

DOI:10.3876/j.issn.1000-1980.2026.05.006

基于多通道偏振与双分支 Transformer 的水下图像复原

许馨月, 霍冠英, 孔伟丰, 彭超, 苏勇

(河海大学信息科学与工程学院)

摘要: 针对水下浑浊环境引起的图像对比度低、色彩失真及细节模糊问题, 提出一种基于多通道偏振与双分支 Transformer 的水下图像复原网络。该网络采用基于 Stokes 矢量的偏振特征自适应融合模块提取水体偏振物理特征, 然后通过构建空间-偏振双分支架构, 利用动态门控机制自适应融合偏振强度、偏振度和偏振角等多源信息与空间特征, 克服单一模态信息表征不全面的局限, 且网络主干采用集成多尺度特征交互、通道注意力及多尺度空间 Transformer 模块的编解码结构, 强化退化表征与全局细节还原能力。实验结果表明, 该方法在结构还原、细节保持和色彩校正方面优势显著(SSIM 为 0.8540, PSNR 为 32.93 dB, 感知指标 LPIPS 为 0.3633), 能显著提升复原图像的清晰度、对比度和色彩保真度。

关键词: 水下图像复原; 多通道偏振; Transformer; 自适应融合; 门控机制; 偏振物理先验

中图分类号: TP391.41

文献标志码: A

文章编号: 1000-1980(2026)05-0049-11

Underwater image restoration based on multi-channel polarization and dual-branch Transformer

Xu Xinyue, Huo Guanying, Kong Weifeng, Peng Chao, Su Yong

(College of Information Science and Engineering, Hohai University)

Abstract: To address the problems of low image contrast, color distortion, and blurred details caused by turbid underwater environments, an underwater image restoration network based on multi-channel polarization and a dual-branch Transformer was proposed. A polarization feature adaptive fusion module based on the Stokes vector was used to extract the physical polarization features of water. A spatial-polarization dual-branch architecture was constructed, and a dynamic gating mechanism was utilized to adaptively fuse multi-source information, such as polarization intensity, degree of polarization, and polarization angle, with spatial features, overcoming the limitation of incomplete representation by single-modality information. An encoder-decoder structure integrating multi-scale feature interaction, channel attention, and multi-scale spatial Transformer modules was adopted for the network backbone to enhance the capabilities of degradation representation and global detail restoration. The experimental results show that this method has significant advantages in structure restoration, detail preservation, and color correction (SSIM: 0.8540, PSNR: 32.93 dB, perceptual metric LPIPS: 0.3633), and can significantly improve the clarity, contrast, and color fidelity of the restored images.

Key words: underwater image restoration; multi-channel polarization; Transformer; adaptive fusion; gating mechanism; polarization physical prior

水下光学成像广泛应用于目标检测、三维重建、环境监测和自动驾驶等领域^[1]。然而, 由于水下光线衰减以及散射现象严重, 水下成像系统所采集到的图像存在严重的色彩失真、曝光度不足和图像模糊等问题^[2], 极大地限制了后续高层视觉任务的准确性和鲁棒性。因此, 如何有效提升浑浊水下环境的图像质量, 实现高保真度图像复原, 已成为当前水下视觉领域亟须解决的关键科学问题。

目前, 水下图像复原的主流方法分为基于物理模型和基于深度学习两类。基于物理模型的复原方法通过优化像素映射关系提升效果^[3]。例如: Yi 等^[4]提出自适应色彩校正、耦合性解耦增强及暗通道加速架构,

基金项目: 国家重点研发计划项目(2023YFB3907203); 国家自然科学基金“叶企孙”科学基金项目(U2441254); 中央高校基本科研业务费专项资金项目(B250201257)

作者简介: 许馨月(2000—), 女, 硕士研究生, 主要从事水下图像处理研究。E-mail: hhuxxy@hhu.edu.cn

通信作者: 霍冠英(1979—), 男, 教授, 博士, 主要从事水下声呐图像处理研究。E-mail: huoguanying@hhu.edu.cn

协同优化色彩保真度、边缘细节与实时性;许丽等^[5]通过设计梯度驱动多尺度自适应算法平衡颜色保真与细节增强;Sun 等^[6]利用隐式表征协同机制抑制暗区伪影与色偏。基于深度学习的复原方法通过学习退化特征映射改善复原质量。例如:Liu 等^[7]在 Retinex-NET 嵌入通道注意力机制,联合去噪损失和颜色损失,提出了一种改进的弱光照图像增强算法;Jiang 等^[8]设计了三阶段级联 DEANET++架构治理退化模式与噪声残留;杨海潮等^[9]采用双分支 CycleGAN 进行解耦内容与亮度增强;Cui 等^[10]基于可分离卷积实现轻量化 ConvIR 架构,降低了 75%的计算负载;Ouyang 等^[11]提出了 Transformer-卷积混合的 UformerGAN 范式实现端到端重建。上述方法虽各有突破,但物理模型受限于复杂退化建模能力,单一特征深度学习方法缺乏物理光学信息,难以区分真实结构与噪声,复原效果均有局限。同时,现有方法未充分融入场景光学特性,复原质量仍有提升空间,设计高效鲁棒的多源信息融合高保真复原方法是当前亟待解决的关键问题。

鉴于此,本文在传统卷积模型与单一特征输入网络的基础上进行改进^[12],构建了基于多通道偏振与双分支 Transformer 的融合网络:设计了基于 Stokes 矢量^[13]的偏振特征自适应融合模块,构建了以多通道 Stokes 矢量为核心输入的空间-偏振双分支架构,以实现端到端的水下退化建模与复原;同时,提出动态门控自适应融合机制,利用可学习的门控函数^[14]动态加权融合偏振物理特征与空间分支特征,以解决多源信息干扰与单一模态表征不全面的问题;此外,引入全局-多尺度特征增强机制,通过多头自注意力、多尺度特征交互与通道注意力的协同设计,集成全局依赖建模、局部细节感知与通道信息筛选能力,以强化退化表征与全局细节还原能力。

1 研究方法

1.1 网络总体结构

本文提出的水下图像复原网络以多通道偏振特征为输入,采用双分支编码-解码架构,由偏振特征自适应融合模块、空间分支、门控融合模块、编解码模块、全局-多尺度特征增强模块构成,整体结构如图 1 所示,图中 B 为批量样本数, H 、 W 分别为特征图的高度与宽度。

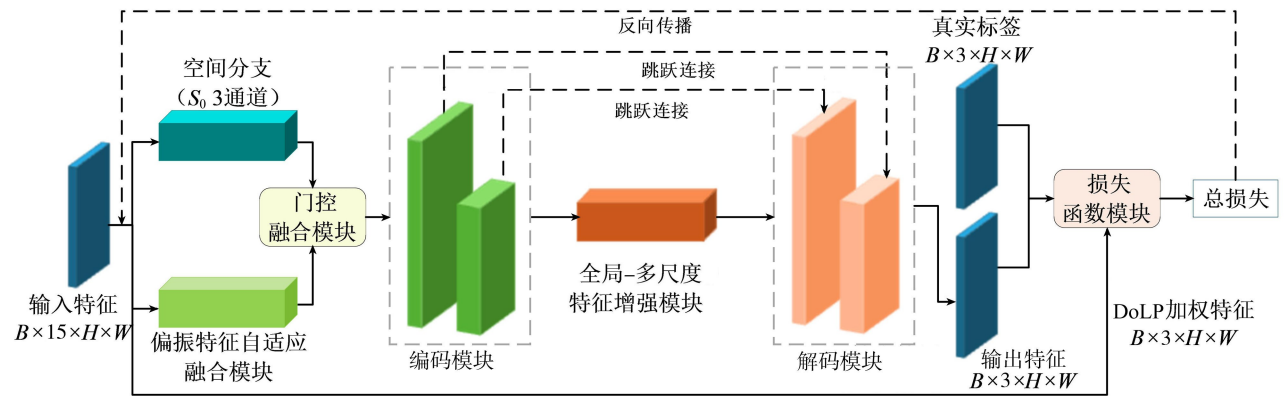


图 1 网络架构

Fig. 1 Network architecture diagram

图像复原核心流程为:①基于 0° 、 45° 、 90° 和 135° 四角度偏振图像计算得到多维度偏振物理特征,构建 15 通道高维输入张量(表 1);②通过偏振特征自适应融合模块完成偏振信息的特征提炼,同时由空间分支提取基础空间结构特征;③双分支特征经编码后送入门控融合模块,实现偏振物理信息与空间信息的自适应加权融合;④融合特征经全局-多尺度特征增强模块进一步强化,从而提升全局依赖建模与多尺度上下文表达能力;⑤通过解码模块逐步恢复空间分辨率,输出复原图像。网络采用端到端方式训练,通过多分量损失函数约束全局结构一致性与细节还原效果。

1.2 偏振特征自适应融合模块

为缓解浑浊水体下偏振信息利用不充分的问题,本文设计了偏振特征自适应融合模块(图 2)。模块以

表 1 15 通道特征张量构成

Table 1 Composition of 15-channel feature tensor

通道索引	特征	物理意义	数据维度
0~2	S_0	总光强度,包含 RGB 三通道	3 通道
3~5	S_1	水平与垂直偏振分量的差异	3 通道
6~8	S_2	45° 与 135° 偏振分量的差异	3 通道
9~11	DoLP	线偏振度,反映偏振强度	3 通道
12~14	AOP	偏振角度,反映偏振方向	3 通道

0° 、 45° 、 90° 和 135° 偏振方向的水下图像的辐射光强 I_0 、 I_{45} 、 I_{90} 和 I_{135} 作为输入,通过式(1)~(3)计算得到 Stokes 参数(S_0 、 S_1 、 S_2)以及线偏振度(D_{oLP})与偏振角(A_{OP})。

$$S = \begin{pmatrix} \hat{e}_0 S_0 \\ \hat{e}_1 S_1 \\ \hat{e}_2 S_2 \end{pmatrix} = \begin{pmatrix} \hat{e}_0 (I_0 + I_{90}) \\ \hat{e}_1 (I_0 - I_{90}) \\ \hat{e}_2 (I_{45} - I_{135}) \end{pmatrix} \quad (1)$$

$$D_{oLP} = \sqrt{S_1^2 + S_2^2} / S_0 \quad (2)$$

$$A_{OP} = \frac{1}{2} \arctan(S_2 / S_1) \quad (3)$$

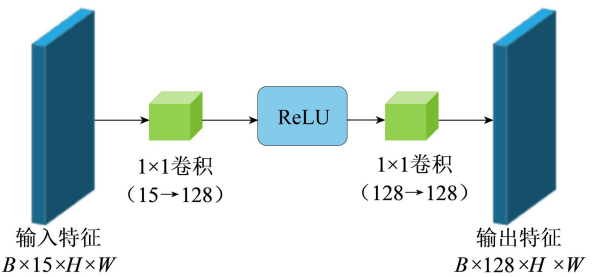


图 2 偏振特征自适应融合模块

Fig. 2 Polarization feature adaptive fusion module

偏振特征自适应融合模块的核心运算是将上述所有物理量连接形成高维张量 $X \in \mathbf{R}^{C \times H \times W}$,之后再通过两级 1×1 卷积与 ReLU 激活函数的非线性变换处理获得高维特征表达。 1×1 卷积核的权重在训练过程中会根据水下图像复原效果不断调整,从而使得输出特征在高维空间中能够更好地表达与任务相关的偏振信息。ReLU 激活函数则进一步引入非线性,有助于提升模型的表达能力和鲁棒性。最终,经过该模块处理后的特征不仅融合了多种偏振信息,还具备了更强的判别能力,为后续的门控融合等模块提供了高质量的输入。

1.3 门控融合模块

为实现空间分支与偏振特征自适应融合模块的深度互补融合,本文设计了门控融合模块(图 3)。该模块通过引入门控机制,能够根据输入特征的空间分布和内容差异,自适应调整两路特征的融合比例,从而提升网络对多源信息的利用效率和特征选择能力。

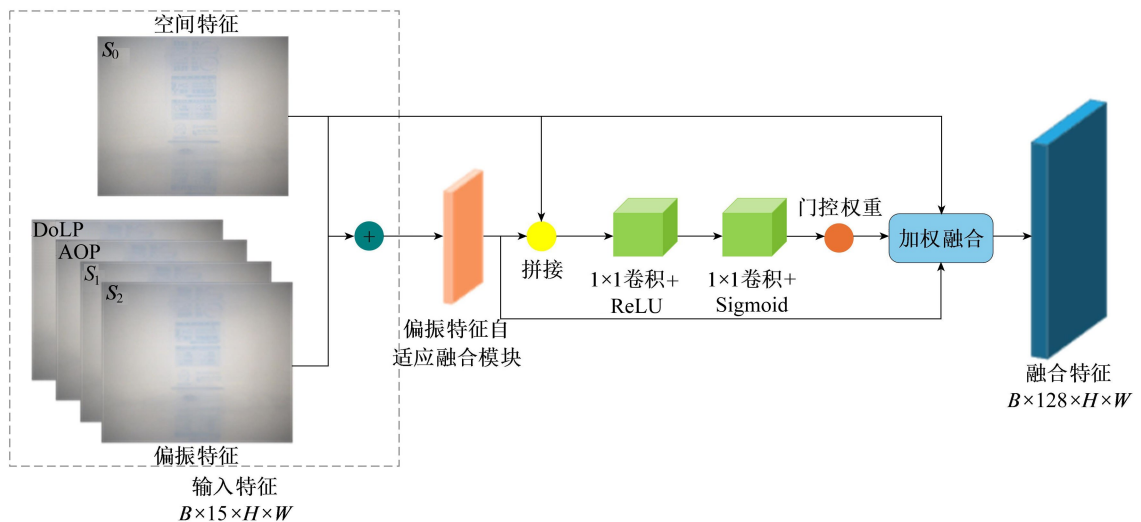


图 3 门控融合模块

Fig. 3 Gated fusion module

设空间分支特征为 $F_{\text{spatial}} \in \mathbf{R}^{C \times H \times W}$ (C 为空间分支特征图通道数),偏振特征自适应融合输出 $F_{\text{polar}} \in \mathbf{R}^{C \times H \times W}$ 。首先,将 F_{spatial} 和 F_{polar} 在通道维度拼接,获得融合特征 F_{cat} 。然后,依次通过 1×1 卷积和 ReLU 激活、 1×1 卷积和 Sigmoid 函数^[15],生成门控权重图 $G \in \mathbf{R}^{1 \times H \times W}$ 。门控权重 G 是通过卷积和 Sigmoid 激活动态生成的,能针对不同输入自适应分配信息流,取值范围为 $[0, 1]$,反映了在不同空间位置对偏振特征的偏好程度,用于自适应调节两路特征的融合。最后,融合输出特征 F_{fuse} 由门控加权相加获得,即:

$$F_{\text{fuse}} = G \odot F_{\text{polar}} + (1 - G) \odot F_{\text{spatial}} \quad (4)$$

式中: $\odot$ 表示逐元素乘法。

门控融合模块的设计核心在于其自适应融合机制。加权方式由 G 和 $1-G$ 动态调节,能够根据特征图的不同空间位置,自适应选择更优的特征来源,可有效实现空间特征信息 F_{spatial} 与偏振物理信息 F_{polar} 的互补与融合。门控融合模块通过可学习的门控权重 G ,提升网络对多源特征的选择和表达能力。相较简单拼接或

元素加和,该机制能依据输入局部特性灵活调整融合策略,增强网络对复杂水下退化场景的适应性与鲁棒性。

1.4 编解码模块

为实现退化图像从特征提取到结果重建的完整过程,本文构建编解码模块,该模块主要用于提取多尺度特征并逐步恢复图像空间分辨率。

1.4.1 编码模块

编码模块的设计如图 4 所示,是本文主干网络中用于多尺度特征提取和空间信息压缩的关键结构。该模块通过多层残差卷积块与下采样操作的交替堆叠,实现对输入特征的逐步抽象和高层表达。每个残差卷积块后均接入 2×2 平均池化实现空间下采样。编码模块通过多层残差卷积块与下采样的交替堆叠,可有效提升网络对复杂水下场景的特征抽象和多尺度表达能力。

残差卷积块结构如图 5 所示,包含两层 3×3 卷积、批归一化、ReLU 激活和残差连接,能够有效缓解梯度消失,提升特征表达能力,其具体流程为:输入特征先经第一层 3×3 卷积提取特征,依次通过批归一化与 ReLU 激活得到中间特征;再经第二层 3×3 卷积与批归一化运算后,与原始输入特征完成残差加和,最终经 ReLU 激活输出。若输入输出通道数不匹配,可通过 1×1 卷积完成维度适配。

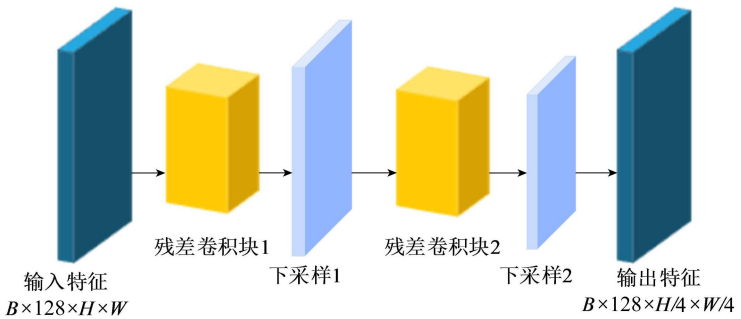


图 4 编码模块
Fig. 4 Encoder module

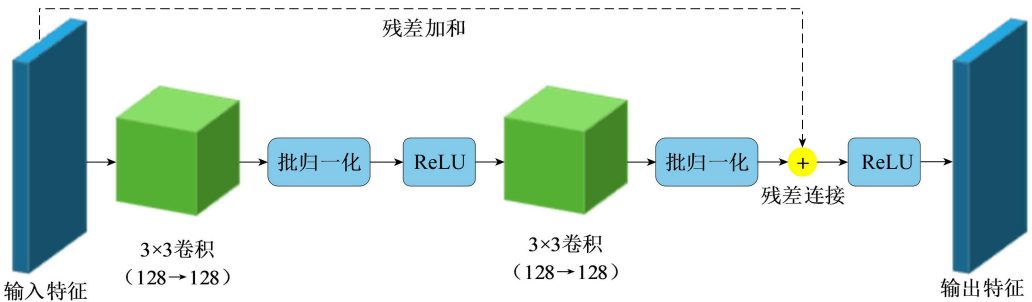


图 5 残差卷积块结构
Fig. 5 Structure of residual convolutional block

1.4.2 解码模块

解码模块的设计如图 6 所示,是本文主干网络实现逐步恢复空间分辨率和细节信息的关键结构,其核心是通过多层上采样与残差卷积块的交替堆叠,并结合编码模块的跳跃连接,实现高分辨率特征的有效重建与多尺度信息的融合。解码采用两级上采样拼接架构,残差卷积块 1 将上采样 1、拼接编码器特征后的 256 通道特征压缩至 128 通道,完成第一级特征重建;输出特征经上采样 2 恢复至原始图像空间尺寸,再次与对应层级的编码器跳跃连接特征拼接,得到 256 通道融合特征后送入残差卷积块 2,处理后输出 128 通道全分辨

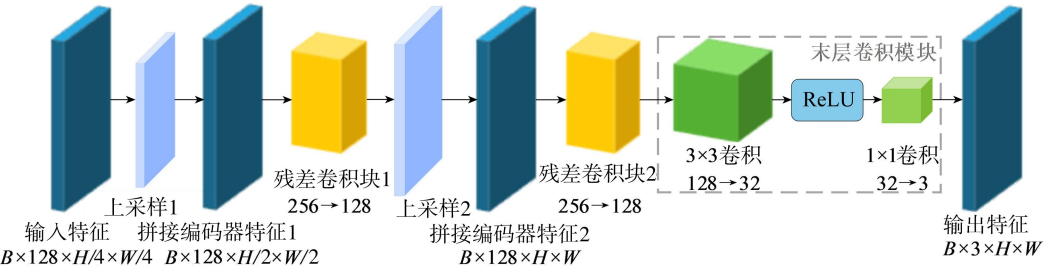


图 6 解码模块
Fig. 6 Decoder module

率特征。随后特征进入末层卷积模块,先通过 3×3 卷积将通道数从 128 降至 32,经 ReLU 激活后,再由 1×1 卷积映射为 3 通道,最终输出复原图像。

综上,编解码模块通过多尺度特征提取与逐步重建,实现了退化图像从高层语义到空间细节的有效恢复,从而为复原结果的整体质量提升提供了主体支撑。

1.5 全局-多尺度特征增强模块

为进一步改善复杂浑浊场景下的细节模糊问题,本文设计全局-多尺度特征增强模块(GMFEM),如图 7 所示。该模块主要通过增强不同尺度特征交互,兼顾全局上下文与局部细节信息,从而提高对目标轮廓、边缘区域和纹理细节的恢复能力。具体流程为:输入特征首先经过位置编码后展平为二维序列,随后输入 Transformer 模块完成全局信息交互。在 Transformer 模块内部,特征通过多头自注意力机制^[16]建立全局依赖关系,再经由前馈网络与残差连接进一步强化特征表达能力,每一层均配置层归一化^[17]以保障训练过程的稳定性与特征传递效率,Transformer 模块输出的特征会被还原为原始空间结构,以适配后续的卷积操作。

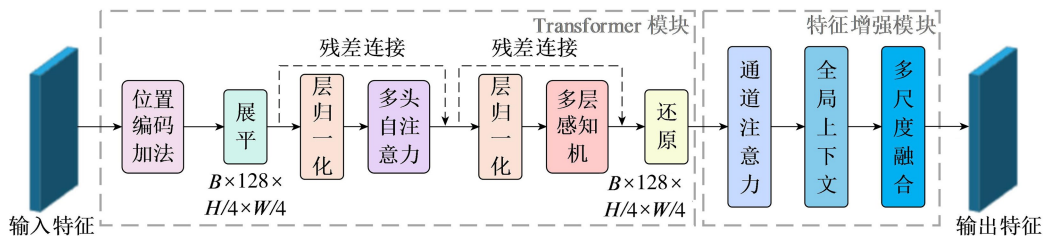


图 7 全局-多尺度特征增强模块

Fig. 7 Global-multi-scale feature enhancement module

完成全局建模后的特征进入多分支特征增强模块,该模块包含通道注意力、全局上下文与多尺度融合 3 个并行的特征增强支路。通道注意力支路通过全局平均池化聚合空间信息,再经由两层全连接层自适应生成各通道的权重,以此突出关键特征通道并抑制冗余信息。全局上下文支路利用全局池化结合卷积运算提取全局语义信息,生成对应的权重图后与原特征相乘,实现全局信息的自适应融合,强化特征的全局感知能力。多尺度融合支路则采用 1×1 、 2×2 、 4×4 尺度的自适应池化操作,将多尺度下采样得到的特征统一上采样至原始空间尺寸后进行拼接,最终通过 1×1 卷积完成多尺度特征的融合,提升模型对不同水下浑浊区域的环境适应性。经过全局-多尺度特征增强模块处理后的特征同时具备全局建模能力与多尺度上下文信息,能够为后续解码器提供更丰富且判别性更强的特征表达。

1.6 损失函数模块

为保证网络训练过程中复原结果在颜色恢复和细节保持之间取得较好平衡,本文设置损失函数模块,如图 8 所示。该模块通过 L_1 损失(L_{L1})、基于结构相似性指数(structural similarity index measure, SSIM)的损失以及基于线偏振度(DoLP)的加权机制对网络进行监督,以引导模型学习更加稳定的图像复原映射关系。

设第 i 个像素位置的预测值为 $\hat{Y}_i$,真实值为 Y_i , N 表示参与计算的像素总数,则 L_1 像素级损失 L_{L1} :

$$L_{L1} = \sum_{i=1}^N |\hat{Y}_i - Y_i| / N \quad (5)$$

SSIM 损失(L_{SSIM}),用于衡量预测图像与真实图像在结构上的一致性,其定义为:

$$L_{SSIM} = 1 - M(\hat{Y}, Y) \quad (6)$$

式中: $M(\cdot)$ 为结构相似性指数,取值范围为 $[0, 1]$,值越接近 1 表示结构越相似。由于训练目标是最小化损失,因此采用其互补形式,此时越接近 0 表示结构越相似。

DoLP 特征由输入特征第 9~11 通道得到,记为 $D = \text{mean}(F_{9-11})$ 。由于 DoLP 能够反映目标区域的偏振强度,对比度较高的区域往往对应强反射表面。因此,相比 S_0 、 S_1 、 S_2 等基础分量,DoLP 在区分目标反射与

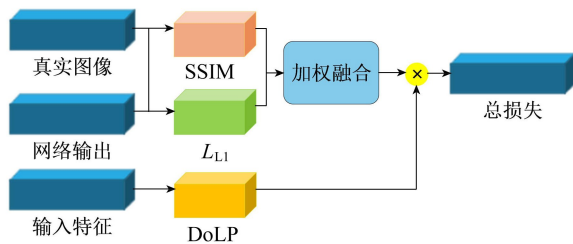


图 8 损失函数模块

Fig. 8 Loss function module

后向散射方面更具鲁棒性和信息优势。基于这一物理先验,本文选用第 9~11 通道 DoLP 特征作为损失加权因子,以引导网络更加关注物理信息丰富的区域。加权损失定义为 L_{weighted} :

$$L_{\text{weighted}} = (\alpha L_{\text{L}_1} + \beta L_{\text{SSIM}})(1 + 0.5\text{mean}(D)) \tag{7}$$

式中: $\text{mean}(D)$ 为对 DoLP 权重图进行全局平均; α, β 为 L_1 损失与 SSIM 损失的平衡权重,本文统一取 $\alpha=\beta=0.5$ 。综上,本文总损失函数为 $L_{\text{total}} = L_{\text{weighted}}$ 。损失函数模块通过联合多种约束项对网络进行监督,使模型在训练过程中能够兼顾颜色恢复、结构保持和细节重建,从而提升了复原结果的稳定性和整体质量。

2 实例验证

2.1 数据集获取与实验设置

通过搭建的水下偏振成像平台采集退化的水下偏振图像,实验在一个小型玻璃水箱中进行,水箱内注满水,并放置目标物体和封装好的偏振相机。为模拟真实的水下环境,向水箱中加入适量牛奶以制造浑浊效果。为确保清晰水下图像与浑浊水下图像之间的一一对应关系,实验过程中严格保持目标物体和相机的位置不变。共拍摄了 50 组图像,涵盖了多个浑浊度等级。每组图像的采集流程如下:①拍摄一组清晰的水下偏振图像,作为真值(GT);②通过多次缓慢加入少量牛奶溶液并充分搅拌,逐步增加水体的浑浊度。每次调整浑浊度后,均使用工具确保水体浊度均匀,然后拍摄当前条件下的水下偏振图像。实验使用的偏振相机型号为 LUCIDTRI050S-QC,配备索尼 IMX250MYR 传感器、微透镜阵列和偏振透镜,能够同时采集 4 个偏振方向($0^\circ, 45^\circ, 90^\circ, 135^\circ$)的图像,分辨率为 1224×1024 像素,像素尺寸为 $3.45 \mu\text{m} \times 3.45 \mu\text{m}$ 。通过该平台,共采集了 362 组清晰-浑浊图像对,并按照 8 : 2 的比例划分为训练集和测试集。

硬件平台为 CPU 为 Intel(R) Xeon(R) Gold 6152CPU@2.10GHzCPU 和 NVIDIA RTX4090GPU 设备,深度学习框架采用 PyTorch1.12.1+CU113,CUDA 版本为 11.3,并启用 AMP 混合精度加速训练。优化器使用 AdamW(adaptive moment estimation with weight decay),初始学习率为 1×10^{-3} ,权重衰减系数设为 1×10^{-4} ,批处理大小为 1,考虑到模型中包含 Transformer 结构,为保证模型充分收敛,设定训练 300 个 epoch,并通过验证集早停策略选取最佳模型。数据预处理包括图像除以 255 归一化至 $[0,1]$;随机水平/垂直翻转、 90° 倍数旋转和颜色抖动增强;基于偏振物理原理计算 Stokes 参数,并获得 DoLP 与 AOP,最终形成 $S_0, S_1, S_2, \text{DoLP}, \text{AOP}$ 各三通道的 15 通道输入张量。

2.2 评价指标

本文采用 SSIM、峰值信噪比(peak signal-to-noise ratio, PSNR)和学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)评估复原效果。SSIM 综合亮度、对比度与结构信息衡量图像间的结构相似程度,SSIM 越接近 1,表示复原图像与参考图像在结构分布上越一致,结构保持效果越好,对结构失真的反映能力越强。PSNR 是基于像素级均方误差计算得到的经典全参考保真指标,反映复原图像与参考图像之间的误差水平,通常以分贝(dB)为单位,PSNR 值越高,说明复原图像与参考图像之间的像素误差越小,整体失真越弱、保真度越高。LPIPS 基于 AlexNet 提取高层深度特征,从感知层面衡量图像语义差异,输入经归一化与尺寸对齐后计算,LPIPS 值越低表示两图感知相似度越高,对细节质量评价更敏感。

2.3 水下图像复原效果分析

为了验证本文所提方法的有效性,从测试集中选取 6 组涵盖不同目标材质的典型水下偏振图像为测试样本,将本文方法与近年来应用广泛的水下图像复原方法(USID-Net^[18]、HALP^[19]、PSD-NET^[20]、YOLY^[21]、SwinIR^[22]、U²PNet^[23]、DA-NET^[24]和 DBGAN^[25])进行比较,结果如图 9 和表 2 所示。

由图 9 和表 2 可知,各方法在浑浊水下图像复原中的表现呈现明显差异。USID-Net、HALP 和 SwinIR 仅能在一定程度上改善图像能见度,对严重浑浊区域的处理效果有限,且部分结果存在颜色过饱和或失真问题;YOLY、PSD-Net 和 U²PNet 的画面清晰度提升更为显著,但仍残留部分浑浊感,同时在纹理细节处易出现过锐化或伪影现象;DA-NET 和 DBGAN 展现出较强的水下复原能力,但在部分图像上存在轻度颜色偏移,或在平滑区域产生不自然纹理。本文方法在视觉效果上最接近真实清晰水下图像,不仅有效去除了严重浑浊干扰,恢复了清晰的物体轮廓和细节,并且色彩还原更为真实自然,整体画面通透、细节丰富;量化指标同样全面最优,SSIM 达到 0.8540,PSNR 达 32.93 dB,LPIPS 低至 0.3633,大幅优于次优方法,表明其在结构完整性、整体保真度、复杂纹理保真度及目标轮廓清晰度等方面均实现了最佳复原效果。

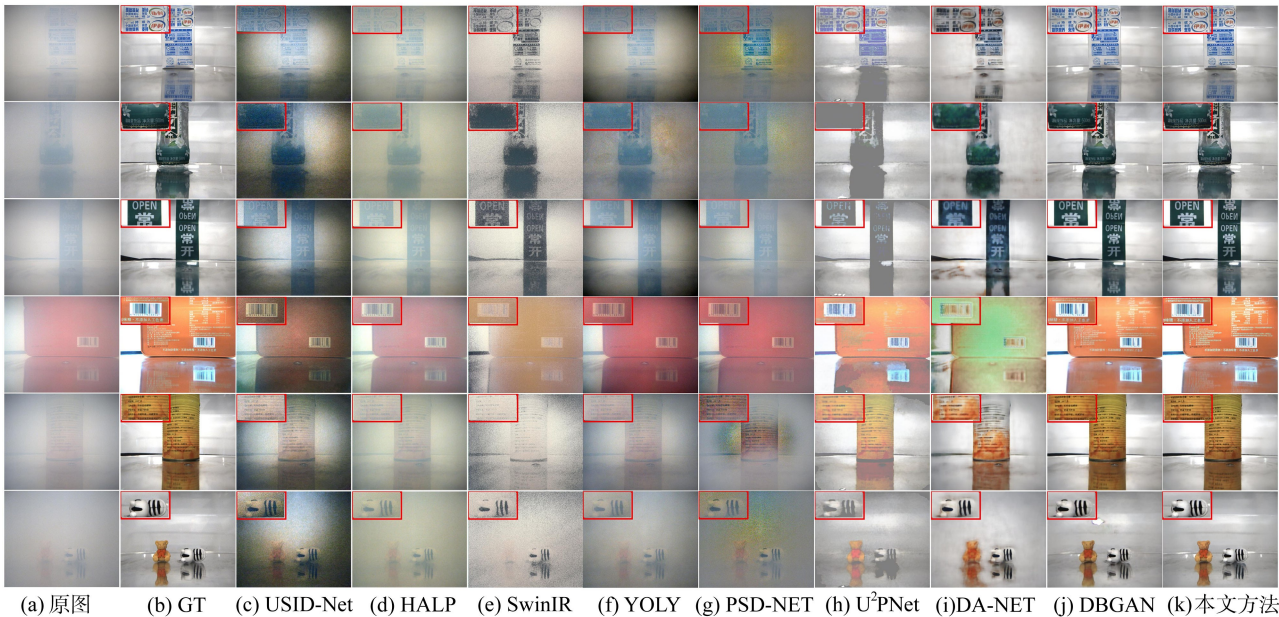


图 9 不同方法在测试集上的复原效果对比

Fig. 9 Comparison of restoration performance of different methods on test set

表 2 不同方法的效率与视觉质量评估指标对比

Table 2 Comparison of efficiency and visual quality evaluation metrics of different methods

方法	参数量/ 10^6 个	FLOPs/ 10^{12} 次	推理时间/s	显存占用/G	SSIM	PSNR/dB	LPIPS
USID-Net	14.40	0.139	2.08	14.80	0.3503	11.50	0.7040
HALP	0	0.121	0.86	0.72	0.5463	18.44	0.4991
SwinIR	11.87	4.709	90.50	20.40	0.5495	19.46	0.6367
YOLY	10.52	0.198	34.32	8.30	0.6650	17.10	0.7282
PSD-Net	7.64	0.238	42.36	8.90	0.6886	13.54	0.7370
U ² PNet	7.24	0.374	39.28	9.50	0.7511	23.91	0.5325
DA-NET	0.79	0.137	3.24	1.33	0.7606	21.99	0.4769
DBGAN	6.07	0.672	7.73	0.95	0.8255	30.45	0.4591
本文方法	2.55	1.479	34.91	5.83	0.8540	32.93	0.3633

此外,尽管本文方法在输入模态上包含偏振特征,但通过 RGB-only 对照与等信息量消融实验验证,性能提升主要来源于偏振物理信息的有效利用,而非输入维度差异,充分验证了本文方法在浑浊水下区域的优越性能。需要说明的是,HALP 是基于 MATLAB 的传统先验方法,不依赖可训练权重,所以其参数量记为 0,而 FLOPs 和时间消耗反映的是算法复杂度,而非神经网络推理。

2.4 模型效率与参数量分析

为评估模型在实际部署中的效率与可扩展性,本文分析了模型在不同输入尺寸下的推理效率特性,使用数据集中原始分辨率为 $1\,224\times1\,024$ 像素的偏振图像,通过双线性插值将输入特征调整至不同尺寸,测试分辨率涵盖 256×256 像素至 $1\,920\times1\,080$ 像素共 9 种常见尺寸。测试过程中,每个分辨率进行 50 次推理取平均值,测试结果仅包含模型前向传播时间,不包括数据加载与预处理开销,结果见表 3。

由表 3 可知,模型的 FLOPs 与输入图像的像素数量呈现出高度线性的增长关系。当分辨率从 256×256 像素增至 $1\,920\times1\,080$ 像素(像素数约增加到原来的 31.6 倍)时,FLOPs 从 77.31×10^9 次同步增长至 $2\,446.27\times10^9$ 次(约 31.6 倍),表明模型成功避免了标

表 3 不同输入尺寸下推理时间和效率对比

Table 3 Comparison of inference time and efficiency under different input sizes

图像大小/像素	FLOPs/ 10^9 次	推理时间/s	显存占用/MB
256×256	77.31	1.62	325.60
512×512	309.26	6.93	1 239.85
642×480	362.41	8.09	1 448.60
800×600	566.27	13.05	2 255.15
1 024×768	927.77	21.33	3 669.85
1 024×1 024	1 237.03	28.56	4 883.85
1 224×1 024	1 478.63	34.94	5 834.10
1 280×720	1 087.23	25.28	4 296.10
1 920×1 080	2 446.27	60.77	9 641.13

准 Transformer 架构中常见的计算复杂度二次方增长问题,具备了良好的可扩展性基础。推理时间的变化与计算复杂度趋势基本一致,但增长曲线更为平缓。分辨率提升至原来的 31.6 倍时,推理时间从 1.62 s 增加至 60.77 s(约为 37.5 倍),时间缩放效率(像素数增幅/时间增幅)约为 0.84。在 512×512 像素及 642×480 像素等中等分辨率下,推理时间分别为 6.93 s 和 8.09 s;而在 1024×768 像素分辨率下为 21.33 s,表明模型在当前实现阶段适用于对实时性要求不高的离线图像处理任务。若要满足更高帧率的实时视频处理,则需通过进一步的模型优化或硬件加速来实现。显存占用方面,本文方法同样表现出优异的可管理性。显存需求从 325.60 MB(256×256 像素)增长至 9 641.13 MB(1 920×1 080 像素)(约 29.6 倍),略低于像素数量的增长比例,体现了其亚线性的显存增长特性。

综上,本文方法在不同输入尺寸下均展现出可预测且可控的计算资源增长规律,在效率和性能之间实现了更优平衡,为在不同计算资源约束下的实际应用提供了良好的可扩展性。

2.5 消融实验

为分析本文方法中不同设计对复原性能的影响,在统一训练配置下设计两组消融实验:一组对网络结构设计进行消融分析,另一组针对损失函数设计进行消融分析。

2.5.1 网络结构设计

网络结构设计主要包括输入特征消融和核心模块消融,实验设置了基线模型、基线模型+偏振自适应融合模块、基线模型+门控融合模块、基线模型+全局-多尺度特征增强模块及所有模块联合使用(本文方法),结果如图 10 所示。单 RGB 图像作为输入模型时,复原结果整体偏暗、细节模糊,红框区域内文字与条形码均难以辨认,对应 SSIM、PSNR 和 LPIPS 分别为 0.721 5、24.38 dB 和 0.487 2,3 项指标均为最差,表明单 RGB 输入在浑浊水体场景下难以起到理想的复原效果。图 10(d)对应基线模型,整体亮度与可视性较单 RGB 输入有所改善,但细节与色彩恢复仍有限,红框内文字、条形码依旧模糊,对应 3 项指标分别为 0.743 6、25.59 dB 和 0.436 7,说明偏振多通道信息可有效补充特征、提升基础复原性能,但基础卷积结构仍难以应对浑浊水体场景。引入偏振自适应融合模块后,图 10(e)的图像质量提升显著,红框内文字边缘更清晰、条形码辨识度明显提高,该模块带来的量化提升幅度同样最大,SSIM 升至 0.799 2,PSNR 升至 28.51 dB, LPIPS 降至 0.385 3,验证了其在特征重组与动态加权上的优势,可有效增强边缘提取与细节恢复能力。引入门控融合模块后,图 10(f)清晰度较基线模型进一步提升,但罐头标签“OPEN”字符边缘仍有轻微扭曲,对应 SSIM、PSNR 和 LPIPS 分别为 0.766 0、26.63 dB 和 0.419 4,说明门控机制通过选择性信息传递优化了特征表达,但细粒度结构保护仍有优化空间。引入全局-多尺度特征增强模块后,图 10(g)细节恢复与边缘锐利度表现突出,但绿茶瓶底部仍存在倒影噪声,对应 SSIM 为 0.773 8、PSNR 为 26.52 dB、LPIPS 为 0.405 7,表明多尺度特征融合强化了细节复原能力,但对真实边缘与噪声的区分能力仍需加强。整合全部核心模块后,图 10(h)取得了最佳复原效果:视觉层面,图像整体清晰自然、色彩真实,既消除了字符边缘畸变,也显著抑制了倒影噪声,牛奶盒纹理、瓶身光泽、泰迪熊毛发等细节还原准确;量化层面,本文方法呈现出显著的协同增益效应,SSIM、PSNR 和 LPIPS 分别达到 0.854 0、32.93 dB 和 0.363 3,3 项指标均为最优,充分验证了各模块在特征融合、信息筛选、多尺度建模上的互补性与有效性。综上,本文方法能够稳定生成高质量复原图像,实现了网络输出质量与整体性能的显著提升。

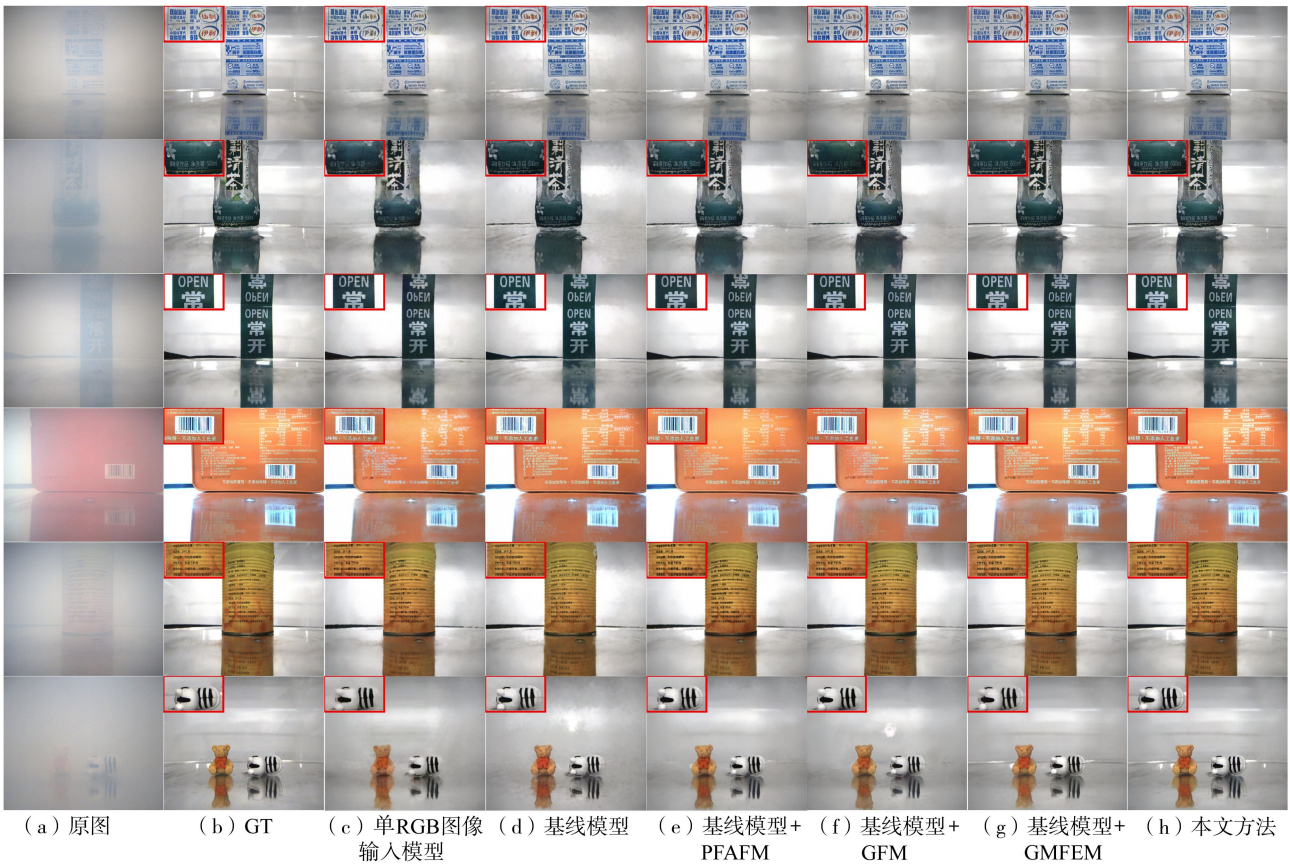
2.5.2 损失函数设计

为分析损失函数构造对模型复原性能的影响,本文围绕损失权重设置和 DoLP 加权策略开展消融实验。由表 4 可知,不同权重组合会对模型的复原性能产生明显影响。当 $\alpha=\beta=0.5$ 时,模型在 SSIM、PSNR 和 LPIPS 3 项指标上分别达到 0.854 0、32.93 dB 和 0.363 3,综合表现优于 $\alpha=0.3, \beta=0.7$ 和 $\alpha=0.7, \beta=0.3$ 两组配置。该结果表明,像素级重建约束与结构相似性约束在等权重设置下能够取得较好的平衡,从而使模型在保持图像整体保真度的同时兼顾结构一致性。若过度偏向某一项损失,容易导致模型对另一类约束的关注不足,进而影响复原结果的综合质量。

表 4 权重敏感性分析结果

Table 4 Results of weight sensitivity analysis				
α	β	SSIM	PSNR/dB	LPIPS
0.7	0.3	0.850 1	32.67	0.368 9
0.5	0.5	0.854 0	32.93	0.363 3
0.3	0.7	0.848 0	32.45	0.371 2

在 $L_{L1}+L_{SSIM}$ 的基础上引入 DoLP 加权后,模型在 SSIM、PSNR 和 LPIPS 3 项指标上均有所改善,SSIM 由



注:PFAFM 为偏振自适应融合模块,GFM 为门控融合模块,GMFEM 为全局-多尺度特征增强模块,下同。

图 10 输入特征与核心模块消融实验对比

Fig. 10 Comparison of ablation experiments on input features and core modules

0.832 1 提升至 0.854 0,PSNR 由 31.85 dB 提升至 32.93 dB,LPIPS 由 0.385 6 降至 0.363 3。该结果说明,DoLP 加权策略能够有效增强损失函数对偏振物理信息的利用能力,使网络在训练过程中更加关注物理信息丰富且结构特征明显的区域。

综上,合理的权重设置能够协调像素级重建与结构保持之间的关系,而 DoLP 加权策略则增强了模型对偏振先验的响应能力,二者共同作用,增强了损失函数对图像复原过程的约束能力。

2.6 模型可解释性

为深入解析模型的决策机制并验证网络架构的有效性,实验在测试集中选取 3 组具有不同材质特性的图像样本,从通道重要性、Grad-CAM 热力图及 Transformer 注意力分布 3 个维度,对模型开展可解释性分析,以系统评估模型在多样化场景下的感知一致性。

2.6.1 通道重要性分析

为评估各通道对最终性能的贡献,提取通道注意力模块的权重分布,并对测试样本的通道权重取平均。表 5 为权重排名前 10 的通道,Top10 通道的平均权重达到 0.73,表明约 30%的通道对复原性能具有显著影响。通道 15、78、24 的权重分别为 0.787、0.748、0.746,显著高于其他通道,说明通道注意力机制能够有效识别关键通道,自适应地增强重要特征并抑制噪声。

2.6.2 Grad-CAM 决策区域分析

为识别模型决策所关注的关键区域,采用 Grad-CAM 生成热力图。通过计算特征图对输出的梯度权重,与特征图加权求和并经 ReLU 激活后上采样,叠加至输入图像的 S_0 通道。如图 11 所示,热力图高响应区域

表 5 通道重要性统计表(前 10)

Table 5 Channel importance statistics (Top 10)

排名	通道编号	平均注意力权重	排名	通道编号	平均注意力权重
1	15	0.787	6	2	0.709
2	78	0.748	7	10	0.701
3	24	0.746	8	83	0.692
4	122	0.729	9	28	0.684
5	25	0.722	10	8	0.680

主要覆盖产品文字信息、物体整体轮廓与边缘细节,背景响应强度低,可有效抑制噪声。模型决策区域与人类视觉感知基本一致,能够从模糊输入图像中准确识别和定位关键特征。

2.6.3 Transformer 注意力可视化

为解析 Transformer 的全局建模机制,提取多头自注意力权重,经平均、重构与上采样生成热力图,如图 12 所示。

Transformer 注意力集中于产品文字标签、物体边缘轮廓等关键结构区域,即使输入模糊仍可实现精准定位,同时显著抑制背景区域注意力以突出关键信息。实验结果表明,Transformer 能在浑浊度变化显著区域建立长程空间依赖,聚焦复原效果的关键部位,对全局一致性恢复具有重要作用。

可解释性分析从机制层面验证了本文模型设计的合理性与有效性;通道注意力机制可自适应筛选关键特征通道,凸显特征优化能力;Grad-CAM 结果表明模型决策与人类视觉感知相契合,解释了感知质量指标 LPIPS 的提升原因;Transformer 可建立长程空间依赖,聚焦关键复原区域。各模块协同作用是实现良好复原性能的根本原因,也为后续方法的进一步优化提供了理论依据。

3 结 语

本文提出了一种面向偏振图像复原的多分支融合深度网络方法。针对偏振图像多通道特性与复原任务的实际需求,设计偏振特征自适应融合模块以提升偏振信息利用率,构建双分支结构并引入动态门控融合模块实现偏振与空间分支特征的自适应加权融合,集成 Transformer 全局建模与多尺度特征融合模块以弥补卷积网络局部感受野不足的问题,结合通道注意力与全局上下文建模增强特征关注与全局一致性,提出偏振感知损失函数保障偏振特征还原与物理一致性。实验表明,该方法有效提升了复原图像的视觉质量和物理一致性,显著优于传统方法和基础网络结构;但在深海弱光、非均匀强浑浊环境等超复杂水下场景的性能表现及低算力设备的实时应用场景中仍需进一步验证,未来将持续丰富多场景偏振图像的采集与标注以提升模型场景普适性,探索轻量化网络架构与高效推理策略,并尝试与水下三维重建等任务的融合以拓展技术应用边界。

参考文献:

[1] Song Wei,Liu Yaling,Huang Dongmei,et al. From shallow sea to deep sea:research progress in underwater image restoration [J]. Frontiers in Marine Science,2023,10:1163831.

[2] Guo Yecai,Li Hanyu,Zhuang Peixian. Underwater image enhancement using a multiscale dense generative adversarial network [J]. IEEE Journal of Oceanic Engineering,2020,45(3):862-870.

[3] Kong Weifeng,Huo Guanying. Application research of polarization imaging technology in inhomogeneous media/material;a review [J]. Optics and Lasers in Engineering,2025,194:109113.

[4] Yi Xunpeng,Xu Han,Zhang Hao,et al. Diff-Retinex:rethinking low-light image enhancement with a generative diffusion model [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris:IEEE,2023:12268-12277.

[5] 许丽,陆桂明,邱贞光. 结合细节信息的自适应 Retinex 算法水下图像增强[J]. 计算机工程与应用,2022,58(11):224-233. (Xu Li, Lu Guiming, Qiu Zhenguang. Adaptive Retinex algorithm based on detail selection used in underwater image enhancement[J]. Computer Engineering and Applications,2022,58(11):224-233. (in Chinese))

[6] Sun Yue,Jin Yutao,Chen Xiaoyan,et al. Unsupervised detail and color restorer for Retinex-based low-light image enhancement [J]. Engineering Applications of Artificial Intelligence,2025,153:110867.

[7] Liu Mingzhu, Chen Junyu, Han Xiaofei. Research on Retinex algorithm combining with attention mechanism for image

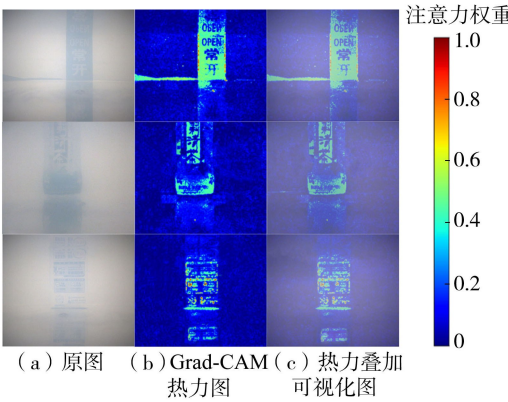


图 11 Grad-CAM 注意力热力图

Fig. 11 Grad-CAM attention heatmap

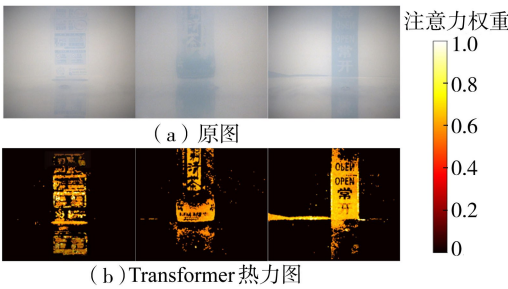


图 12 Transformer 注意力热力图

Fig. 12 Transformer attention heatmap

- enhancement[J]. Electronics, 2022, 11(22): 3695.
- [8] Jiang Yonglong, Zhu Jiahe, Li Liangliang, et al. A joint network for low-light image enhancement based on Retinex[J]. Cognitive Computation, 2024, 16(6): 3241-3259.
- [9] 杨海潮, 李新凯, 张宏立, 等. 结合亮度约束的双分支结构暗光图像增强算法[J]. 计算机工程与应用, 2025, 61(8): 250-259. (Yang Haichao, Li Xinkai, Zhang Hongli, et al. Dual-branch structure low-light image enhancement algorithm combined with brightness constraint[J]. Computer Engineering and Applications, 2025, 61(8): 250-259. (in Chinese))
- [10] Cui Yuning, Ren Wenqi, Cao Xiaochun, et al. Revitalizing convolutional network for image restoration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 9423-9438.
- [11] Ouyang Xu, Chen Ying, Zhu Kaiyue, et al. Image restoration refinement with Uformer GAN[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2024: 5919-5928.
- [12] 白岗岗, 侯精明, 韩浩, 等. 基于深度学习的道路积水智能监测方法[J]. 水资源保护, 2021, 37(5): 75-80. (Bai Ganggang, Hou Jingming, Han Hao, et al. Intelligent monitoring method for road inundation based on deep learning[J]. Water Resources Protection, 2021, 37(5): 75-80. (in Chinese))
- [13] 曾诚, 尹雨然, 周婕, 等. 基于二阶规则波模拟的数值波浪水槽优化[J]. 河海大学学报(自然科学版), 2023, 51(4): 33-40. (Zeng Cheng, Yin Yuran, Zhou Jie, et al. Optimization of a numerical wave tank based on simulations of 2nd order regular waves[J]. Journal of Hohai University (Natural Sciences), 2023, 51(4): 33-40. (in Chinese))
- [14] 朱跃龙, 赵群, 余宇峰, 等. 基于时空特征挖掘的流量过程智能模拟方法[J]. 河海大学学报(自然科学版), 2021, 49(1): 7-12. (Zhu Yuelong, Zhao Qun, Yu Yufeng, et al. Intelligent simulation method of runoff process based on spatiotemporal feature mining[J]. Journal of Hohai University (Natural Sciences), 2021, 49(1): 7-12. (in Chinese))
- [15] 王红瑞, 魏豪杉, 胡立堂, 等. 基于遗传算法的 SVM-AR 改进模型与应用[J]. 河海大学学报(自然科学版), 2020, 48(6): 488-497. (Wang Hongrui, Wei Haoshan, Hu Litang, et al. Improved model and application of SVM-AR based on genetic algorithm [J]. Journal of Hohai University (Natural Sciences), 2020, 48(6): 488-497. (in Chinese))
- [16] 李凡, 刘晨阳, 孙智博, 等. 基于端到端深度学习网络的大容量图像隐写算法[J]. 北京航空航天大学学报, 2026, 52(7): 2621-2629. (Li Fan, Liu Chenyang, Sun Zhibo, et al. A high-capacity image steganography algorithm based on end-to-end deep learning networks[J]. Journal of Beijing University of Aeronautics and Astronautics, 2026, 52(7): 2621-2629. (in Chinese))
- [17] 张珂, 牛杰帆, 李曦, 等. 洪水预报智能模型在中国半干旱半湿润区的应用对比[J]. 水资源保护, 2021, 37(1): 28-35. (Zhang Ke, Niu Jiefan, Li Xi, et al. Comparison of artificial intelligence flood forecasting models in China's semi-arid and semi-humid regions[J]. Water Resources Protection, 2021, 37(1): 28-35. (in Chinese))
- [18] Li Jiafeng, Li Yaopeng, Zhuo Li, et al. USID-Net: unsupervised single image dehazing network via disentangled representations [J]. IEEE Transactions on Multimedia, 2023, 25: 3587-3601.
- [19] He Yufeng, Li Cuili, Li Xu. Remote sensing image dehazing using heterogeneous atmospheric light prior[J]. IEEE Access, 2023, 11: 18805-18820.
- [20] Shi Yingjie, Guo Enlai, Bai Lianfa, et al. Polarization-based haze removal using self-supervised network[J]. Frontiers in Physics, 2022, 9: 789232.
- [21] Li Boyun, Gou Yuanbiao, Gu Shuhang, et al. You only look yourself: unsupervised and untrained single image dehazing neural network[J]. International Journal of Computer Vision, 2021, 129(5): 1754-1767.
- [22] Liang Jingyun, Cao Jiezhong, Sun Guolei, et al. SwinIR: image restoration using Swin transformer[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal: IEEE, 2021: 1833-1844.
- [23] Shen Linghao, Xia Haisheng, Zhang Xun, et al. U²PNet: an unsupervised underwater image-restoration network using polarization [J]. IEEE Transactions on Cybernetics, 2024, 54(9): 5164-5177.
- [24] Kim N, Choi I S, Han S S, et al. DA-Net: dual attention network for haze removal in remote sensing image[J]. IEEE Access, 2024, 12: 136297-136312.
- [25] Desai C, Tabib R S, Reddy S S, et al. RUIG: realistic underwater image generation towards restoration[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Nashville: IEEE, 2021: 2181-2189.

(收稿日期: 2025 - 07 - 31 编辑: 刘晓艳)