

## 基于多尺度注意力和对比学习的水下图像增强

王悦<sup>1,2,3</sup>, 范慧杰<sup>1,2\*</sup>, 刘世本<sup>1,2,3</sup>, 唐延东<sup>1,2</sup><sup>1</sup>中国科学院沈阳自动化研究所机器人学国家重点实验室, 辽宁 沈阳 110016;<sup>2</sup>中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016;<sup>3</sup>中国科学院大学, 北京 100049

**摘要** 由于水体对光的吸收和散射作用, 水下图像普遍存在颜色失真和细节模糊这两种退化问题。为了得到色彩鲜明、细节清晰的水下图像, 提出了一个基于多尺度注意力和对比学习的水下图像增强算法模型。该模型采用编码-解码结构作为基础框架, 为了提取到更细粒度的特征, 在编码部分设计了一个多尺度通道像素注意力模块, 利用 3 个并行支路提取图像中不同层次的特征, 然后将 3 条支路提取的特征进行融合输入到下一层编码器以及对应的解码层, 提高网络特征提取以及细节增强的能力。最后, 为了进一步提高增强图像的质量, 引入对比学习训练网络。大量实验证明, 本文算法增强后的图像色彩鲜明且细节信息保留较好。峰值信噪比和结构相似性指数的平均值最高可达到 25.46 和 0.8946, 与其他方法相比至少提高了 4.4% 和 2.8%; 水下彩色图像质量指标和信息熵的平均值最高为 0.5802 和 7.6668, 与其他方法相比均至少提高了 2%; 特征点匹配平均比原始图像多 24 个。

**关键词** 图像增强; 注意力; 多尺度; 对比学习

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP223047

## Underwater Image Enhancement Based on Multi-Scale Attention and Contrast Learning

Wang Yue<sup>1,2,3</sup>, Fan Huijie<sup>1,2\*</sup>, Liu Shibin<sup>1,2,3</sup>, Tang Yandong<sup>1,2</sup><sup>1</sup>State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, Liaoning, China;<sup>2</sup>Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, Liaoning, China;<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** The two common degradations of underwater images are color distortion and blurred detail due to the absorption and dispersion of light by water. We propose an underwater image-enhancement algorithm model based on multi-scale attention and contrast learning to acquire underwater images with bright colors and clear details. The model adopts the encoding-decoding structure as the basic framework. To extract more fine-grained features, a multi-scale channel pixel attention module is designed in the encoder. The module uses three parallel branches to extract features at different levels in the image. In addition, the extracted features by the three branches are fused and introduced to the subsequent encoder and the corresponding decoding layer to improve the ability to extract network features and enhance details. Finally, a contrast-learning training network is introduced to improve the quality of enhanced images. Several experiments prove that the enhanced image by the proposed algorithm has vivid colors and complete detailed information. The average values of the peak signal-to-noise ratio and structural similarity index are up to 25.46 and 0.8946, respectively, and are increased by 4.4% and 2.8%, respectively, compared with the other methods. The average values of the underwater color image quality index and information entropy are 0.5802 and 7.6668, respectively, and are increased by at least 2% compared with the other methods. The number of feature matching points is increased by 24 compared to the original images.

**Key words** image enhancement; attention; multi-scale; contrast learning

收稿日期: 2022-10-17; 修回日期: 2022-11-24; 录用日期: 2022-12-07; 网络首发日期: 2023-01-04

基金项目: 国家自然科学基金(62273339, 61991413)、中国科学院青年创新促进会项目(2019203)

通信作者: \*fanhuijie@sia.cn

## 1 引言

海洋资源是地球资源的重要组成部分,海洋资源的开发、利用和保护已经成为各个国家的发展战略。由于海洋环境的特殊性,人类通常无法直接进行探索和研究,需要依靠水下设备来获取相关信息。水下图像作为水下信息的重要呈现形式,在海底探测、水下考古、水下目标检测等重要研究中具有不可替代的作用<sup>[1]</sup>。但与大气环境不同,水体对光的吸收和散射作用,会造成水下图像出现颜色失真、模糊等退化现象。所以获得清晰的水下图像来更好地了解水下世界变得尤为重要。

目前水下图像的增强方法有传统方法和基于深度学习的方法。传统方法主要有基于直方图的算法<sup>[2]</sup>,多尺度视网膜增强<sup>[3]</sup>和基于融合的方法<sup>[4]</sup>。此外还有基于先验的方法,例如:针对水下环境的特殊性,Drews等<sup>[5]</sup>提出了一种基于暗通道先验(DCP)的水下传输估计方法,将DCP<sup>[6]</sup>的方法用在蓝色和绿色2个通道;Song等<sup>[7]</sup>提出了一种基于水下光衰减先验(ULAP)的场景深度估计模型来帮助恢复清晰水下图像。上述方法虽然都在一定程度上提高了视觉效果,但是通常会造成增强过度或校正不彻底的现象。

近年来由于深度学习技术的迅速发展以及深度神经网络对特征强大的学习能力,深度学习被广泛应用于各类视觉任务中,其中就包括水下图像增强。方明等<sup>[8]</sup>提出了一个基于注意力的多尺度水下图像增强网络。Li等<sup>[9]</sup>构建了融合不同特征的Water-Net,Li等<sup>[10]</sup>还提出了一种通过介质传输引导的多颜色空间嵌入的水下图像增强网络,称为Ucolor。为了将网络部署在便携式设备中,Naik等<sup>[11]</sup>提出了一种浅层神经网络结构Shallow-UWnet。除了上述基于CNN的方法外,Guo等<sup>[12]</sup>为了更好地利用特征恢复水下图像,设计了一个多尺度密集生成对抗网络。Fabbri等<sup>[13]</sup>提出了一种利用生成对抗网络(GANs)来提高水下视觉场景质量的方法。Islam等<sup>[14]</sup>基于条件生成对抗网络(cGAN)建立了一个多模态目标函数来训练模型。

深度学习强大的学习能力,使得水下图像的质量得到了一定程度的提升,但仍存在色偏校正不彻底甚至引入其他色偏,或校正后图像相对模糊,细节信息不明显的现象。因此本文提出了一个基于多尺度注意力和对比学习的水下图像增强网络,与现有的多尺度模型不同的是,所提的多尺度模块采用了多重注意力机制来帮助网络关注更重要的特征,并将多尺度模块同时放置于模型的浅层和深层,帮助网络获得更丰富的浅层颜色、纹理特征和深层语义特征。最后通过对比学习从颜色、纹理以及内容等方面约束网络训练,进一步提高增强图像质量。

## 2 相关算法

### 2.1 注意力机制

注意力机制的基本思想就是帮助网络学会关注特征图中的重点信息。根据关注点的不同,视觉注意力可以分为空间注意力<sup>[15]</sup>、通道注意力<sup>[16]</sup>、像素注意力<sup>[16]</sup>等。其中:空间注意力通过学习不同区域位置的权重,最终增强关键区域特征表达。Niu等<sup>[17]</sup>将空间注意力机制应用于超分辨率任务中,学习每个特征图中不同位置的重要程度;Li等<sup>[18]</sup>为了平衡图像去雾任务中存在的过度校正和欠校正问题,引入空间注意力来自适应调整图像中不同位置的权重。通道注意力帮助网络获取特征图中每个通道的重要程度,网络根据其重要程度学习权重,达到关注重要特征通道、抑制对当前任务作用较小的特征通道的目的。Li等<sup>[19]</sup>将通道注意力机制应用于图像去雨任务中,让网络更好地关注雨图特征;Li等<sup>[10]</sup>为了在水下图像增强任务中更好地利用不同颜色空间的特征,在网络中加入通道注意力机制学习不同颜色空间特征的相关性。像素注意力关注的是不同像素位置的特征重要程度,通过像素注意力网络可以更好地关注到对当前任务更重要的像素位置。Qin等<sup>[16]</sup>针对去雾任务中雾分布不均匀的问题,提出了像素注意力机制帮助网络关注更模糊的像素位置;Liu等<sup>[20]</sup>将像素注意力应用于水下图像增强任务中,利用像素注意力关注图像中的光照和颜色特征。

### 2.2 对比学习

对比学习是一种自监督学习方法,近年来已经被广泛应用于自然语言处理以及高级视觉任务中<sup>[21]</sup>。对比学习的核心思想就是选取一个锚点,构建一个正样本,一个负样本;在表征空间中将选取的锚点拉近正样本,远离负样本<sup>[22]</sup>。Chen等<sup>[23]</sup>在去雨任务中加入对比引导分支,让相似的特征分布更接近而不相似的更远,从而更好地去除雨水,恢复图像。最近,对比学习也被用于有监督任务中,学习不同域样本之间的差别。Khosla等<sup>[24]</sup>的实验证明使用有监督的对比学习训练模型可以使模型从样本中学习到更好的特征表示,提高模型的泛化性。Wu等<sup>[22]</sup>提出了新的有监督对比学习方式来生成更好的去雾图像。

## 3 本文算法

模型的主干框架为编码-解码结构,编码器模块提取退化图像特征,然后将特征输入解码器重建图像来获得清晰水下图像。为了提高网络提取特征的能力,设计了一个多尺度通道像素注意力模块,将其添加在编码器中。多尺度通道像素注意力模块可以帮助网络提取更丰富的视觉特征,更好地适应不同的水下环境,恢复图像的细节信息。此外,还通过增强后图像、参考图像以及退化图像之间的对比学习,充分利用正、负样本信息进一步提高增强图像的质量。网络整体结构如图1所示。

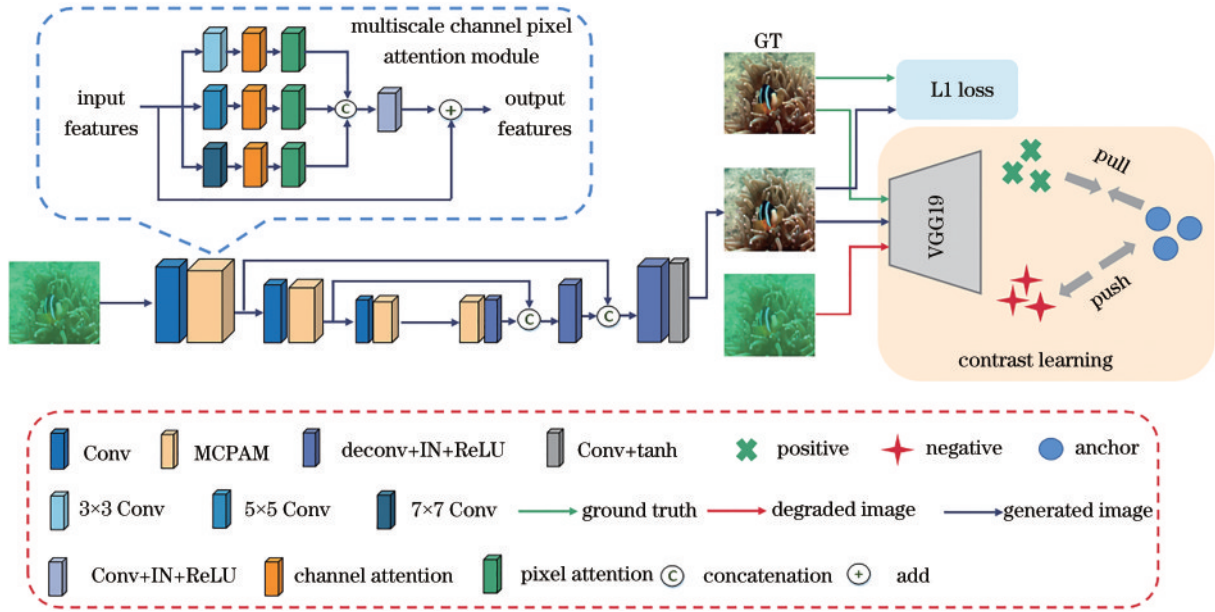


图 1 整体网络结构

Fig. 1 Overall network structure

### 3.1 多尺度通道像素注意力模块

由于水下图像的退化程度不同,同一幅退化图像中的不同位置色偏和模糊程度也不同,为了提取更丰富的全局特征和局部特征,设计了一个多尺度的通道像素注意力模块,如图 2 所示。模块的多尺度通过并列使用 3×3、5×5、7×7 不同大小的卷积实现,一般来说,卷积核大,感受野大,网络可以更好地学习图像中

的全局特征,反之卷积核小,感受野小,网络可以提取更好的局部特征。在特征提取部分利用 3×3、5×5、7×7 3 个分支可以更好地提取到图像不同层次的视觉信息,更有利于网络学习退化图像和清晰图像的映射关系;若使用 1×1 分支,空间信息缺少,特征信息提取不完全,不利于图像细节的恢复;若采用更大的卷积核,则会浪费较多计算资源。

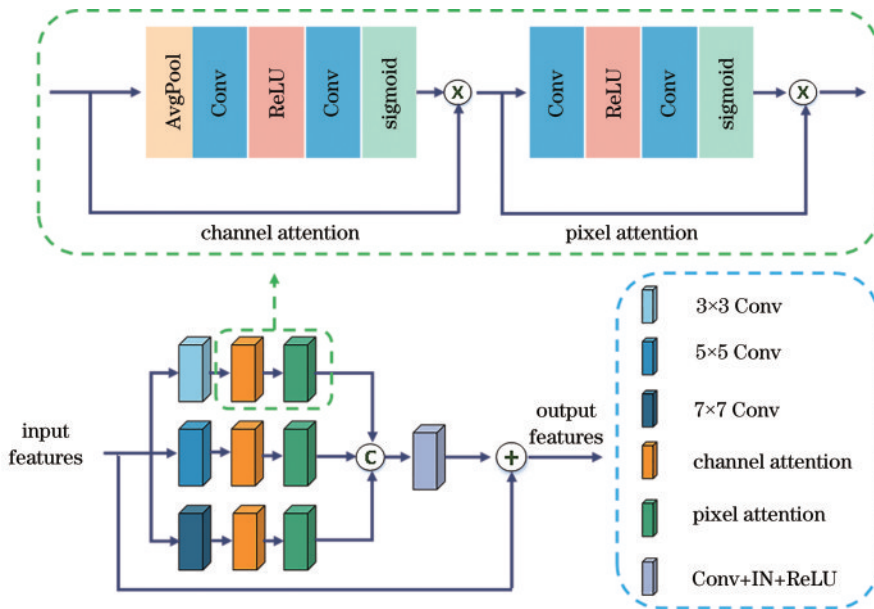


图 2 多尺度通道像素注意力模块

Fig. 2 Multiscale channel pixel attention module

为了进一步关注退化图像中退化较为严重的区域,在每一个卷积后都加入了通道像素注意力,具体结构如图 2 所示。输入图像经过卷积之后会得到多通道的特征,每一个特征通道代表图像中不同的特征信息,并且

不同的特征信息对增强水下图像的重要程度也不相同,所以采用通道注意力来捕捉不同特征通道之间的相关性。具体来说,就是将卷积提取的多通道特征经过自适应平均池化将特征图的全局空间信息压缩到通道描述



符中,再通过卷积-ReLU-卷积-sigmoid,得到不同通道的特征权重矩阵,最后将输入特征与权重矩阵相乘得到输出特征。在通道注意力模块后添加了像素注意力模块,像素注意力模块可以帮助网络关注图像中的信息特征,如高频图像区域和模糊程度较为严重的区域<sup>[22]</sup>。具体来说,就是将经过通道注意力加权后的特征作为像素注意力模块的输入,再经过卷积-ReLU-卷积-sigmoid,得到不同像素值的权重,然后将输入特征与权重相乘得到新的特征。最后为了充分地利用不同尺度的特征信息,将 $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  3个分支的输出特征进行拼接,并利用卷积对3个特征进行融合,然后输入下一个编码层以及对应的解码层。同时,为了减少特征信息的丢失,采用残差连接的方式将输入特征加到多尺度通道像素注意力模块的输出,帮助网络学习到更稳定的特征。

### 3.2 对比学习

水下图像增强的目的就是让增强后的图像在颜色和细节上都更接近其参考图像,远离退化图像,符合对比学习思想,因此采用像素级对比学习训练网络,将网络增强后的图像定义为锚点,将参考图像(GT)定义为正样本,将输入的退化图像作为负样本。为了更好地从颜色、纹理以及内容等方面约束增强图像,从预训练模型VGG19的1、3、5、10、14层分别提取了锚点,正样本以及负样本特征进行计算。同时为了增加训练的稳定性,采用了Smooth L1损失分别计算锚点与正样本、锚点和负样本之间的损失;最后2个损失的比值即为模型的对比约束。通过对比学习约束网络训练,提升网络学习退化图像和清晰图像映射关系的能力,进一步提高增强图像质量。

### 3.3 损失函数

采用了L1损失和对比正则化作为网络的优化函数。其中L1损失是输出图像与真实值之间的最小绝对值偏差。对比正则化利用预训练的VGG19网络分别提取增强后图像(锚点)、正样本(参考图像)和负样本(退化图像)的特征进行像素级的对比学习,对比正则化 $L_c$ 可表示为

$$L_c = \sum_{i=1}^n \omega_i \frac{L_{sm}\{\varphi_i(I_c) - \varphi_i[G(I_d)]\}}{L_{sm}\{\varphi_i(I_d) - \varphi_i[G(I_d)]\}}, \quad (1)$$

式中: $I_c$ 为参考图像; $I_d$ 为退化图像; $G(I_d)$ 为生成图像; $\omega_i$ 为第 $i$ 层特征的权重; $L_{sm}$ 为计算Smooth L1损失; $\varphi_i$ 为从预训练VGG19模型中提取的第 $i$ 个隐藏特征。总损失函数 $L$ 可表示为

$$L = \omega_o L_o + \omega_c L_c, \quad (2)$$

式中: $L_o$ 为L1损失; $\omega_o$ 为L1损失的权重; $\omega_c$ 为对比正则化的权重。

## 4 实验及结果分析

### 4.1 实验设置

1) 数据集:收集了5个可用的水下图像数据集,包

括UIEBD、RUIE、HRCID、EUVP1(scenes)、EUVP2(imagenet)<sup>[9,13,25]</sup>。其中,UIEBD、RUIE、HRCID为真实的水下数据集,EUVP1、EUVP2为合成的数据集。在UIEBD中随机选择了90张图像作为测试集,剩余800张作为训练集,RUIE、HRCID为非配对数据集,分别选取了100张作为测试集。在EUVP1中随机选择了185张图像作为测试集,剩余2000张作为训练集,在EUVP2中随机选择了300张图像作为测试集,剩余3400张作为训练集。

2) 训练设置:采用Python 3.6和PyTorch 1.6完成模型搭建,所有实验均在NVIDIA TITAN XP(12 G)上进行。模型的输入、输出图像尺寸均为 $256 \text{ pixel} \times 256 \text{ pixel}$ ,设置的批处理数量为8,学习率为0.0003, $\omega_o=1$ 和 $\omega_c=0.1$ ,共进行200次迭代。

3) 评价方式:采用客观和主观2种方式对所提的模型进行评价。对有参考图像的UIEBD、EUVP1、EUVP2数据集而言,采用峰值信噪比(PSNR, $f_{psnr}$ ),结构相似性(SSIM, $f_{ssim}$ )<sup>[26]</sup>以及特征点匹配测试3种客观评价指标。PSNR是衡量图像质量的客观标准(最大值信号和背景噪声的比值),其值越大代表图像失真越小,具体可表示为

$$f_{psnr} = 10 \lg [255^2 / f_{mse}], \quad (3)$$

式中: $f$ 为参考图像; $g$ 为测试图像; $f_{mse}$ 为当前参考图像和测试图像的均方误差。结构相似性从亮度、对比度、结构等3个方面来衡量测试图像和参考之间的相似性,其值越大两幅图像越相似,具体可表示为

$$f_{ssim} = \frac{(2\mu_f \mu_g + c_1)(2\sigma_{fg} + c_2)}{(\mu_f^2 + \mu_g^2 + c_1)(\sigma_f^2 + \sigma_g^2 + c_2)}, \quad (4)$$

式中: $\mu_f$ 、 $\mu_g$ 分别为两幅图像的均值; $\sigma_f$ 、 $\sigma_g$ 分别为两幅图像的方差; $\sigma_{fg}$ 为协方差; $c_1$ 、 $c_2$ 为常量。对RUIE、HRCID测试集而言,采用水下彩色图像质量评估(UCIQE)以及信息熵指标对图像质量进行衡量。UCIQE是评价色度、饱和度、对比度的综合指标,其值越高表示图像视觉质量越好,具体可表示为

$$f_{uciqe} = c_3 \times \sigma_c + c_4 \times c_o + c_5 \times \mu_s, \quad (5)$$

式中: $c_3$ 、 $c_4$ 、 $c_5$ 为加权系数; $\sigma_c$ 为色度标准差; $c_o$ 为亮度的对比度; $\mu_s$ 为饱和度平均值。信息熵是用来衡量图像中所包含信息量的大小,熵越大说明包含的信息越多。主观评价是对图像的视觉效果进行比较分析。

4) 对比方法:传统方法有水下暗通道先验(UDCP)<sup>[5]</sup>、Fusion-based<sup>[4]</sup>、ULAP<sup>[7]</sup>;基于深度学习的方法有Water-Net<sup>[9]</sup>、UWCNN<sup>[27]</sup>、FUNIE-GAN<sup>[14]</sup>、Shallow-UWnet<sup>[11]</sup>、Ucolor<sup>[10]</sup>。Water-Net、FUNIE-GAN、Shallow-UWnet、Ucolor都采用相应参考文献中提供的参数在本文收集的数据集上进行了训练和测试,UWCNN采用的是参考文献中作者提供的预训练模型。

## 4.2 实验结果

### 4.2.1 客观评价

首先将不同的方法在 UIEBD、EUVP1、EUVP2 等 3 个数据集上进行定量比较。对有参考图像的数据集而言,PSNR 越高表明增强图像与 GT 的内容越相

似,SSIM 越接近于 1 表明增强图像的结构和纹理更接近于 GT<sup>[28]</sup>。不同方法的 PSNR、SSIM 指标如表 1 所示。由表 1 可知,所提模型在真实数据集或合成数据集中 2 个指标都最优,证明采用所提的模型增强后的图像更接近真实值。

表 1 不同方法在 UIEBD、EUVP、EUVP2 数据集上的 PSNR 和 SSIM

Table 1 PSNR and SSIM values for the different methods on UIEBD, EUVP1 and EUVP2 datasets

Method	UIEBD		EUVP1		EUVP2	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
UDCP <sup>[5]</sup>	11.13	0.4814	13.57	0.5222	12.31	0.3355
Fusion-based <sup>[4]</sup>	20.66	0.8160	17.54	0.6003	19.53	0.5922
ULAP <sup>[7]</sup>	16.63	0.6048	19.81	0.7334	17.14	0.5046
Water-Net <sup>[9]</sup>	21.10	0.8689	18.70	0.7240	20.88	0.7168
UWCNN <sup>[27]</sup>	12.93	0.4425	16.42	0.5907	17.88	0.6588
FUNIE-GAN <sup>[14]</sup>	19.17	0.7728	24.38	0.8704	22.62	0.7250
Shallow-UWnet <sup>[11]</sup>	18.27	0.7072	23.86	0.8375	22.40	0.6892
Ucolor <sup>[10]</sup>	20.69	0.8031	23.96	0.8607	23.78	0.7699
Ours	<b>23.83</b>	<b>0.8866</b>	<b>25.46</b>	<b>0.8946</b>	<b>24.39</b>	<b>0.7952</b>

此外将所提的模型与 Fusion-based、Water-Net、Ucolor 一起进行了特征点匹配测试,匹配结果如图 3 所示。表 2 为不同方法匹配的特征点数,可以看出在相同特征相似度阈值条件下,所提的模型能匹配到更多特征点,比 Water-Net 方法平均高出 12 个特征点,比

原始图像平均高出 24 个特征点。一般来说,图像越清晰,匹配的特征点越多,在特征提取等应用中效果更好<sup>[29]</sup>,这表明所提的模型增强后的图像质量更好,在后续的应用中效果也较好。

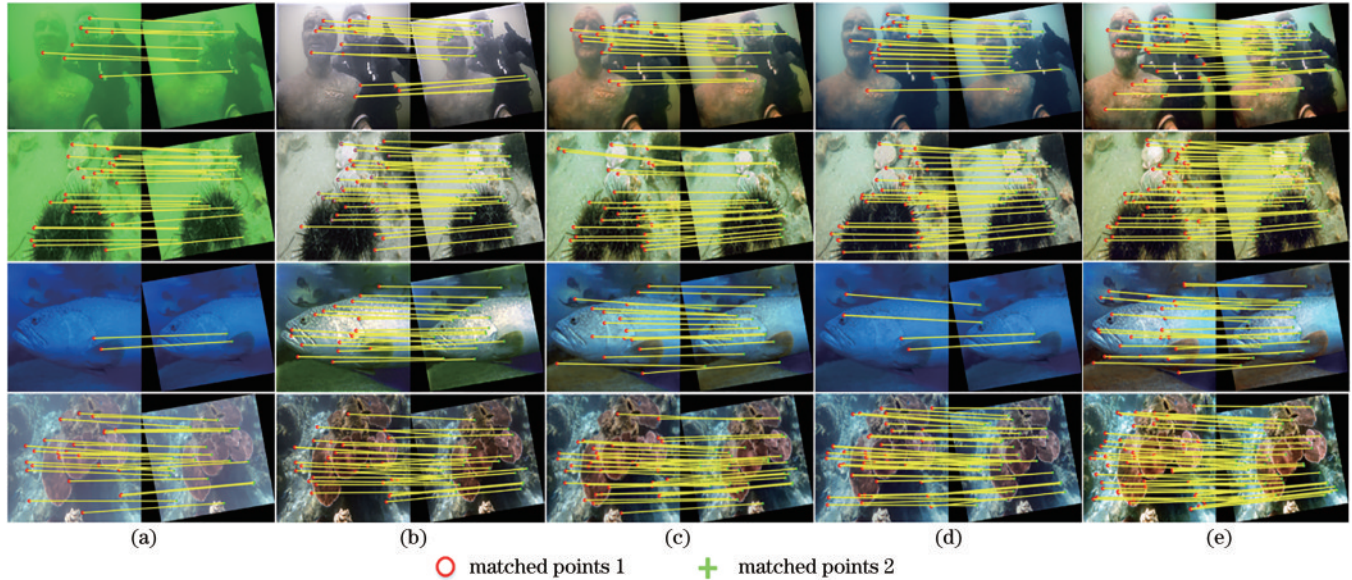


图 3 特征匹配实验结果。(a)原始图像;(b)Fusion-based;(c)Water-Net;(d)Ucolor;(e)本文算法

Fig. 3 Feature matching experimental results. (a) Original image; (b) Fusion-based; (c) Water-Net; (d) Ucolor; (e) ours

### 4.2.2 主观评价

为了进一步验证模型的有效性,对不同方法进行了色卡恢复实验,实验结果如图 4 所示。由图 4 可知,ULAP 引入了红色色偏,色卡整体偏红。FUNIE-GAN 也在结果中引入了黄色色偏。Water-Net 去除色偏的效果较好,但图像中仍然存在模糊的问题。

Shallow-UWnet 恢复的结果整体偏白色,色卡的颜色不鲜明。Ucolor 恢复的结果水体部分存在偏粉色的现象,色卡部分存在模糊的现象。与上述方法相比,所提的模型恢复的结果图像更加清晰,色卡颜色更鲜明,有较好的视觉效果。

除了色卡恢复实验,将所有方法在 UIEBD、



表 2 不同方法的特征点匹配数

Table 2 Number of feature point matches for the different methods

Image No.	Original image	Fusion-based <sup>[4]</sup>	Water-Net <sup>[9]</sup>	Ucolor <sup>[10]</sup>	Ours
1	9	16	23	21	<b>39</b>
2	32	29	32	35	<b>46</b>
3	3	21	20	6	<b>26</b>
4	25	31	43	44	<b>56</b>
Average	17.25	24.25	29.50	26.50	<b>41.75</b>

EUVPI、EUVPI2 数据集上进行视觉效果比较。不同方法在真实数据集 UIEBD 上的比较,如图 5 所示,展示了不同方法对水下图像常见的蓝绿色偏和模糊 2 种退化现象的处理结果。由图 5 可知,本文的模型校正色偏、去除模糊的效果更好,更接近于参考图像。UDCP、ULAP、UWCNN、Shallow-UWnet 以及 Ucolor 去除色偏的效果较差,且 UDCP 引入了严重的深蓝色色偏。Fusion-based 出现了过度增强的现象,FUNIE-GAN 的结果存在较为明显的噪声点。

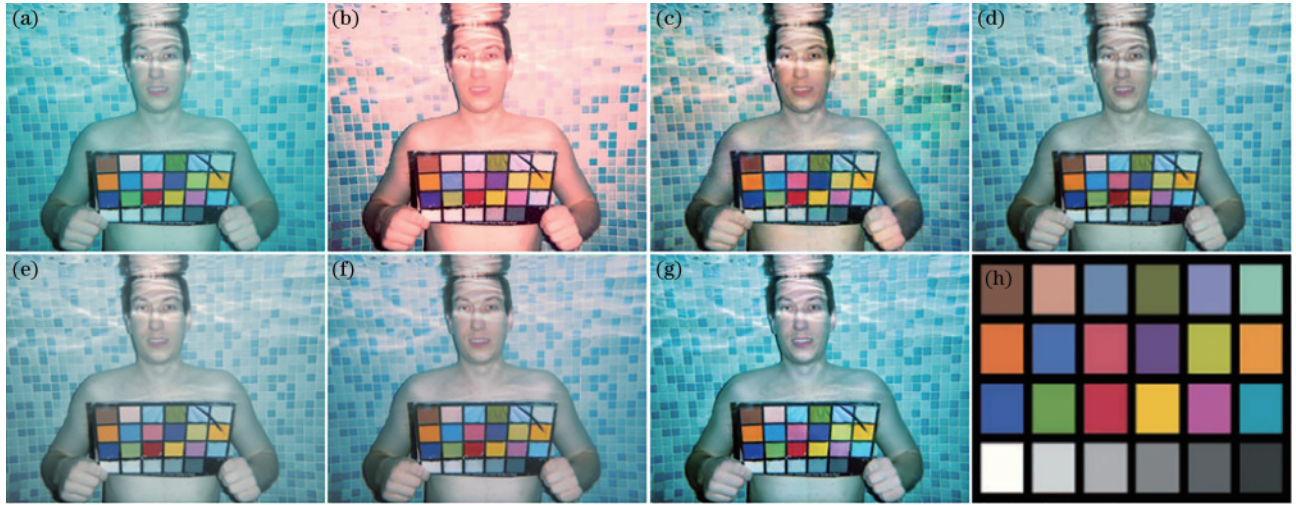


图 4 标准色卡恢复实验结果。(a)原始图像; (b)ULAP; (c)FUNIE-GAN; (d)Water-Net; (e)Shallow-UWnet; (f)Ucolor; (g)本文算法; (h)标准色卡

Fig. 4 Results of the standard color card recovery experiments. (a) Original image; (b) ULAP; (c) FUNIE-GAN; (d) Water-Net; (e) Shallow-UWnet; (f) Ucolor; (g) ours; (h) standard color card

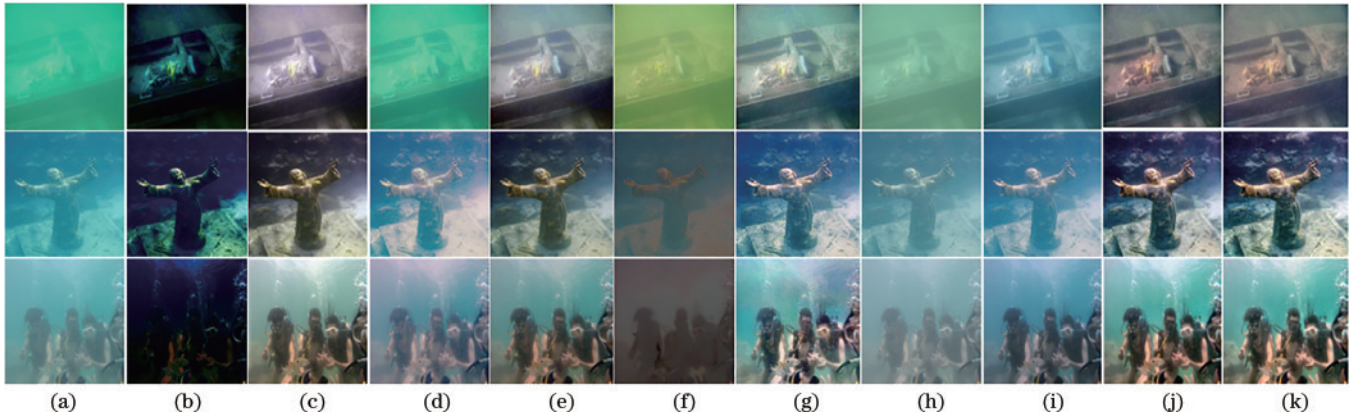


图 5 不同方法在 UIEBD 上的增强结果。(a)原始图像; (b)UDCP; (c) Fusion-based; (d) ULAP; (e)Water-Net; (f) UWCNN; (g) FUNIE-GAN; (h) Shallow-UWnet; (i) Ucolor; (j) 本文算法; (k) 参考图像

Fig. 5 Enhancement results of the different methods on UIEBD. (a) Original image; (b)UDCP; (c) Fusion-based; (d) ULAP; (e) Water-Net; (f) UWCNN; (g) FUNIE-GAN; (h) Shallow-UWnet; (i) Ucolor; (j) ours; (k) GT

不同方法在合成数据集 EUVPI、EUVPI2 上的比较,如图 6 所示。其中,第 1 行和第 2 行是在 EUVPI 上的比较,可以看出所提模型对颜色的校正更加彻底,并且增强结果在颜色和细节上都更接近于参考图像。UDCP、ULAP、Water-Net 都引入了较为严重的其他色偏,Fusion-based 存在校正过度的现象,其他方法的

校正不彻底,结果仍然存在较为明显的蓝绿色偏。第 3 行和第 4 行是在 EUVPI2 上的比较,从第 3 行可以看出,与其他方法相比所提模型去除了色偏,也很好保留了图像原本的色彩。第 4 行图像,只有所提模型结果更接近于 GT,其他方法的结果都存在颜色校正不彻底或引入其他颜色偏差的现象。



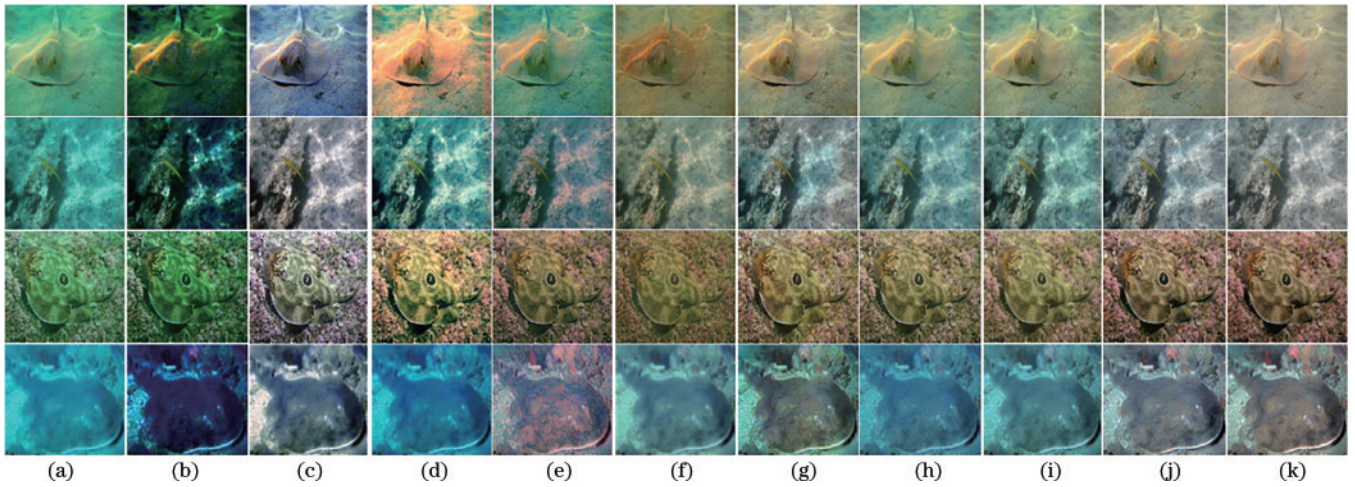


图 6 不同方法在 EUVP1 和 EUVP2 上的增强结果。(a)原始图像; (b)UDCP; (c) Fusion-based; (d) ULAP; (e)Water-Net; (f) UWCNN; (g) FUNIE-GAN; (h) Shallow-UWnet; (i) Ucolor; (j)本文算法; (k)参考图像

Fig. 6 Enhancement results of the different methods on EUVP1 and EUVP2. (a) Original image; (b)UDCP; (c) Fusion-based; (d) ULAP; (e)Water-Net; (f) UWCNN; (g) FUNIE-GAN; (h) Shallow-UWnet; (i) Ucolor; (j) ours; (k) GT

从 3 个数据集的对比结果来看,本文的网络在校正色偏、去除模糊、恢复图像细节 3 个方面都要优于其他网络,这得益于本文的网络提取利用了多尺度特征,并利用通道像素注意力机制使网络更加关注图像中退化较为严重的区域,而且引入对比学习也使本文模型的增强结果更接近于参考图像。此外,从对比结果还可以看出UDCP加重了图像的色偏,这可能是由于基于先验的方法对水下环境较为敏感。Fusion-based 容易出现校正过度的现象,这可能是由于忽略了水下图像的成像机制。而其他基于深度学习的方法,也都存在校正不彻底或引入色偏以及噪声的现象,这可能是因为这些方法在增强过程中特征学习不够稳定,因此在增强水下图像时容易产生不稳定的结果。

#### 4.2.3 泛化性实验

为了验证模型的泛化性,分别在非配对的真实水下数据集 RUIE 和 HRCID 中各选取 100 张退化图像作为测试集。利用在 UIEBD 数据集训练好的模型直接进行测试,采用 UCIQE 和信息熵 2 种图像质量评价指标来定量比较结果,同时通过比较视觉效果进一步验证所提模型的有效性。表 3 为不同方法在 RUIE 和 HRCID 测

试集上的客观比较。对无参考图像而言,UCIQE 越高表明图像的视觉质量越好,信息熵越大表明图像中包含的信息越丰富。由表 3 可知,RUIE 数据集中,本文模型的 UCIQE 和信息熵结果都是最优的,HRCID 数据集中,本文模型的 UCIQE 仅比 FUNIE-GAN 方法低 0.0092,信息熵高于所有方法,这表明本文模型的结果拥有较好的视觉效果,还保留了更丰富的信息。

在 RUIE 和 HRCID 测试集上的主观比较结果,如图 7 所示。其中,第 1 行和第 2 行为 RUIE 数据集,第 3 行和第 4 行为 HRCID 数据集,图中的方框则是对不同方法增强结果的局部细节比较。由图 7 可知,UWCNN 和 Shallow-UWnet 的结果仍然存在明显的色偏以及模糊。FUNIE-GAN 虽然去除了色偏,但引入了较为严重的其他色偏。Ucolor 在 HRCID 数据集中去除绿色色偏的效果相对较好,但在 RUIE 数据集中仍然存在较为明显的蓝色色偏。Water-Net 和本文方法都有较好的去除色偏的效果,但 Water-Net 的结果细节较为模糊。通过比较可知,所提模型在去除色偏的同时也更好地保留了图像的细节信息。综合以上客观和主观比较的结果可以看出,所提模型泛化性更好。

#### 4.2.4 消融实验

为了进一步验证网络的有效性,在 UIEBD 数据集上进行了消融实验,验证了多尺度注意力模块以及对比损失的有效性,如表 4 所示。其中,Model 1 指在不添加多尺度注意力模块和对比损失的情况下进行训练测试,可以看出其指标是最低的。Model 2、Model 3、Model 4 分别将多尺度注意力模块中的  $7 \times 7$ 、 $5 \times 5$ 、 $3 \times 3$  支路去掉再进行训练和测试,由表 4 可知,同时采用 3 个尺度的特征时模型的结果最优。Model 2(add Conv  $7 \times 7$ )、Model 3(add Conv  $5 \times 5$ )、Model 4(add Conv  $3 \times 3$ ) 分别将  $7 \times 7$ 、 $5 \times 5$ 、 $3 \times 3$  支路中的注意力

表 3 不同方法在 RUIE、HRCID 数据集上的 UCIQE、信息熵

Table 3 UCIQE and entropy values for the different methods on RUIE and HRCID datasets

Method	RUIE		HRCID	
	UCIQE	Entropy	UCIQE	Entropy
Water-Net <sup>[9]</sup>	0.5680	7.5092	0.5207	7.4918
UWCNN <sup>[27]</sup>	0.4481	7.2463	0.4554	6.9498
FUNIE-GAN <sup>[14]</sup>	0.5535	7.4984	<b>0.5579</b>	7.4125
Shallow-UWnet <sup>[11]</sup>	0.3835	6.6503	0.3872	6.3581
Ucolor <sup>[10]</sup>	0.5258	7.3653	0.4766	7.2015
Ours	<b>0.5802</b>	<b>7.6668</b>	0.5487	<b>7.6461</b>



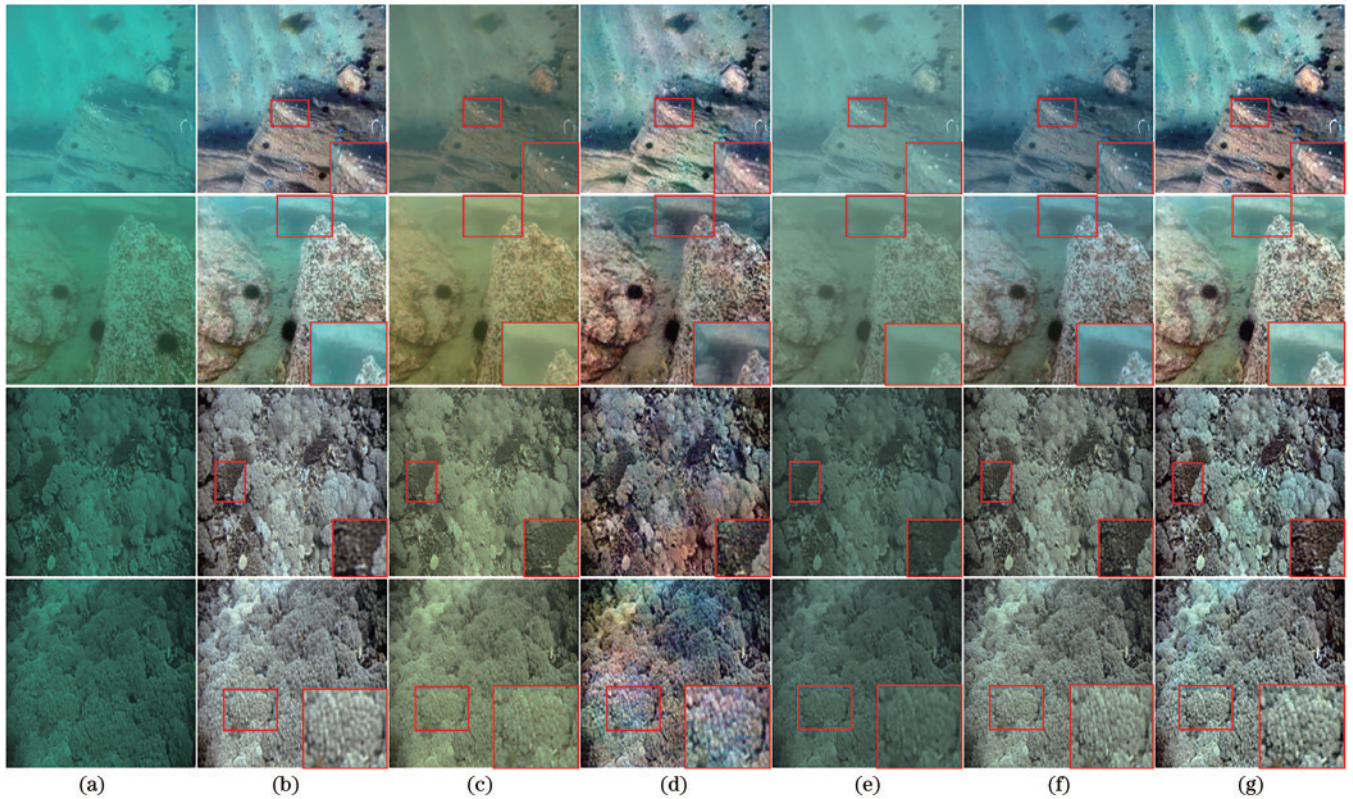


图 7 不同方法在 RUIE 和 HRCID 上的增强结果。(a)原始图像；(b) Water-Net；(c) UWCNN；(d) FUNIE-GAN；(e) Shallow-UWnet；(f) Ucolor；(g) 本文算法

Fig. 7 Enhancement results of the different methods on RUIE and HRCID . (a) Original image; (b) Water-Net; (c) UWCNN; (d) FUNIE-GAN; (e) Shallow-UWnet; (f) Ucolor; (g) ours

表 4 消融实验结果

Table 4 Results of the ablation experiments

Model	PSNR	SSIM
Model 1	23.13	0.8711
Model 2	23.16	0.8824
Model 2(add Conv 7×7)	23.69	0.8847
Model 3	23.49	0.8835
Model 3(add Conv 5×5)	23.51	0.8845
Model 4	23.39	0.8845
Model 4(add Conv 3×3)	23.56	0.8846
Model 5	23.45	0.8796
Model 6	23.30	0.8783
Ours	<b>23.83</b>	<b>0.8866</b>

机制去掉,只保留卷积层再进行训练测试,实验结果证明了添加注意力机制的有效性;同时与去除整条支路的模型相比,保留卷积层的模型效果更好,进一步证明了采用不同尺寸卷积提取特征的有效性。Model 5 指不采用对比损失,仅使用 L1 损失来训练网络,表 4 的结果显示,当去掉对比损失时模型的效果下降。Model 6 指不添加多尺度注意力模块,仅使用对比损失来训练网络,表 4 中 Model 1 和 Model 6 的实验结果,证明加入对比损失时模型效果提升。综合表 4 的实验结果,可以看出所提模型指标最优。

## 5 结 论

针对水下图像色偏和模糊的问题,提出了一个基于多尺度注意力和对比学习的水下图像增强网络。该网络通过采用不同大小的卷积核来提取多尺度特征,并通过通道像素注意力机制来自适应调整、融合提取到的特征,使增强后的图像具有更多细节信息。为了进一步提高增强后图像质量,引入了对比损失,联合 L1 损失训练网络。大量实验证明,本文方法的 PSNR、SSIM、UCIQE 和信息熵指标整体优于其他对比方法,特征点匹配平均比其他方法多 12~17 个,证明所提模型可以更好地增强水下图像、去除色偏、恢复更多的细节信息。由于真实水下成对图像难以获取,在未来的工作中将考虑设计一个无监督网络来增强水下退化图像。

## 参 考 文 献

- [1] 丛润民, 张禹墨, 张晨, 等. 深度学习驱动的水下图像增强与复原研究进展[J]. 信号处理, 2020, 36(9): 1377-1389. Cong R M, Zhang Y M, Zhang C, et al. Research progress of deep learning driven underwater image enhancement and restoration[J]. Journal of Signal Processing, 2020, 36(9): 1377-1389.
- [2] Hummel R. Image enhancement by histogram transformation [J]. Computer Graphics and Image Processing, 1977, 6



- (2): 184-195.
- [3] Jobson D J, Rahman Z, Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes[J]. *IEEE Transactions on Image Processing*, 1997, 6(7): 965-976.
- [4] Ancuti C O, Ancuti C, de Vleeschouwer C, et al. Color balance and fusion for underwater image enhancement[J]. *IEEE Transactions on Image Processing*, 2018, 27(1): 379-393.
- [5] Drews P, Jr, do Nascimento E, Moraes F, et al. Transmission estimation in underwater single images [C]//2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 825-830.
- [6] He K M, Sun J, Tang X O. Single image haze removal using dark channel prior[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(12): 2341-2353.
- [7] 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[M]//Hong R C, Cheng W H, Yamasaki T, et al. *Advances in multimedia information processing-PCM 2018*. Lecture notes in computer science. Cham: Springer, 2018, 11164: 678-688.
- [8] 方明, 刘小晗, 付飞蚨. 基于注意力的多尺度水下图像增强网络[J]. *电子与信息学报*, 2021, 43(12): 3513-3521. Fang M, Liu X H, Fu F R. Multi-scale underwater image enhancement network based on attention mechanism[J]. *Journal of Electronics & Information Technology*, 2021, 43(12): 3513-3521.
- [9] 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.
- [10] Li C Y, Anwar S, Hou J H, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding[J]. *IEEE Transactions on Image Processing*, 2021, 30: 4985-5000.
- [11] 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.
- [12] 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.
- [13] Fabbri C, Islam M J, Sattar J. Enhancing underwater imagery using generative adversarial networks[C]//2018 IEEE International Conference on Robotics and Automation, May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE Press, 2018: 7159-7165.
- [14] 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.
- [15] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018*. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [16] Qin X, Wang Z L, Bai Y C, et al. FFA-net: feature fusion attention network for single image dehazing[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 11908-11915.
- [17] Niu B, Wen W L, Ren W Q, et al. Single image super-resolution via a holistic attention network[M]//Vedaldi A, Bischof H, Brox T, et al. *Computer vision-ECCV 2020*. Lecture notes in computer science. Cham: Springer, 2020, 12357: 191-207.
- [18] Li P Y, Tian J D, Tang Y D, et al. Deep retinex network for single image dehazing[J]. *IEEE Transactions on Image Processing*, 2021, 30: 1100-1115.
- [19] Li P Y, Tian J D, Tang Y D, et al. Model-based deep network for single image deraining[J]. *IEEE Access*, 2020, 8: 14036-14047.
- [20] Liu S B, Fan H J, Lin S, et al. Adaptive learning attention network for underwater image enhancement[J]. *IEEE Robotics and Automation Letters*, 2022, 7(2): 5326-5333.
- [21] He K M, Fan H Q, Wu Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 9726-9735.
- [22] Wu H Y, Qu Y Y, Lin S H, et al. Contrastive learning for compact single image dehazing[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 10546-10555.
- [23] Chen X, Pan J S, Jiang K, et al. Unpaired deep image deraining using dual contrastive learning[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 2007-2016.
- [24] Khosla P, Teterwak P, Wang C, et al. Supervised contrastive learning[EB/OL]. (2020-04-23)[2022-05-06]. <https://arxiv.org/abs/2004.11362>.
- [25] Han J L, Shoeiby M, Malthus T, et al. Underwater image restoration via contrastive learning and a real-world dataset[J]. *Remote Sensing*, 2022, 14(17): 4297.
- [26] Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM [C]//2010 20th International Conference on Pattern Recognition, August 23-26, 2010, Istanbul, Turkey. New York: IEEE Press, 2010: 2366-2369.
- [27] Li C Y, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement [J]. *Pattern Recognition*, 2020, 98: 107038.
- [28] 王丹, 张子玉, 赵金宝, 等. 基于场景深度估计的自然光照水下图像增强方法[J]. *机器人*, 2021, 43(3): 364-372. Wang D, Zhang Z Y, Zhao J B, et al. An enhancement method for underwater images under natural illumination based on scene depth estimation[J]. *Robot*, 2021, 43(3): 364-372.
- [29] 林森, 白莹, 李文涛, 等. 基于修正模型与暗通道先验信息的水下图像复原[J]. *机器人*, 2020, 42(4): 427-435, 447. Lin S, Bai Y, Li W T, et al. Underwater image restoration based on the modified model and dark channel prior[J]. *Robot*, 2020, 42(4): 427-435, 447.