

基于视觉 Transformer 的多损失融合水下图像增强网络

丛晓峰¹, 桂杰¹, 章军²

(1. 东南大学网络空间安全学院, 江苏 南京 210000;

2. 安徽大学人工智能学院, 安徽 合肥 230000)

摘要: 由于水中存在光的吸收和散射现象, 水下机器人拍摄到的图像存在颜色失真和对比度降低的问题。针对水下图像存在的质量退化现象, 提出了一种基于视觉 Transformer 的多损失融合的方式训练水下图像增强网络。图像增强网络采用编码与解码的结构, 可以采用端到端的方式进行训练。将多损失的线性组合作为总体优化目标, 有效地更新水下图像增强网络的参数, 包括像素损失、结构损失、边缘损失和特征损失。在两个大型水下数据集上进行了量化实验, 并与 7 种水下图像增强算法进行对比。以峰值信噪比和结构相似性为有参考评估指标, 以水下评估指标为无参考评估指标进行实验。实验结果表明, 提出的水下图像增强网络能够有效地解决图像的颜色失真与对比度降低问题。

关键词: 水下图像; 质量增强; 视觉 Transformer; 神经网络

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202252

Underwater image enhancement network based on visual Transformer with multiple loss functions fusion

CONG Xiaofeng¹, GUI Jie¹, ZHANG Jun²

1. School of Cyber Science and Engineering, Southeast University, Nanjing 210000, China

2. School of Artificial Intelligence, Anhui University, Hefei 230000, China

Abstract: Due to the absorption and scattering of light in water, the images captured by underwater robots suffer from color distortion and reduced contrast. Aiming at alleviating the quality degradation phenomenon of underwater images, an underwater image enhancement network based on vision Transformer that be trained with multiple losses fusion strategy was proposed. The image enhancement network adopted an encoder-decoder architecture, and could be trained in an end-to-end manner. In order to effectively update the parameters of the network for enhancing underwater images, a linear combination of various losses was adopted as the overall optimization objective, including pixel loss, structure loss, edge loss and feature loss. Quantitative experiments were carried out on two large underwater datasets, and the proposed underwater image enhancement network was compared with 7 underwater image enhancement algorithms. The full reference evaluation metrics peak signal-to-noise ratio and structural similarity were calculated in experiment, and the non-referenced metric underwater image quality measure was also computed. The experimental results showed that the proposed underwater image enhancement network could effectively deal with color distortion and contrast reduction.

Key words: underwater image, quality enhancement, vision Transformer, neural network

收稿日期: 2022-09-05; 修回日期: 2022-10-25

通信作者: 桂杰, guijie@seu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62172090)

Foundation Item: The National Natural Science Foundation of China (No.62172090)

0 引言

海洋中含有丰富的资源, 针对海洋的探索任务是水下资源开发的重要环节^[1]。水下机器人是探索海洋的重要工具, 可用于水下图像采集的任务。水下机器人采集到的图像可用于后续的其他操作。然而, 由于不同波长的光在海水中会出现不同程度的衰减, 因此水下机器人拍摄到的图像存在颜色偏移和对比度降低等问题^[1]。水下图像增强算法的作用是对失真的图像进行质量增强, 此项任务是计算机视觉和水下机器人领域的重要研究方向之一。

当前水下图像增强算法主要可分为非基于深度学习的增强算法与基于深度学习的增强算法。非基于深度学习的增强算法一般将统计观测假设作为参数估计的依据, 并且通过显式建模的方式对水下图像进行增强与颜色校正。非基于深度学习的增强算法包括水下暗通道先验 (underwater dark channel prior, UDCP)^[2-3]、图像模糊与亮度吸收 (image blurriness and light absorption, IBLA)^[4]以及水下光衰减先验 (underwater light attenuation prior, ULAP)^[5]等。由于水下场景的复杂性, 非基于深度学习的增强算法存在一定的参数估计失准问题。基于深度学习的增强算法通常需要充足的数据用于增强过程, 采用特定的损失对用于质量增强的深度神经网络进行优化。基于深度学习的增强算法包括水下残差网络 (underwater residual network, UResNet)^[6]、浅层水下网络 (shallow underwater network, UWNNet)^[7]以及水下卷积神经网络 (underwater convolution neural network, UWCNN)^[8]等。当前基于深度学习的增强算法已经在一定程度上解决了水下图像的失真问题, 但是仍存在质量复原效果不好的问题。

针对水下拍摄的图像存在的颜色偏移、色彩不真实和对比度降低问题^[9], 本文结合视觉 Transformer 网络^[10]设计了一种有效的水下图像质量增强模型。训练出可用于水下图像增强的 Transformer 是一个新兴研究领域, 本文采用多种损失组合的方式进行训练, 损失的选取过程充分考虑了图像的像素信息、结构信息、边缘信息与特征信息, 该模型可以采用完全端到端的方式进行训练, 充分利用水下图像数据存在的统计规律。

1 相关工作

水下图像增强研究的主要目的是对失真的水

下图像进行增强, 复原色彩信息与对比度信息等^[1]。本节首先探讨水下图像增强的研究现状, 然后介绍本文提出的增强模型的基础——视觉 Transformer 网络^[10-12]的相关研究。

1.1 水下图像增强

严浙平等^[1]指出水下成像的机制要比陆地上更加复杂, 光线在水中与陆地上的传播衰减不同, 水体中存在的溶解物会影响成像过程, 而且当水下深度较深时, 拍摄过程中需要辅助光源。因此, 水下图像的增强是具有挑战性的研究问题^[13]。当前的水下图像增强研究可分为非基于深度学习与基于深度学习两种。第一种是非基于深度学习的算法。Drews Jr P 等人^[2-3]基于图像去雾领域的 DCP 提出了 UDCP 算法。Peng Y T 等人^[4]基于图像模糊和亮度吸收, 提出了一种针对水下图像的深度估计算法, 并进一步用于水下图像的质量复原过程, 该方法被称为 IBLA。Song W 等人^[5]同样从水下图像的深度估计角度出发, 基于 ULAP 设计了一种快速的深度预测模型。常戡等人^[14]利用导向滤波和自适应算子设计了水下图像增强算法。

神经网络已经被应用于多种视觉任务中^[15-17]。第二种水下图像增强算法是基于深度学习的算法。Liu P 等人^[6]使用均方误差损失和边缘差分损失对卷积神经网络参数进行更新, 提出了端到端的 UResNet 增强算法。Naik A 等人^[7]利用常规的卷积设计了 UWNNet, 结果表明, 其在公开数据集上能获得有效的增强效果。Li C Y^[8]提出了采用残差学习策略的 UWCNN 算法, 并利用像素损失和结构损失进行网络训练。陈龙等人^[18]将残差学习和注意力机制结合, 提出了端到端的神经网络增强模型。范新南等人^[19]利用生成对抗网络的建模能力和对称型网络的图像重建能力设计了基于特征融合的增强网络。Li C Y 等人^[20]设计了水下图像增强网络 (underwater image enhancement network, WaterNet), 融合了白平衡、直方图均衡和伽马校正算法。Guo Y C 等人^[21]利用密集连接和跳跃连接设计了基于生成对抗网络的增强模型, 采用对抗损失和梯度损失对增强模型进行优化。Fabbri C 等人^[22]利用生成对抗网络构建数据集, 并设计了具有全卷积编码解码结构的水下生成对抗网络 (underwater generative adversarial network, UGAN)。已有的增强算法主要采用卷积神经网络进行水下图像的增强, 缺少对局部与全局特征的

高质量建模以及对多种优化目标的考虑。Boudiaf A 等人^[23]采用预训练的 Transformer，初步将 Transformer 用于水下图像增强任务。本文将含有多头自注意力（multi-head self-attention）的 Transformer^[10]模块作为增强模型的基础组件构建具有编码-解码结构的去雾网络，并利用多种损失组合提高增强后的水下图像的质量。

1.2 视觉 Transformer 网络

深度学习领域包含众多有效的通用骨干网络，Transformer 网络^[10,24]受到了研究人员的广泛关注。在自然语言处理领域中，Transformer^[24]是一种被广泛应用的架构，研究人员设计了有效的改进策略，将语言任务中的 Transformer 模型应用到计算机视觉任务中^[10,12]。Transformer 网络采用了多头自注意力机制和多层感知机（multi-layer perceptron）等技术，可以用于多种视觉任务。Swin Transformer^[10]架构在图像分类、目标检测和语义分割任务上验证了层级式 Transformer 网络的性能，所设计的分阶段网络可以作为多种计算机视觉任务的基础网络，基于窗口的多头自注意力机制如式（1）所示：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (1)$$

其中， Q 、 K 和 V 分别为查询、键和值， d 为特征维度， B 为相对位置偏差。Restormer 算法^[11]主要用于图像去雨和图像去噪等任务，采用了编码-解码的结构，能够获取局部和全局的特征表示。DehazeFormer 算法^[25]用于图像去雾任务，该方法通过去掉部分网络层的注意力模块，构建了具有不同参数容量的去雾网络，适用于计算能力不同的平

台。对 Transformer 架构的研究证明了它对计算机视觉任务的有效性，本文设计的编码-解码结构的水下图像增强模型将 Swin Transformer 架构中的基本 Transformer 模块作为基础网络块。

2 水下图像增强网络

本文提出的水下图像增强网络由 Transformer 模块^[10]和常规卷积模块组成，模型总体上采用编码与解码的结构。为了提升水下图像增强模型的效果，模型采用了 4 种损失组合的方式进行训练，不同的损失具有不同的作用。为了便于分析，将本文提出的模型记为 UWT-Net（underwater Transformer network）。

2.1 图像增强网络

UWT-Net 采用了编码-解码的结构方式，包括编码的降维过程和解码的升维过程。编码过程负责学习特征表示，解码过程负责增强图像的像素重建。下面给出 UWT-Net 的整体结构及前向计算的原理，并分别对网络的各个模块进行介绍。

（1）整体网络架构

编码-解码结构的 Transformer 网络如图 1 所示。其中，阶段 1、阶段 2 和阶段 3 是编码过程，阶段 4 和阶段 5 是解码过程。网络中的组件包括分块、Transformer 模块、下采样模块、上采样模块以及块投射。阶段 1~阶段 3 是降维过程，输入数据的高度 H 、宽度 W 和通道数 C 。输入数据经过阶段 1 以后获得的特征图尺寸为 $H \times W \times L$ ，其中 L 代表网络前向过程的通道数，后面 4 个阶段的通道数均是 L 的整数倍。阶段 2 和阶段 3 采用具有残差连接的双路模式，输出特征维度分别是

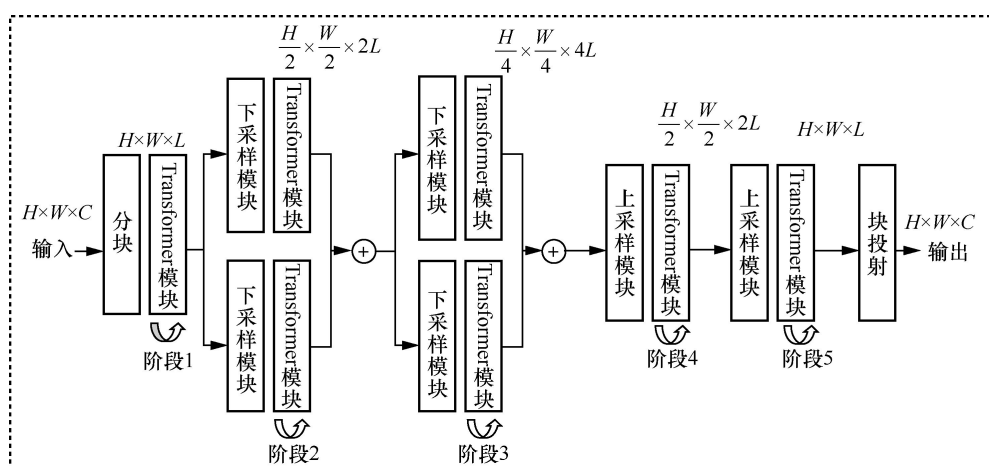


图1 编码-解码结构的 Transformer 网络

$\frac{H}{2} \times \frac{W}{2} \times 2L$ 和 $\frac{H}{4} \times \frac{W}{4} \times 4L$ ，这种设计的作用是降低特征图的尺寸并增加通道数。阶段4~阶段5是解码过程，分别将特征图的尺寸恢复到 $\frac{H}{2} \times \frac{W}{2} \times 2L$ 和 $H \times W \times L$ 。UWT-Net 阶段1输出通道数被设定为24，因此阶段2和阶段3的输出通道数分别为48和96。

(2) 并行连接编码模块

图1的阶段2和阶段3作为编码过程，由并行连接的 Transformer 模块（parallel connection Transformer block, PCTB）构成，该编码过程的作用是提升增强网络的特征提取能力。PCTB 将前一层输入特征分两个支路进行处理，然后对获得的特征进行求和，作为下一层的输入。

(3) 分块与块投射

图1中输入端的分块与输出端的块投射均采用核尺寸为 3×3 且步长为1的卷积实现。分块的作用是将输入的图像投射到特征空间，获取通道数为 L 的特征图。块投射用于将尺寸为 $H \times W \times L$ 的特征图投射为尺寸为 $H \times W \times C$ 的图像。

(4) 下采样模块与上采样模块

在编码过程中，Transformer 模块前使用了下采样模块降低特征图尺寸，下采样模块通过步长为2且核尺寸为 3×3 的卷积实现。在解码过程中，Transformer 模块前使用了上采样模块提升特征图尺寸，并且采用步长为1的卷积及像素重组（pixel shuffle）操作将特征图尺寸提升一倍。

(5) Transformer 模块

基本的 Transformer 模块主要包含两个核心组件，第一个是多头自注意力模块，采用的是基于窗口的模式，第二个是多层感知机模块。为了实现信息的有效传递，Transformer 模块一般采用层归一化（layer norm）模块对信息流进行处理，第 i 层的计算方式如式（2）~式（3）所示：

$$z_i^* = \phi(\omega(z_{i-1})) + z_{i-1} \quad (2)$$

$$z_i = \varphi(z_i^*) + z_i^* \quad (3)$$

其中， z_i 为当前层的输出， z_{i-1} 为增强网络中前一层的输出， z_i^* 为中间过程获得的特征， ω 为层归一化， ϕ 为式（1）所示的多头自注意力模块， φ 为多层感知机。

Transformer 模块结构如图2所示，重复使用 Transformer 模块可以构建 UWT-Net 中的5个阶段，每个阶段含有的模块数量见第3节。

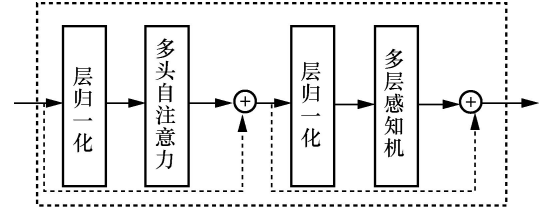


图2 Transformer 模块结构

2.2 多损失组合优化

为了对增强网络进行有效优化，设计了多种损失组合的方式对网络的参数进行更新。损失的设计从以下4个方面展开。

- 像素损失（pixel loss）：像素损失在像素层面面对网络进行优化，使增强后图像的像素值趋近参考图像。

- 结构损失（structural loss）：为了精确地对水下图像的结构信息进行复原，根据结构相似性原理使用结构损失优化网络。

- 边缘损失（edge loss）：水下图像的边缘信息是图像质量的重要组成部分，因此将基于梯度的边缘损失作为总体损失中的一项。

- 特征损失（feature loss）：基于大规模图像数据集训练得到的模型可以表征图像的统计规律，因此使用特征损失对增强网络进行优化可以提高增强效果。

下面对4种损失进行介绍，给出每种损失的设计原理和计算过程，并说明最终损失的构成方式。 τ 代表失真图像，即增强网络的输入数据； y 代表真实标签，即水下数据集中的参考图像； $\mathcal{H}(\cdot)$ 代表 UWT-Net，则网络获得的增强结果为 $\mathcal{H}(\tau)$ 。

(1) 像素损失

逐像素计算的损失是图像复原任务的基础损失，对增强网络的优化采用 L1 损失，则像素损失计算如式（4）所示：

$$L_p = \frac{1}{w \times h} \sum_{i=1}^h \sum_{j=1}^w |\mathcal{H}(\tau(i, j)) - y(i, j)| \quad (4)$$

其中， w 和 h 分别为图像的长和宽， i 和 j 分别为增强结果和参考图像的像素位置。

(2) 结构损失

参考文献[26]证明了结构相似性度量可以作为图像复原任务的损失，为了提升水下图像增强网络

对结构信息的复原能力, 将结构相似性度量作为优化目标之一。将增强网络预测结果 $\mathcal{H}(\tau)$ 记为 η , 则结构相似性 (structural similarity, SSIM) [18] 的计算方式如式 (5) 所示:

$$\text{SSIM}(p) = \frac{2u_\eta u_y + C_1}{u_\eta^2 + u_y^2 + C_1} \cdot \frac{2\sigma_{\eta y} + C_2}{\sigma_\eta^2 + \sigma_y^2 + C_2} \quad (5)$$

其中, u_η 和 u_y 分别为增强后图像和参考图像的均值, σ_η 和 σ_y 分别为增强后图像和参考图像的方差, $\sigma_{\eta y}$ 为协方差, C_1 与 C_2 为常数, P 为像素。SSIM 的值越大, 代表两张图像的相似程度越高, 所以结构损失的定义如式 (6) 所示:

$$L_s = 1 - \frac{1}{N} \sum_{p \in P} \text{SSIM}(p) \quad (6)$$

其中, N 为图像块 P 中的像素数量。

(3) 边缘损失

水下的失真图像存在模糊现象, 导致图像的边缘强度降低。边缘信息是图像的重要组成部分[27], 为了提升增强后的图像的边缘质量, 使用图像的梯度属性设计了边缘损失, 通过计算预测结果 $\mathcal{H}(\tau)$ 与参考图像 y 的梯度信息, 可以获得边缘损失如式 (7) 所示:

$$L_e = \|G_h(\mathcal{H}(\tau)) - G_h(y)\|_2 + \|G_v(\mathcal{H}(\tau)) - G_v(y)\|_2 \quad (7)$$

其中, G_h 和 G_v 分别为水平和垂直方向的梯度值。

(4) 特征损失

基于大规模图像数据集训练得到的预训练模型[28]能够辅助基于 Transformer 的水下图像增强网络的训练, 将这种在特征层面计算的损失称为特征损失。记预训练模型为 $\psi(\cdot)$, 则特征损失的计算方式如式 (8) 所示:

$$L_f = \|\psi(\mathcal{H}(\tau)) - \psi(y)\|_2 \quad (8)$$

特征损失的计算需要选择预训练模型, 在此将 VGG19 网络[29]作为特征提取网络, 并将网络最后的全连接层去掉, 以全连接层前面的池化层输出为特征输出。

(5) 总体损失与训练流程

上述 4 种用于水下图像增强网络训练的损失需要在每次迭代中被同时计算, 并用于指导网络参数的更新, 损失的组合方式为线性加权, 具体如式 (9) 所示:

$$L = L_p + \lambda_s L_s + \lambda_e L_e + \lambda_f L_f \quad (9)$$

其中, λ_s 、 λ_e 与 λ_f 分别为对应损失的权重值, 具体

的设置见第 3 节。对所设计的增强网络和组合型损失, 采用梯度下降算法进行网络的参数更新。增强网络的训练流程伪代码见算法 1。

算法 1 增强网络的训练流程伪代码

网络的参数记为 $\theta_{\mathcal{H}}$, 学习率记为 lr , 对数据集的总迭代轮数记为 E , 数据集中数据总量记为 N , 每次迭代的输入批数据量 (batch size) 记为 b , 则总迭代次数为 $\lfloor N/b \rfloor \times E$, 其中 $\lfloor N/b \rfloor$ 代表向下取整。

初始化

for $\lfloor N/b \rfloor \times E$ steps do

利用参数 $Y = \{y_1, y_2, y_3, \dots, y_b\}$ 采样 $U =$

$\{\tau_1, \tau_2, \tau_3, \dots, \tau_b\}$

计算增强图像 $\mathcal{H}(U)$

计算损失 $L((U), Y) = L_p + \lambda_s L_s + \lambda_e L_e + \lambda_f L_f$

更新参数

$\theta_{\mathcal{H}} = \theta_{\mathcal{H}} - lr \times \nabla_{\theta_{\mathcal{H}}} (L_p + \lambda_s L_s + \lambda_e L_e + \lambda_f L_f)$

end for

3 实验与评估

3.1 实验设置

为了公平地对比不同的水下图像增强算法, 实验将两个大型公开的水下图像数据集作为训练和测试基准数据集。UIEB[20]数据集含有 890 组水下图像与对应的参考图像, 根据 UIEB 中的设置将 800 组图像作为训练数据, 其余 90 组图像作为测试数据。此外, 在实验中使用了 EUVP[30]数据集下的水下场景 (underwater-scenes) 数据子集, 将此水下场景数据子集记为 EUVP-US, EUVP-US 含有 2 185 组有参考数据, 随机选取其中 1 967 组作为训练数据集, 其余 218 组作为测试数据集。UIEB 和 EUVP-US 的失真图像和参考图像如图 3 所示。

实验采用 PyTorch 1.8 深度学习框架。优化器选择 Adam, 学习率设置为 0.0002, 批数据量设置为 4; 在 UIEB 和 EUVP-US 数据集上的迭代轮数为 400, 在训练过程中将图像尺寸统一为 256×256 ; 损失的权重 λ_s 、 λ_e 与 λ_f 理论上可以选取大于 0 的全体实数, 因此是无法枚举所有情况的, 根据经验与实际调试将 λ_s 、 λ_e 与 λ_f 的比例设置为 1:2:1。阶段 1~阶段 5 的 Transformer 模块数量分别设置为 4、8、8、4、4, 每个阶段的注意力模块比例为 $\frac{1}{4}$ 。

作为对比试验的评估算法包括 UDCP[31]、

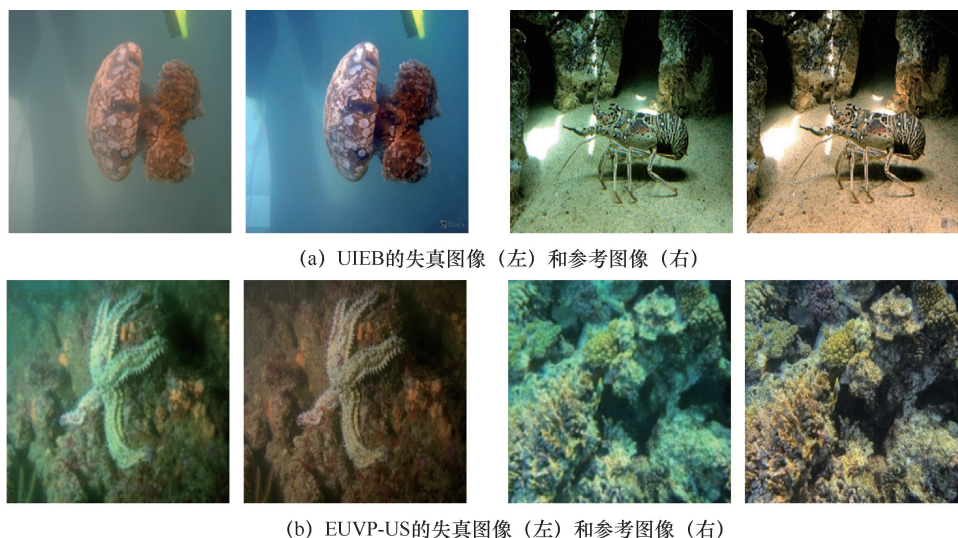


图3 UIEB 和 EUVP-US 的失真图像和参考图像

IBLA^[4]、ULAP^[5]、绿色蓝色增强（green blue enhancement, GB）^[31]、UResNet^[6]、UWNet^[7]与WaterNet^[20]。非基于深度学习的增强算法实现来自参考文献[32]；基于深度学习的对比算法用到的批数据量、学习率和迭代轮数等设置均与UWT-Net保持一致。

评估指标包括峰值信噪比（peak signal-to-noise ratio, PSNR）、SSIM 与无参考水下图像质量度量（underwater image quality measure, UIQM）^[33]。根据参考文献[18]及其中提供的均方误差（mean square error, MSE）计算方式，PSNR 的计算方式如式（10）所示：

$$\text{PSNR}(X, Y) = 10 \times \lg \left(\frac{(2^k - 1)^2}{\text{MSE}(X, Y)} \right) \quad (10)$$

其中， k 为比特数，对于当前数据集， $k=8$ 。SSIM 是结构信息复原效果的评估指标。UIQM 被广泛用于水下图像增强的质量评估任务，它主要包括 3 个角度的评估，第 1 是水下图像的色彩评估，第 2 是水下图像的锐度评估，第 3 是水下图像的对比度评估，将 3 种评估值进行线性求和可以获得最终的 UIQM 值。

3.2 UIEB 数据集定量与视觉评估

在 UIEB 数据集上进行了定量实验与增强效果

的视觉评估，定量实验的结果见表 1。表 1 结果表明了 UWT-Net 在 SSIM 和 UIQM 上获得的指标都是最高的，分别为 0.870 与 3.084，PSNR 值为第二高，该结果证明了算法获得的增强图像与参考图像较为接近。

图 4 展示了不同算法在 UIEB 数据集上获得的增强效果，其中图 4 (a) 为待增强的失真图，图 4 (j) 为参考图。为了便于分析，将含有鱼的图像记为场景 1，含有虾的图像记为场景 2。UDCP 导致场景整体偏绿色；IBLA 造成图像偏淡蓝色，场景 1 中的鱼和场景 2 中的虾颜色明显失真；ULAP 对场景 2 的增强效果较好，但是处理后的场景 1 明显偏红；GB 获取的场景 1 相对接近参考图像，但是场景 2 与参考图像具有一定的色彩差异；UResNet 处理后的场景颜色偏白，场景的细节变得模糊；UWNet 对场景 1 和场景 2 的处理均出现一定的对比度降低现象；WaterNet 处理后的场景 1 与场景 2 与参考图像存在一定的色彩差异；UWT-Net 获得的增强图像与参考图像在颜色 and 对比度方面最为接近，同时保持了图像细节的完整性。

3.3 EUVP-US 数据集定量与视觉评估

EUVP-US 数据集上的定量实验结果见表 2，表 2 结果显示，UWT-Net 能够获得最高的 PSNR

表 1

UIEB 数据集上的定量实验结果

指标	UDCP	IBLA	ULAP	GB	UResNet	UWNet	WaterNet	UWT-Net
PSNR	12.368	17.299	17.601	15.078	18.684	18.913	22.694	21.163
SSIM	0.526	0.678	0.692	0.608	0.754	0.784	0.854	0.870
UIQM	1.842	1.783	1.878	2.272	2.772	2.746	2.883	3.084

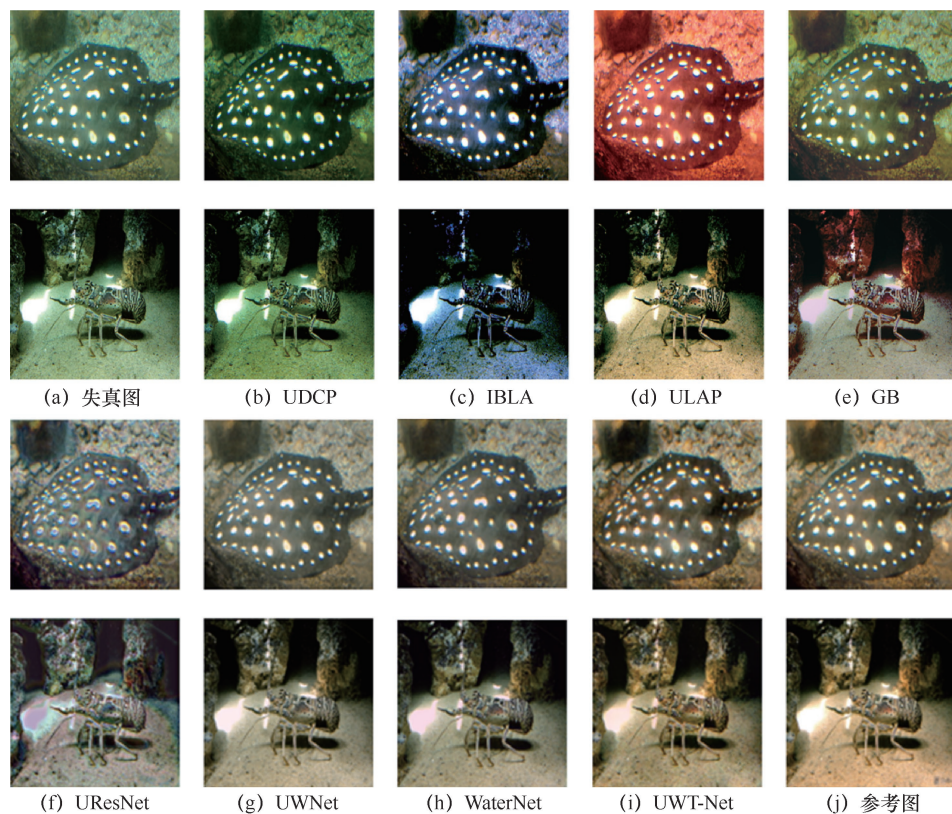


图4 UIEB测试集的增强效果

表2

EUVP-US数据集上的定量实验结果

指标	UDCP	IBLA	ULAP	GB	UResNet	UWNet	WaterNet	UWT-Net
PSNR	14.994	22.138	21.897	16.553	27.638	27.727	23.722	28.343
SSIM	0.570	0.721	0.761	0.606	0.839	0.832	0.818	0.871
UIQM	2.114	2.214	2.369	2.583	2.867	2.888	2.889	2.806

和 SSIM 值, 分别为 28.343 和 0.871, 同时 UIQM 也相对较高。总体上看, UWT-Net 的定量实验结果是最好的。

EUVP-US 测试集的增强效果如图 5 所示, 图 5 的上图和下图分别记为场景 3 和场景 4。UDCP 处理后的两个场景整体颜色偏绿, 场景 4 中的海星与环境的对比度降低; IBLA 获得的增强后的场景 3 明显偏绿且亮度降低; ULAP 处理后的图像相比参考图像存在一定的色差; GB 获得的增强结果明显偏红, 颜色失真明显; UResNet 与 UWT-Net 获得的视觉效果较为接近, 在色彩方面与参考图像保持一致; UWNet 处理后的场景 3 上半部分颜色偏绿, 与参考图像差异明显; WaterNet 处理后的图像整体偏绿; UWT-Net 获得的结果最接近参考图像。EUVP-US 测试集的视觉增强效果证明了 UWT-Net 的有效性。

3.4 消融实验

针对 UWT-Net 的消融实验分为两部分: 第一部分是 UWT-Net 训练过程所用损失的对比研究, 第二部分是网络结构的对比研究。消融实验使用的是 UIEB 数据集。

(1) 损失的消融实验

实验中采用多损失组合优化的方式对神经网络的参数进行更新, 分别为像素损失、结构损失、边缘损失和特征损失。像素损失是主体损失, 其余 3 种损失是辅助损失。为了验证所选择的损失是有效的, 需要对损失组合情况下获得的水下图像增强效果进行量化分析。针对损失的消融实验分为以下 4 组。

- 设置 1: L1 损失+结构损失+边缘损失+特征损失。
- 设置 2: L1 损失+结构损失+边缘损失。
- 设置 3: L1 损失+边缘损失+特征损失。

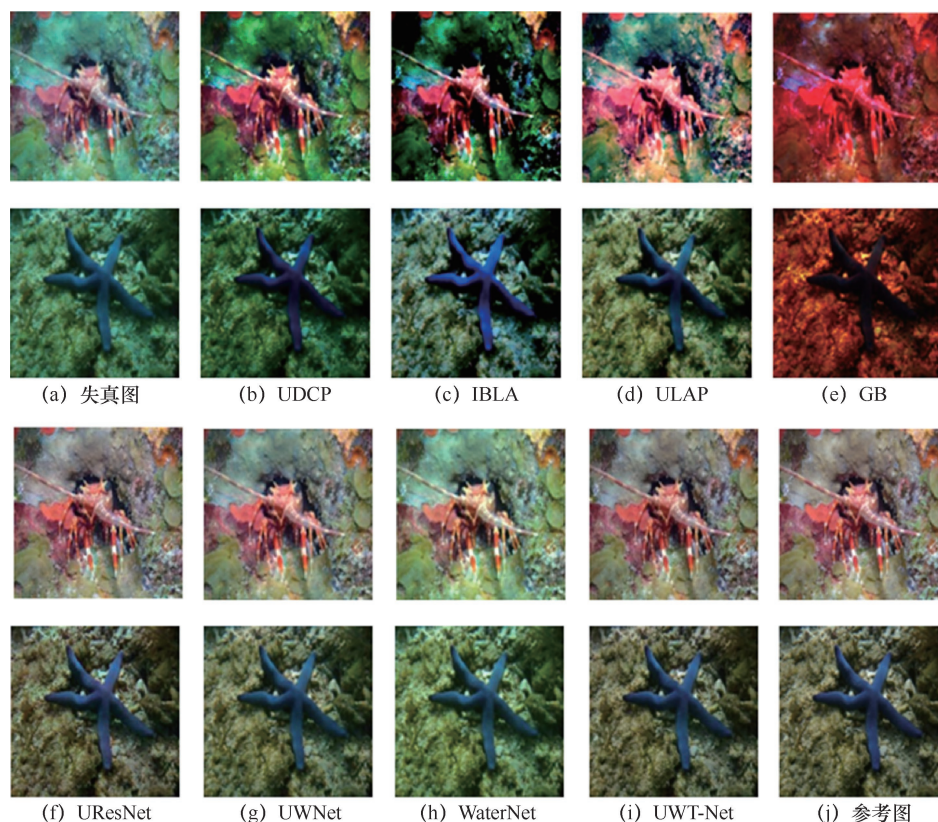


图5 EUVP-US 测试集的增强效果

- 设置4: L1 损失+结构损失+特征损失。

设置1包含了所有损失项,设置2、设置3和设置4中分别在设置1的基础上去掉一种损失,目的是验证去掉的损失对水下图像增强任务是否具有促进作用。

L1 损失消融实验结果见表3,从PSNR、SSIM与UIQM这3种指标值可以看出,使用设置1获得的增强性能最好,此时3种辅助损失都被使用。实验结果表明,获得的最高的PSNR值为21.163,最高的SSIM值为0.870,最高的UIQM值为3.084。

表3 L1 损失消融实验结果

指标	设置1	设置2	设置3	设置4
PSNR	21.163	19.244	20.577	19.973
SSIM	0.870	0.830	0.853	0.851
UIQM	3.084	2.900	3.041	3.053

(2) 网络结构的消融实验

UWT-Net 采用了并行连接的 Transformer 模块。为了验证 PCTB 是否能够提升模型的性能,对比了是否使用 PCTB 获得的表现。根据表1,

使用 PCTB 获得的3种指标值分别为21.163、0.870与3.084。将 PCTB 换成单层下采样模块加 Transformer 模块后,测得的3种指标值分别为20.595、0.857与3.021,证明了 PCTB 能够提升增强结果。

3.5 损失收敛分析

为了证明 UWT-Net 的收敛性,对水下图像增强网络的训练过程中各项损失值进行记录,包括整体损失 L 、像素损失 L_p 、结构损失 L_s 、边缘损失 L_e 和特征损失 L_f 。整体损失 L 的收敛情况如图6所示,从图6可以看出,损失值在迭代300轮时已经趋近于收敛,当迭代轮数达到400时基本处于稳定收敛状态。水下图像复原任务将数据分为训练集和测试集,损失的计算是在训练集上进行的,针对训练集的过度优化可能会导致模型在测试集上的效果降低,这种情况即机器学习中的过拟合现象。为了防止模型出现过拟合现象,在400轮时需要停止对 UWT-Net 的训练。4种损失的收敛情况如图7所示,像素损失收敛最快,边缘损失与结构损失的收敛趋势比较接近,特征损失基本趋近于0。从图6和图7的损失收敛情况

可以看出,增强网络的训练过程稳定,损失的收敛趋势明显。

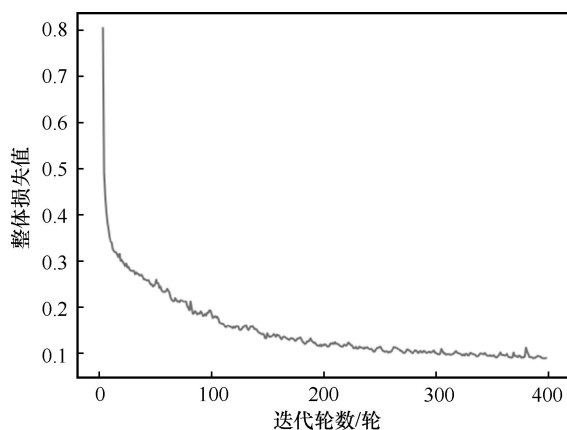


图6 整体损失 L 的收敛情况

4 结束语

本文针对水下图像的颜色失真与对比度降低问题,将视觉 Transformer 模型作为基础网络模块,

构建了端到端的水下图像增强模型 UWT-Net。模型的训练过程将4种不同类型的损失组合作为图像增强网络的整体损失,4种损失分别从像素、结构、边缘和特征的角度设计,能够有效地对增强网络的参数更新过程进行指导。本文在两个公开的水下图像数据集上进行了定量和定性的实验,将提出的水下图像增强模型与其他水下图像增强模型进行对比与分析,实验结果表明,本文设计的模型能够获得较高的峰值信噪比,并保证增强后的水下图像与参考图像具有较高的结构相似性。

参考文献:

- [1] 严浙平, 曲思瑜, 邢文. 水下图像增强方法研究综述[J]. 智能系统学报, 2022, 17(5): 860-873.
YAN Z P, QU S Y, XING W. An overview of underwater image enhancement methods[J]. CAAI Transactions on Intelligent Systems, 2022, 17(5): 860-873.
- [2] DREWS JR P, DO NASCIMENTO E, MORAES F, et al. Transmission estimation in underwater single images[C]//Proceedings of 2013

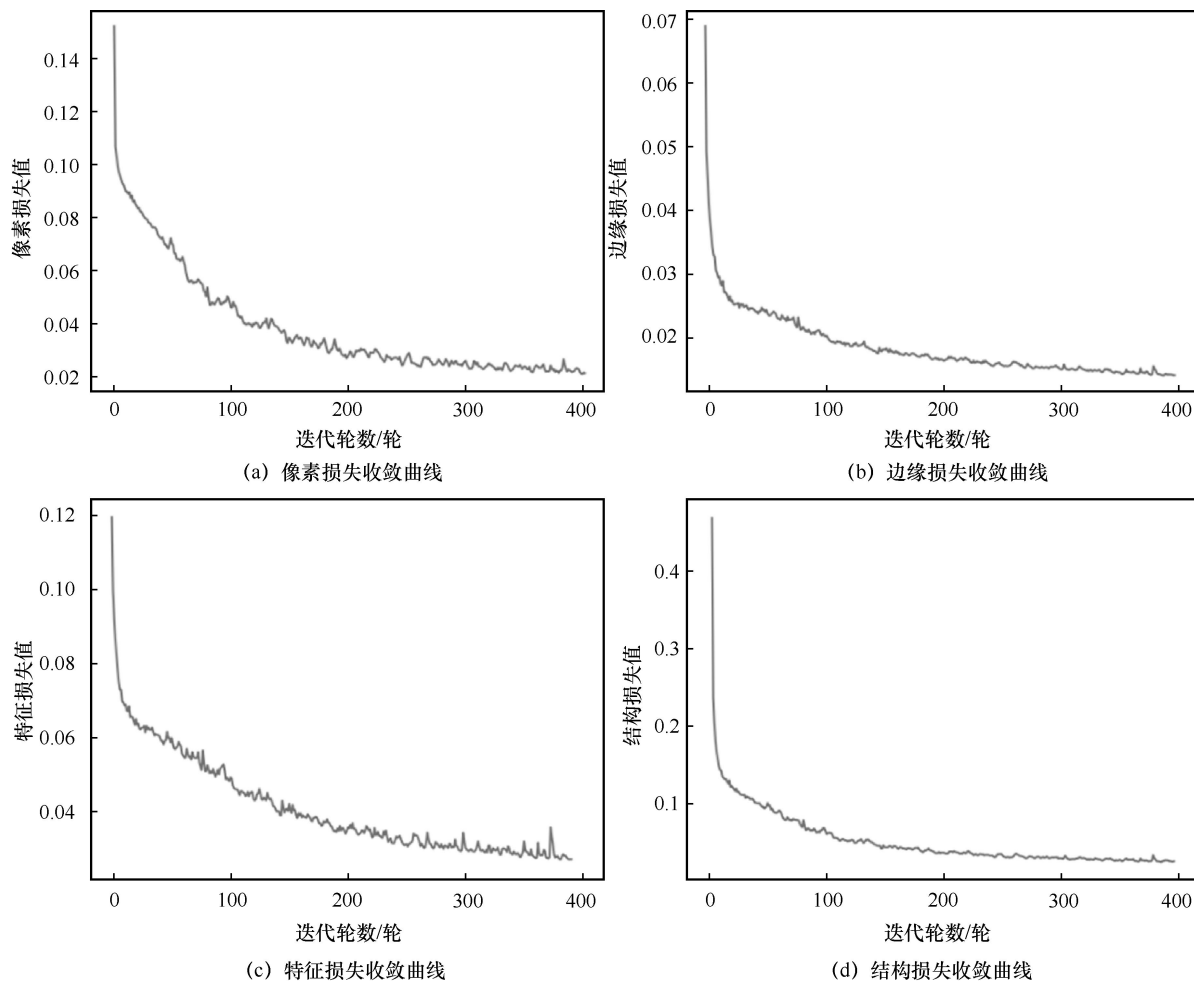


图7 4种损失的收敛情况

- IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE Press, 2013: 825-830.
- [3] DREWS JR P, NASCIMENTO E R, BOTELHO S S C, et al. Underwater depth estimation and image restoration based on single images[J]. IEEE Computer Graphics and Applications, 2016, 36(2): 24-35.
- [4] PENG Y T, COSMAN P C. Underwater image restoration based on image blurriness and light absorption[J]. IEEE Transactions on Image Processing, 2017, 26(4): 1579-1594.
- [5] SONG W, WANG Y, HUANG D M, et al. A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration[C]//Proceedings of 2018 Advances in Multimedia Information Processing. Cham: Springer, 2018: 678-688.
- [6] LIU P, WANG G Y, QI H, et al. Underwater image enhancement with a deep residual framework[J]. IEEE Access, 2019, 7: 94614-94629.
- [7] 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.
- [8] LI C Y. Underwater scene prior inspired deep underwater image and video enhancement[J]. Pattern Recognition, 2020, 98.
- [9] BERMAN D, LEVY D, AVIDAN S, et al. Underwater single image color restoration using haze-lines and a new quantitative dataset[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2822-2837.
- [10] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 9992-10002.
- [11] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 5718-5729.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[J]. arXiv preprint, 2020, arXiv:2010.11929.
- [13] LI J, SKINNER K A, EUSTICE R M, et al. WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images[J]. IEEE Robotics and Automation Letters, 2018, 3(1): 387-394.
- [14] 常骞, 韩旭. 结合导向滤波与自适应算子的水下图像增强[J]. 计算机工程与应用, 2021, 已录用.
- CHANG J, HAN X. Underwater image enhancement combining guide filtering with adaptive operator[J]. Computer Engineering and Applications, 2021, Accepted.
- [15] 李颖, 陈龙, 黄钊宏, 等. 基于多尺度卷积神经网络特征融合的植株叶片检测技术[J]. 智能科学与技术学报, 2021, 3(3): 304-311.
- LI Y, CHEN L, HUANG Z H, et al. Plant leaf detection technology based on multi-scale CNN feature fusion[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 304-311.
- [16] 王禾扬, 杨启鸣, 朱旗. 基于深度卷积集成网络的视网膜多种疾病筛查和识别方法[J]. 智能科学与技术学报, 2021, 3(3): 259-267.
- WANG H Y, YANG Q M, ZHU Q. Retinal multi-disease screening and recognition method based on deep convolution ensemble network[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 259-267.
- [17] 邵虹, 张鸣坤, 崔文成. 基于分层卷积神经网络的皮肤镜图像分类方法[J]. 智能科学与技术学报, 2021, 3(4): 474-481.
- SHAO H, ZHANG M K, CUI W C. Classification method of dermoscopic image based on hierarchical convolution neural network[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(4): 474-481.
- [18] 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 L, DING D D. Joint multi-residual learning for underwater image enhancement[J]. Journal of Image and Graphics, 2022, 27(5): 1577-1588.
- [19] 范新南, 杨鑫, 史鹏飞, 等. 特征融合生成对抗网络的水下图像增强[J]. 计算机辅助设计与图形学学报, 2022, 34(2): 264-272.
- FAN X N, YANG X, SHI P F, et al. Underwater image enhancement based on feature fusion generative adversarial networks[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(2): 264-272.
- [20] LI C Y, GUO C L, REN W Q, et al. An underwater image enhancement benchmark dataset and beyond[J]. IEEE Transactions on Image Processing, 2020, 29: 4376-4389.
- [21] GUO Y C, LI H Y, ZHUANG P X. Underwater image enhancement using a multiscale dense generative adversarial network[J]. IEEE Journal of Oceanic Engineering, 2020, 45(3): 862-870.
- [22] FABBRI C, ISLAM M J, SATTAR J. Enhancing underwater imagery using generative adversarial networks[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 7159-7165.
- [23] BOUDIAF A, GUO Y H, GHIMIRE A, et al. Underwater image enhancement using pre-trained transformer[C]//Proceedings of International Conference on Image Analysis and Processing. [S.l.:s.n.], 2022: 480-488.
- [24] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv preprint, 2017, arXiv:1706.03762.
- [25] SONG Y D, HE Z Q, QIAN H, et al. Vision transformers for single image dehazing[J]. arXiv preprint, 2022, arXiv:2204.03883.
- [26] ZHAO H, GALLO O, FROSIO I, et al. Loss functions for image restoration with neural networks[J]. IEEE Transactions on Computa-

- tional Imaging, 2017, 3(1): 47-57.
- [27] ZHANG H, PATEL V M. Densely connected pyramid dehazing network[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3194-3203.
- [28] JOHNSON J, ALAHI A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]//Proceedings of 2016 European Conference on Computer Vision. [S.l.:s.n.], 2016: 694-711.
- [29] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint, 2014, arXiv:1409.1556.
- [30] ISLAM M J, XIA Y Y, SATTAR J. Fast underwater image enhancement for improved visual perception[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 3227-3234.
- [31] LI C Y, QUO J, PANG Y W, et al. Single underwater image restoration by blue-green channels dehazing and red channel correction[C]//Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2016: 1731-1735.
- [32] WANG Y, SONG W, FORTINO G, et al. An experimental-based review of image enhancement and image restoration methods for underwater imaging[J]. IEEE Access, 7: 140233-140251.
- [33] PANETTA K, GAO C, AGAIAN S. Human-visual-system-inspired underwater image quality measures[J]. IEEE Journal of Oceanic Engineering, 2016, 41(3): 541-551.

[作者简介]



丛晓峰(1997-)，男，东南大学网络空间安全学院博士生，主要研究方向为水下图像处理、生成式算法、图像去雾等。



桂杰(1982-)男，博士，东南大学网络空间安全学院教授、博士生导师，主要研究方向为生成式算法、图像去雾、对抗机器学习以及自监督学习等。



章军(1971-)男，博士，安徽大学人工智能学院教授、博士生导师，主要研究方向为模式识别、智能信息处理等。