

基于时域-频域协同引导的海洋机器人视觉感知增强方法

纪勋^{1,2} 戴成菴^{1,2} 李晓琳³ 赵昊中^{1,2} 张云泽¹ 郝立颖¹

(1. 大连海事大学船舶电气工程学院, 大连 116026; 2. 大连海事大学水路交通控制全国重点实验室, 大连 116026;
3. 中国飞行试验研究院测试技术研究所, 西安 710089)

摘要: 在海洋机器人执行水下探测与作业任务过程中,受水体对光线吸收与散射效应的影响,采集到的水下图像普遍存在亮度不足、细节模糊及色彩失真等问题,严重制约了后续视觉感知与智能决策任务的性能。针对现有方法多侧重于时域信息建模而忽略频域特征在结构保持与细节恢复中重要作用的不足,提出了一种时域-频域协同引导的海洋机器人视觉感知增强模型。从整体结构上看,所提出的模型主要包括时域-频域特征提取与多维度特征细化两个阶段。在时域-频域特征提取阶段,分别构建时域分支与频域分支并引入跨域协同模块实现多域特征的交互融合,从而增强对水下图像结构与纹理信息的综合表征能力。在多维度特征细化阶段,进一步设计多种特征增强模块以提升重建质量,首先提出偏振残差模块,通过多方向特征提取与自适应融合,在突出有效结构信息的同时抑制噪声干扰,实现细节与对比度的协同增强;其次,构建双注意力耦合模块,联合通道注意力与空间注意力机制,在保持全局结构一致性的基础上强化局部关键细节;最后,设计双重感受野的多尺度残差模块,结合多尺度卷积、注意力调节及可学习缩放参数,实现对关键特征的自适应强调,在细节保真与整体视觉效果之间取得平衡。在多个公开水下数据集上的实验结果表明,本文方法在主观视觉效果上均优于现有主流视觉增强方法,并在复杂水下环境中表现出更好的鲁棒性与泛化能力,为海洋机器人视觉感知与下游任务提供有效支撑。

关键词: 海洋机器人; 感知增强; 卷积网络; 特征融合; 特征提取; 深度学习

中图分类号: TP391 文献标识码: A 文章编号: 1000-1298(2026)19-0095-08 OSID: 

Time-frequency Collaborative Guidance-based Visual Perception Enhancement for Marine Robots

Ji Xun^{1,2} Dai Chengsong^{1,2} Li Xiaolin³ Zhao Haozhong^{1,2} Zhang Yunze¹ Hao Liying¹

(1. School of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China
2. State Key Laboratory of Maritime Technology and Safety, Dalian Maritime University, Dalian 116026, China
3. Testing Technology Research Institute, Chinese Flight Test Establishment, Xi'an 710089, China)

Abstract: Marine robots often capture severely degraded images during underwater exploration and operation tasks due to light absorption and scattering in water, which leads to low brightness, blurred details, and significant color distortion. These degradations substantially hinder subsequent visual perception and intelligent decision tasks. To solve the limitations of existing approaches that mainly rely on time-domain modeling and neglect the role of frequency-domain features in structure preservation and detail recovery, the temporal-frequency collaborative guided network was proposed for marine robots. The proposed network consisted of two stages: time-frequency feature extraction and multi-dimensional feature refinement. In the time-frequency feature extraction stage, a time domain branch and a frequency domain branch were employed to capture complementary representations, and a cross domain collaboration module was introduced to explicitly fuse multi domain features, thereby strengthening the representation of underwater image structures and textures. In the multi-dimensional feature refinement stage, several enhancement modules were incorporated to further improve reconstruction quality. Specifically, polarization residual module extracted multi-directional features and performed adaptive fusion were

收稿日期: 2026-01-26 修回日期: 2026-02-09

基金项目: 国家自然科学基金项目(52501429、52171292)、辽宁省自然科学基金项目(2025-BS-0210)和辽宁省教育厅高校基本科研业务项目(LJ212410151006)

作者简介: 纪勋(1992—),男,副教授,博士,主要从事海洋机器人技术研究,E-mail: jixun@dlmu.edu.cn

通信作者: 郝立颖(1982—),女,教授,博士,主要从事海洋机器人技术研究,E-mail: haoliying@dlmu.edu.cn

proposed to emphasize effective structural information while suppressing noise, which enabled simultaneous enhancement of image details and contrast. In addition, dual attention coupling module combined channel attention and spatial attention was introduced to reinforce local salient details while preserving global structural consistency. Finally, multi-scale residual module with dual receptive fields integrated multi-scale convolutions, attention modulation, and learnable scaling parameters was proposed to adaptively highlight critical features and balance fine-detail preservation with overall visual quality. Extensive experiments on multiple public underwater image datasets showed that the proposed method consistently outperformed state-of-the-art visual enhancement approaches in subjective visual quality, and demonstrated strong robustness and generalization capability in complex underwater environments, effectively supporting marine robot visual perception and downstream tasks.

Key words: marine robot; perception enhancement; convolutional network; feature fusion; feature extraction; deep learning

0 引言

随着人类对海洋蕴含的巨大资源与能源进行持续深入的探索,海洋机器人作为海洋观测、作业与采集的重要装备,其相关技术也面临着更高的要求,以进一步推动海洋通信、感知与信息处理能力的发展^[1]。在众多感知方式中,光学图像因能够直观呈现环境结构与目标特征,被视为获取海洋信息的重要方式,对揭示海洋世界的奥秘具有重要意义^[2]。然而,与空气介质下的常规视觉图像不同,水下图像在光线传播过程中不可避免地会受到散射与吸收的影响,从而引发严重的颜色失真、低对比度和颜色偏移等^[3]。此外,水体中的悬浮颗粒、气泡以及生物体等多种因素,也会导致图像出现噪声与模糊^[4]。这些因素严重阻碍后续的海洋机器人任务的发展,例如目标检测、水下自主定位、深度估计等^[5]。因此,如何有效提升水下图像质量,已成为推动海洋机器人智能感知能力持续发展的关键研究方向。

为解决上述问题,大量搭载在海洋机器人平台上的智能感知增强方法不断涌现,其可大致分为两类:传统非学习方法与深度学习方法^[6]。传统的海洋机器人视觉感知增强方法分为物理模型驱动策略与像素校正策略^[7]。基于物理模型的驱动策略通过建立水下光传播的物理成像模型,来描述光在水下传输时的散射和吸收过程。然后,利用物理模型的逆过程来恢复图像的真实色彩^[8]。然而由于海洋环境复杂多变,该方法在复杂场景下存在一定的局限性。另一类传统方法主要依赖对图像像素或统计特性直接处理来改善图像质量,无需借助水下物理模型^[9]。虽然这类方法也能够一定程度上提升图像质量,改善对比度,但其本质上缺乏学习能力,增强手段相对单一,难以适应复杂多变的海洋环境^[10]。因此,这类方法往往容易过度增强或锐化图像,从而引发噪声增加、条纹出现或色彩失真等问

题,使图像显得不自然。

与传统的非学习方法相比,基于深度学习的方法在海洋机器人视觉感知增强方面表现出显著的优势。深度学习模型能够通过数据驱动的方法,自适应地提取图像特征,并支持端到端的训练,从而实现复杂环境下的海洋机器人视觉感知增强^[11]。此外,深度学习方法不仅能够有效恢复退化图像的细节与色彩,还具备更强的特征表达能力,可以在一定程度上克服传统方法在建模灵活性和适应性方面的不足。

现有的海洋机器人视觉感知增强策略大多侧重于时域特征建模,往往忽略频域信息在结构细节恢复与质量提升中的重要作用,致使模型在复杂海洋环境下的鲁棒性与普适性仍存在不足。为此,本文提出一种时域-频域协同引导的海洋机器人视觉感知增强网络 TF-Net,通过联合建模时域与频域特征,提升复杂海洋环境下的视觉感知质量。主要工作如下:提出 TF-Net,融合时域特征与频域信息,以缓解水下图像颜色失真、散射干扰和低对比度等问题;设计跨域协同模块,通过门控机制自适应分配时域与频域特征权重;设计偏振残差模块,利用多方向卷积提取细节特征,并通过残差连接增强结构信息、抑制噪声;设计双注意力耦合模块,结合通道注意力与空间注意力突出关键特征、抑制冗余响应;设计多尺度残差模块,通过多尺度卷积捕获上下文信息,并结合注意力机制实现自适应增强。

1 网络模型

1.1 整体框架

海洋环境中复杂的光学传播特性使得海洋机器人获取的视觉信息普遍存在色彩失真、对比度降低以及细节模糊等退化问题。这类退化现象严重制约了海洋机器人在目标识别、环境理解与自主导航等任务中的可靠性。现有视觉感知增强方法多侧重于

单一时域特征建模,往往难以兼顾全局结构保持与局部纹理恢复,对复杂海洋退化场景的适应能力仍然有限。基于上述问题,本文从跨域特征互补与多维度细节建模的角度出发,探讨如何通过联合时域与频域信息,对关键特征进行有针对性的细化处理。

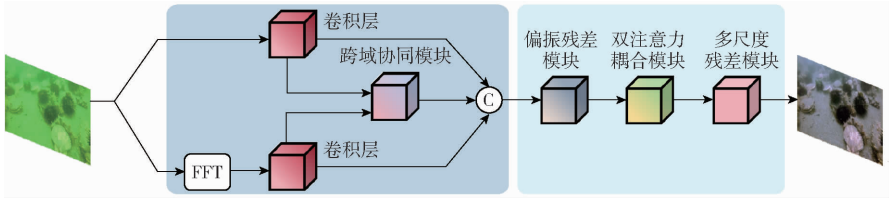


图 1 TF-Net 整体框架

Fig. 1 Overall architecture of proposed TF-Net

(1)在时域-频域特征提取阶段,网络通过并行的时域处理分支与频域处理分支对输入的水下退化图像 $X \in \mathbf{R}^{3 \times H \times W}$ 进行特征提取,随后引入时域-频域协同互补机制实现跨域信息的有效融合。具体而言,在时域分支中,网络通过两层 3×3 卷积对时域特征进行逐步提取,并将输入通道数由 3 映射至 64 以得到时域特征图 X_t ,进而使网络拥有更丰富的时域特征表示能力;在频域分支中,首先利用二维傅里叶变换将输入特征映射至频域以得到频域特征图 X_f ;随后,通过两层 3×3 卷积进行频域特征提取,并将通道数扩展至 64。所提出频域分支的优势在于引入了傅里叶变换,其能够有效保留图像的相位信息,因此有助于提升网络在局部纹理细节提取方面的能力,同时为后续的时域-频域融合提供更加全面的信息支持。再分别从时域与频域两方面对输入图像进行特征提取后,将得到的 X_t 与频域特征图 X_f 同步送入到跨域协同模块中以进行自适应交互,进而生成协同特征图 X_m ,该部分有助于提升网络在联合建模全局结构与局部细节特征方面的表现,从而更准确地恢复水下图像的整体结构与纹理信息。

(2)在多维度特征细化阶段,网络进一步对跨域协同后的特征图 X_m 进行特征细化处理,具体包括方向特征细化、注意力感知细化以及双重感受野细化。在方向特征细化部分,通过设计偏振残差模块将协同特征图 G_{in} 沿水平、垂直及对角线 3 个方向进行分解,并通过小型卷积层块进行局部增强,进而解决水下图像中常见的局部模糊、散射及细节丢失问题;在注意力感知细化部分,通过设计双注意力耦合模块对协同特征图 F_{in} 进行特征权重评估与分配,以突出关键信息并抑制冗余噪声,从而提升水下图像的清晰度和对比度。在双重感受野细化部分,通过设计多尺度残差模块捕获协同特征图 B_{in} 在不同感受野下的特征表示,实现对局部细节与全局结构的多层次增强,从而提升水下图像的视觉感知质量。

在上述研究动机的驱动下,本文提出了一种时域-频域协同引导的海洋机器人视觉感知增强算法,其整体框架如图 1 所示,主要由时域-频域特征提取阶段与多维度特征细化阶段两部分组成。其具体处理流程如下:

1.2 时域-频域特征提取阶段

为应对单一时域信息难以满足复杂海洋机器人视觉感知增强需求的挑战,本文在时域-频域特征提取阶段采用了一种双分支处理策略,以实现时域与频域特征的深度融合与交互。具体而言,将输入的水下退化图像 X 分为两个分支,一条分支通过傅里叶变换映射至频域,另一条分支在时域中直接进行特征表征。随后,在各分支分别引入两层 3×3 卷积,得到时域特征 X_t 和频域特征 X_f ,以逐层提取局部结构信息并增强非线性表示。

为进一步促进时域与频域特征的深度交互与融合,本文在双分支策略基础上引入了跨域协同模块。如图 2 所示,该模块遵循“通道合并-卷积压缩-权重生成-通道分离-特征融合”的设计思路,对时域与频域特征进行协同建模。首先,通过 1×1 的卷积对输入特征 X_t 和 X_f 进行初步提取。随后,将两个特征在通道维度进行合并(通道合并),生成融合特征 X_r ,从而整合不同空间域下的特征分布,实现多域信息的初步交互。在此基础上,对 X_r 使用 1×1 的卷积对融合特征进行通道压缩(卷积压缩),将通道数从 128 降至 64,再进一步降至 2,并结合 ReLU 激活函数,以进一步提取特征并提升训练稳定性。随后,将得到的增强特征使用 Sigmoid 函数归一化后生成特征权重(权重生成),并使用通道分离将分离出来的权重 X_{r1} 和 X_{r2} 分别作用于时域特征和频域特征,得到加权特征 X'_{t1} 和 X'_{f2} 。通过权重相乘,网络能够自适应地突出不同空间域下的重要信息。最

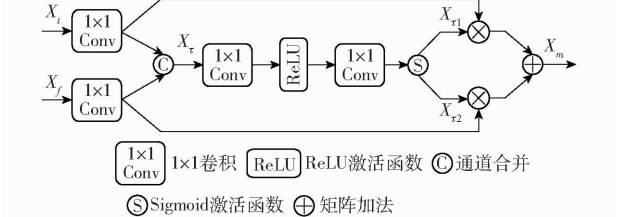


图 2 跨域协同模块

Fig. 2 Architecture of cross domain collaboration module

后,再次进行通道合并得到跨域融合特征 X_m ,实现跨域特征的高效整合。

综上所述,跨域协同模块通过结构化的权重建模机制,有效实现了时域与频域特征的互补与增强,在保留时域分支对局部结构敏感性的同时,充分发挥了频域分支对高频细节的表征优势,从而提升了整体特征表示的鲁棒性。

1.3 多维度特征细化阶段

本文所提出的多维度特征细化阶段可以进一步分为 3 部分,包括方向特征细化、注意力感知细化以及双重感受野细化。

1.3.1 方向特征细化

为增强网络对不同空间方向特征的关注能力,本文在方向特征细化部分提出了偏振残差模块,如图 3 所示。该模块主要包括 3 条方向卷积分支、残差增强单元以及自适应权重融合机制。具体而言,对时域-频域阶段输出的特征进行方向分解,分别提取水平方向、垂直方向和对角方向的特征。其中,水平方向分支采用 1×3 卷积提取水平纹理特征得到 G_{hor} ,垂直方向分支采用 3×1 卷积提取垂直纹理特征得到 G_{ver} ,而对角线方向分支则采用卷积核大小为 3×3 且空洞率为 2 的空洞卷积进行特征提取并得到 G_{dia} ,从而能够捕获更大感受野的方向信息。随后,利用 3×3 的卷积层对提取到的各方向特征进行精细增强,以充分挖掘不同方向上的局部结构信息。不同方向特征对图像内容的贡献存在差异。为此,本文在特征增强的基础上引入自适应权重融合机制,首先对输入特征 G_{in} 进行全局平均池化,得到反映通道整体响应的全局特征 G_0 ;随后,利用卷积层将该特征映射至通道为 3 的特征图,并通过 Softmax 归一化得到 3 个方向分支的自适应权重值。接着,将权重分别作用于对应的方向特征分支,并将加权后的 3 路特征进行逐元素相加得到融合特征,实现多方向信息的自适应整合。最终,模块输出通过残差连接的方式与原始输入特征相加得到输出 G_{out} ,一方面保留了输入特征的细节信息,另一方面进一步提升了多方向特征的表达能力。

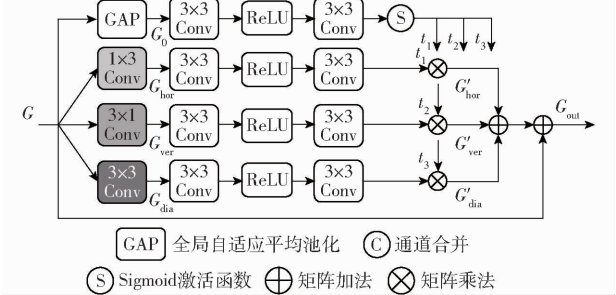


图 3 偏振残差模块

Fig. 3 Architecture of polarization residual module

总而言之,偏振残差模块能够从水平、垂直及对角 3 个方向提取并细化特征,有效捕获水下图像中受光照不均呈现的多方向纹理信息。通过自适应权重融合,模块能够根据图像内容动态调整不同方向特征的重要性,从而突出有效结构、抑制噪声。最终,残差连接既保证了输入细节的完整性,又增强了多方向特征的表达能力,有助于改善水下图像的对比度,提升结构保真度。

1.3.2 注意力感知细化

为增强网络对重要特征的关注,本文在注意力感知细化部分提出了双注意力耦合模块,该模块在通道注意力和空间注意力的基础上进行设计,实现对特征信息的协同优化与精细化增强。其结构如图 4 所示。该模块首先通过一个 3×3 卷积层捕捉图像的局部细节,并为后续注意力提供基础。在通道注意力部分,将 F_{in} 分别经过全局最大池化和全局平均池化得到 F_a 和 F_b ,并对其进行 1×1 卷积层处理,先将通道数降低至 4,再通过激活函数引入非线性,随后通过卷积升维至 64。将增强后的 F_a 和 F_b 相加后,通过 Sigmoid 函数生成注意力权重 F_m ,与原始特征相乘得到 F_n ,从而使网络更加关注局部细节,增强特征表达能力。在空间注意力部分,对增强后的特征沿通道维度分别使用 Mean 和 Max 操作,并在通道维度上进行拼接,生成一个通道为 2 的特征图。随后,利用卷积将通道数从 2 调整到 64,并通过 Sigmoid 函数生成空间注意力权重,与输入特征 F_n 相乘得到增强后的特征 F_o 。使网络能够聚焦于显著细节的区域,从而提升局部结构的感知能力。最后,将原始输入特征与增强后的特征进行相加得到 F_{out} ,实现特征的增强与融合。

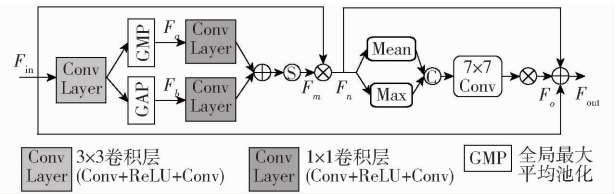


图 4 双注意力耦合模块

Fig. 4 Architecture of dual attention coupling module

双注意力耦合模块通过融合通道注意力和空间注意力,在保留全局语义的同时强化了局部细节。通道注意力负责建模各通道特征的重要性,空间注意力突出关键区域并抑制冗余信息,残差连接确保特征传递的完整性。由此,模块输出的特征兼具全局结构与细节对比度的提升,从而增强了网络在复杂海洋环境中的视觉感知能力。

1.3.3 双重感受野细化

为增强网络对不同尺度特征的感知能力,本文

在双重感受野细化部分提出了多尺度残差模块,其结构如图 5 所示。该模块通过多感受野机制以提升多尺度表征能力。具体而言,输入特征 B_{in} 分别通过 3×3 、 5×5 和 7×7 卷积核进行特征提取,分别得到 B_0 、 B_1 和 B_2 ,从而捕获不同感受野下的局部结构信息。随后,利用自适应平均池化压缩空间维度,在保留通道语义的同时获得全局上下文,再经过 1×1 卷积与 Sigmoid 激活函数生成通道注意力权重 B'_i 。将该权重与增强后的特征 B_i 进行通道乘法得到 B''_i ,实现对各通道的自适应加权,从而使网络能够关注重要信息并抑制冗余特征。最后,将加权特征与可学习的缩放参数相乘,并将其与原始特征融合,通过结合可学习参数使网络能够根据水下图像中光照衰减、散射及颜色失真造成的区域差异,自适应地调整增强强度。

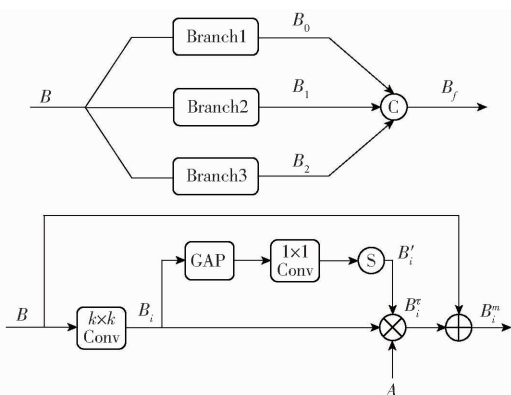


图 5 多尺度残差模块

Fig. 5 Architecture of multi-scale residual module

总体而言,所提出的多尺度残差模块通过多尺度卷积捕获不同感受野下的特征,并结合自适应池化与通道注意力实现对关键特征的自适应强调与冗余特征的抑制。同时,引入的可学习缩放参数进一步赋予网络根据水下退化特性灵活调整增强强度的能力,从而在提升局部细节保真度的同时兼顾全局视觉效果。

2 实验结果与分析

2.1 实验设置

2.1.1 数据集

为验证所提出方法的有效性,本文选取了 3 个高质量的海洋环境光学图像数据集,具体包括 LSUI^[12]、EUVP^[13] 和 UFO^[14]。LSUI 数据集包含 4 279 幅高分辨率水下图像,涵盖多种水体环境和光照条件,场景丰富多样,非常适合用于训练和评估海洋机器人视觉感知增强模型。本文从中选取 3 879 幅图像作为训练集,其余 400 幅图像用于测试。EUVP 数据集提供了大量水下图像,包括清晰图像与退化图像的成对数据,其显著优势在于高质量的

图像配对,便于对比评估图像增强方法的性能。本文直接采用 EUVP 数据集的测试集进行跨数据集验证,以评估所提方法在不同水下环境下的泛化能力。UFO 测试集包含 120 幅低能见度水下图像,主要集中于弱光和浑浊水体环境,可用于检验方法在极端条件下的增强效果。通过在多数据集上的联合测试,本文能够更充分地分析不同方法在实际应用场景中的稳定性与鲁棒性。

2.1.2 实验设置与评价指标

为确保公平比较,本文所有基于深度学习的海洋机器人视觉感知增强模型均使用 PyTorch 框架实现。网络优化采用 Adam 优化器,并使用默认参数设置。训练批次大小设为 8,总训练轮数为 140 轮,初始学习率为 0.001,每 20 轮将学习率乘以 0.04 进行衰减。所有实验均在 NVIDIA RTX 4090 GPU 的环境中进行。该环境不仅能够保证模型训练过程中的高效计算性能,同时也为多组对比实验提供了统一的运行条件,从而确保实验结果的可重复性和公平性。

为了验证本文方法的优越性,本文选取了多种评价指标进行全面实验。具体而言,采用了包括峰值信噪比 (Peak signal-to-noise ratio, PSNR)、结构相似性指数 (Structural similarity index measure, SSIM)、感知质量指标 (Perceptual quality index, PCQI) 以及感知相似性指数 (Learned perceptual image patch similarity, LPIPS) 在内的常用指标,以对方法性能进行多维度评估。其中,所选指标兼顾了全参考评价与感知质量评价,能够全面反映图像增强效果的客观质量与视觉感知性能。此外,这些指标覆盖了纹理保真度、结构保持能力以及人眼对视觉细节的主观感知差异,从而能够更准确地揭示模型在不同层面上的优势。同时,多指标评价机制也有效避免了单一指标带来的偏差,使性能分析更加全面客观。

2.2 对比实验

2.2.1 实验结果

将本文所提方法与最近所发表的先进的 9 种深度学习方法进行比较,包括 UIEC² Net、Shallow - UWNet^[15]、UHD - SFNet^[16]、SurroundNet^[17]、ADNet^[18]、LiteEnhanceNet^[19]、CFENet、DSC^[20]、DFF - Net^[21]。为保证比较结果的公平性,进行评估的预测图由原作者提供的源代码生成。表 1 显示了本文方法与 9 个对比方法的定量比较结果。可以看出,本文方法在上述公开数据集上的多项评价指标上均取得了较为优异的结果,整体性能优于现有对比方法。尽管 ADNet 在部分数据集和指标上表现接近本文方法,但从单个数据集及综合性能来看,本文方法在整体上仍保持了更稳定且更具优势的表现。

表 1 在 3 个基准数据集上本文 TF - Net 与其他方法比较结果

方法	年份	LSUI				EUVP				UFO			
		PSNR/dB	SSIM	LPIPS	PCQI	PSNR/dB	SSIM	LPIPS	PCQI	PSNR/dB	SSIM	LPIPS	PCQI
UIEC ² Net	2021	25.060	0.846	0.303	0.795	25.002	0.812	0.323	0.751	23.490	0.740	0.384	0.730
Shallow - UWNet	2021	22.807	0.800	0.333	0.864	26.532	0.833	0.272	0.794	24.793	0.767	0.337	0.775
UHD - SFNet	2022	22.905	0.785	0.368	0.839	26.465	0.819	0.305	0.784	24.603	0.750	0.371	0.762
SurroundNet	2023	24.458	0.849	0.284	0.758	26.132	0.838	0.267	0.747	24.657	0.778	0.336	0.725
ADNet	2024	25.576	0.853	0.275	0.680	26.813	0.834	0.263	0.660	25.091	0.767	0.331	0.616
LiteEnhanceNet	2024	24.685	0.857	0.268	0.796	24.033	0.823	0.276	0.742	22.074	0.758	0.346	0.705
CFENet	2024	25.062	0.849	0.326	0.789	23.462	0.811	0.339	0.716	24.071	0.759	0.385	0.721
DSC	2024	23.044	0.831	0.309	0.792	23.636	0.806	0.332	0.732	21.957	0.736	0.394	0.707
DFF - Net	2025	23.295	0.797	0.327	0.862	26.712	0.835	0.269	0.791	25.138	0.770	0.341	0.771
TF - Net(本文)	—	26.098	0.862	0.264	0.874	27.370	0.845	0.259	0.829	25.609	0.781	0.328	0.809

2.2.2 实验结果分析

实验结果表明,本文方法在 LSUI、EUVP 和 UFO 3 个数据集上均取得了最佳表现。在上述数据集中,本文方法在 PSNR、SSIM、LPIPS 及 PCQI 4 项评价指标上均优于对比方法。其中,PSNR 的提升表明增强图像在像素重建方面具有更高的一致性,SSIM 的提高反映了图像结构信息得到了更有效的保持;LPIPS 指标的降低说明生成结果在感知质量上更符合人类视觉特性,而 PCQI 的提升则表明图像在对比度、清晰度及整体视觉质量方面均得到了显著改善。综合来看,多项指标的一致领先充分验

证了本文方法在海洋机器人任务中的有效性与先进性,表明其在不同水下场景和复杂成像条件下均具有良好的鲁棒性和泛化能力。如图 6 所示,与对比方法相比,本文方法在多种典型水下场景中均能有效缓解颜色偏移问题,恢复更加自然合理的色彩分布,同时在目标边缘和纹理区域保持了较好的结构完整性,避免了过增强或细节模糊等现象。这些视觉效果上的优势与定量实验中 PSNR、SSIM、LPIPS、PCQI 等指标的最优结果保持一致,从定性和定量两个层面共同验证了本文方法在海洋机器人任务中的有效性。

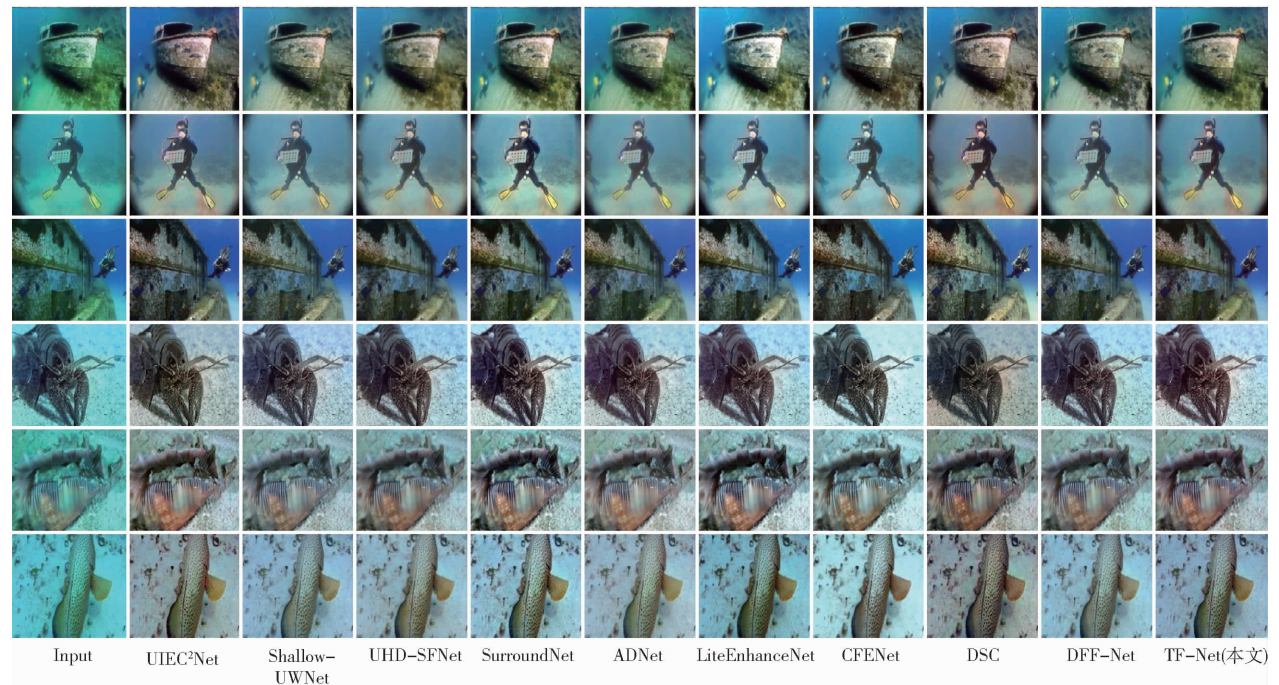


图 6 TF - Net 和其他算法在 UFO(前 3 行)和 EUVP(最后 3 行)数据集上的比较

Fig. 6 Comparison of TF - Net with other methods on UFO dataset (first three rows) and EUVP dataset (last three rows)

2.3 消融实验

2.3.1 融合模块有效性消融研究

为验证模型中各模块的有效性,对整体网络进

行了消融实验,通过逐一移除关键组件(如跨域协同模块、偏振残差模块等),并比较不同配置下的性能变化,以分析各模块对最终效果的实际贡献。实

验结果如表 2 所示,其中表格中列出的各项分别对应去除相应模块后的模型结构,由此可以直观地观察到每个模块对于增强性能的影响程度。

表 2 不同模块的消融实验结果

Tab.2 Ablation experimental results of different modules

方法	PSNR/dB	SSIM	LPIPS	PCQI
CDCM	25.490	0.859	0.267	0.707
PRM	26.066	0.861	0.265	0.708
DACM	25.426	0.860	0.271	0.734
MSRM	25.917	0.860	0.275	0.706
TF-Net(本文)	26.098	0.862	0.264	0.874

从实验结果可以看出,当单独移除任一模块时,模型在多项指标上均出现不同程度的下降,这说明各模块在整体网络中均承担着不可替代的功能。而当全部模块共同参与特征提取过程时,网络能够取得最优的整体表现,进一步表明这些模块在应对复杂水下退化问题时具有协同增益作用。此外,消融实验结果也为模块设计的合理性提供了有力支撑,验证了其在提升海洋机器人增强质量方面的重要性和必要性。进一步地,这些分析也揭示了多模块协同建模在处理多样化的水下光学退化场景中的优势。多模块协同在未来更复杂的海洋机器人任务与自主感知系统中具有广泛的应用潜力。

2.3.2 替换模块有效性消融研究

为验证模型双注意力耦合模块的有效性,将该模块中的注意力机制分别替换为 SK^[22]、ECA^[23]、SE^[24]以及 CA^[25]等方法。如表 3 所示,不同注意力机制在相同实验设置下表现出明显差异,所提出的双注意力耦合模块在各项评价指标上均取得了最优或次优结果。

表 3 替换双注意力耦合模块的消融实验结果

Tab.3 Ablation experimental results of replacing dual attention coupling module

方法	PSNR/dB	SSIM	LPIPS	PCQI
SK	21.215	0.771	0.298	0.752
ECA	24.244	0.857	0.264	0.746
SE	25.981	0.862	0.262	0.717
CA	26.115	0.859	0.264	0.687
TF-Net(本文)	26.098	0.862	0.264	0.874

2.4 应用测试

为进一步评估本文方法与下游任务的兼容性,将不同海洋机器人视觉感知增强方法生成的增强图像应用于两种典型的水下应用场景,具体包括水下边缘检测和水下显著性目标检测。通过对增强图像在这些下游任务中的表现进行对比分析,可更全面地验证本文方法在海洋机器人视觉感知中的实用价值与适应性。

2.4.1 水下边缘检测应用测试

图 7 水下边缘检测结果展示了基于 Canny 算法的检测结果。可以观察到,不同方法在物体结构保持与细节还原方面存在明显差异。部分方法虽然生成的图像在整体亮度上有所改善,但边缘检测结果中出现了较多断裂,反映出在细节恢复与结构保持方面仍存在不足。相比之下,本文方法在增强后的图像中能够更好地保留目标的边缘轮廓与细节信息,边缘检测结果更加清晰、完整且连续,有效避免了过度平滑或细节丢失的问题。

2.4.2 水下显著性应用测试

另一方面,图 7 水下显著性检测结果通过 HDNet 检测算法展示了不同海洋机器人视觉感知增

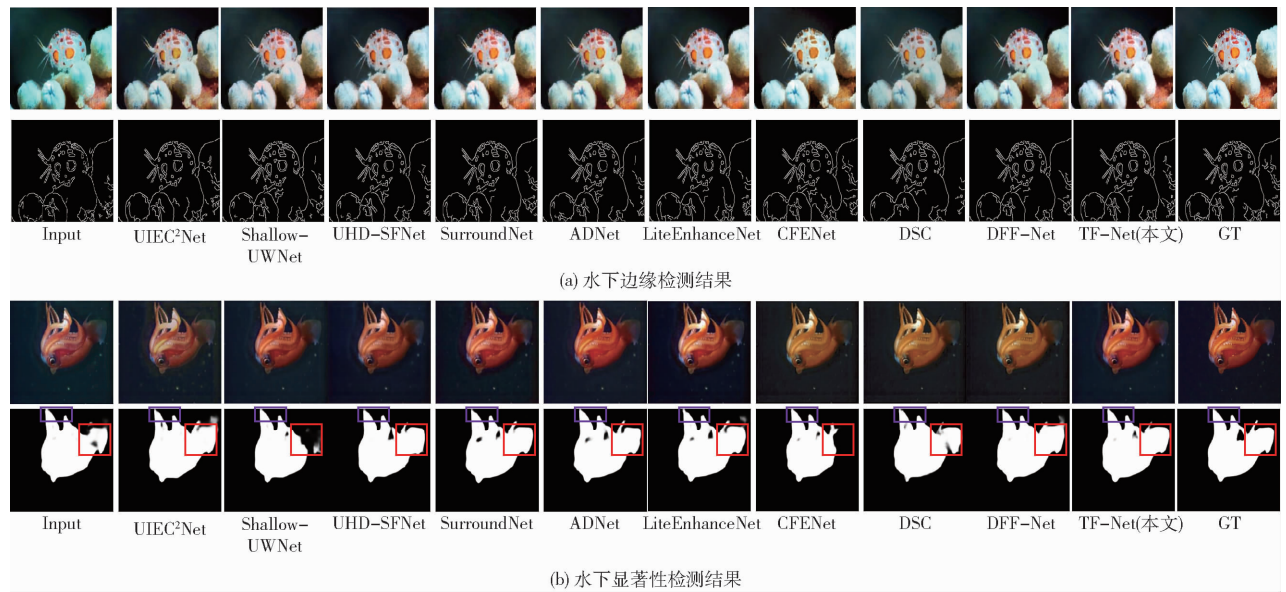


图 7 TF-Net 和其他的对比算法在 UFO 数据集上的水下边缘检测结果和水下显著性检测结果

Fig.7 Underwater edge detection and saliency detection results of TF-Net and other comparative algorithms on UFO dataset

强方法在显著性目标检测任务中的表现。结果表明,本文方法生成的增强图像能够显著提升检测效果,不仅在大尺度目标(红框中白色区域)的轮廓识别上更为准确,同时在小尺度目标(紫框中白色区域)的边缘细节检测上也具有更高的鲁棒性。

综上所述,本文方法在水下边缘检测与显著性目标检测任务中均展现出优越的性能,进一步证明了其在提升下游任务表现方面的有效性和在实际水下应用中的广阔潜力。

3 结束语

本文提出了一种基于时域-频域协同引导的海

洋机器人视觉感知增强网络,通过时域-频域特征提取阶段和多维度特征细化阶段,实现了对水下图像细节与整体质量的有效提升。在时域-频域特征提取阶段,本文分别对时域与频域进行特征提取并引入跨域协同模块对时频信息进行交互,随后,多维度特征细化阶段通过引入偏振残差模块、双注意力耦合模块以及多尺度残差模块,实现了对水下图像方向特征、注意力信息及多尺度感受野的精细增强,从而提升了图像的局部细节、全局结构及整体视觉感知质量。综上所述,本文方法能够在复杂水下环境中缓解光学退化问题,提升视觉质量,为水下目标识别、定位等关键任务提供更加可靠的感知输入。

参 考 文 献

[1] Esmailzadeh A, Ou Y, Ahmad M O, et al. DMML: deep multi-prior and multi-discriminator learning for underwater image enhancement[J]. IEEE Transactions on Broadcasting, 2024, 70(2): 637 – 653.

[2] Li C, Guo C, Ren W, et al. An underwater image enhancement benchmark dataset and beyond[J]. IEEE Transactions on Image Processing, 2019, 29: 4376 – 4389.

[3] Lu Y, Yang D, Gao Y, et al. AoSRNet: all-in-one scene recovery networks via multi-knowledge integration[J]. Knowledge-Based Systems, 2024, 294: 111786.

[4] Ren S, Bao X, Wang T, et al. UIEVUS: an underwater image enhancement method for various underwater scenes[J]. Signal Processing: Image Communication, 2025, 135: 117264.

[5] Wang B, Xu H, Jiang G, et al. UIE-Conformer: underwater image enhancement based on convolution and feature fusion transformer[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(2): 1952 – 1968.

[6] 郭继昌,李重仪,郭春乐,等. 水下图像增强和复原方法研究进展[J]. 中国图象图形学报, 2017, 22(3): 273 – 287.

Guo Jichang, Li Chongyi, Guo Chunle, et al. Research progress of underwater image enhancement and restoration methods[J]. Journal of Image and Graphics, 2017, 22(3): 273 – 287. (in Chinese)

[7] 陈龙,丁丹丹. 多残差联合学习的水下图像增强[J]. 中国图象图形学报, 2022, 27(5): 1577 – 1588.

Chen Long, Ding Dandan. Joint multi-residual learning for underwater image enhancement[J]. Journal of Image and Graphics, 2022, 27(5): 1577 – 1588. (in Chinese)

[8] Zhang D, Zhou J, Zhang W, et al. ReX – Net: a reflectance-guided underwater image enhancement network for extreme scenarios[J]. Expert Systems with Applications, 2023, 231: 120842.

[9] Ji X, Wang X, Leng N, et al. Dual-branch underwater image enhancement network via multiscale neighborhood interaction attention learning[J]. Image and Vision Computing, 2024, 151: 105256.

[10] 岳成海,徐会希,吕凤天,等. 基于光照补偿与金字塔融合的水下图像增强方法[J]. 水下无人系统学报, 2025, 33(1): 46 – 55.

Yue Chenghai, Xu Huixi, Lü Fengtian, et al. Underwater image enhancement method based on illumination compensation and pyramid-based blending[J]. Journal of Unmanned Undersea Systems, 2025, 33(1): 46 – 55. (in Chinese)

[11] Cong X, Zhao Y, Gui J, et al. A comprehensive survey on underwater image enhancement based on deep learning[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2026, 10(4): 2783 – 2801.

[12] Peng L, Zhu C, Bian L. U-shape transformer for underwater image enhancement[J]. IEEE Transactions on Image Processing, 2023, 32: 3066 – 3079.

[13] Islam M J, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 3227 – 3234.

[14] Islam M J, Luo P G, Sattar J. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception[C]//Conference on Robotic-Science and Systems(RSS). ELECTRONETWORK: IEEE, 2020.

[15] Naik A, Swarnakar A, Mittal K. Shallow-UWnet: compressed model for underwater image enhancement (student abstract)[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(18): 15853 – 15854.

[16] Wei Y, Zheng Z, Jia X. Uhd underwater image enhancement via frequency-spatial domain aware network[C]//Proceedings of the Asian Conference on Computer Vision, 2022: 299 – 314.

[17] Zhou F, Sun X, Dong J, et al. SurroundNet: towards effective low-light image enhancement[J]. Pattern Recognition, 2023, 141: 109602.

[17] Thakur P S, Khanna P, Sheorey T, et al. Vision transformer for plant disease detection: PlantViT[C] // International Conference on Computer Vision and Image Processing. Cham: Springer International Publishing, 2021: 501 – 511.

[18] 陈永年,马骏,袁哲明,等. 春甘蓝外叶和球叶损失对产量的影响及主要食叶害虫防治行动阈值[J]. 湖南农业大学学报(自然科学版), 2002(4): 308 – 313.
Chen Yongnian, Ma Jun, Yuan Zheming, et al. Effects of outer leaf and bulb loss on yield and thresholds for action against major leaf-feeding pests in spring kale[J]. Journal of Hunan Agricultural University (Natural Science Edition), 2002(4): 308 – 313. (in Chinese)

[19] Balakrishnan B, Chelliah R, Venkatesan M, et al. Comparative study on various architectures of yolo models used in object recognition[C] // 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, 2022: 685 – 690.

[20] 李振波,刘欣,竺明昊,等. 面向种植农业复杂场景的 YOLO 算法应用综述[J]. 农业机械学报,2026,57(13):45 – 67.
Li Zhenbo, Liu Xin, Zhu Minghao, et al. A review of YOLO algorithm applications in complex scenarios of crop farming[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(13): 45 – 67. (in Chinese)

[21] Han X, Chang J, Wang K. You only look once: unified, real-time object detection[J]. Procedia Computer Science, 2021, 183(1): 61 – 72.

[22] Mo H, Wei L. Tomato yellow leaf curl virus detection based on cross-domain shared attention and enhanced BiFPN[J]. Ecological Informatics, 2025, 85: 102912.

[23] Doherty J, Gardiner B, Kerr E, et al. BiFPN-yolo: one-stage object detection integrating bi-directional feature pyramid networks[J]. Pattern Recognition, 2025, 160: 111209.

[24] Park H, Yoo Y, Seo G, et al. C3: concentrated-comprehensive convolution and its application to semantic segmentation[PP/OL]. V3. arXiv (2019 – 7 – 28) [2025 – 6 – 15]. <https://arxiv.org/abs/1812.04920>.

[25] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021: 13713 – 13722.

[26] Liang H, Wang H, Mao L, et al. Bone stick image classification study based on C3CA attention mechanism enhanced deep cascade network[J]. IEEE Access, 2023, 11: 94057 – 94068.

[27] Varghese R, Sambath M. Yolov8: a novel object detection algorithm with enhanced performance and robustness[C] // 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024: 1 – 6.

(上接第 102 页)

[18] Park C W, Eom I K. Underwater image enhancement using adaptive standardization and normalization networks[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107445.

[19] Zhang S, Zhao S, An D, et al. LiteEnhanceNet: a lightweight network for real-time single underwater image enhancement[J]. Expert Systems with Applications, 2024, 240: 122546.

[20] Lu X, Yuan Y, Liu X, et al. Low-light salient object detection by learning to highlight the foreground objects[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(8): 7712 – 7724.

[21] Zhang H, Wang L, Zou Q, et al. DFF – Net: deep feature fusion network for low-light image enhancement[J]. Image and Vision Computing, 2025, 161: 105645.

[22] Li X, Wang W, Hu X, et al. Selective kernel networks[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 510 – 519.

[23] Wang Q, Wu B, Zhu P, et al. ECA – Net: efficient channel attention for deep convolutional neural networks[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11534 – 11542.

[24] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7132 – 7141.

[25] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.