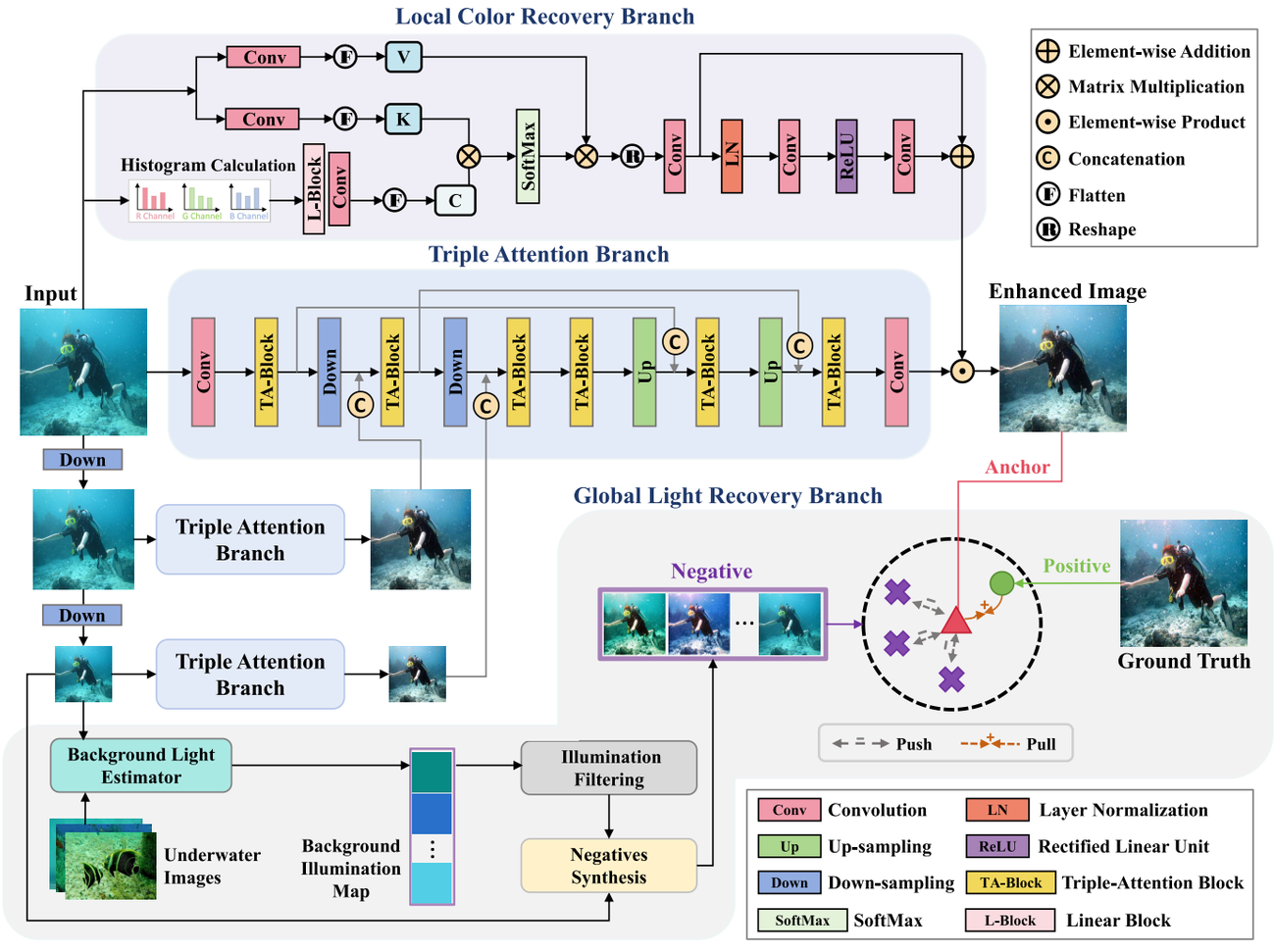


Fig. 4. Framework of proposed BCTA-Net for UIE.

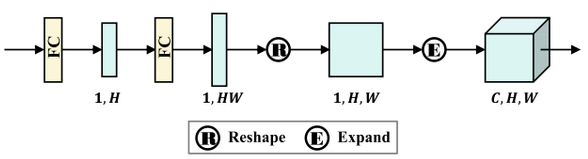


Fig. 5. Diagram of linear block. FC denotes the fully connected layer. The blue rectangle and blue cube represent the feature.

LCR Branch to be element-wise multiplied with the feature maps from the TA block.

2) Color-Based Attention Mechanism: The self-attention mechanism enables adaptive correction by

$$\text{Attention}(\boldsymbol{C}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{C}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{2}$$

where $\boldsymbol{K} \in \mathbb{R}^{C\times HW}$ and $\boldsymbol{V} \in \mathbb{R}^{C\times HW}$ are obtained by transforming input image $\mathcal{I}$ through a convolution layer. d_k is the dimension of the key vectors. Then, a series of deep learning layers are applied to the output of the attention operation, producing adaptive color correction maps $\boldsymbol{O}_{\text{LCR}} = f_{\text{LCR}}(\mathcal{I}) \in \mathbb{R}^{C\times H\times W}$. These maps aim to provide accurate, color-guided weights that enhance the model's capability for UIE.

B. TA Branch

The TA branch $f_{\text{TA}}(\cdot)$ integrates TA-blocks in a U-shaped manner with skip-connections, forming a hierarchical structure that enhances crucial image regions and contextual dependencies through a series of downsampling and upsampling operations. By progressively applying TA-blocks and downsampling the input to obtain high-level semantic features, followed by upsampling to recover fine-grained details, it ensures both global and local feature refinement for images. In addition, we obtain two inputs at different scales through two additional downsamplings and feed these two inputs into two separate TA branches, forming a multiscale fusion structure with a total of three TA branches, as shown in Fig. 4. In this structure, the output of the second TA branch is fused with the output features after the first downsampling in the top TA branch, and the output of the third TA branch is fused with the output features after the second downsampling in the top TA branch. This approach captures features at different resolutions, which is crucial for dealing with objects and details at varying scales in underwater images. The output $\boldsymbol{O}_{\text{TA}} = f_{\text{TA}}(\mathcal{I}) \in \mathbb{R}^{C\times H\times W}$ is multiplied with the features of the same size from the LCR branch in an element-wise manner.

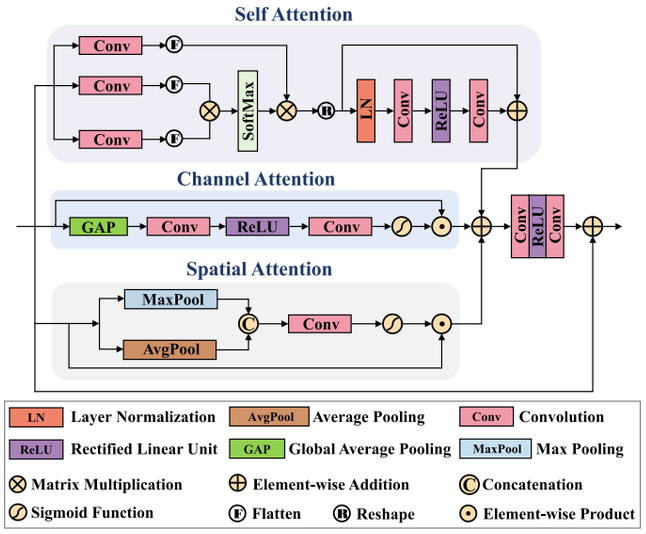


Fig. 6. Diagram of TA block.

The TA-Block is a content-aware module designed to enhance features in underwater images by integrating three attention mechanisms: spatial attention, channel attention, and self-attention. The TA-Block architecture is illustrated in Fig. 6.

1) Spatial Attention: The spatial attention module identifies and enhances important spatial regions, crucial for UIE due to uneven lighting and visibility conditions in underwater images. By focusing on significant spatial regions, the TA-branch highlights important features like marine organisms and underwater structures. Given an input feature map $\boldsymbol{F} \in \mathbb{R}^{C\times H\times W}$, the process is as follows:

$$\begin{aligned}\boldsymbol{M}_s(\boldsymbol{F}) &= \sigma\left(f^{7\times 7}\left(\left[\text{AvgPool}(\boldsymbol{F}); \text{MaxPool}(\boldsymbol{F})\right]\right)\right)\\ &= \sigma\left(f^{7\times 7}\left(\left[\boldsymbol{F}^s_{\text{avg}}; \boldsymbol{F}^s_{\max}\right]\right)\right)\end{aligned} \tag{3}$$

where σ denotes the sigmoid function and $f^{7\times 7}$ represents a convolution operation with the filter size of 7×7. AvgPool and MaxPool are two pooling operations across the channel. The spatially attended feature map $\boldsymbol{F}_s$ is then obtained by element-wise multiplication of $\boldsymbol{M}_s \in \mathbb{R}^{1\times H\times W}$ and each channel map of $\boldsymbol{F}$.

2) Channel Attention: The channel attention is introduced to emphasize the most informative channels, balancing color channels and enhancing key spectral features, essential for correcting color distortions caused by water absorption and scattering. Given the input feature map $\boldsymbol{F}$, the channel attention vector $\boldsymbol{M}_c \in \mathbb{R}^C$ is computed as

$$\boldsymbol{M}_c = \sigma\left(f^{1\times 1}_{\text{ch}}\left(\delta\left(f^{1\times 1}_{\text{ch}}(\text{GAP}(\boldsymbol{F}))\right)\right)\right) \tag{4}$$

where GAP denotes global average pooling, and $f^{(1\times 1)}_{\text{ch}}$ is a convolution operation with a 1×1 filter, and σ and δ denote the sigmoid and ReLU activation functions, respectively. GAP is global average pooling which takes the channel-wise global spatial information into a channel descriptor

$$\text{GAP}(\boldsymbol{F}) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{F}_c(i,j) \tag{5}$$

where $\boldsymbol{F}_c\ (i, j)$ stands for the value of cth channel X_c at position(i, j). The channel-attended feature map $\boldsymbol{F}_c$ is obtained by element-wise multiplication of $\boldsymbol{F}$ and the replication of $\boldsymbol{M}_c$ on the spatial dimension.

3) Self-Attention: The self-attention module captures long-range dependencies, allowing the network to understand the global context. This ensures coherence across the entire scene, making enhancements contextually consistent throughout the underwater image. Given $\boldsymbol{F}$, the self-attention map $\boldsymbol{M}_{\text{self}} \in \mathbb{R}^{C\times C}$ is computed as

$$\boldsymbol{M}_{\text{self}} = \text{Softmax}\left(\boldsymbol{Q}_s\boldsymbol{K}_s^{\text{T}}\right) \tag{6}$$

where $\boldsymbol{Q}_s = f_Q(\boldsymbol{F}) \in \mathbb{R}^{C\times HW}$ and $\boldsymbol{K}_s = f_K(\boldsymbol{F}) \in \mathbb{R}^{C\times HW}$ are the query and key projections, respectively, and f_Q, f_K, and f_V are transformations, which include one convolution followed by a flatten operation, enabling the branch to ultimately obtain features with dimensions of $\mathbb{R}^{C\times C}$. The self-attended feature map $\boldsymbol{F}_{\text{self}}$ is obtained by applying the self-attention mask $\boldsymbol{M}_{\text{self}}$ to the value projection $\boldsymbol{V}_s = f_V(\boldsymbol{F}) \in \mathbb{R}^{C\times HW}$, followed by additional deep learning layers for further refinement.

4) Integration: The final output of the TA-Block is obtained by effectively combines the spatially attended, channel-attended, and self-attended features. By using concatenation followed by two convolutional layers and one ReLU layer, the block can reduce the channel dimension and learn a fused representation that enhances the overall feature quality.

C. GLR Branch

The GLR branch employs contrastive learning to adjust the global illumination characteristics of underwater images. By processing pairs of images with simulated different lighting conditions, the network learns robust features to approximate the true solution space of background illumination and apply corrective global adjustments. The GLR Branch is primarily designed to extract the global ambient light map from the dataset images and replace the input global ambient light map based on the underwater image formation model, creating more negative samples with varying global ambient light maps. It also employs the designed Contrastive Learning to make the enhanced images close to the ground truth and as far away from the negative samples as possible. The designed background light estimation is mainly used for extracting the global ambient light map, while the purpose of Illumination Filtering is to prevent the replaced global ambient light map from being similar to the ground truth global ambient light map. Ultimately, this enables the network to generate enhanced images with satisfactory global illumination characteristics. Detailed introductions to background light estimation, illumination filtering, negative synthesis, and contrastive learning are as follows.

1) Background Light Estimation: The underwater image formation model [20] describes how light is absorbed and scattered as it travels through water, impacting the color and clarity of captured images. The background light estimation is based on this model, which is expressed as

$$\boldsymbol{I}(x) = \boldsymbol{J}(x)\boldsymbol{t}(x) + \boldsymbol{A}(\boldsymbol{1} - \boldsymbol{t}(x)) \tag{7}$$

where x is the pixel index, $\mathbf{1}$ is a matrix with the proper size of all 1 s, $\boldsymbol{I}(x)$ is the inherent scene radiance at pixel x, representing the true colors and intensity of the scene as it would appear in clear water or in air. $\boldsymbol{A}$ is the global ambient light map with all a, which accounts for the background light or veiling light in the water. This light is usually scattered light from the water itself and other particulate matters. $\boldsymbol{t}(x)$ is the medium transmission at pixel x, which quantifies the amount of light that reaches the camera without being scattered away. It is influenced by the distance light travels through water and the water's properties. By transformation, we can get

$$\boldsymbol{A} = (\boldsymbol{I}(x) - \boldsymbol{J}(x)\boldsymbol{t}(x))/(\mathbf{1} - \boldsymbol{t}(x)). \tag{8}$$

2) Illumination Filtering: For the ith distorted underwater image $\mathcal{I}_i$ and its ground truth $\mathcal{I}'_i$, the global background light $\boldsymbol{A}'_i$ of $\mathcal{I}'_i$ can be computed using (8). To generate negatives that significantly differ from the ground truth in terms of background light, the background lights for available data are initially calculated and stored in a bank $\mathcal{L}^i = [\boldsymbol{L}^i_1, \ldots, \boldsymbol{L}^i_N]$ for $\mathcal{I}_i$. From this bank, illumination filtering is applied to effectively construct the set of negatives $\mathcal{B}^i = [\boldsymbol{B}^i_1, \ldots, \boldsymbol{B}^i_M]$ for $\mathcal{I}_i$.

For illumination filtering, the Euclidean distance d_n is calculated between each $\boldsymbol{L}^i_n$ and $\boldsymbol{A}'_i$. Then, $\boldsymbol{L}^i_n$ whose l_n is more than the threshold λ and others will be eliminated for effective negative samples. The calculation process for d_n is as follows:

$$d_n = \sqrt{(r_n - r)^2 + (g_n - g)^2 + (b_n - b)^2} \tag{9}$$

where r_n, g_n, and b_n denote three rgb dimension of same b_m values in $\boldsymbol{L}^i_n$, r, g, and b represent three rgb dimension of same a' of $\boldsymbol{A}'_i$.

3) Negatives Synthesis: Given ith underwater distorted image $\mathcal{I}_i$, the scene radiance map I_i and medium transmission map t_i can be calculated based on (8). Then, M negatives for $\mathcal{I}_i$ can be calculates as

$$\boldsymbol{Z}^i_m(x) = \boldsymbol{J}(x)\boldsymbol{t}(x) + \boldsymbol{B}^i_m(\mathbf{1} - \boldsymbol{t}(x)) \quad \forall m \in \{1, \ldots, M\}. \tag{10}$$

4) Contrastive Learning: We use the ground-truth $\tilde{\mathcal{I}}$ and abnormal-light images $\boldsymbol{Z}$ as the positive samples and negative samples to restore the global light of underwater images, respectively. The goal of contrastive learning is to learn a representation to pull together "positive" pairs in the latent feature space and push apart the representation between "negative" pairs. In our method, positive pairs are formed by ground-truth, negative pairs are created from distorted, newly added negative samples, and an enhanced image.

D. Loss Function

Given a set of images $\{\mathcal{I}_i\}_{i=1}^N$ and their ground truths $\{\tilde{\mathcal{I}}_i\}_{i=1}^N$. Let $\{\hat{\mathcal{I}}_i\}_{i=1}^N$ denote the objective images predicted by our model. The training loss consists of three components. First, L_{hist} is designed to correct the potential color deviations in the restored image for LCR and TA branches by comparing histograms by

$$L_{\text{hist}} = \sum_{i=1}^{N} \left(|\mathcal{H}(\hat{\mathcal{I}}_i) - \mathcal{H}(\tilde{\mathcal{I}}_i)|_1 \right) \tag{11}$$

where $\mathcal{H}(\cdot)$ is histogram calculation operation in the LCR branch. To better preserve the sharpness of edges and details in enhanced images and follow [47], [68], we use the SSIM [69] to minimize the structure error of the enhancements, and combined it with L_1 function, denoted as L_{recons}. The contrastive learning loss $L_{\text{cl}} = \sum_{i=1}^{N} d_i$, where d_i is defined as

$$\left\{ \left| \psi(\hat{\mathcal{I}}_i) - \psi(\tilde{\mathcal{I}}_i) \right| - \left(\sum_{m=1}^{M} \left| \psi(\hat{\mathcal{I}}_i) - \psi(\boldsymbol{Z}^i_m) \right| \right) \Big/ M + \alpha, 0 \right\} \tag{12}$$

where α is the hyper-parameter to adjust the margin in the triplet loss, and it is set to 0.04 experimentally. $\psi(\cdot)$ is VGG-16 network [70] pretrained on the ImageNet dataset, used to extract features in latent perceptual feature space of positive-anchor and negative-anchor pairs for contrasting. L_{cl} aims to guarantee that the distance between features of restored images and GT is smaller than the distance between features of restored images ours and negatives. Then, the total loss function used to supervise the generated image, is expressed as

$$\mathcal{L} = \lambda_1 L_{\text{hist}} + \lambda_2 L_{\text{recons}} + \lambda_3 L_{\text{cl}} \tag{13}$$

where λ_1, λ_2, and λ_3 are three hyper-parameters for balancing three losses.

IV. Experiments

A. Experimental Settings

1) Datasets: To comprehensively evaluate the performance of the proposed method, we conducted experiments on several widely adopted benchmark datasets for UIE, including real-world datasets (UIEB [71], LSUI [17], MABLS [32], U45 [72], and SUIM [73]) and a synthetic dataset (EUVP [15]). Specifically, EUVP was synthesized using a CycleGAN-based model. UIEB and LSUI datasets have reference images, which are selected from the results generated by dive+[1] and fusion-based [39]. Another three real-world underwater datasets MABLS, U45, and SUIM are real-world datasets without reference images. For convenience, we summarize the detailed information of the above datasets in the following Table I. In our experiments, we test the EUVP test set using weights trained on the EUVP training set, the LSUI test set using weights trained on the LSUI training set, and the UIEB test set as well as the three reference-free datasets (MABLS, U45, and SUIM) using weights trained on the UIEB training set.

2) Methods for Comparisons: To validate the effectiveness and generalization ability of the proposed network and its components, we conducted comparisons with several outstanding UIE methods. The compared methods include traditional methods (i.e., GDCP [74], HLRP [75], MLLE [30], and UNTV [76]) and recent deep learning-based approaches (i.e., Water-Net [77], FUnIE-GAN [15], UWnet [78],

TABLE I

SUMMARY OF THE TRAINING/TEST IMAGE NUMBER AND UNDERWATER IMAGE TYPES OF SIX DIFFERENT DATASETS USED IN THIS WORK

Dataset	Image Num.		Underwater Image Types
	Training	Test	
EUVP	11435	515	synthetic [15]
UIEB	800	90	real-world [71]
LSUI	4500	504	real-world [17]
MABLS	–	319	real-world [32]
U45	–	45	real-world [72]
SUIM	–	110	real-world [73]

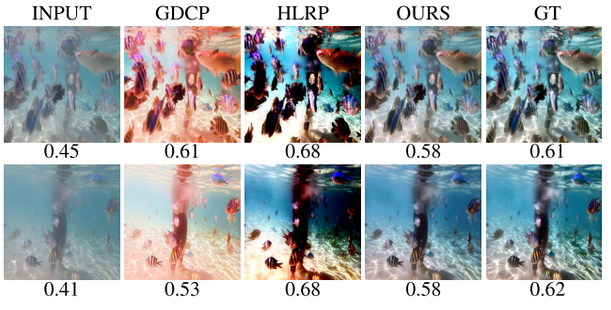


Fig. 7. Two sets of visualization results and UCIQE [86] scores for different methods on the UIEB test set.

USUIR [49], PUIE-Net [79], GUPDM [51], SGUIE [50], U-SHAPE [17], PUGAN [16], EAU-GAN [80], DAE-GAN [81], ULD-CycleGAN [82], LUUW-GAN [83], and CECF [52]). Whenever applicable, we quote the results of these methods from existing literature; otherwise, we retrain them using the same data as ours. A dash (-) in the table signifies that results are unavailable for comparison with a particular method on a dataset. In addition, methods having no results for a specific dataset will be excluded from the relevant tables and figures.

3) Evaluation Metrics: Following [69], we adopt both reference-dependent and nonreference evaluation metrics to comprehensively assess the performance of our model. For reference-dependent metrics, we select peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The results of full-reference image quality evaluation using the reference images can provide realistic feedback on the performance of different methods to some extent, although the real ground truth images might differ from the reference images. A higher PSNR score denotes that the result is closer to the reference image in terms of image content, while a higher SSIM score means the result is more similar to the reference image in terms of image structure and texture. As for the nonreference metrics, we employ underwater image quality measure (UIQM [84]), which reflects the visual quality of the image from a human perspective. A higher UIQM score indicates that the result is more consistent with human visual perception. We also adopt the natural image quality evaluator (ILNIQE [85]), which evaluates the naturalness of generated images. A lower ILNIQE score indicates that the image is more natural. The UCIQE [86] is not considered due to the observation shown in Fig. 7. It can be observed that restored images of GDCP [74] with severe color distortion have excellent scores, GTs with better visual performance have worse scores than results restored by HLRP [75].

TABLE II

QUANTITATIVE COMPARISON IN FULL-REFERENCE METRICS ON THE REAL-WORLD DATASET: UIEB, AND THE SYNTHETIC DATASET: EUVP. THE RED AND BLUE DIGITS INDICATE THE BEST, AND THE SECOND BEST RESULTS, RESPECTIVELY

Method	Source	UIEB	EUVP
		PSNR(dB) / SSIM	PSNR(dB) / SSIM
GDCP [74]	TIP2018	13.72 / 0.83	13.35 / 0.71
HLRP [75]	TIP2022	12.17 / 0.32	11.32 / 0.64
MLLE [30]	TIP2022	18.82 / 0.74	15.06 / 0.78
UNTV [76]	TCSVT2022	16.57 / 0.79	17.50 / 0.66
Water-Net [77]	TIP2019	19.81 / 0.82	20.58 / 0.79
FUnIE-GAN [15]	RAL2020	18.07 / 0.83	23.53 / 0.86
UWnet [78]	AAAI2021	15.99 / 0.70	22.65 / 0.80
USUIR [49]	AAAI2022	20.31 / 0.84	18.05 / 0.82
PUIE-Net [79]	ECCV2022	21.19 / 0.88	19.01 / 0.78
GUPDM [51]	MM2023	21.53 / 0.86	25.48 / 0.86
SGUIE [50]	TIP2022	21.76 / 0.89	24.80 / 0.84
U-SHAPE [17]	TIP2023	20.46 / 0.79	25.80 / **0.88**
PUGAN [16]	TIP2023	21.67 / **0.90**	24.05 / 0.87
EAU-GAN [80]	CCDC2024	21.36 / 0.87	– / –
DAE-GAN [81]	MDPI2024	– / –	**26.33** / 0.78
ULD-CycleGAN [82]	JOE2024	**22.32** / 0.62	– / –
LUUW-GAN [83]	ICWOC2024	– / –	24.76 / 0.84
CECF [52]	AAAI2024	21.82 / 0.89	25.16 / 0.86
BCTA-Net	Proposed	**22.44** / **0.93**	**26.95** / **0.90**

TABLE III

QUANTITATIVE COMPARISON IN FULL-REFERENCE METRICS ON THE REAL-WORLD DATASET: LSUI. THE RED AND BLUE DIGITS INDICATE THE BEST, AND THE SECOND BEST RESULTS, RESPECTIVELY

Method	Source	LSUI	
		PSNR(dB)↑	SSIM↑
GDCP [74]	TIP2018	13.31	0.78
HLRP [75]	TIP2022	13.84	0.28
MLLE [30]	TIP2022	18.12	0.68
UNTV [76]	TCSVT2022	17.35	0.61
Water-Net [77]	TIP2019	17.73	0.82
UWnet [78]	AAAI2021	18.32	0.80
USUIR [49]	AAAI2022	19.35	0.85
PUIE-Net [79]	ECCV2022	19.61	0.86
SGUIE [50]	TIP2022	22.32	0.83
PUGAN [16]	TIP2023	**24.42**	**0.88**
BCTA-Net	Proposed	**24.61**	**0.92**

4) Implementation Details: Our BCTA-Net for UIE is implemented in PyTorch [87] and trained on an NVIDIA RTX 3090 GPU. Training configuration includes: batch size of 8, AdamW optimizer with initial learning rate $1e^{-4}$ (decreased on plateau), 300 epochs, and the number of negative samples and the threshold λ in illumination filtering is set to 20 and 26, respectively. The loss function weights are set to $\lambda_1 = 0.02$, $\lambda_2 = 0.96$, and $\lambda_3 = 0.02$.

B. Comparison With State-of-the-Art Methods

1) Quantitative Comparison:

a) In terms of full-reference metrics: Table II summarizes the PSNR and SSIM results on the UIEB and EUVP datasets, while Table III summarizes the PSNR and SSIM results on the LSUI dataset; these three datasets provide ground truth images for quantitative evaluation. The performance of traditional methods is generally inferior to deep learning-based techniques, owing to the limited

TABLE IV

QUANTITATIVE COMPARISON IN NONREFERENCE METRICS ON FIVE DATASETS. THE RED AND BLUE DIGITS INDICATE THE BEST, AND THE SECOND BEST RESULTS AMONG DEEP LEARNING-BASED METHODS, RESPECTIVELY

Method	Source	UIEB		EUVP		MABLS		U45		SUIM	
		UIQM↑	ILNIQE↓	UIQM↑	ILNIQE↓	UIQM↑	ILNIQE↓	UIQM↑	ILNIQE↓	UIQM↑	ILNIQE↓
GDCP [74]	TIP2018	2.67	33.27	2.43	64.82	2.38	57.91	2.41	53.45	2.22	51.81
HLRP [75]	TIP2022	1.99	49.33	2.41	60.93	2.73	57.95	1.86	52.16	2.30	59.09
MLLE [30]	TIP2022	2.65	42.91	2.28	51.38	2.32	59.43	2.26	58.78	2.04	57.05
UNTV [76]	TCSVT2022	2.94	39.09	2.47	51.41	2.40	59.89	2.97	54.23	2.25	58.05
Water-Net [77]	TIP2019	2.82	**41.80**	2.77	63.95	3.02	**52.13**	3.17	47.78	–	–
UWnet [78]	AAAI2021	2.68	67.45	2.92	72.93	2.62	74.00	2.73	65.62	2.43	66.11
USUIR [49]	AAAI2022	2.87	44.95	2.38	64.83	2.92	54.29	**3.38**	46.42	2.65	**45.36**
PUIE-Net [79]	ECCV2022	2.68	47.68	**2.98**	67.54	2.86	55.24	3.25	47.24	2.66	53.99
U-SHAPE [17]	TIP2023	2.95	41.87	2.75	63.40	2.89	52.95	3.06	49.40	–	–
PUGAN [16]	TIP2023	**3.28**	44.02	2.94	**61.09**	**3.03**	54.27	3.27	**46.09**	**2.68**	48.95
BCTA-Net	Proposed	**3.12**	**41.73**	**3.00**	**58.61**	**3.05**	**50.82**	**3.36**	**45.99**	**2.73**	**46.24**

representational capability of handcrafted features. Among the deep learning approaches, our proposed method achieves the best PSNR and SSIM scores on both the UIEB and EUVP datasets, outperforming the second-best methods by 0.5% and 2.4% in PSNR on UIEB and EUVP, respectively, and by 3.3% and 2.3% in SSIM. It achieves the best PSNR and SSIM scores on the LSUI dataset. The significant performance gains demonstrate the strength of our approach in effectively handling the challenges of underwater image degradation.

b) In terms of no-reference metrics: Table IV presents the quantitative results of different methods on the five datasets as assessed by the UIQM and ILNIQE metrics. Our proposed BCTA-Net achieves the best UIQM scores on the EUVP, MABLS, and SUIM datasets, and second-best on UIEB and U45, demonstrating its superiority in enhancing key attributes like colorfulness, sharpness, and contrast. In addition, BCTA-Net ranks first among learning-based approaches on UIEB, EUVP, MABLS, and U45 in terms of the ILNIQE metric, which measures the perceptual quality of natural images. It ranks second on SUIM. These comprehensive results validate that our method reliably outperforms existing techniques, as quantified by both UIQM and ILNIQE metrics. The consistent top performance of BCTA-Net across diverse datasets affirms its effectiveness in holistically enhancing myriad aspects of underwater image quality.

2) Qualitative Comparisons: As shown in the visual comparison of background light estimation in Fig. 8, the low-light enhancement results in Fig. 9, and the heatmap visualization comparison in Fig. 10, BCTA-Net demonstrates superior enhancement effects. The qualitative comparison results in Figs. 11 and 12 also indicate that BCTA-Net outperforms other methods in terms of visual quality.

a) Color distortion correction: Color distortion, a prevalent underwater image degradation, stems from the wavelength-dependent absorption of light in water, significantly impacting visual appeal. Effective enhancement necessitates prioritizing color correction. First, our BCTA-Net plays a significant role in background light correction. As shown in Fig. 8, we can effectively correct the different color casts caused by background light in various environments. Second, Our BCTA-Net demonstrates superior performance in producing natural, realistic colors,

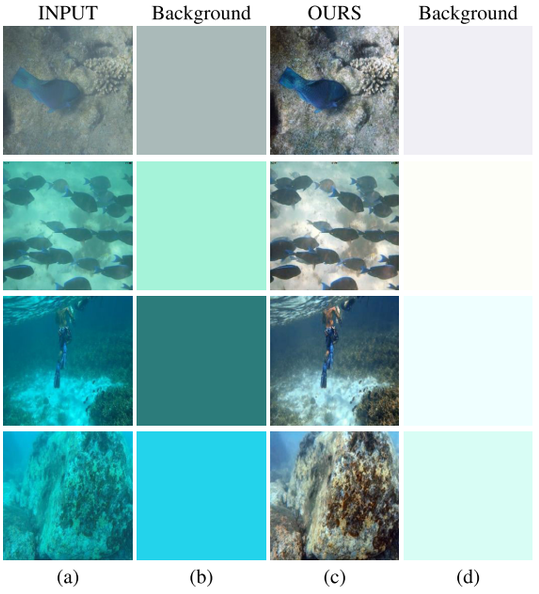


Fig. 8. Visual comparison of background light estimation on the MABLS dataset (a) input images, (b) input background maps, (c) enhanced outputs, and (d) enhanced background maps.

as evidenced by visual comparisons using randomly sampled images from reference datasets (UIEB and EUVP in Fig. 11) and reference-free datasets (MABLS, U45, and SUIM in Fig. 12). For quantitative validation, we have added feature activation maps in Fig. 11 showing how our color-aware attention mechanism identifies and processes different color components. For example, in Fig. 11 (images in the fourth row), while USUIR, PUGAN, and UWnet leave residual greenish distortion, our BCTA-Net produces more natural colors. Similarly, for bluish distortion correction (see Fig. 11, images in the second row), although PUIE-Net and PUGAN show improvements, our method achieves superior color recovery in critical regions, as demonstrated by the progressive color correction visualization in Fig. 11. The feature maps clearly show how our content-aware approach adaptively handles diverse color distortions, restoring more natural and vivid colors compared to current techniques.

b) Low light enhancement: Forward scattering often induces low contrast and insufficient lighting in underwater images, requiring enhancement methods to increase color con-

Fig. 9. Low-light enhancement results across multiple underwater datasets. Original low-light scenes (top row) and their enhanced results (bottom row) from MABLS (first column), U45 (second column), and UIEB (third and fourth columns) datasets.

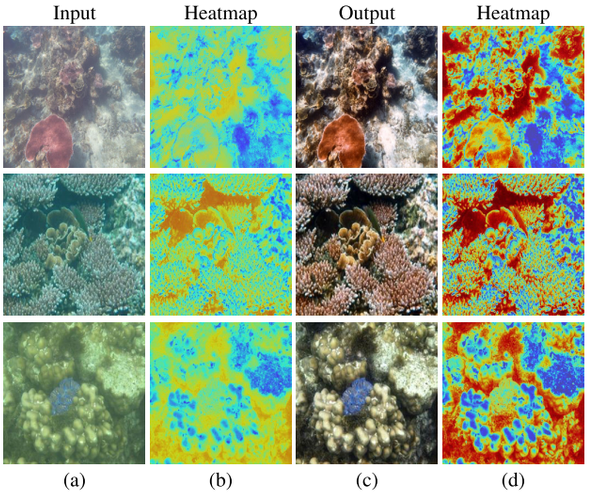


Fig. 10. Heatmap visualization comparison. (a) Input images with (b) their corresponding heatmaps compared to (c) our enhanced outputs and (d) their heatmaps. Results are shown from U45 (first row) and MABLS (second and third rows) datasets.

trast while removing blur. As shown in Fig. 9, our method can effectively enhance brightness in low-light scenes and achieve satisfactory restoration results. As shown in Fig. 12, deep learning methods may fail to sufficiently improve brightness without color distortions, like UWnet and USUIR, respectively. Our approach effectively handles both low-contrast and low-light images without introducing distortions. It excels in detail recovery, as evidenced across images in Fig. 12. The results validate our method's strength in adaptively increasing brightness and contrast to restore clarity, while avoiding artifacts such as over-enhancement, color deviations, and loss of details that affect current techniques. Our content-aware approach reliably enhances low-contrast, low-light underwater images.

c) Complex scenes restoration: In the complex, coral-rich scene of Fig. 11 (fifth image), BCTA-Net demonstrates superior performance. Despite the challenging restoration task posed by numerous corals with diverse colors and depths, our method excels where others falter. PUIE-Net inaccurately renders coral colors. UWnet and USUIR struggle with restoring red coral hues. Although PUGAN performs reasonably well, it lacks clarity in detailed areas. In contrast, BCTA-Net achieves results closest to the ground truth, accurately

TABLE V

QUANTITATIVE RESULTS OF THE ABLATION STUDY IN TERMS OF AVERAGE PSNR AND SSIM VALUES ON THE UIEB AND EUVP DATASETS

Setting	Detail	UIEB	EUVP
		PSNR (dB) / SSIM	PSNR (dB) / SSIM
Full model	-	22.44 / 0.93	26.95 / 0.90
No.1	w/o LCR	21.72 / 0.89	25.86 / 0.86
No.2	w/o TA	20.89 / 0.85	25.38 / 0.81
No.3	w/o SA & SelfA	21.52 / 0.88	26.27 / 0.86
No.4	w/o CA & SA	21.78 / 0.87	26.33 / 0.84
No.5	w/o CA & SelfA	21.60 / 0.87	26.34 / 0.85
No.6	w/o GCR	22.14 / 0.91	26.08 / 0.89
No.7	w/o L_{hist}	22.21 / 0.90	26.63 / 0.89
No.8	w/o MS	22.30 / 0.90	26.58 / 0.89

adjusting colors and preserving definition for both corals and background elements. In the heatmap of Fig. 10, we can observe that the heatmap of input has minimal color variation and the edge information is not well-defined. In contrast, the output result from our method displays a heatmap with noticeable changes in different regions of the complex scene, and the edge information is much more pronounced.

C. Ablation Study

We conducted extensive ablation studies to validate the effectiveness of different modules in our network. Based on the structure of our model, we divided the experimental validations into four parts: ablation of the LCR branch, ablation of the TA branch, ablation of the GCR branch, and ablation of the loss function, including a total of seven designs.

1) *No.1:* "w/o LCR" means the LCR branch is removed.
2) *No.2:* "w/o TA" means the TA branch is removed.
3) *No.3:* "w/o SA & SelfA" means the spatial attention and the self-attention in TA-Block are removed.
4) *No.4:* "w/o CA & SA" means the channel attention and the spatial attention in TA-Block are removed.
5) *No.5:* "w/o CA & SelfA" means the channel attention and the self-attention in TA-Block are removed.
6) *No.6:* "w/o GCR" means the GCR branch is removed.
7) *No.7:* "w/o L_{hist}" means the loss function L_{hist} is removed.
8) *No.8:* "w/o MS" means the multiscale inputs are removed.

As shown in Table V, the full model outperforms other ablation models in terms of PSNR and SSIM scores. We also show the visual results of the ablation experiments in Fig. 13.

a) Effectiveness of LCR branch: The LCR branch integrates color histograms for precise color information extraction and employs adaptive weights to address varying color distributions in underwater images. Experimental results in Table V demonstrate that removing LCR leads to performance degradation across diverse underwater datasets, underscoring its effectiveness in color correction. Visual comparisons in Fig. 13 further illustrate LCR's impact: without it, significant color distortions are evident, such as on a turtle's back (first row, second column), an overall light blue tint (second row, second column), and a blue-green bias affecting a shark's grayish-white coloration

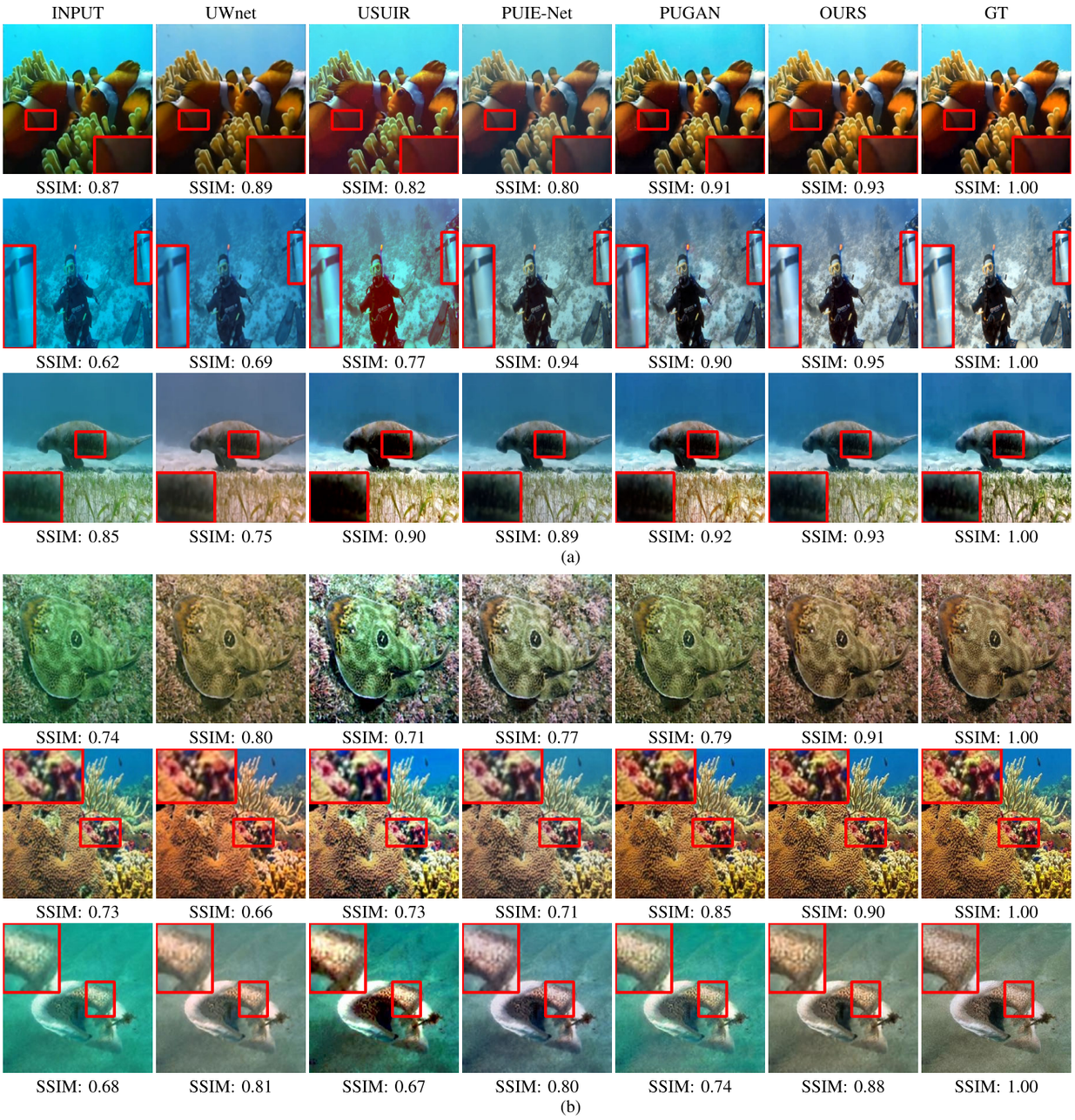


Fig. 11. Visual results of different methods on the real-world dataset with references: UIEB (first three rows) and synthetic dataset: EUVP (last three rows). (a) Visual comparisons on the UIEB dataset. (b) Visual comparisons on EUVP dataset.

(third row, second column). In contrast, our full model incorporating LCR achieves well-restored colors across these challenging scenarios, as shown in the tenth column of Fig. 13, highlighting its ability to adaptively correct nonuniform color distortions in complex underwater environments.

b) Effectiveness of TA branch: The TA branch enhances focus on critical regions, essential channels, and global information with long-range dependencies, improving detailed texture restoration and reducing blurriness. Experimental results in Table V show that removing either the entire TA or individual components degrades performance across various underwater datasets, confirming its efficacy. Visual comparisons in Fig. 13 further illustrate TA's impact: without TA (third column), texture on the turtle's back blurs and artifacts appear on the shark's back; partial TA removal (columns 4–6) results in unsatisfactory color and texture on the turtle, reduced shark clarity, and artifacts (e.g., sixth column, second row). In contrast, the full model with TA achieves satisfactory restoration of texture and color, improved clarity, and reduced artifacts, highlighting TA's crucial role in restoring spatially varying local details and enhancing texture precision in complex underwater scenes.

c) Effectiveness of GCR branch: The GCR branch leverages underwater image formation principles to enhance negative samples and mitigate the influence of ambient light. Experimental results in Table V demonstrate that removing

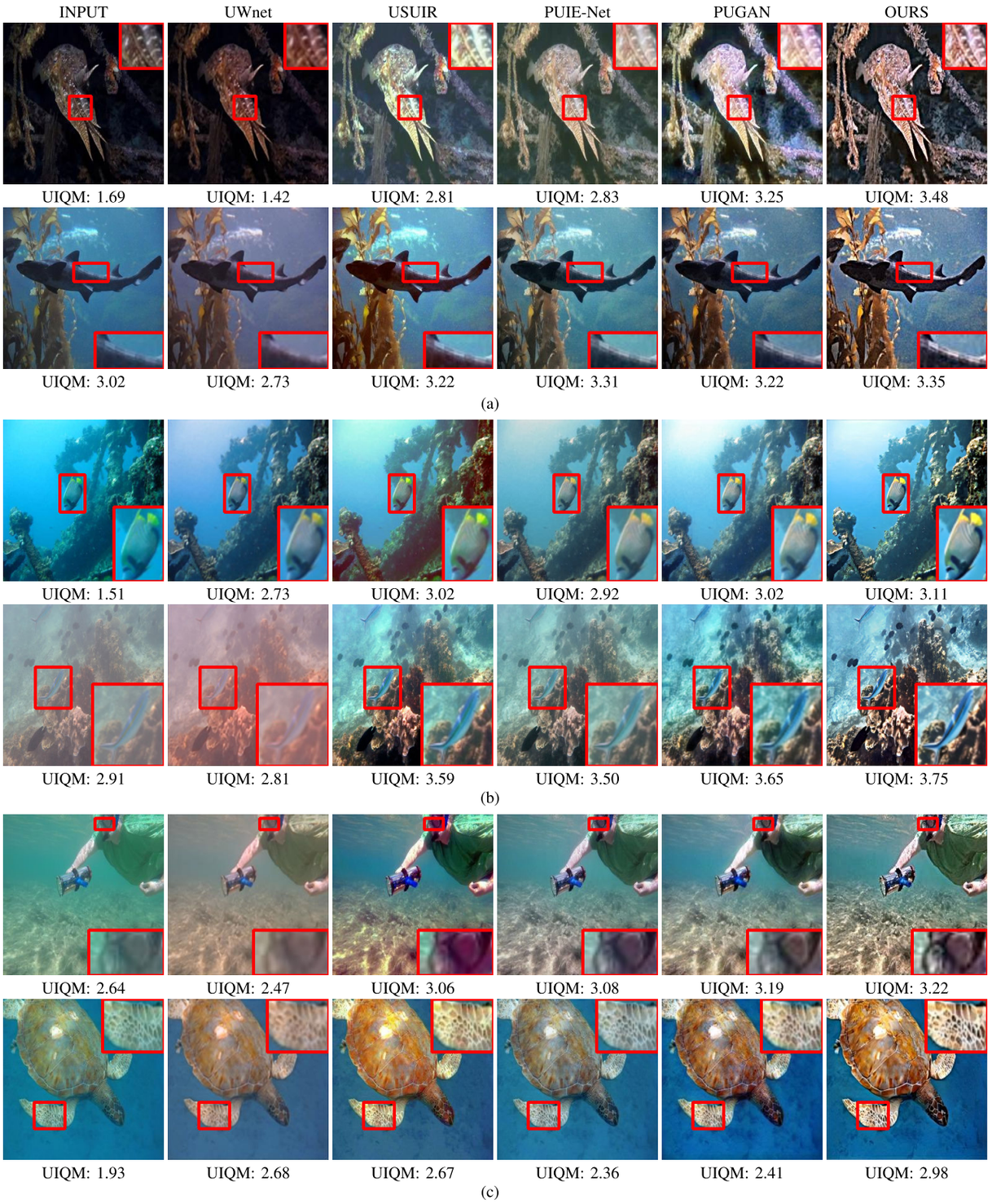


Fig. 12. Visual results and UIQM values of different methods on real-world datasets: MABLS (first two rows), U45 (from the third to the fourth row), and SUIM (last two rows), where a higher UIQM value is better. (a) Visual comparisons on the MABLS dataset. (b) Visual comparisons on the U45 dataset. (c) Visual comparisons on SUIM dataset.

GCR leads to performance degradation across various underwater datasets, confirming its effectiveness in ambient light correction. Visual comparisons in Fig. 13 illustrate GCR's impact: without it (seventh column), images exhibit an overall light blue hue (second row) and a slight blue-green tint on the shark (third row), indicative of ambient light influence. In contrast, the full model incorporating GCR (tenth column) successfully eliminates these ambient light effects, resulting in more accurate color representation. These findings highlight GCR's crucial role in global color correction, particularly in

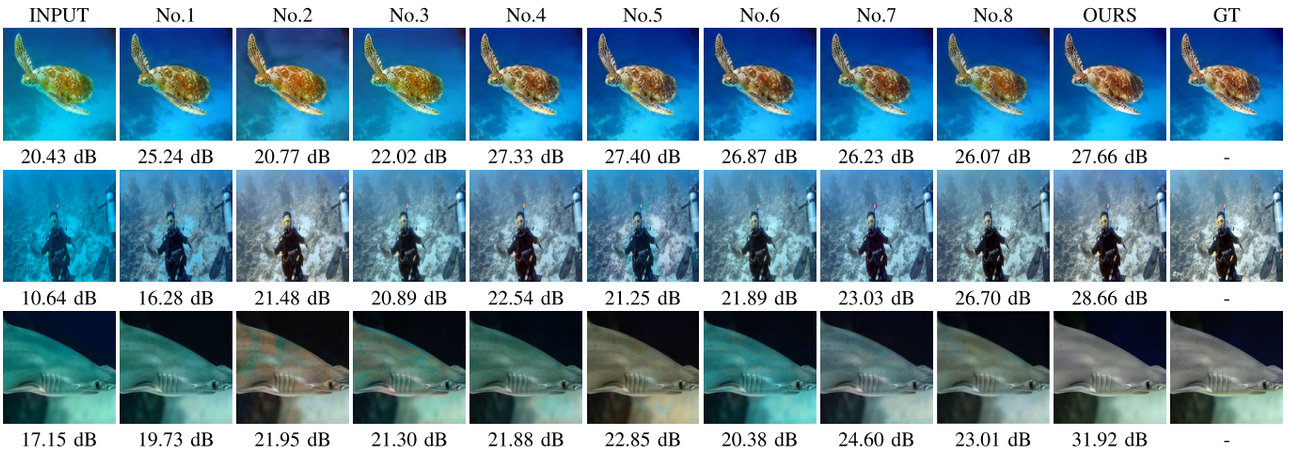


Fig. 13. Visual comparisons with PSNR values for ablation study of different settings.

addressing uneven illumination issues in complex underwater environments.

d) Effectiveness of loss L_{hist}: By comparing the color histogram information of the enhanced images and the ground truth, L_{hist} minimizes the difference in color distribution, enabling our model to neutralize color interference and better address color distortion issues. Comparing the images in column 8 with those in column 10 in Fig. 13, it is evident that the results without L_{hist} exhibit some color distortion. In Table V, it can be observed that the metrics on both datasets decreased compared to the full model after removing L_{hist}, demonstrating the effectiveness of L_{hist}.

e) Effectiveness of the multiscale inputs: The multiscale inputs help capture features at different resolutions, enabling the network to better handle objects and details of varying scales in underwater images. Comparing the images in columns 9 and 10 of Fig. 13, it can be seen that the results without using multiscale inputs do not satisfactorily restore detail textures. In Table V, it can be observed that the metrics on both datasets decreased compared to the full model after removing the multiscale inputs, demonstrating the effectiveness of MS.

D. Complexity Comparison

The model complexity is evaluated in terms of the number of parameters (Params) and the number of multiply-accumulate operations (#MACs). The recent methods including Water-Net [77], UWnet [78], USUIR [49], PUIE-Net [79], and PUGAN [16] with available codes are selected for comparison. See Table VI for the results. In terms of Params, we outperform other methods, and we rank as the second best on #MAC, which means the proposed model is lightweight and deployable.

E. Visualization of Estimated Color Map

The LCR branch employs color histograms to capture detailed color information and adaptive weighting to address diverse color distributions in underwater images. Fig. 14 illustrates the efficacy of our color map in representing both

TABLE VI
COMPLEXITY RESULTS OF DIFFERENT METHODS

Method	#Params(M)	#MACs(G)
Water-Net	2.23	71.53
UWnet	2.19	21.71
USUIR	2.26	14.88
PUIE-Net	16.12	30.08
PUGAN	10.19	67.69
BCTA-Net (Ours)	2.17	20.59

Fig. 14. Outputs of the LCR branch. (a) Shake. (b) Person.

global and local color information. In Fig. 14(a), the color map distinctly delineates the chromatic attributes of key elements—shark, seawater, and seafloor—demonstrating its capacity to establish the global color tone. Fig. 14(b) exemplifies the preservation of local color details, evidenced by the clear representation of water reflections and the sharp definition of the human silhouette. This dual-scale color mapping contributes significantly to BCTA-Net's performance in UIE. The color map's ability to capture color distributions while preserving local details enables more precise, context-aware color enhancement. This approach effectively addresses the challenges of nonuniform color distortions and detail preservation in underwater imagery, as corroborated by our experimental results.

F. Application to Down-Stream Object Detection Tasks

To validate the effectiveness of our method in enhancing downstream tasks, we conducted extensive experiments using a web application based on state-of-the-art models YOLO-WORLD [88] for object detection and EfficientSAM [89] for segmentation. The experiments were performed on a

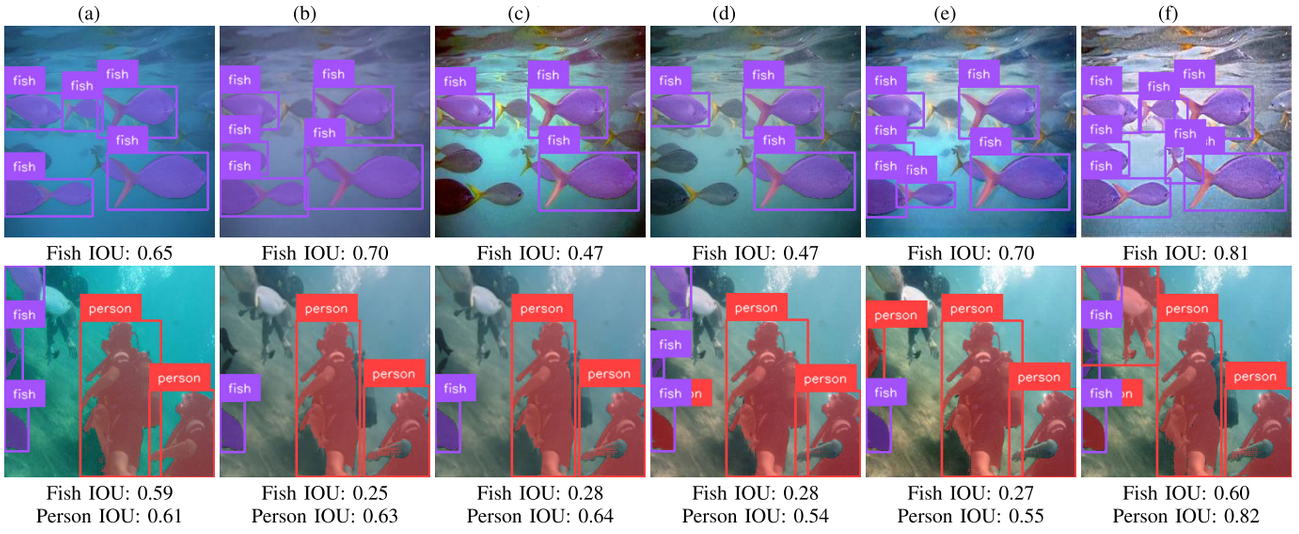


Fig. 15. Object detection and segmentation results with IoU of underwater images in MABLS dataset and their enhanced images. (a) INPUT. (b) UWnet. (c) USUIR. (d) PUIE-Net. (e) PUGAN. (f) OURS.

diverse set of images, including inputs, and images processed by UWnet, USUIR, PUIE-Net, PUGAN, and our proposed method, as shown in Fig. 15. The intersection-over-union (IoU) is also calculated for detected and segmented results of each enhanced image for intuitive quantitative comparisons. From the visualization and quantitative comparisons, our method consistently outperforms the baseline input and competing approaches in target identification, achieving more precise recognition and segmentation results.

V. Conclusion

This study presents BCTA-Net, an adaptive bi-level color-based network for UIE. The proposed architecture effectively addresses color distortions and reduced contrast in underwater imagery through integrated global and local restoration strategies. BCTA-Net's dual-branch structure, combining a color-aware attention mechanism and a TA module, successfully mitigates local issues of contrast, color fidelity, and detail precision. The novel application of color histograms for adaptive correction and contrastive learning for global color distortion correction represents significant advancements. Comprehensive experiments demonstrate BCTA-Net's superior performance over existing methods, particularly in image quality and detail recovery. These results establish a new benchmark in adaptive image enhancement for challenging underwater environments and open avenues for future research in complex visual scenarios.

References

[1] C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2019.

[2] J.-C. Zhou, D.-H. Zhang, and W.-S. Zhang, "Classical and state-of-the-art approaches for underwater image defogging: A comprehensive survey," *Frontiers Inf. Technol. Electron. Eng.*, vol. 21, no. 12, pp. 1745–1769, Dec. 2020.

[3] N. Wang, T. Chen, S. Liu, R. Wang, H. R. Karimi, and Y. Lin, "Deep learning-based visual detection of marine organisms: A survey," *Neurocomputing*, vol. 532, pp. 1–32, May 2023.

[4] N. Jiang, W. Chen, Y. Lin, T. Zhao, and C.-W. Lin, "Underwater image enhancement with lightweight cascaded network," *IEEE Trans. Multimedia*, vol. 24, pp. 4301–4313, 2022.

[5] G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.

[6] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, Dec. 1977.

[7] Y.-C. Liu, W.-H. Chan, and Y.-Q. Chen, "Automatic white balance for digital still camera," *IEEE Trans. Consum. Electron.*, vol. 41, no. 3, pp. 460–466, Aug. 1995.

[8] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6723–6732.

[9] C.-Y. Li, J.-C. Guo, R.-M. Cong, Y.-W. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016.

[10] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[11] X. Ding, Y. Wang, J. Zhang, and X. Fu, "Underwater image dehaze using scene depth estimation with adaptive color correction," in *Proc. OCEANS*, Jun. 2017, pp. 1–5.

[12] A. Ignatov, N. Kobyshev, R. Timofte, and K. Vanhoey, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3277–3285.

[13] Y. Wang, J. Guo, H. Gao, and H. Yue, "UIEC^2-Net: CNN-based underwater image enhancement using two color space," *Signal Process., Image Commun.*, vol. 96, Aug. 2021, Art. no. 116250.

[14] P. Hambarde, S. Murala, and A. Dhall, "UW-GAN: Single-image depth estimation and image enhancement for underwater images," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[15] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3227–3234, Apr. 2020.

[16] R. Cong et al., "PUGAN: Physical model-guided underwater image enhancement using GAN with dual-discriminators," *IEEE Trans. Image Process.*, vol. 32, pp. 4472–4485, 2023.

[17] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023.

[18] Y. Tang, T. Iwaguchi, H. Kawasaki, R. Sagawa, and R. Furukawa, "AutoEnhancer: Transformer on U-Net architecture search for underwater image enhancement," in *Proc. Asian Conf. Comput. Vis.*, 2022, pp. 1403–1420.

[19] Z. Huang, J. Li, Z. Hua, and L. Fan, "Underwater image enhancement via adaptive group attention-based multiscale cascade transformer," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–18, 2022.

[20] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.

[21] M. Li, Y. Lin, L. Shen, Z. Wang, K. Wang, and Z. Wang, "Human perceptual quality driven underwater image enhancement framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5634415.

[22] C. Zhao, W. Cai, C. Dong, and Z. Zeng, "Toward sufficient spatial-frequency interaction for gradient-aware underwater image enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 3220–3224.

[23] B. Wang et al., "UIE-convformer: Underwater image enhancement based on convolution and feature fusion transformer," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 8, no. 2, pp. 1952–1968, Apr. 2024.

[24] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and M. Sbert, "Color channel compensation (3C): A fundamental pre-processing step for image enhancement," *IEEE Trans. Image Process.*, vol. 29, pp. 2653–2665, 2020.

[25] A. Pramanick, S. Sarma, and A. Sur, "X-CAUNET: Cross-color channel attention with underwater image-enhancing transformer," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 3550–3554.

[26] H. F. Tolie, J. Ren, and E. Elyan, "DICAM: Deep inception and channel-wise attention modules for underwater image enhancement," *Neurocomputing*, vol. 584, Jun. 2024, Art. no. 127585.

[27] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[28] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5728–5739.

[29] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16259–16268.

[30] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, 2022.

[31] S. Raveendran, M. D. Patil, and G. K. Birajdar, "Underwater image enhancement: A comprehensive review, recent trends, challenges and applications," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 5413–5467, Oct. 2021.

[32] W. Song, Y. Wang, D. Huang, A. Liotta, and C. Perra, "Enhancement of underwater images with statistical model of background light and optimization of transmission map," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 153–169, Mar. 2020.

[33] H.-H. Chang, C.-Y. Cheng, and C.-C. Sung, "Single underwater image restoration based on depth estimation and transmission compensation," *IEEE J. Ocean. Eng.*, vol. 44, no. 4, pp. 1130–1149, Oct. 2019.

[34] M. Jha and A. K. Bhandari, "CBLA: Color-balanced locally adjustable underwater image enhancement," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.

[35] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof, and Z. Bachok, "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *Proc. Int. Conf. Comput. Appl. Technol. (ICCAT)*, Jan. 2013, pp. 1–5.

[36] S. M. Pizer, "Contrast-limited adaptive histogram equalization: Speed and effectiveness Stephen M. Pizer, R. Eugene Johnston, James P. Ericksen, Bonnie C. Yankaskas, Keith E. Muller medical image display research group," in *Proc. Conf. Visualizat. Biomed. Comput.*, 1990, p. 1.

[37] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2017.

[38] S. Gao, M. Zhang, Q. Zhao, X. Zhang, and Y. Li, "Underwater image enhancement using adaptive retinal mechanisms," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5580–5595, Nov. 2019.

[39] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 81–88.

[40] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A Retinex-based enhancing approach for single underwater image," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 4572–4576.

[41] S. Zhang, T. Wang, J. Dong, and H. Yu, "Underwater image enhancement via extended multi-scale Retinex," *Neurocomputing*, vol. 245, pp. 1–9, Jul. 2017.

[42] P. Drews Jr., E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 825–830.

[43] Y. Wang, H. Liu, and L.-P. Chau, "Single underwater image restoration using adaptive attenuation-curve prior," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 3, pp. 992–1002, Mar. 2018.

[44] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.

[45] D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1682–1691.

[46] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, 2021.

[47] J. Hu, Q. Jiang, R. Cong, W. Gao, and F. Shao, "Two-branch deep neural network for underwater image enhancement in HSV color space," *IEEE Signal Process. Lett.*, vol. 28, pp. 2152–2156, 2021.

[48] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.

[49] Z. Fu et al., "Unsupervised underwater image restoration: From a homology perspective," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, 2022, pp. 643–651.

[50] Q. Qi, K. Li, H. Zheng, X. Gao, G. Hou, and K. Sun, "SGUIE-Net: Semantic attention guided underwater image enhancement with multi-scale perception," *IEEE Trans. Image Process.*, vol. 31, pp. 6816–6830, 2022.

[51] P. Mu, H. Xu, Z. Liu, Z. Wang, S. Chan, and C. Bai, "A generalized physical-knowledge-guided dynamic model for underwater image enhancement," in *Proc. 31st ACM Int. Conf. Multimedia*, Oct. 2023, pp. 7111–7120.

[52] X. Cong, J. Gui, and J. Hou, "Underwater organism color fine-tuning via decomposition and guidance," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, Mar. 2024, pp. 1389–1398.

[53] C. Liu, X. Shu, L. Pan, J. Shi, and B. Han, "Multiscale underwater image enhancement in RGB and HSV color spaces," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.

[54] J. Zhou, S. Wang, Z. Lin, Q. Jiang, and F. Sohel, "A pixel distribution remapping and multi-prior retinex variational model for underwater image enhancement," *IEEE Trans. Multimedia*, vol. 26, pp. 7838–7849, 2024.

[55] J. Wu, X. Liu, N. Qin, Q. Lu, and X. Zhu, "Two-stage progressive underwater image enhancement," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–18, 2024.

[56] H. Qiang, Y. Zhong, Y. Zhu, X. Zhong, Q. Xiao, and S. Dian, "Underwater image enhancement based on multichannel adaptive compensation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–10, 2024.

[57] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107038.

[58] X. Fu and X. Cao, "Underwater image enhancement with global–local networks and compressed-histogram equalization," *Signal Process., Image Commun.*, vol. 86, Aug. 2020, Art. no. 115892.

[59] P. Sharma, I. Bisht, and A. Sur, "Wavelength-based attributed deep neural network for underwater image restoration," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 1, pp. 1–23, Jan. 2023.

[60] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 387–394, Jan. 2018.

[61] C. Li, J. Guo, and C. Guo, "Emerging from water: Underwater image color correction based on weakly supervised color transfer," *IEEE Signal Process. Lett.*, vol. 25, no. 3, pp. 323–327, Mar. 2018.

[62] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 7159–7165.

[63] W. Xia, Y. Zhang, Y. Yang, J.-H. Xue, B. Zhou, and M.-H. Yang, "GAN inversion: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3121–3138, Mar. 2023.

[64] Y. Guo, H. Li, and P. Zhuang, "Underwater image enhancement using a multiscale dense generative adversarial network," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 862–870, Jul. 2020.

[65] S. Islam et al., "A comprehensive survey on applications of transformers for deep learning tasks," *Expert Syst. Appl.*, vol. 241, May 2024, Art. no. 122666.

[66] T. Ren et al., "Reinforced Swin-Convs transformer for simultaneous underwater sensing scene image enhancement and super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4209616.

[67] Z. Shen, H. Xu, T. Luo, Y. Song, and Z. He, "UDAformer: Underwater image enhancement based on dual attention transformer," *Comput. Graph.*, vol. 111, pp. 77–88, Apr. 2023.

[68] B. Sun, Y. Mei, N. Yan, and Y. Chen, "UMGAN: Underwater image enhancement network for unpaired image-to-image translation," *J. Mar. Sci. Eng.*, vol. 11, no. 2, p. 447, Feb. 2023.

[69] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[70] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[71] B. Li et al., "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2018.

[72] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," 2019, *arXiv:1906.06819*.

[73] M. J. Islam et al., "Semantic segmentation of underwater imagery: Dataset and benchmark," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 1769–1776.

[74] Y. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.

[75] P. Zhuang, J. Wu, F. Porikli, and C. Li, "Underwater image enhancement with hyper-Laplacian reflectance priors," *IEEE Trans. Image Process.*, vol. 31, pp. 5442–5455, 2022.

[76] J. Xie, G. Hou, G. Wang, and Z. Pan, "A variational framework for underwater image dehazing and deblurring," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3514–3526, Jun. 2022.

[77] M. A. Syariz, C.-H. Lin, M. V. Nguyen, L. M. Jaelani, and A. C. Blanco, "WaterNet: A convolutional neural network for Chlorophyll-A concentration retrieval," *Remote Sens.*, vol. 12, no. 12, p. 1966, Jun. 2020.

[78] A. Naik, A. Swarnakar, and K. Mittal, "Shallow-UWnet: Compressed model for underwater image enhancement (student abstract)," in *Proc. 35th AAAI Conf. Artif. Intell. (AAAI)*, vol. 18, 2021, pp. 15853–15854.

[79] Z. Fu, W. Wang, Y. Huang, X. Ding, and K.-K. Ma, "Uncertainty inspired underwater image enhancement," in *Proc. IEEE Conf. Eur. Conf. Comput. Vis.*, Oct. 2022, pp. 465–482.

[80] H. Zhang, H. Xu, X. Yu, J. Wang, and C. Wu, "Efficient attentional underwater image enhancement generative adversarial network," in *Proc. 36th Chin. Control Decis. Conf. (CCDC)*, May 2024, pp. 3813–3818.

[81] M. Gao, Z. Li, Q. Wang, and W. Fan, "DAE-GAN: Underwater image super-resolution based on symmetric degradation attention enhanced generative adversarial network," *Symmetry*, vol. 16, no. 5, p. 588, May 2024.

[82] G. Zhang, C. Li, J. Yan, and Y. Zheng, "ULD-CycleGAN: An underwater light field and depth map-optimized CycleGAN for underwater image enhancement," *IEEE J. Ocean. Eng.*, vol. 49, no. 4, pp. 1275–1288, Oct. 2024.

[83] X. Wang et al., "Generative adversarial network with lightweight U-Net for underwater optical image enhancement," in *Proc. 12th Int. Conf. Intell. Comput. Wireless Opt. Commun. (ICWOC)*, Jun. 2024, pp. 29–34.

[84] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.

[85] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.

[86] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.

[87] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2019, pp. 8024–8035.

[88] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, "YOLO-world: Real-time open-vocabulary object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 16901–16911.

[89] Y. Xiong et al., "EfficientSAM: Leveraged masked image pretraining for efficient segment anything," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 16111–16121.

证书号第8429779号

专利公告信息

发明专利证书

发明名称：基于夜间物理感知和灰度世界的夜间去雾方法及装置

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

发 明 人：梁云;肖新杰;周子涵;李亮辉;彭世杰;肖雨婷

专　利　号：ZL 2024 1 0939864.3　　授权公告号：CN 119048373 B

专利申请日：2024年07月15日　　授权公告日：2025年10月31日

申请日时申请人：华南农业大学

申请日时发明人：梁云;肖新杰;周子涵;李亮辉;彭世杰;肖雨婷

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，并予以公告。

专利权自授权公告之日起生效。专利权有效性及专利权人变更等法律信息以专利登记簿记载为准。

局长
申长雨

国家知识产权局
2025年10月31日

证书号第8422125号

专利公告信息

发明专利证书

发明名称：基于颜色感知融合注意力和背景光驱离对比学习的水下图像增强方法及装置

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

发 明 人：梁云;李亮辉;周子涵;肖雨婷;肖新杰

专　利　号：ZL 2024 1 1366554.3　　授权公告号：CN 119444593 B

专利申请日：2024年09月29日　　授权公告日：2025年10月31日

申请日时申请人：华南农业大学

申请日时发明人：梁云;李亮辉;周子涵;肖雨婷;肖新杰

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，并予以公告。

专利权自授权公告之日起生效。专利权有效性及专利权人变更等法律信息以专利登记簿记载为准。

局长
申长雨

国家知识产权局
2025年10月31日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导赵齐荣获第十五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：021574745

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年6月2日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导钟紫妍荣获第十五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组优秀奖，被评为优秀指导教师。

特发此证，以资鼓励。

证书编号：021580516

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年6月2日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导赵齐荣获第十六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：1602114740

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年6月23日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导赵齐荣获第十五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B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：021549371

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导钟紫妍荣获第十五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B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：021549697

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导杨立创荣获第十五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：021548148

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导林依宁荣获第十五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：021548245

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导钟家力荣获第十五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：021548431

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导冯锦玉荣获第十五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：021548596

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导陈宣睿荣获第十五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：021549556

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导廖凯颖荣获第十五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：021547279

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导林瑞敏荣获第十五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：021548506

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导甘智杰荣获第十五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：021548972

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导吕慧娴荣获第十五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：021549097

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导吴思铜荣获第十五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：021549571

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导何睿涛荣获第十五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：021550145

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导苏子鑫荣获第十五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：021548340

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2024年4月29日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导赵齐荣获第十六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B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：1602058342

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导林依宁荣获第十六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：1602058311

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导林瑞敏荣获第十六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：1602058532

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导林瑞敏荣获第十六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：1602058532

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导江穗湘荣获第十六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：1602058727

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导钟家力荣获第十六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：1602057374

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导何卿荣获第十六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：1602058492

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导吴思铜荣获第十六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：1602058438

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

蓝桥杯大赛

获奖证书

华南农业大学周子涵：

指导甘智杰荣获第十六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：1602057952

证件号码：342601199602060028

工业和信息化部
人才交流中心

蓝桥杯大赛组委会

2025年5月26日

传智杯

荣誉证书

华南农业大学 周子涵

在2024-2025年度全国大学生计算机应用能力与数字素养大赛，暨第七届传智杯全国IT技能大赛中
您指导的学生荣获程序设计挑战赛研究生组初赛二等奖

授予您：优秀指导教师

学生姓名：王鸿宇

证书编号：CZB10001S027911

传智杯大赛组委会

全国高等院校计算机基础教育研究会

2024年12月31日


GDCA
广东省计算机学会
Computer Academy of Guangdong
广东省计算机学会优秀论文奖
证 书
为表彰2023年度广东省计算机学会
优秀论文奖获奖者，特颁发此证书。
项目名称：Image Quality Assessment Using
Kernel Sparse Coding
奖励等级：二等奖
获奖单位：华南农业大学
获 奖 者：周子涵 李静 许若涛 全宇晖
粤计学证：【2023】57号 二〇二三年十二月
项目编号：2023-J2014


GDCA
广东省计算机学会
Computer Academy of Guangdong
广东省计算机学会
优秀论文奖
证书
为表彰2025年度广东省计算机学会
优秀论文奖获奖者，特颁发此证书。
项目名称：Deep blind image quality assessment using dynamic neural model with dual-order statistics
奖励等级：二等奖
获奖单位：华南农业大学
获 奖 者：周子涵 李静 钟德祥 许勇 Patrick Le Callet
项目编号：2025-J2-100
广东省计算机学会
二〇二五年十二月

奖状

周子涵 同志：

在2024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

2025年1月

广东省科学技术协会

2025-2026年度广东省科协青年科技人才培育计划人选名单公示

按照2025-2026年度广东省科协青年科技人才培育计划立项要求，各立项单位共遴选出2025-2026年度广东省科协青年科技人才培育计划人选563人。根据《广东省科学技术协会青年科技人才培育计划管理办法》规定，现对人选名单予以公示，公示期为5个工作日，从2025年10月16日至10月22日止。

在公示期间，社会各界均可通过来电、来访等形式，向省科协反映公示对象的有关情况和问题。反映情况须客观真实，以单位名义反映情况的材料需加盖单位公章，以个人名义反映情况的材料应署实名并提供有效的联系方式。

附件：2025-2026年度广东省科协青年科技人才培育计划人选名单

广东省科学技术协会

2025年10月15日

（受理部门：省科协组织联络部，地址：广州市连新路171号，邮政编码：510040，联系电话：020-83551764、83546294，邮箱：）

附件

2025-2026年度广东省科协青年科技人才培育计划人选名单

一、省级学会（37个单位，培育对象251人）

序号	姓名	性别	实施单位	工作单位
1	王鹏	男	广东省力学学会	深圳大学
2	杨奎坚	男	广东省力学学会	中山大学
3	黄帅	男	广东省力学学会	中山大学
4	肖越	女	广东省力学学会	中山大学
5	虞鹏鹏	男	广东省地质学会	中山大学
6	梁银秀	女	广东省环境科学学会	广东省科学院微生物研究所
7	许明熠	男	广东省环境科学学会	华南理工大学
8	高方舟	男	广东省环境科学学会	华南师范大学
9	林楠	女	广东省环境科学学会	清华大学深圳国际研究生院
10	唐斌	男	广东省环境科学学会	生态环境部华南环境科学研究所
11	陈思琪	女	广东省白蚁学会	广东省科学院动物研究所
12	陈鑫	女	广东省生物物理学会	华南理工大学附属第二医院（广州市第一人民医院）

序号	姓名	性别	实施单位	工作单位
13	石玉娇	女	广东省生物物理学会	华南师范大学
14	常好才	男	广东省生物物理学会	华南师范大学
15	申琪	女	广东省生物物理学会	华南师范大学
16	刘凯华	男	广东省农业机械学会	广东省现代农业装备研究院
17	温翔宇	男	广东省农业机械学会	广东省现代农业装备研究院
18	曾岚	女	广东省非开挖技术协会	暨南大学
19	谈继勇	男	广东省电子学会	电子科技大学（深圳）高等研究院
20	王磊	男	广东省电子学会	工业和信息化部电子第五研究所
21	庄庆威	男	广东省计算机学会	测绘遥感信息工程全国重点实验室深圳研发中心
22	袁成哲	男	广东省计算机学会	广东技术师范大学
23	姜思羽	女	广东省计算机学会	广东外语外贸大学
24	王宁	女	广东省计算机学会	广州大学
25	汤非易	男	广东省计算机学会	广州职业技术大学
26	吴宇琳	女	广东省计算机学会	哈尔滨工业大学（深圳）
27	周子涵	女	广东省计算机学会	华南农业大学
28	林荣华	男	广东省计算机学会	华南师范大学
29	郑立彬	男	广东省计算机学会	中山大学
30	曾丹	女	广东省计算机学会	中山大学人工智能学院

序号	姓名	性别	实施单位	工作单位
31	臧建波	男	广东省土木建筑学会	广东省建筑科学研究院集团股份有限公司
32	万军	男	广东省土木建筑学会	广东省源天工程有限公司
33	张亚飞	男	广东省土木建筑学会	广州建筑湾区智造科技有限公司
34	胡方鑫	男	广东省土木建筑学会	华南理工大学
35	苏慧慧	女	广东省食品学会	广东省科学院生物与医学工程研究所
36	朱振军	男	广东省食品学会	暨南大学
37	杨琼琼	女	广东省食品学会	汕头大学
38	陈肯	男	广东省材料研究学会	东莞理工学院
39	王猛	男	广东省材料研究学会	广东腐蚀科学与技术创新研究院
40	李群洋	男	广东省材料研究学会	广东省科学院生物与医学工程研究所
41	柳颖	女	广东省材料研究学会	广东省科学院生物与医学工程研究所
42	周晟昊	男	广东省材料研究学会	广东省科学院新材料研究所
43	高硕洪	男	广东省材料研究学会	广东省科学院新材料研究所
44	李文波	男	广东省材料研究学会	广东省科学院新材料研究所
45	范秀娟	女	广东省材料研究学会	广东省科学院新材料研究所

序号	姓名	性别	实施单位	工作单位
260	何昕	男	中山大学附属第一医院科协	中山大学附属第一医院
261	曾雪贞	女	中山大学附属第一医院科协	中山大学附属第一医院
262	郭玥	女	中山大学附属第一医院科协	中山大学附属第一医院
263	李洁	女	中山大学附属第一医院科协	中山大学附属第一医院
264	涂剑	男	中山大学附属第一医院科协	中山大学附属第一医院
265	刘贻豪	男	中山大学附属第一医院科协	中山大学附属第一医院
266	谢宝树	男	中山大学附属第一医院科协	中山大学附属第一医院

项目编号：SKXRC2025____　　合同编号：2025QNRC____

项目合同书

项目名称：2025-2026 年度广东省科协青年科技人才培育计划

培育对象：周子涵

实施单位：广东省计算机学会

工作单位：华南农业大学

经办部室：

签署日期：2025 年 9 月 30 日

填 写 说 明

一、本合同书为项目工作的主要依据之一，项目实施单位必须保证其真实性和严肃性，请严格按照表中要求认真填写。

二、合同书应为A4开本，具体报送要求请参照通知中有关要求执行。

三、项目编号和合同编号请填写培育对象各自对应的编号。

四、法人/委托代理人、项目联系人须本人签字或盖签名章。

五、各栏目如填写内容较多，可另加附页。

一、项目实施单位基本情况			
单位名称	广东省计算机学会	单位性质	社会团体
单位地址	广州市越秀区连新路171号科学馆大院3号楼104房	邮政编码	510000
单位负责人	黄轩	职称/职务	副理事长兼秘书长
单位联系人	杨惜爱	职称/职务	副秘书长
办公电话	020-83561784	手机	13824468800
电子邮箱	83533315@qq.com		
项目起止时间	2025年1月-2026年12月	立项资助经费（万元）	3

二、指导老师基本情况					
姓　名	梁云	性别	女	出生年月	1981.06
政治面貌	汉族	学　位	工学博士	研究领域	计算机视觉、人工智能
专业技术职称	教授	手机号码	13760698353	身份证号	371328198106204540
毕业院校	中山大学		工作单位	华南农业大学	

三、培育对象基本情况					
姓名	周子涵	性别	女	民族	汉
出生年月	1996.02	政治面貌	群众	研究领域	计算机视觉
籍贯	皖巢	学历	博士	学位	工学博士
专业技术职称	讲师	身份证号码	342601199602060028		

工作单位及职务	华南农业大学，无		
单位性质	☑高等院校 □科研院所 □省级学会 □政府机关 □其他事业单位 □国有企业 □民营企业 □外资企业 □其他		
通信地址	广州市天河区五山路382号		
联系电话	18819472687	手机	18819472687
电子邮箱	zhouzihan@scau.edu.cn	邮政编码	510000

三、项目总目标 （资助培育期内个人计划预期）

基于申请人在图像处理与视觉质量评价领域的学术积淀和突出研究潜力，制定科学合理的学术荣誉申报路线图，争取指导申请人冲击广东省杰出青年基金等。

四、主要工作任务和绩效考核指标（对项目总目标的细化和量化）

（一）工作任务（不少于100字）

基于申请人在图像处理与视觉质量评价领域的学术积淀，深化"内容-退化解耦、动态统计表征与跨模态协同"等核心技术路线，2年内在国内外期刊发表高水平论文2-5篇，参加国内外学术交流会议2-5场，显著提升国际学术声誉。结合广东省人工智能、农业工程等优势产业需求，推进科技成果转化和推广，申请专利1-2项。

（二）考核指标（4个以上，量化指标个数需占指标总个数比例达2/3以上，不应设置预算完成率、管理有效性等共性指标，每个绩效指标设置相应的指标预期值，量化指标需明确计算方法。须与财政支出项目绩效目标申报表。）

1. 产出指标：

指标1：在国内外期刊/会议发表高水平论文1-5篇；

指标2：参加国内外学术交流会议2-5场；

指标3：申请专利1-2项；

指标4：培养硕士研究生1-3名；

……

2. 效益指标：（通过开展项目产生的经济、社会、生态、可持续、科普等方面的效益）

无

五、经费支出预算明细（须与财政支出项目绩效目标申报表和申请表一致）

编号	支出内容	金额（万元）	备注
1	参加学术会议等	2.4万	会议注册费等
2	申请专利、发表论文等	0.6万	版面费等
3			
……			
合计	3万元		

六、共同条款

（一）项目实施单位积极提供服务，加强与培育对象之间的联络。项目实施单位和培育对象要及时向省科协上报培养工作进展情况。

（二）项目实施单位主要任务：1. 指导帮助培育对象制定培养方案，签订项目合同书；2. 为培育对象搭建培养平台；3. 与培育对象建立长效联系机制，实时掌握培育对象发展情况，保障项目按计划实施，并做好项目的总结工作；4. 接受省科协的监督，并按要求提供项目相关材料。

（三）培育对象的主要任务：1. 制定个人成长发展规划及经费使用计划；2. 积极主动落实培养方案；3. 及时反馈个人成长情况；4. 按要求完成省科协和项目实施单位布置的有关工作；5.积极开展科普相关工作；6.积极参加各级科协组织的活动。

（四）项目实施单位应依据经费使用范围及本单位财务规定，制定经费签报程序，帮助指导培育对象完成经费执行。资助经费只能用于培育对象学术成长过程中所发生的各项直接支出，主要包括:1. 参加国际性学术会议、国际交流合作项目、短期培训差旅费、注册费等相关支出；2. 申请专利、发表论文、出版自然科学范围内的原创性科技、科普类著作等相关支出；3.开展课题研究和技术攻关的相关直接支出；4.其他与培育计划工作相关的支出。

（五）项目经费的使用管理应遵守国家有关法律规定和财务制度。资助经费要专款专用，不得截留或挪用，不得用于项目实施单位的基本建设、对外投资、罚款、捐赠、工作人员工资及福利等，禁止使用现金方式结算。对由于各种原因合同终止的项目，已拨付经费的，项目实施单位应退还尚未使用和使用不符合规定的财政经费，未拨付经费的将不予拨付。

（六）项目经费的使用必须接受财政、审计及省科协等部门的监督与检查。对存在弄虚作假、虚报业绩等违规行为的项目实施单位进行通报，并责令退还项目经费，终止培育计划。对违反国家有关规定，截留、侵占、挪用、挥霍专项经费的单位或个人，按有关法律法规进行查处。

（七）受资助方完成的论文、出版的刊物等知识产权，省科协享有无偿使用权（注：如双方协商确定不享有的，则本条款删除）。

（八）合同书签订各方均负有相应的责任。若有争议或纠纷时，按有关法规和管理办法处理。

（九）本合同书一式四份，实施单位保留两份，培育人保留两份，具有同等法律效力。

（十）合同书协议的其他条款如下：

1._____；

2._____；

3._____。

七、合同签署各方

项目实施单位名称：广东省计算机学会

项目实施单位单位法人/委托代理人（签字）：

（单位公章）

2025年 10 月 27 日

培育对象（签字）：

2025 年 9 月 30 日