

# DRHA-UIE: 基于双重残差混合注意力模块的水下图像增强方法

王 欣, 石 慧

(吉林大学计算机科学与技术学院, 吉林长春 130000)

**摘 要:** 在水下环境中, 悬浮的颗粒会对光造成散射和波长相关的衰减, 使水下图像呈现出颜色失真、对比度低等问题. 针对上述问题, 本文提出一种基于双重残差混合注意力模块的水下图像增强方法 (Dual Residual Hybrid Attention Underwater Image Enhancement method, DRHA-UIE). 该方法采用改进的双重残差块完成特征学习, 并通过在双重残差块中引入混合注意力模块, 沿通道和空间两个维度对特征进行注意力权重推断, 以捕获显著特征. 本文设计结合像素级损失、结构相似性损失和内容感知损失的联合特征损失函数, 以获得具有更精细纹理的增强图像. 此外, 本文应用水下图像形成模型 (Image Formation Model, IFM) 对水下图像基准数据集 (Underwater Image Enhancement Benchmark dataset, UIEB) 进行优化, 获得了具有更高视觉质量的 N-UIEB (New-UIEB) 数据集, 实验表明, 相较于 UIEB 数据集, 基于 N-UIEB 训练得到的增强图像具有更自然的颜色和更清晰的细节. 为验证本文方法的有效性, 将本文方法与主流的 10 种方法进行测试和比较, 结果表明, DRHA-UIE 方法有效提高了图像对比度并恢复了图像颜色, 在与水下图像增强算法的定量比较中获得了最优的性能.

**关键词:** 水下图像增强; 图像处理; 计算机视觉; 深度学习; 机器学习

**基金项目:** 吉林省教育厅科学技术研究项目 (No. JJKH20211100KJ)

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 0372-2112(2023)09-2398-10

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20211619

## DRHA-UIE: An Underwater Image Enhancement Method Based on Dual Residual Hybrid Attention Block

WANG Xin, SHI Hui

(College of Computer Science and Technology, Jilin University, Changchun, Jilin 130000, China)

**Abstract:** In seawater, light suffers from scattering and wavelength-related attenuation, which makes underwater images exhibit color distortion and low contrast. In this paper, we propose a robust method called dual residual hybrid attention underwater image enhancement method (DRHA-UIE). The proposed method uses an improved dual residual block for features learning. To capture the significant features, a hybrid attention mechanism is introduced in the dual residual block, which infers the attention weights on the features along the channel and spatial dimensions. To train the proposed method end-to-end, a joint feature loss function that consists of pixel-level loss, structural similarity loss, and content-aware loss is designed, then enhanced images with finer textures are obtained. In addition, a revised underwater image formation model (IFM) is applied to optimize the underwater image enhancement benchmark dataset (UIEB), and a dataset with higher visual quality, new-underwater image enhancement benchmark dataset (N-UIEB), is obtained. Experiments show that the enhanced images trained by N-UIEB have more natural colors and sharper details than those trained by the UIEB dataset. The proposed model is extensively evaluated on the above two datasets. Results show that the DRHA-UIE method effectively enhances the image quality subjectively and outperforms the other 10 state-of-the-art methods in quantitative comparisons.

**Key words:** underwater image enhancement; image processing; computer vision; deep learning; machine learning

**Foundation Item(s):** Scientific and Technological Research Foundation of the Education Department of Jilin Province, China (No. JJKH20211100KJ)

## 1 引言

近年来,水下视觉在水下资源探索领域发挥了重要作用,研究人员利用水下机器人或飞行器拍摄的图像来研究水下环境和目标<sup>[1,2]</sup>。由相机直接捕获的图像,因受到水中悬浮颗粒对光线的吸收和散射影响,往往存在可见度低、对比度差、偏色等问题<sup>[3]</sup>,不仅给观察者带来较差的视觉感受,还会严重影响水下机器人等搭载的计算机视觉系统的正常工作。因此,许多研究者运用图像增强技术对失真图像进行处理,以提高图像质量,便于后续识别、分割等任务的展开。图1为水下图像示例,由于不同水下环境中光的衰减和散射程度不同,虽然图像均呈现出亮度低、细节模糊等特点,但在偏色和模糊程度方面仍存在差异,难以统一处理,给水下图像增强算法的研究带来了巨大的挑战。



图1 不同水体环境下的水下图像

目前,水下图像增强算法主要分为三类,分别是基于非物理模型的增强方法<sup>[4,5]</sup>、基于物理模型的增强方法<sup>[6-11]</sup>以及基于深度学习的增强方法<sup>[12-18]</sup>。

基于非物理模型的增强方法不考虑水下图像成像过程,直接通过直方图拉伸等方法对各颜色模型中的像素进行修改。由于未考虑水下图像退化的物理过程,此类方法获得的增强图像中往往存在伪影或过度增强,且图像中的噪声被放大,无法取得满意的效果。

基于物理模型的增强方法的主要思想是针对水下图像退化过程构建物理成像模型,这类方法遵循以下步骤:(1)建立退化的物理模型;(2)估计模型中的未知参数;(3)将参数代入模型并解决逆问题。水下成像物理模型的建立依赖于先验知识和统计特性,且参数的估计过程难以全面考虑不同水体环境,缺乏较强的普适性。

深度学习因其强大的特征学习和非线性映射能力,近年来被广泛应用于各种图像处理任务<sup>[19,20]</sup>。基于学习的水下图像增强方法通常需要包含原始水下图像和其对应理想参考图像的数据集,但水下理想参考图像获取困难<sup>[12]</sup>,因此现有深度学习增强方法如 WaterGAN (Water Generative Adversarial Networks)<sup>[13]</sup>、UGAN (Underwater GAN)<sup>[14]</sup>以及 FUnIE-GAN (Fast Underwater Image Enhancement GAN)<sup>[15]</sup>均使用生成对抗网络 (Generative Adversarial Networks, GAN),在合成数据集

的同时,实现水下图像的增强。而 GAN 方法无法控制生成样本的模式,尽管可以使用条件生成对抗网络 (Conditional Generative Adversarial Networks, CGAN) 来对生成的样本加以约束,但实验结果表明,合成水下数据集和真实水下数据集之间仍存在差异,基于 GAN 的水下图像增强方法在鲁棒性和泛化方面表现不够理想<sup>[12]</sup>。

针对生成对抗网络在合成数据集中存在的问题, Li 等<sup>[12]</sup>构建了一个具有参考图像的水下图像基准数据集 (Underwater Image Enhancement Benchmark dataset, UIEB),该数据集针对每一张真实水下图像,提供了参考增强图像,同时基于此数据集提出 Water-Net 方法。除 Water-Net 方法外,其他基于卷积神经网络的水下图像增强方法如 Shallow-UWnet<sup>[16]</sup>、UWCNN<sup>[17]</sup>以及 UWCNN-SD<sup>[18]</sup>等也相继被提出,但这些方法均无法提供能够处理水下图像多样性的通用方案。

基于上述分析,本文针对现有水下图像增强方法存在的局限性,提出一种基于双重残差混合注意力模块的水下图像增强方法 (Dual Residual Hybrid Attention Underwater Image Enhancement method, DRHA-UIE),该方法以低质量水下图像为输入,采用双重残差混合注意力模块完成特征学习,有效捕捉空间和通道信息,最终得到能够有效增强水下图像的增强模型。实验结果表明,该方法在定性和定量比较中均得到较优的结果。本文的主要贡献如下:

(1) 提出一种新的基于改进双重残差块的水下图像增强方法。采用具有配对操作的双重残差块完成特征学习,并在其中嵌入混合通道注意力模块,对特征信息进行沿通道维度和空间维度的注意力加权。获得了颜色自然、对比度高的增强图像,有效提高了水下图像的质量。

(2) 提出一个新的联合特征损失函数用于网络学习。设计了一个综合考虑  $l_1$  损失、结构相似性损失和内容感知损失的联合特征损失函数,在有效恢复图像颜色和提高图像对比度的同时,获得了更精细的图像纹理。

(3) 为满足深度学习训练数据集的需求,构建了新的数据集 N-UIEB。使用 Sea-thru<sup>[11]</sup>方法生成新的参考图像,并通过定量和定性比较,将优于原参考图像的新参考图像加入数据集,使数据集中的参考图像更为符合人的视觉感知,从而激励产生颜色更自然、细节更清晰的增强图像。

## 2 双重残差混合注意力网络

本文针对水下图像的颜色失真和对比度低问题,提出了一种基于双重残差混合注意力模块的网络模型,有效提高了视觉质量,最终获得了颜色自然、对比度高的增强图像。本节将对模型结构进行详细阐述。

## 2.1 网络整体结构

DRHA-UIE 的结构如图 2 所示. 网络以尺寸为  $256 \times 256$  的 3 通道图像为输入, 首先使用 3 个卷积层对图像进行基础特征提取. 其中第 1 个卷积层 conv1 使用核尺寸为  $7 \times 7$ , padding 为 3 的卷积操作完成特征通道数的变换, 输出特征通道数为 64, conv2 和 conv3 采用相同结构, 使用卷积核大小为  $3 \times 3$ , padding 为 1, 步长为 2 的卷积操作完成下采样, 其输出层尺寸为输入层的 2 倍下采样. 然后将提取到的基础特征  $F$  送入双重残差混合注意力 (Dual Residual Hybrid Attention, DRHA) 模块中, 以进行深层特征提取, 对于输入的特征, DRHA 模块为强化更重要的通道和空间信息, 引入混合注意力模块使网络更有效地捕捉对结果贡献更大的特征, 得到经通道和空间加权后的特征  $F_w$ . 最后, 完成图像重建, 使用 PixelShuffle<sup>[21]</sup> 和卷积操作对特征  $F_w$  进行上采样和通道恢复, 最终输出尺寸为  $256 \times 256$  的 3 通道图像, 即增强结果图像. 为减少梯度反向传播时可能出现的梯度消失或梯度爆炸, 设置从输入到输出的残差连接.

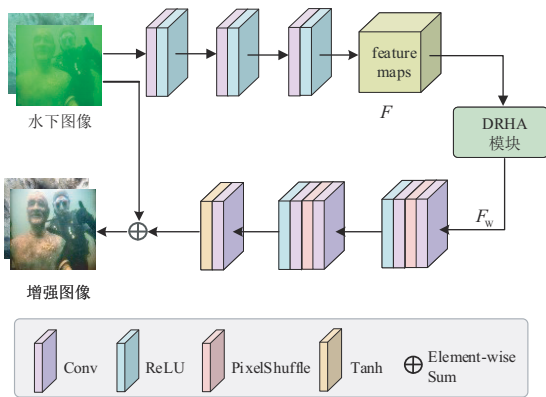


图 2 DRHA-UIE 图像增强模型

## 2.2 DRHA 模块

DRHA 模块基于配对操作. 2018 年, Saganuma 等<sup>[22]</sup>提出一种用于图像恢复任务搜索卷积自编码器 (Convolutional Auto-Encoder, CAE) 架构的进化算法, 实验表明, 反复执行一对卷积核大小不同的配对卷积层, 能够更好地实现图像去噪. Liu 等<sup>[23]</sup>以 Saganuma 等人的工作为基础, 进一步探索了配对操作在图像恢复任务中的应用, 提出基于配对操作的双重残差块. 图 3 展示了配对操作间三种不同的连接方式, 其中,  $f_i$  表示配对操作中的第一个操作,  $g_i$  表示配对操作中的第二个操作, Liu 等设计的双重残差块间采用了图 3 中第三种连接方式.

### 2.2.1 DRHA 模块结构

DRHA 模块采用 [(Conv. + PixelShuffle), 混合注意力模块 + Conv.] 的配对操作. 对于输入到 DRHA 模块的基础特征  $F \in \mathbf{R}^{C \times H \times W}$ , 首先通过两个卷积操作来学习特

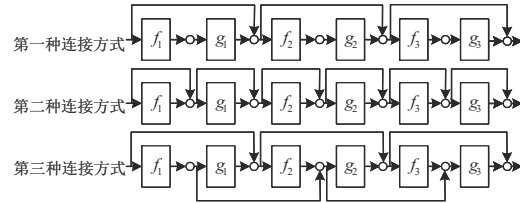


图 3 双重残差结构

征, 为获得更多的特征信息, 本文对特征执行 PixelShuffle 将特征尺寸扩大为原来的 2 倍, 通道数保持不变, 得到处理后的特征  $F' \in \mathbf{R}^{C \times 2H \times 2W}$ . 然后将其送入设计的混合注意力模块, 获得经混合注意力加权后的特征图  $X_w \in \mathbf{R}^{C \times 2H \times 2W}$ , 混合注意力模块结合了通道注意和空间注意机制, 使得网络更加专注失真严重的像素和更重要的通道, 同时增加了网络的灵活性. 最后使用一个卷积操作对特征尺寸进行恢复, 将特征尺寸从  $2H \times 2W$  恢复至  $H \times W$ , 得到最终的特征  $F_w$ . DRHA 模块在卷积操作、配对操作、混合注意力模块均设置残差连接, 部分残差连接使网络能够选择性舍弃低频信息, 进一步提高网络性能和训练的稳定性. DRHA 采用如图 3 所示的第一种连接方式, 将配对结构作为一个整体执行跳跃连接, 其主要原因为, 考虑到水下图像增强算法需以水下机器人或航行器为载体, 可用内存通常较小, 本文为有效减少模型参数量, 仅采用一个 DRHA 模块, 无法实现 DRHA 模块间的相互作用, 且配对操作中的 (Conv. + PixelShuffle) 操作改变了特征尺度, 无法直接连接. 本文设计的 DRHA 模块的结构如图 4 所示, 实验结果表明, 采用此连接方式的双重残差块在水下增强任务中表现优异.

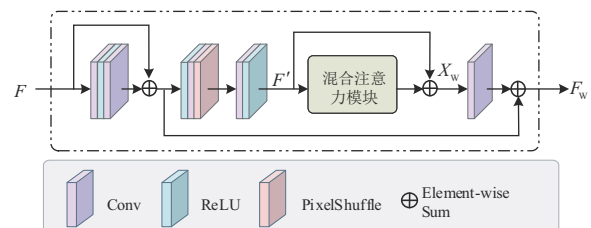


图 4 DRHA 模块结构

### 2.2.2 混合注意力模块

目前, 注意力机制在多项视觉任务中展示出优越性, 该机制使网络模型在训练过程中聚焦整个视觉空间中的重点区域, 以获得更多关注目标的细节信息, 同时抑制其他无用信息. 为更有效的学习特征, 本文设计了混合注意力模块, 对特征进行通道和空间两个维度的注意力权重推断, 以加强对增强任务贡献更大的特征. 混合注意力模块的结构如图 5 所示, 其中黄色区域表示通道注意力模块, 蓝色区域表示空间注意力模块. 对于经 PixelShuffle 处理后的基础特征  $F' \in \mathbf{R}^{C \times 2H \times 2W}$ , 混合注意力模块首先应用两个卷积操作完成特征学习, 得到通道注意力模块的输入  $X \in \mathbf{R}^{C \times 2H \times 2W}$ .

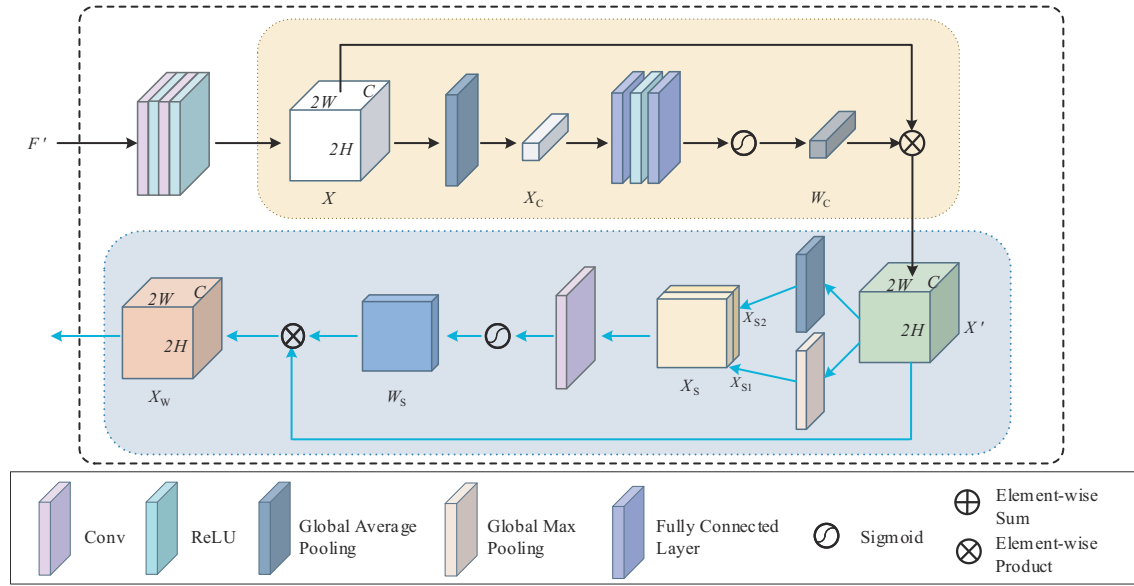


图5 混合注意力模块

通道注意力机制通过学习的方式来自动获取每个特征通道的重要程度,然后依照重要程度去提升相对重要的通道的权重并抑制重要程度较低的通道权重.在本文中,首先对通道注意力模块的输入特征  $X$  进行全局平均池化,将每个通道上的空间信息编码为一个全局特征,得到通道描述符  $X_c \in \mathbf{R}^{C \times 1 \times 1}$ . 然后,为学习不同通道之间的非线性关系,将  $X_c$  送入由两个全连接层组成的网络中,此过程描述如下:

$$W_c = \sigma(W_2 \delta(W_1 \text{AvgPool}(X))) \quad (1)$$

其中,  $W_c$  表示与输入特征  $X$  具有相同通道数的通道注意力权重,  $\delta(\cdot)$  表示 ReLU 激活函数,  $\sigma(\cdot)$  表示 sigmoid 激活函数,  $W_1$ 、 $W_2$  分别表示第一个和第二个全连接层的权重,  $W_1 \in \mathbf{R}^{C/r \times C}$ ,  $W_2 \in \mathbf{R}^{C \times C/r}$ , 为了降低模型复杂度,第一个全连接层对模型进行降维,  $r$  表示降维系数,在 Hu 等人<sup>[24]</sup>和 Woo 等人<sup>[25]</sup>提出的注意力机制中,设置  $r$  为 16,考虑到在水下增强任务中,较高的降维系数会损失特征细节,因此本文设置  $r$  为 2,实验表明,降维系数的调整有利于获得更高质量的增强图像.最后,将输入特征  $X$  与通道注意力权重  $W_c$  按通道维度相乘,获得通道加权后的特征图  $X' \in \mathbf{R}^{C \times 2H \times 2W}$ ,如式(2)所示,其中  $i = 1, 2, \dots, C$  表示特征的不同通道:

$$X'^i = W_c^i \cdot X^i \quad (2)$$

特征重要程度的差异不仅存在于不同通道之间,还存在于不同的空间位置之间,空间注意力机制通过捕获特征不同位置的权重,使得模型更加关注对任务贡献更大的位置.对于空间注意力的输入特征  $X'$ ,首先应用全局平均池化和全局最大池化分别获得通道降维后的空间描述符  $X_{s1} \in \mathbf{R}^{1 \times 2H \times 2W}$  和  $X_{s2} \in \mathbf{R}^{1 \times 2H \times 2W}$ ,然后对两者执行通道拼接后得到空间描述符

$X_s \in \mathbf{R}^{2 \times 2H \times 2W}$ ,最后经一个核为  $7 \times 7$  的卷积层降维得到通道数为 1 的特征,此过程描述如下:

$$W_s = \sigma(\text{Conv}^{7 \times 7}([\text{AvgPool}(X'); \text{MaxPool}(X')])) \quad (3)$$

其中,  $W_s$  表示与输入特征  $X'$  具有相同尺寸,通道数为 1 的空间注意力权重,  $\sigma(\cdot)$  表示 sigmoid 激活函数.将输入特征  $X'$  与通道注意力权重  $W_s$  按位置相乘,获得经混合注意力机制加权后的特征图  $X_w \in \mathbf{R}^{C \times 2H \times 2W}$ ,如式(4)所示,其中  $m = 1, 2, \dots, 2H$ ,  $n = 1, 2, \dots, 2W$  表示特征图中的不同位置:

$$X_w = W_s^{m,n} \cdot X'^{m,n} \quad (4)$$

在网络训练过程中,常使用批归一化层(Batch Normalization, BN)来对特征进行归一化,以加速网络收敛, Hu 等人<sup>[24]</sup>和 Woo 等人<sup>[25]</sup>提出的注意力机制中均使用了 BN 层.本文通过实验发现,在水下增强任务中,使用 BN 层的网络模型表现较差,其主要原因为, BN 层在归一化过程使用的参数是每个 mini-batch 的一阶和二阶统计量,这便要求每个 mini-batch 的统计量彼此之间满足近似同分布,并能够近似代表整体统计量,但本文分析实验数据发现,不同 mini-batch 间的数据分布存在一定差异,图 6 展示了对不同 mini-batch 分别统计 R、G、B 颜色通道得到的数值分布结果, batch 大小设置为 40.

如图 6 所示,对于 G 通道和 B 通道,不同 mini-batch 数据间基本满足近似同分布,但各 mini-batch 间 R 通道分布差距较大.其原因为,在水下图像形成过程中,长波光的衰减程度要大于短波光,使得红色等长波光更容易受到衰减的影响,因此 R 通道在不同图像中呈现出更多样的分布.不同 mini-batch 统计量间的分布差异使得 BN 层表现较差,并对网络的训练造成负担,因此在本文设计的混合注意力模块中,并未使用 BN 层.

实验结果表明,不使用BN层的混合注意力模块能够得到具有更高质量的增强图像。

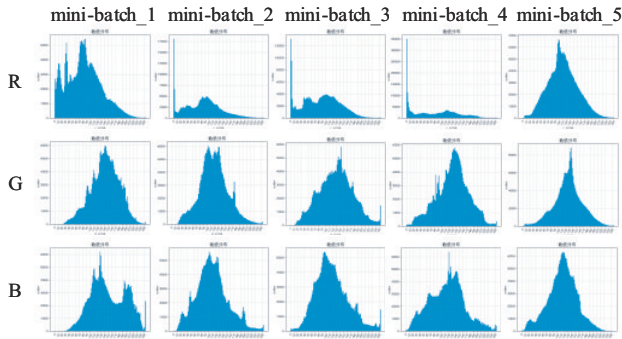


图6 不同 mini-batch 数值统计

### 2.3 损失函数

在空气中图像的增强任务中,常使用 $l_1$ 损失函数来衡量图像间的像素级差距,SSIM (Structural Similarity)<sup>[26]</sup>损失函数来获得与目标图像具有相似结构的增强图像,而在水下图像增强任务中,为减少像素级损失造成的纹理过度平滑,常使用内容感知损失<sup>[15]</sup>.为使获得的增强图像在像素、结构及纹理方面均无限接近目标图像,本文设计了一个联合特征损失函数,该损失函数由三个损失项组成,分别为 $l_1$ 损失、结构相似性损失和内容感知损失,并在实验中使用此联合特征损失函数对DRHA-UIE进行训练,有效提高了图像质量.本小节将对该损失函数进行详细介绍,其中使用 $y'$ 表示网络的输出图像, $y$ 表示数据集中提供的参考图像.

#### (1) $l_1$ 损失

使用 $l_1$ 损失来衡量参考图像与网络训练结果间的像素级损失,其公式如下:

$$\text{Loss}_{l_1}(y', y) = \sum_{p=1}^N |y'(p) - y(p)| \quad (5)$$

其中, $p$ 表示图像中的像素点, $N$ 表示像素点的数量.

#### (2) 结构相似性损失

结构相似性指数从图像组成的角度将结构信息定义为独立于亮度、对比度的信息,SSIM损失从亮度、对比度和结构三个角度来衡量图像间的结构相似程度.其公式如下:

$$\text{Loss}_{\text{SSIM}}(y', y) = [l(y', y)^\alpha \cdot c(y', y)^\beta \cdot s(y', y)^\gamma] \quad (6)$$

其中:

$$l(y', y) = \frac{2\mu_{y'}\mu_y + c_1}{\mu_{y'}^2 + \mu_y^2 + c_1}, \quad c(y', y) = \frac{2\sigma_{y'}\sigma_y + c_2}{\sigma_{y'}^2 + \sigma_y^2 + c_2},$$

$$s(y', y) = \frac{\sigma_{y'y} + c_3}{\sigma_{y'}\sigma_y + c_3} \quad (7)$$

$\mu_{y'}, \mu_y$ 分别为图像 $y'$ 与图像 $y$ 的均值, $\sigma_{y'}^2, \sigma_y^2$ 分别为图像 $y'$ 与图像 $y$ 的方差, $c_i (i=1, 2, 3)$ 为避免除数为0

引入的常数项. SSIM作为损失函数从亮度 $l$ 、对比度 $c$ 、结构 $s$ 三个维度来衡量图像 $y'$ 与 $y$ 的相似度.在实际应用中,为简化计算,一般设置 $\alpha = \beta = \gamma = 1, c_3 = c_2/2$ ,则SSIM表达式简化为

$$\text{Loss}'_{\text{SSIM}}(y', y) = \frac{(2\mu_{y'}\mu_y + c_1)(2\sigma_{y'y} + c_2)}{(\mu_{y'}^2 + \mu_y^2 + c_1)(\sigma_{y'}^2 + \sigma_y^2 + c_2)} \quad (8)$$

#### (3) 内容感知损失

内容感知损失激励产生与目标图像相似的特征表示,以获得具有更精细纹理的增强图像,其公式如下:

$$\text{Loss}_{\text{PER}}^\phi(y', y) = \frac{1}{C_j H_j W_j} \|\phi_j(y') - \phi_j(y)\|_2^2 \quad (9)$$

其中, $C_j$ 表示VGG-19网络中卷积层 $j$ 提取的特征图的通道数, $H_j$ 表示特征图的高度, $W_j$ 表示特征图的宽度, $\phi_j(\cdot)$ 表示内容感知函数,受文献[15]启发,本文指定 $\phi_j(\cdot)$ 为预训练的VGG-19网络block5\_conv2层提取的特征图.

综上,最终的损失函数为 $l_1$ 损失、结构相似性损失和内容感知损失的加权和,其公式如下:

$$\text{Loss} = p_1 \cdot \text{Loss}_{l_1} + p_2 \cdot \text{Loss}'_{\text{SSIM}} + p_3 \cdot \text{Loss}_{\text{PER}}^\phi \quad (10)$$

其中 $p_1, p_2, p_3$ 均表示损失权重系数,决定了不同损失函数对于最终结果的重要程度.为确定损失系数的值,本文首先使用3个损失项分别训练模型,损失函数变化情况如图7所示,模型训练往往使得损失函数最小化,而结构相似性损失其值越大,图像结构越相似,因此训练时设置结构相似性损失项的初始系数为-1,分析得知,损失项的收敛速度比例约等于1:3.4:2.45,得到损失函数基准权重约为1:0.29:0.4,本文为获得更高的SSIM指标值,适度提高结构相似性损失项的权重,并根据Islam等<sup>[15]</sup>及Liu等<sup>[23]</sup>的工作,调整内容感知损失的比例,最终设置 $p_1 = 1.0, p_2 = 0.75, p_3 = 0.3$ .

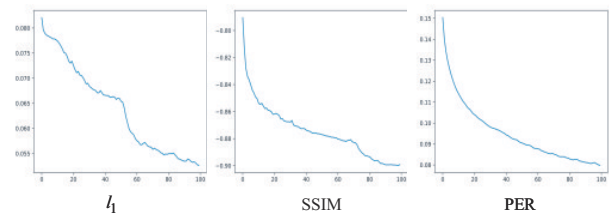


图7 各损失函数项损失函数值变化情况

## 3 实验

### 3.1 新水下数据集的构建

网络模型通过最小化输出图像和Ground Truth图像的损失差值来更新参数,Ground Truth图像的质量将直接影响模型的输出结果,在UIEB数据集中,参考图像即为Ground Truth图像,其质量决定了模型能否得到高质量的增强图像. UIEB数据集使用12种图像增强

方法来生成候选参考图像<sup>[12]</sup>,包括9种水下图像增强方法、2种去雾方法和1项用于增强水下图像的商业应用.对于得到的候选增强图像,选择50名志愿者,其中25名具有图像处理相关经验,另外25名无相关经验,每位志愿者在同一显示器下对每张水下图像的12种增强结果进行两两比较,最终经过多数表决获得最终的参考图像.UIEB数据集共有890对配对图像,包含多样的水下场景和水下对象.

UIEB数据集中的参考图像来自于多种增强算法获得的增强结果的择优,但该数据集仍存在限制,具体来说,在使用现有算法进行水下图像增强的过程中,后向散射的影响仍然无法完全消除,尤其是对于远距离的后向散射,其主要原因为现有算法遵循的图像形成模型或假设不正确.本文使用Sea-thru<sup>[11]</sup>方法对UIEB数据集进行改进,这一方法以Akkaynak等<sup>[27]</sup>提出的水下图像形成模型为理论基础,该水下图像形成模型首次论证了衰减系数和后向散射系数不同且具有不同的依赖条件,Sea-thru方法不仅输入水下图像,还需输入其对应的深度图.深度图包含水下图像中每个点到成像设备的距离信息,由于UIEB数据集中未提供深度图,本文首先使用Godard等<sup>[28]</sup>提出的基于自监督的单目深度估计算法来完成对图像深度图的估计,因缺少实际距离信息,此算法获得的深度只是相对深度,在使用前需进行如下映射:

$$F(x_i, y_i) = A_{\max}(1 - G(x_i, y_i)) + A_{\min}G(x_i, y_i) \quad (11)$$

其中,  $(x_i, y_i)$  表示坐标值,  $F$  为映射后的绝对深度,  $G$  为应用单目深度估计算法获得的相对深度,  $A_{\max}$ 、 $A_{\min}$  分别表示映射后的最远距离与最近距离.由式(11)可知,绝对深度的值不仅取决于获得的相对深度,还与  $A_{\max}$ 、 $A_{\min}$  取值有关.为获得  $A_{\max}$ 、 $A_{\min}$  取值,本文对文献[11]中的RGBD(Red, Green, Blue, Depth)水下数据集进行分析,该数据集包含暗礁和峡谷两种水下场景,其中暗礁场景的  $A_{\max}$  值为2~3.5 m,  $A_{\min}$  值为0 m,峡谷场景  $A_{\max}$  值为3.5~14 m,  $A_{\min}$  值为0 m.为包含更复杂的场景和更多变的拍摄角度,本文适当拓宽距离范围,设置  $A_{\max} \in [2 \text{ m}, 16 \text{ m}]$ , 步长为1 m,  $A_{\min} \in [0 \text{ m}, 1.5 \text{ m}]$ , 步长为0.5 m.对于每张水下图像,遍历  $A_{\max}$ 、 $A_{\min}$  取值,获得  $3 \times 15$  共45张具有不同距离范围的深度图,然后将其与水下图像共同作为Sea-thru方法的输入,获得45张增强后的水下图像.对于每张水下图像获得的45张增强图像,首先采用UIQM(Underwater Image Quality Measure)<sup>[29]</sup>和UCIQE(Underwater Color Image Quality Evaluation)<sup>[30]</sup>两项无参考评价指标进行初步筛选,选出定量表现最优的增强图像作为候选参考图像.然后将其与UIEB中的参考图像进行比较,若候选参考图像在定量和定性比较中均胜出,则用此候选参考图像替换UIEB中的原参考图像.其中定量比较使用

UIQM、UCIQE和信息熵三项指标,采用多数表决策略,若候选参考图像的三项指标结果中至少两种优于原参考图像,则其在定量比较中胜出.在定性比较阶段,组织5名具有图像处理背景的志愿者对两张参考图像进行评价,若三者一致认为候选参考图像优于UIEB中参考图像,则候选参考图像胜出.N-UIEB数据集最终替换参考图像33张,部分示例如图8所示.



图8 参考图像对比

### 3.2 实验设计

本文使用Pytorch框架实现了DRHA-UIE方法,并在UIEB、N-UIEB以及EUVP(Enhancing Underwater Visual Perception)<sup>[15]</sup>数据集上分别进行了训练和测试.因UIEB和N-UIEB数据集数据量较小,本文通过翻转和旋转获取原始图像的3个扩充版本,形成由3560张图像组成的数据集,并将图像尺寸调整为  $256 \times 256$ ,其中随机选取500张图像作为测试集,其余作为训练集.在EUVP数据集上,本文选用EUVP-ImageNet数据集作为训练集,EUVP-Dark数据集作为测试集.实验在2块Nvidia RTX 2080 Ti显卡上完成,使用Adam进行网络优化,每训练100个epoch大约需要2.5 h.批量大小(batch\_size)设置为20,初始学习率为  $10^{-4}$ ,为防止越过全局最优解,每1000轮降低学习率为原来的0.1倍,直到网络收敛.

### 3.3 实验结果与分析

#### 3.3.1 定性比较

将DRHA-UIE与其他方法进行定性比较,部分输出结果如图9所示.其中histogram-prior<sup>[7]</sup>, fusion-based<sup>[4]</sup>, retinex-based<sup>[5]</sup>, UDCP(Underwater Dark Channel Prior)<sup>[8]</sup>均为非学习方法,不需要在配对数据集上进行训练,对于UGAN<sup>[14]</sup>, FUnIE-GAN<sup>[15]</sup>, Water-Net<sup>[12]</sup>, Shallow-UWnet<sup>[16]</sup>, Ucolor<sup>[31]</sup>, 去雾算法FFA-Net(Feature Fusion Attention Network)<sup>[32]</sup>以及本文DRHA-UIE方法,使用N-UIEB数据集进行训练.

如图9所示,UDCP获得的增强图像无法有效去除水下图像中较严重的偏色,如第一行中增强后的图像仍有明显的绿色;histogram-prior、retinex-based以及fusion-based获得的部分增强图像中会引入红色伪影,

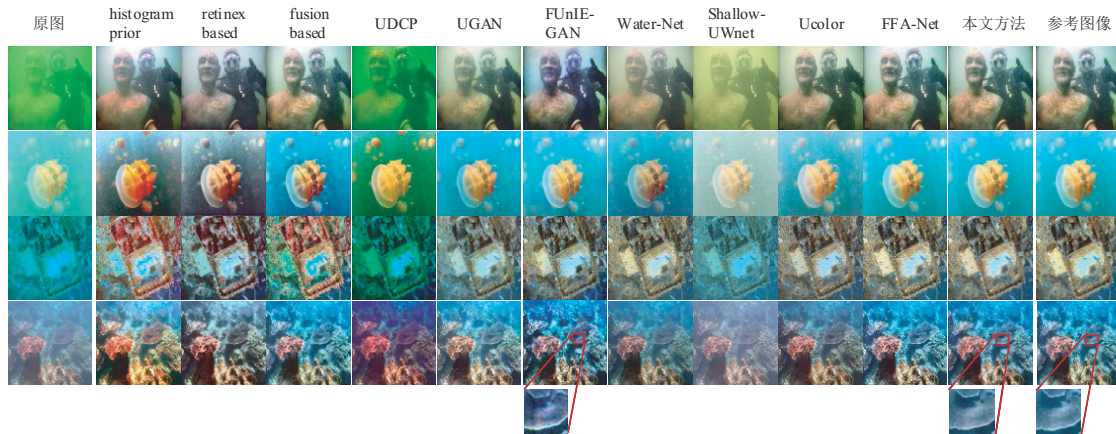


图9 将DRHA-UIE与其他方法进行定性比较

如上数第三行所示;Shallow-UWnet获得的增强图像具有不自然的颜色,且对模糊的去除效果较差;UGAN、Water-Net以及Ucolor获得的部分增强图像对于浑浊水体的处理效果较差,图像具有较明显的模糊;FUnIE-GAN方法获得的增强图像,虽有效提高了图像的对比度,但处理后的图像清晰度下降,且会出现颜色不均匀以及纹理不够精细的问题,如图9中最后一行局部放大图所示。而本文提出的DRHA-UIE方法和去雾算法FFA-Net在颜色校正、细节恢复和对比度提高方面均表现优异,获得了与参考图像视觉上最为接近的增强

图像。

### 3.3.2 客观评价

为测试DRHA-UIE方法的性能,本文采用PSNR (Peak Signal-to-Noise Ratio)和SSIM两种定量评价指标对方法进行客观评价。本文将DRHA-UIE及对比方法在UIEB数据集和N-UIEB数据集上分别进行了训练和测试,对获得的输出图像计算PSNR及SSIM两项指标值,并计算模型的参数量和FLOPs值,最终定量结果如表1所示(此处基于学习的方法呈现的为平均结果,排名前2的数据加粗表示)。

表1 UIEB和N-UIEB数据集增强结果对比

| 方法                             | UIEB            |                | N-UIEB          |                | FLOPs         | 参数量          |
|--------------------------------|-----------------|----------------|-----------------|----------------|---------------|--------------|
|                                | PSNR            | SSIM           | PSNR            | SSIM           |               |              |
| histogram-prior <sup>[7]</sup> | 19.064 0        | 0.787 5        | 18.890 0        | 0.787 0        | —             | —            |
| fusion-based <sup>[4]</sup>    | 22.231 6        | 0.848 9        | 22.237 8        | 0.849 2        | —             | —            |
| retinex-based <sup>[5]</sup>   | 17.398 8        | 0.768 2        | 17.394 3        | 0.768 0        | —             | —            |
| UDCP <sup>[8]</sup>            | 12.936 8        | 0.619 9        | 12.917 0        | 0.619 5        | —             | —            |
| UGAN <sup>[14]</sup>           | 22.742 7        | 0.901 1        | 22.762 8        | 0.900 1        | 18.0 G        | 41.8 M       |
| FUnIE-GAN <sup>[15]</sup>      | 22.768 0        | 0.853 5        | 22.123 9        | 0.831 8        | 10.2 G        | 7.0 M        |
| Water-Net <sup>[12]</sup>      | 20.227 0        | 0.870 0        | 20.233 3        | 0.869 5        | 70.7 G        | 1.1 M        |
| Shallow-UWnet <sup>[16]</sup>  | 20.783 0        | 0.836 0        | 20.836 0        | 0.834 0        | 21.6 G        | <b>0.2 M</b> |
| Ucolor <sup>[31]</sup>         | 22.051 7        | 0.879 0        | 22.048 1        | 0.879 2        | 2805.3 G      | 157.4 M      |
| FFA-Net <sup>[32]</sup>        | <b>34.859 7</b> | <b>0.977 7</b> | <b>34.369 9</b> | <b>0.975 6</b> | 302.7 G       | 4.7 M        |
| 本文方法                           | <b>30.054 9</b> | <b>0.966 7</b> | <b>30.083 9</b> | <b>0.967 0</b> | <b>12.0 G</b> | <b>0.9 M</b> |

由表1可知,本文提出的方法在原数据集UIEB和改进数据集N-UIEB上,PSNR和SSIM两项指标结果均排名第二,大幅度优于对比的水下图像增强方法,且模型参数量较小,并在小于1/5参数量的情况下最大程度接近去雾算法FFA-Net,确保了模型能够在水下机器人或水下航行器等小型设备上运行。

本文同样在EUVP数据集上对DRHA-UIE及对比方法进行了训练和测试,该数据集通过各种相机如Go-Pros、低光USB等,在多种能见度的海洋环境中收集图

像,并使用CycleGAN生成相应的配对图像,EUVP数据集的实验结果如表2所示。由表2所知,本文方法表现优秀,获得了最优的PSNR和SSIM指标值。

### 3.3.3 不同数据集增强结果对比

为验证改进数据集对实验结果的影响,分别对以UIEB或N-UIEB为数据集获得的增强结果进行定量和定性比较。采用UIQM作为定量评价指标,如表3所示,以N-UIEB作为实验数据集获得的增强图像具有更高的UIQM值,这表示图像具有更高的视觉质量。

表 2 EUVP数据集增强结果对比

| 方法                            | PSNR     | SSIM    |
|-------------------------------|----------|---------|
| UGAN <sup>[14]</sup>          | 18.759 6 | 0.785 5 |
| FUnIE-GAN <sup>[15]</sup>     | 17.686 7 | 0.662 2 |
| Water-Net <sup>[12]</sup>     | 15.582 9 | 0.741 4 |
| Shallow-UWnet <sup>[16]</sup> | 19.439 1 | 0.803 6 |
| Ucolor <sup>[31]</sup>        | 19.501 9 | 0.847 0 |
| 本文方法                          | 19.598 3 | 0.848 0 |

表 3 UIEB或N-UIEB数据集增强图像UIQM指标值

| 数据集    | UIQM    |
|--------|---------|
| UIEB   | 4.947 9 |
| N-UIEB | 5.054 1 |

图 10展示了两种数据集分别获得的增强图像. 其中,中间行表示以UIEB为数据集得到的增强结果,最下行表示以N-UIEB为数据集得到的增强结果,观察可知,两者均有效解决了水下图像中存在的对比度低和偏色问题,但相较于UIEB,基于N-UIEB训练得到的增强结果对模糊的去除效果更加明显,得到的图像具有更自然的颜色和更清晰的细节.

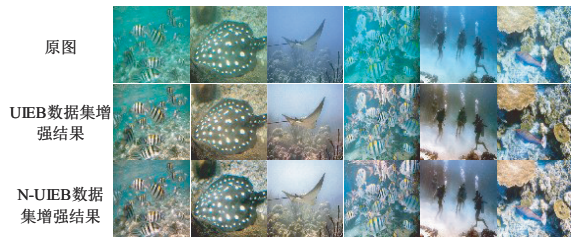


图 10 以UIEB、N-UIEB为训练数据集获得的增强图像

## 4 讨论

### 4.1 损失函数消融实验

为进一步探究损失函数中 $l_1$ 损失项、结构相似性损失项、内容感知损失项分别对实验结果的影响,对各损失项进行消融实验. 在N-UIEB数据集上分别使用不同的损失函数进行训练和测试,最终实验结果如表4所示(此处呈现的为平均结果),排名前二的结果加粗表示. 结果表明,本文使用的联合特征损失函数可以获得最优的PSNR和SSIM指标值.

图 11展示了使用不同损失函数分别训练获得的增强结果,其中最后一行是对第三行图像应用边缘检测

表 4 损失项消融实验

| 损失项                              |                      |                     | PSNR            | SSIM           |
|----------------------------------|----------------------|---------------------|-----------------|----------------|
| Loss <sub><math>l_1</math></sub> | Loss <sub>SSIM</sub> | Loss <sub>PER</sub> |                 |                |
|                                  | √                    | √                   | 28.302 2        | 0.966 5        |
| √                                |                      | √                   | 30.059 7        | 0.953 2        |
| √                                | √                    |                     | <b>30.060 1</b> | <b>0.966 7</b> |
| √                                | √                    | √                   | <b>30.083 9</b> | <b>0.967 0</b> |

的结果,如图 11所示,相较于其他损失函数,本文使用的损失函数能够得到颜色自然、细节丰富的增强结果.

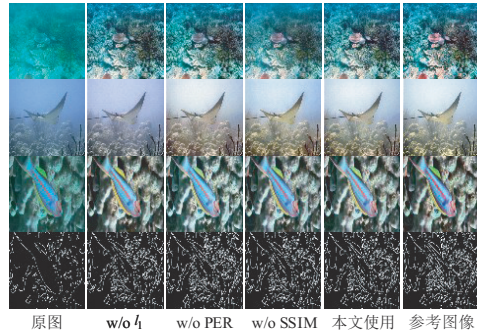


图 11 分别使用不同损失函数获得的增强结果

### 4.2 注意力机制消融实验

为探究本文设计的混合注意力模块在水下增强任务中的作用,将本文提出的混合注意力模块以及Woo等人提出的CBAM(Convolutional Block Attention Module)分别加入DRHA-UIE中,并对本文混合注意力模块中的修改降维系数和舍弃BN层两项操作进行消融研究,以验证不同混合注意力设计方案在DRHA-UIE中的性能. 将它们分别在N-UIEB数据集训练1500轮,得到实验结果如表5所示. 实验表明,本文设计的混合注意力模块能够更有效完成特征信息的注意力权重推断,在保留细节特征的同时,更好的完成了图像增强工作,获得的增强图像在PSNR和SSIM两项指标均优于CBAM.

表 5 注意力机制消融实验

| 注意力机制         | PSNR            | SSIM           |
|---------------|-----------------|----------------|
| CBAM          | 24.044 8        | 0.928 1        |
| CBAM + 舍弃BN层  | 26.737 2        | 0.949 0        |
| CBAM + 修改降维系数 | 24.505 5        | 0.925 5        |
| 本文混合注意力模块     | <b>28.557 6</b> | <b>0.961 5</b> |

### 4.3 基于合成数据集进行实验

为证明本文算法的优秀性能,除基于真实水下数据集UIEB和N-UIEB的实验外,本文还基于合成数据集进行了训练和测试. Li等人<sup>[17]</sup>根据水下场景的物理模型和光学特性,以RGB-D NYU-v2室内数据集为基础,合成了十种类型的合成水下图像,其中包含五种类型的开阔海域(即Type-I、Type-1A、Type-1B、Type-II和Type-III)和五种类型的沿海海域(即Type-1、Type-3、Type-5、Type-7和Type-9). 本文选择了三种典型类型进行实验结果的对比,对比方法为UWCNN<sup>[17]</sup>、UWCNN-SD<sup>[18]</sup>、Shallow-UWnet<sup>[16]</sup>以及Ucolor<sup>[31]</sup>,结果如表6所示,其中排名最高的结果加粗显示,结果表明,本

文方法在三种类型的水下数据集中均获得了最高的 SSIM 值和最高的 PSNR 值。

表 6 合成数据集增强结果对比

| 方法                            | Type-I          |                | Type-5          |                | Type-III        |                |
|-------------------------------|-----------------|----------------|-----------------|----------------|-----------------|----------------|
|                               | PSNR            | SSIM           | PSNR            | SSIM           | PSNR            | SSIM           |
| UWCNN <sup>[17]</sup>         | 21.185 4        | 0.856 3        | 16.635 8        | 0.744 2        | 21.739 3        | 0.854 2        |
| UWCNN-SD <sup>[18]</sup>      | 15.723 4        | 0.739 4        | 13.395 6        | 0.661 8        | 15.705 7        | 0.741 5        |
| Shallow-UWnet <sup>[16]</sup> | 18.763 8        | 0.751 6        | 15.060 1        | 0.590 0        | 19.571 2        | 0.769 9        |
| Ucolor <sup>[31]</sup>        | 22.630 8        | 0.875 9        | 18.521 9        | 0.770 7        | 22.232 3        | 0.872 4        |
| 本文方法                          | <b>25.222 1</b> | <b>0.926 8</b> | <b>19.556 8</b> | <b>0.801 7</b> | <b>24.427 4</b> | <b>0.920 4</b> |

## 5 总结

因水中悬浮颗粒对光线的吸收和散射,水下图像往往存在可见度低、对比度差、偏色等问题,针对水下图像存在的上述问题,本文提出了一个基于双重残差混合注意力模块的鲁棒性水下图像增强方法 DRHA-UIE. 该方法采用基于配对操作的双重残差块来完成特征提取,并引入混合注意力模块,对特征沿通道和空间维度计算注意力权重,以加强有用特征. 使用基于  $l_1$  损失、结构相似性损失和内容感知损失的联合特征损失函数对方法进行端到端训练. 应用修正的水下图像形成模型对通用数据集 UIEB 进行优化,得到数据集 N-UIEB. 将 DRHA-UIE 方法与对比方法在 UIEB、N-UIEB、EUVP 以及合成数据集上分别进行了训练和测试. 实验结果表明,DRHA-UIE 方法不仅有效恢复了图像颜色并提高了图像对比度,且在细节恢复方面表现优异,在与现有方法的定量比较中,DRHA-UIE 方法获得了优秀的比较结果,得到了更接近参考图像的增强图像.

## 参考文献

- [1] JOHNSON-ROBERSON M, BRYSON M, FRIEDMAN A, et al. High-resolution underwater robotic vision-based mapping and three-dimensional reconstruction for archaeology[J]. *Journal of Field Robotics*, 2017, 34(4): 625-643.
- [2] BRYSON M, JOHNSON-ROBERSON M, PIZARRO O, et al. Automated registration for multi-year robotic surveys of marine benthic habitats[C]//2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo: IEEE, 2013: 3344-3349.
- [3] HAN M, LYU Z Y, QIU T, et al. A review on intelligence dehazing and color restoration for underwater images[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 50(5): 1820-1832.
- [4] ANCUTI C, ANCUTI C O, HABER T, et al. Enhancing underwater images and videos by fusion[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 81-88.
- [5] FU X Y, ZHUANG P X, HUANG Y, et al. A retinex-based enhancing approach for single underwater image[C]//2014 IEEE International Conference on Image Processing (ICIP). Paris: IEEE, 2014: 4572-4576.
- [6] GALDRAN A, PARDO D, PICÓN A, et al. Automatic red-channel underwater image restoration[J]. *Journal of Visual Communication and Image Representation*, 2015, 26: 132-145.
- [7] LI C Y, GUO J C, CONG R M, et al. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior[J]. *IEEE Transactions on Image Processing*, 2016, 25(12): 5664-5677.
- [8] DREWS P L J J, 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.
- [9] 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.
- [10] 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.
- [11] AKKAYNAK D, TREIBITZ T. Sea-thru: A method for removing water from underwater images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 1682-1691.
- [12] 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.
- [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] FABRI C, ISLAM M J, SATTAR J. Enhancing underwater imagery using generative adversarial networks[C]//2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE, 2018: 7159-7165.
- [15] 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.
- [16] 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.
- [17] LI C Y, ANWAR S, PORIKLI F. Underwater scene prior inspired deep underwater image and video enhancement [J]. Pattern Recognition, 2020, 98: 107038.
- [18] WU S C, LUO T, JIANG G Y, et al. A two-stage underwater enhancement network based on structure decomposition and characteristics of underwater imaging[J]. IEEE Journal of Oceanic Engineering, 2021, 46(4): 1213-1227.
- [19] 肖进胜, 周景龙, 雷俊锋, 等. 基于霾层学习的单幅图像去雾算法[J]. 电子学报, 2019, 47(10): 2142-2148.  
XIAO J S, ZHOU J L, LEI J F, et al. Single image dehazing algorithm based on the learning of hazy layers [J]. Acta Electronica Sinica, 2019, 47(10): 2142-2148. (in Chinese)
- [20] 盖杉, 王俊生. 基于深度学习的非局部注意力增强网络图像去雨算法研究[J]. 电子学报, 2020, 48(10): 1899-1908.  
GAI S, WANG J S. Image raindrop algorithm research using nonlocal attention enhanced network based on deep learning[J]. Acta Electronica Sinica, 2020, 48(10): 1899-1908. (in Chinese)
- [21] SHI W Z, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1874-1883.
- [22] SUGANUMA M, OZAY M, OKATANI T. Exploiting the potential of standard convolutional autoencoders for image restoration by evolutionary search[EB/OL]. (2018-03-01). <https://arxiv.org/abs/1803.00370>.
- [23] LIU X, SUGANUMA M, SUN Z, et al. Dual residual networks leveraging the potential of paired operations for image restoration[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 7000-7009.
- [24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 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. Berlin: Springer, 2018: 3-19.
- [26] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [27] AKKAYNAK D, TREIBITZ T. A revised underwater image formation model[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6723-6732.
- [28] GODARD C, MAC AODHA O, FIRMAN M, et al. Digging into self-supervised monocular depth estimation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 3827-3837.
- [29] PANETTA K, GAO C, AGAIAN S. Human-visual-system-inspired underwater image quality measures[J]. IEEE Journal of Oceanic Engineering, 2015, 41(3): 541-551.
- [30] YANG M, SOWMYA A. An underwater color image quality evaluation metric[J]. IEEE Transactions on Image Processing, 2015, 24(12): 6062-6071.
- [31] 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.
- [32] 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.

#### 作者简介



王 欣 女, 1975年9月出生于吉林省白山市. 现为吉林大学计算机科学与技术学院副教授、硕士生导师. 在国内外发表学术论文40余篇.

E-mail: w\_x@jlu.edu.cn



石 慧 女, 1997年1月出生于山东省德州市. 现为吉林大学计算机科学与技术学院硕士. 主要研究方向为数字图像处理.

E-mail: shihui19@mails.jlu.edu.cn