面向水下图像增强的内容与风格双域改善方法

刘春晓，曹越，王成骅，周子翔

(浙江工商大学计算机科学与技术学院 杭州 310018)
(cxliu@mail.zjgsu.edu.cn)

摘 要：水下图像增强技术对于海洋资源开发具有重要意义。针对现有的水下图像增强方法难以应对现实世界复杂的水下环境，尤其对于困难样本上增强效果不佳的问题，提出一种基于内容与风格双域改善的水下图像增强方法。首先设计一个简单有效的框架，将水下图像解耦为内容特征与风格特征并重建；对内容域设计了双域自适应高频调整模块（dual-domain adaptive high-frequency adjust block，DAHAB），增强原始水下图像内容特征中的细节并抑制噪声；对风格域建立风格库，并在推理过程中匹配最优风格实现风格的变换；最后通过构建内容一致性损失和重建损失，监督模型完成水下图像的重建与增强。实验结果表明，在水下图像增强领域公开的 UIEB、EUVP 和 RUIE 数据集上，所提方法的主观视觉效果和客观评价指标（PSNR、SSIM、UIQM、UCIQE 和 MUSIQ）均优于对比方法，尤其在增强困难样本方面展现了明显的优势。

关键词：水下图像增强；解耦重建；高频调整；内容与风格
中图法分类号：TP391.41 **DOI**：10.3724/SP.J.1089.2024-00106

Content and Style Dual-Domain Improvement Method for Underwater Image Enhancement

Liu Chunxiao, Cao Yue, Wang Chenghua, and Zhou Zixiang

(*School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou* 310018)

Abstract: Underwater image enhancement techniques are crucial for the development of marine resources. To address the difficulties faced by existing underwater image enhancement methods in dealing with the complex underwater environments of the real world, especially in effectively enhancing challenging samples, we propose a method based on dual-domain improvement in content and style. Firstly, a simple and effective framework is designed to decouple underwater images into content features and style features, followed by reconstruction. For the content domain, a dual-domain adaptive high-frequency adjustment block is designed to enhance detail and suppress noise in the content features of the original underwater images. For the style domain, a style library is established to enable style transformation by matching the optimal style during inference. Finally, the content consistency loss and the reconstruction loss are used to supervise the model in completing the reconstruction and enhancement of underwater images. Experimental results on the publicly available underwater image datasets UIEB, EUVP, and RUIE show that our method outperforms compared methods in both subjective visual effects and objective evaluation metrics (PSNR, SSIM, UIQM,

收稿日期：2024-02-28；修回日期：2024-07-23。**基金项目**：国家自然科学基金（61976188）；浙江省自然科学基金（LY24F020004，LZ23F020004）；浙江省重点研发计划（2023C01039）；浙江工商大学“数字+”学科建设项目（SZJ2022B016）；浙江省大学生科技创新活动计划暨新苗人才计划（2023R408035，2023R408072）。**刘春晓**（1979—），男，博士，副教授，硕士生导师，CCF 会员，主要研究方向为计算机视觉、计算机图形学、机器学习、智能系统、视觉安全与隐私保护；**曹越**（1999—），男，硕士研究生，主要研究方向为图像处理、深度学习；**王成骅**（2002—），男，在校学生，主要研究方向为图像处理、深度学习；**周子翔**（2002—），男，在校学生，主要研究方向为图像处理、深度学习。

UCIQE, and MUSIQ). Notably, it demonstrates significant advantages in enhancing challenging samples.

Key words: underwater image enhancement; decoupling and reconstruction; high-frequency adjustment; content and style

随着人类对海洋资源的进一步探索,海洋石油钻探、水下考古、生物研究、救生打捞等各类水下作业越来越频繁,相关工作均需要稳定、清晰的水下图像作为辅助分析的手段才能顺利开展。然而,由于水下环境的特殊性,获取高质量的水下图像通常面临着多重挑战:如水体对不同色光的吸收和散射程度的不同导致颜色失真,水体浑浊导致对比度降低和视觉信息的损失等。因此研发能够有效应对各类水下退化类型的水下图像增强方法,具有非常重要的意义。

已有的水下图像增强方法可大致分为 2 类:基于图像处理的传统方法和基于深度学习的现代方法。传统方法易受水域类型和光照条件等因素的影响,泛化性能有限。与传统方法相比,基于深度学习的方法通常性能更优,但它们也面临着限制与挑战,其中最显著的是多数方法依赖于成对的训练数据,即它们需要同时拥有原始水下图像和对应的高质量参考图像。当前的水下图像增强基准(underwater image enhancement benchmark,UIEB)[1]通过同时使用多种已有水下图像增强方法,并人工挑选最优结果的方式构建,然而其较小的数据量(仅 890 对原始图像和参考图像)限制了深度学习方法的性能。已有方法无法同时应对各类水下复杂退化,如模糊、偏色和低光照等。尽管水下场景启发的方法[2]尝试使用合成图像扩充数据集,然而合成图像域与真实图像域之间的差异使其性能提升有限。

为了解决上述问题,本文提出内容与风格双域改善的水下图像增强方法。首先提出基于内容与风格解耦重建的水下图像增强新框架,利用内容一致性损失保证内容特征与风格特征能够完全分离,对解耦得到的内容特征和风格特征分别进行增强,并通过融合内容特征与风格特征对水下图像进行重建,以得到更好的增强结果;然后提出双域自适应高频调整模块(dual-domain adaptive high-frequency adjust block,DAHAB),同时关注内容特征的频率域与空间域,通过学习水下图像内容特征与参考图像内容特征的互相映射,自适应地提取和调整内容特征的高频信息,帮助模型更好地解决水下退化导致的图像细节模糊、缺失等问题;提出基于建立风格库和最佳匹配的推理方法,预先收集优质水下图像的风格特征建立风格库,根据待增强的水下图像内容特征从风格库中匹配最佳的风格特征,将两者融合重建得到增强结果,利用匹配最优风格代替直接对水下风格进行增强,提升了模型对复杂水下环境的适应能力。

1 相关工作

1.1 基于图像处理的传统方法

(1)常见的基于图像处理的方法是通过估计各个颜色通道的衰减并补偿,实现水下图像增强。Berman 等[3]提出基于霾线技术的方法,通过估计蓝-红和蓝-绿颜色通道衰减比来校正水下图像的偏色;类似地,Galdran 等[4]则专注于恢复水下图像中的红色通道,通过分析水下环境的光学特性来补偿红色通道在水下的衰减。(2)基于直方图的水下图像增强方法。Hitam 等[5]提出对比度受限自适应直方图均衡化(contrast limited adaptive histogram equalization,CLAHE)技术和混合模式处理。(3)基于暗通道先验(dark channel prior,DCP)[6]和基于 Retinex 理论[7]的方法。Peng 等[8]扩展了 DCP 方法,使其不再局限于去雾,通过分析图像中的暗色区域来预测和补偿导致图像质量降低的因素;Zhang 等[9]对传统的 Retinex 理论进行扩展,利用多个尺度上的 Retinex 处理原始水下图像,每个尺度针对图像的不同特征以得到更好的结果。(4)Fusion[10]和 MMLE[11]等方法也取得了一定的成功。然而,传统方法易受水域类型和光照条件等因素的影响,通常难以应对复杂的真实场景。

1.2 基于深度学习的现代方法

近年来,越来越多的研究将深度学习应用于水下图像增强。Li 等[1]构建一个具有 950 幅水下图像的增强基准数据集,并在此基础上提出基于卷积神经网络的水下图像增强网络(underwater image enhancement network,Water-Net);Islam 等[12]提出基于条件生成对抗网络的实时水下图像增强模型(conditional GAN based model for real-time underwater image enhancement,FGAN),通过全局内容、颜色、局部纹理和风格信息评估感知图像质量,并以此建立目标函数监督对抗训练;Chen 等[13]提出

一种通过内容和风格分离的水下图像增强（underwater image enhancement via content and style separation，UIESS）方法，但其未对内容与风格解耦的过程进行监督，且仅对风格分支进行增强，未对内容分支做任何处理；Huang 等[14]提出一种基于均值教师模型的半监督水下图像恢复（semi-supervised underwater image restoration，Semi-UIR）框架，缓解了水下图像成对数据短缺的问题；Fu 等[15]提出一种基于空间和通道的网络（spatial and channel dimensions network，SCNet），通过关注水类型不敏感表示来增强水下图像；Li 等[16]提出首个基于比较学习框架的水下图像增强（comparative learning framework for under water enhancement，CLUIE）方法，在增强过程中对比使用多个增强参考，并提出区域质量优势判别网络（regional quality-superiority discriminative network，RQSD- Net）作为参考图像的质量判别器。

然而，上述基于深度学习的方法都存在一定的局限性。Water-Net、FGAN 和 CLUIE 等方法均难以有效地改善水下图像的偏色问题；UIESS、Semi-UIR 和 SCNet 等虽然能在一定程度上改善偏色，但对于偏色严重的水下图像，UIESS 的增强结果呈现暗红色，Semi-UIR 的增强结果依然存在偏色，SCNet 的增强结果存在偏色的同时还较为模糊。此外，这些方法几乎都不能对低光照的水下图像进行增强。

1.3　内容风格解耦

近年来，风格迁移的成功表明图像的内容和风格是可分离的。从输入图像中学习内容和风格的独立表示称为内容风格的解耦，该技术目前已被应用在图像到图像的转换[17]、视频内容转换[18]和 3D 动画[19]等多种任务。本文将水下图像解耦为内容特征和风格特征，并分别从这 2 个角度处理，最终实现水下图像的有效增强。

2　本文方法

本文方法整体框架如图 1 所示，分为左右 2 个部分：左部分将水下图像解耦为风格特征和内容特征，并对水下图像的内容特征进行增强，对参考图像的内容特征施加退化；右部分则是将内容特征和风格特征进行融合并重建图像。

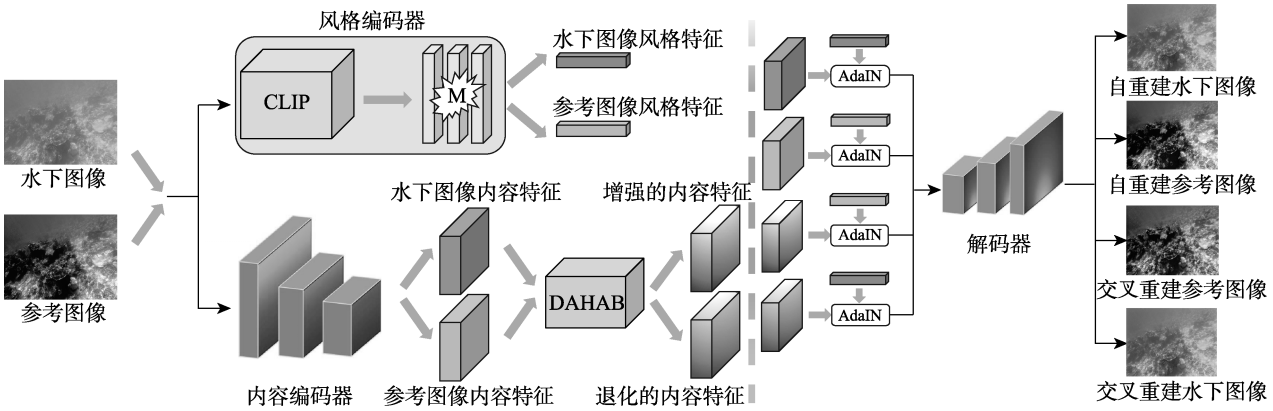


图 1　本文方法的整体框架

2.1　内容编码器与风格编码器

给定一对水下图像和参考图像，内容编码器与风格编码器分别将它们映射到内容域和风格域，公式为

$$C_{\text{raw}},C_{\text{ref}}=E^{\text{C}}\left(I_{\text{raw}},I_{\text{ref}}\right),$$

$$S_{\text{raw}},S_{\text{ref}}=E^{\text{S}}\left(I_{\text{raw}},I_{\text{ref}}\right)。$$

其中，$\{I_{\text{raw}},I_{\text{ref}}\}$ 分别表示原始水下图像和对应的参考图像；$\{E^{\text{C}},E^{\text{S}}\}$ 分别表示内容编码器和风格编码器；$\{C_{\text{raw}},C_{\text{ref}}\}$ 和 $\{S_{\text{raw}},S_{\text{ref}}\}$ 分别表示内容特征和风格特征，内容特征包含该图像所有的空间结构信息，而风格特征则与结构信息无关，仅代表图像的视觉风格。水下图像和参考图像共用一组内容编码器和风格编码器。

风格编码器由 CLIP（contrastive language-image pre-training）图像编码器[20]与风格网络 M 组成，其中，M 由一系列线性层和 ReLU 激活函数构成。在整个训练过程中，CLIP 图像编码器参数保持固定。CLIP 图像编码器具有极佳的零样本学习能力，在面对各类未接触过的复杂水下图像时，依然能够保持强大的特征提取与编码能力，可以有效地提取和区分不同图像的深层特征，不但有助于

M 获取图像对应的风格特征，还减轻了基准数据集规模较小带来的影响。

内容编码器是一个多尺度输入/输出编码器，其结构如图 2 所示。将输入图像进行 2 次下采样，分别得到分辨率为原来 1/4 和 1/16 的图像。对于原始分辨率尺度，首先将图像输入卷积层，使其通道数扩展为 32，再将其送入多尺度残差块[21]中。多尺度残差块的结构如图 2 右部所示：首先将输入的特征在通道方向上切分为 2 部分，前 16 个通道的特征分别经过下采样、残差块和上采样，后 16 个通道的特征则直接输入残差块，最后将二者的结果进行拼接，得到第 1 个尺度的内容特征。由于多尺度残差块有多个尺度的视野，因此能够更好地提取内容特征，并且还能在一定程度上减少网络的参数。对于另外 2 个尺度，首先将对应分辨率的图像输入卷积层，使其通道数分别扩展到 64 和 128；然后将其与上一个尺度的内容特征经过卷积层处理后的特征拼接；最后送入残差块中，得到对应尺度的内容特征。由此，内容编码器将得到 3 个尺度的内容特征，在提取到足够高级的特征同时，保证细节不会丢失。

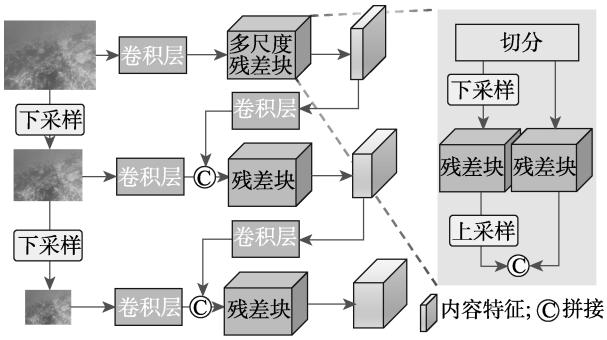


图 2 内容编码器

2.2 DAHAB

对于一对水下图像和参考图像，内容域的低频信息是不变的，但水下图像到参考图像的增强过程使得它们内容特征的高频信息发生了一些变化。为了实现后续的图像重建，本文提出 DAHAB，对它们的内容特征进行调整。

DAHAB 将学习 2 种内容特征的映射。首先学习水下图像内容特征到参考图像内容特征的增强过程，即 $M_{\text{enhance}}: C_{\text{raw}} \to C_{\text{ref}}$；其次学习参考图像内容特征到水下图像内容特征的退化过程，即 $M_{\text{degrade}}: C_{\text{ref}} \to C_{\text{raw}}$。记调整后的内容特征分别为 $C_{\text{raw}}^{\text{en}}$ 和 $C_{\text{ref}}^{\text{deg}}$。

DAHAB 的结构如图 3 所示。首先沿空间维度压缩输入的内容特征，然后通过线性层对每个通道预测 2 组参数，数学表示为

$$\sigma, w = \text{MLP}\left(P_{\text{avg}}^{\text{S}}(F) + P_{\text{max}}^{\text{S}}(F)\right)。$$

其中，MLP 是多层感知机（multilayer perceptron，MLP）；$P_{\text{avg}}^{\text{S}}$ 和 $P_{\text{max}}^{\text{S}}$ 分别表示沿空间方向的平均池化和最大池化；F 表示输入的内容特征。预测的参数均为三元组，即 $\sigma = \{\sigma_i\}, w = \{w_i\}, i = 1,2,3$，且 $\sigma_i > 0, \sum_{i=1}^{3} w_i = 1$。这些参数将用于初始化平滑层，以自适应地筛选高频信息。对于输入特征的每个通道，平滑层都会设置核大小为 3、5 和 7 这 3 种高斯核，它们的方差则由 $\sigma = \{\sigma_i\}$ 决定。当 3 种高斯核得到 3 种不同的平滑结果后，根据 $w = \{w_i\}$ 加权得到最终的平滑结果。内容特征经过该平滑层后，其高频信息会被筛去，仅保留低频信息，记作 F_{smooth}；然后 DAHAB 对内容特征沿通道方向进行压缩，并通过卷积层得到特征图，公式为

$$\tilde{F} = \text{Conv}\left(\left[P_{\text{avg}}^{\text{C}}(F), P_{\text{max}}^{\text{C}}(F)\right]\right)。$$

其中，Conv 是核大小为 3×3 的卷积层；$[\cdot,\cdot]$ 表示特征拼接操作；$P_{\text{avg}}^{\text{C}}$ 和 $P_{\text{max}}^{\text{C}}$ 分别表示沿通道方向的平均池化和最大池化；$\tilde{F} \in \mathbb{R}^{H \times W \times 1}$ 表示不同空间位置需要关注的程度。最终，DAHAB 的输出结果可表示为

$$F_{\text{out}} = F + \tilde{F} \times \left(F - F_{\text{smooth}}\right)。$$

其中，$F - F_{\text{smooth}}$ 表示高频信息。通过与 $\tilde{F}$ 相乘再与 F 相加，DAHAB 自适应地对不同空间位置进行不同程度的高频信息调整。由于 $\tilde{F}$ 求取过程中没有使用 ReLU 激活函数，且保留了正负值的差异，使得 DAHAB 能区分细节与噪声，并且能够同时学习 M_{enhance} 和 M_{degrade} 这 2 种映射关系。

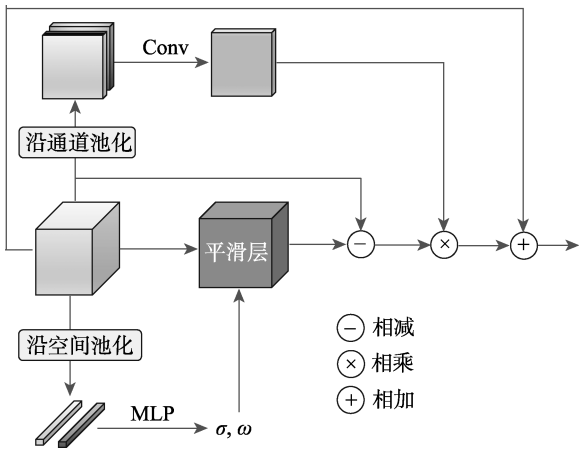


图 3 DAHAB 的结构

2.3　水下图像增强

如图 1 左部分所示，获得内容特征与风格特征后，模型将重新融合它们并重建对应的图像，共有 4 个分支：使用 C_{raw} 和 S_{raw} 自重建水下图像 $I_{\text{raw}}^{\text{self}}$，使用 C_{ref} 和 S_{ref} 自重建参考图像 $I_{\text{ref}}^{\text{self}}$，使用 $C_{\text{raw}}^{\text{en}}$ 和 S_{ref} 交叉重建参考图像 $I_{\text{ref}}^{\text{cross}}$，使用 $C_{\text{ref}}^{\text{deg}}$ 和 S_{raw} 交叉重建水下图像 $I_{\text{raw}}^{\text{cross}}$。它们的数学表达式分别为

$$I_{\text{raw}}^{\text{self}} = D\left(A\left(C_{\text{raw}}, S_{\text{raw}}\right)\right),$$

$$I_{\text{ref}}^{\text{self}} = D\left(A\left(C_{\text{ref}}, S_{\text{ref}}\right)\right),$$

$$I_{\text{ref}}^{\text{cross}} = D\left(A\left(C_{\text{raw}}^{\text{en}}, S_{\text{ref}}\right)\right),$$

$$I_{\text{raw}}^{\text{cross}} = D\left(A\left(C_{\text{ref}}^{\text{deg}}, S_{\text{raw}}\right)\right)。$$

其中，D 表示与内容编码器结构相对应的多尺度解码器；A 表示 Huang 等[22]提出的自适应实例归一化（adaptive instance normalization，AdaIN）。AdaIN 能够有效地融合风格特征与内容特征，其数学表达式为

$$\text{AdaIN}\left(x, a, b\right) = a\frac{x-\mu\left(x\right)}{\sigma\left(x\right)} + b。$$

其中，x 表示内容特征；$\mu\left(x\right)$ 和 $\sigma\left(x\right)$ 分别表示 x 的均值和方差；a 和 b 表示风格特征输入线性层后动态产生的拉伸和偏置系数。4 个分支共同使用同一个解码器 D。

通过对水下图像和参考图像的重建，模型能够提取图像的所有信息。在交叉重建 $I_{\text{ref}}^{\text{cross}}$ 的分支中，模型完成了从水下图像到参考图像的转换，即实现了水下图像的增强。需要注意的是，在实际推理过程中并没有成对的参考图像提供风格。因此，本文采用 2 种不同推理方式获取参考图像风格：（1）训练一个能够将水下图像风格域映射为参考图像风格域的风格映射网络；（2）预先收集优质水下图像的风格特征建立风格库，对于待增强的水下图像，根据其内容特征从风格库中匹配最佳的风格特征。由于文献[13,23]已对第 1 种推理方式有详细的介绍，此处不再赘述。本文重点讨论第 2 种推理方式，并将在实验部分验证其有效性。

建立风格库首先要收集优质的水下图像。本文采用水下图像质量测量指标（underwater image quality measure，UIQM）[24]和水下彩色图像质量评价综合指标（underwater color image quality evaluation，UCIQE）[25]对 UIEB 数据集的所有参考图像进行筛选。若参考图像的 UIQM 与 UCIQE 均大于整个数据集的平均值则保留该图像，反之则舍弃。经过上述操作后，共得到 242 幅参考图像；使用预训练的风格编码器提取它们的风格特征以构建风格库。由于直接使用包含 242 种风格特征的风格库会大幅降低内容特征匹配最佳风格特征的速度，因此需对其进行第 2 次筛选。

使用预训练的内容编码器提取 UIEB 数据集中原始水下图像的内容特征，将其与风格库中所有风格特征进行融合并重建，即对于一幅水下图像共生成 242 种增强结果；然后对每种增强结果计算 UIQM 和 UCIQE，并根据 1∶1 加权作为评分。评分最高的增强结果对应的风格特征被称为最佳风格。对 UIEB 数据集中每幅水下图像执行上述操作，并统计风格库中各个风格特征成为最佳风格的次数。经统计，在 242 种风格特征中有 77 种成为了最佳风格，其中，频率最高的 16 种风格特征占据 98%内容特征的最佳风格。本文采用这 16 种风格特征构建最终的风格库，完成第 2 次筛选。

在实际推理过程中，只需要直接读取预先编码的 16 种风格特征，将输入的水下图像内容特征复制 16 份后与之融合，批量生成 16 个增强结果，并计算它们的 UIQM 和 UCIQE 总分以保留一个最优结果。由于整个过程都是依靠批量计算进行的，因此只需要一次前向过程，与第 1 种推理方式相比，不需要加载 CLIP 图像编码器实时编码风格，进一步加快了推理速度，因此整个推理过程依然是高效的。

2.4　损失函数

（1）内容一致性损失

为了保证内容编码器提取的特征是去风格化的，应当使水下图像的内容特征和参考图像的内容特征尽量相似。分别计算各个尺度下水下图像内容特征和参考图像内容特征的 L_1 距离，公式为

$$L_{\text{content}} = \alpha\sum_{i=1}^{3}\left\|C_{\text{raw}}, C_{\text{ref}}\right\|。$$

其中，$\alpha = 0.1$ 表示权重参数。

（2）重建损失

为了使内容特征与风格特征融合后能重建对应的图像，对每个分支均使用 L_1 损失进行约束，公式为

$$L_{\text{rec}} = \sum_{i=1}^{3}\left\|I_{\text{raw}}^{\text{self}}, S\left(I_{\text{raw}}\right)\right\| + \sum_{i=1}^{3}\left\|I_{\text{ref}}^{\text{self}}, S\left(I_{\text{ref}}\right)\right\| + \sum_{i=1}^{3}\left\|I_{\text{ref}}^{\text{cross}}, S\left(I_{\text{raw}}\right)\right\| + \sum_{i=1}^{3}\left\|I_{\text{raw}}^{\text{cross}}, S\left(I_{\text{ref}}\right)\right\|。$$

其中，$S(\cdot)$ 表示缩放到对应尺度。

3 实验结果与分析

3.1 实验设置

本文选用 UIEB[1]、EUVP[11]和 RUIE[26]这 3 个数据集进行实验，这些数据涵盖了各种水下场景、水域类型和光照条件。其中，UIEB 数据集包含 890 对水下图像和参考图像，以及 60 幅无参考水下图像（称为 UIEB-challenge 子集）；EUVP 和 RUIE 数据集分别包含 330 幅和 300 幅无参考水下图像。将本文方法与 7 种主流水下增强方法进行比较，包括 2 种传统方法（GDCP[8]、Fusion[10]），5 种基于深度学习的方法（FGAN[12]、UIESS[13]、Semi-UIR[14]、SCNet[15]和 CLUIE[16]）。对 UIEB 数据集进行随机分割，取 800 对水下图像和参考图像用于训练/验证模型，记作 UIEB-train；剩余的 90 对图像使用第 1 种推理方式进行测试，记作 UIEB-test。在 UIEB-challenge，EUVP 和 RUIE 这 3 个无参考数据集上，使用第 2 种推理方式进行测试，验证本文方法的有效性。

本文方法基于 Pytorch 框架实现，并在 NVIDIA RTX 3090 GPU 上进行实验。采用 Adam 优化器，设初始学习率为 0.000 1，通过余弦退火算法[27]逐渐降低到 10^{-6}，设批量处理大小为 8。整个训练过程中，输入的图像都会被缩放为 256×256 像素大小，并采用 224×224 像素的随机裁剪和概率为 0.5 的水平/垂直翻转进行数据增强。

3.2 主观对比

从 UIEB-test、UIEB-challenge、EUVP 和 RUIE 数据集中，共选取 10 幅具有代表性退化的困难样本。将这些样本按照退化类型进行编号排序：（1）图①~图④为整体偏蓝的图像；（2）图⑤~图⑧为整体偏绿的图像；（3）图⑨~图⑩为低光照图像。此外，这些图像均呈现不同程度的模糊。其中，图①、图③和图⑥来自 UIEB-test，图④、图⑨和图⑩来自 UIEB-challenge，图②和图⑤来自 EUVP，图⑦和图⑧来自 RUIE。表 1 所示为在 4 个数据集上 8 种方法视觉效果对比。

从表 1 可以看出，存在偏色的水下图像中，GDCP[8]的增强结果多数过暗，如图①、图⑥~图⑧等，在部分增强结果上呈现过亮或过饱和，如图③和图⑤；Fusion[10]整体的增强结果均较为暗淡，视觉效果不佳；FGAN[12]无法有效地改善水下图像的偏色，甚至加重偏色，且部分结果产生了菱形伪影，如图②；CLUIE[16]略优于 FGAN[12]，能够改善部分整体偏绿的水下图像，但对于多数整体偏蓝的图像无法有效增强；Semi-UIR[14]和 SCNet[15]只能够改善偏色而不能彻底解决，如图②、图④和图⑤等，其增强结果依然偏蓝或绿，其中，SCNet[15]的增强结果偏色更严重且更模糊；UIESS[13]虽然能解决部分偏色问题，但对于偏色较为严重的图像，其增强结果偏暗偏红，如图④和图⑥；本文方法能够较好地校正这些存在偏色的困难样本，且所有增强结果的亮度和饱和度均在一个适当的区间。在图⑨和图⑩中，只有 GDCP[8]和本文方法能有效增强这类困难样本。实验结果表明，本文方法在面对困难样本时具有较大优势。

3.3 客观对比

采用峰值信噪比（peak signal-to-noise ratio，PSNR）和结构相似性指数（structural similarity index measure，SSIM）[28]作为评价指标，在 UIEB-test 数据集上对 8 种方法进行客观对比，结果如表 2 所示。在 UIEB-challenge、EUVP 和 RUIE 等无参考数据集上，选择 UIQM[24]、UCIQE[25]和 MUSIQ[29]指标对 8 种方法进行评估，结果如表 3 所示。其中，UIQM 包含水下图像色彩度量（underwater image colorfulness measure，UICM）、清晰度度量（underwater image sharpness measure，UISM）和对比度度量（underwater image contrast measure，UIConM）这 3 部分；UCIQE 则是色度、饱和度和对比度的线性组合：它们都是专门为水下图像设计的无参考指标。MUSIQ 是一种基于 Transformer 架构的自然图像评估模型。

从表 2 可以看出，本文方法的 PSNR 取得了次优的结果，比第 1 名 Semi-UIR[14]仅落后 0.13，比第 3 名 SCNet[15]则高出 1.58，本文方法和 Semi-UIR[14]的 PSNR 大大优于其他方法；本文方法的 SSIM 取得了最优结果，比第 2 名 Semi-UIR[14]领先 0.009，比第 3 名 SCNet[15]领先 0.039。需要注意的是，尽管本文方法的 PSNR 略逊于 Semi-UIR[14]，但正如文献[30-31]所述，与 PSNR 相比，SSIM 作为衡量标准更准确客观，因为它基于亮度、对比度和结构作为人类感知模拟，而不像 PSNR 那样以非归一化绝对散度作为度量。

从表 3 可以看出，本文方法在 9 个结果中取得了 6 个最优和 3 个次优的优良成绩；尽管在部分指标结果上低于 UIESS[13]和 GDCP[8]，但综合所有结果看，本文方法依然具备优势；此外，图④~图⑧均表明，本文方法的增强结果在视觉上优于 UIESS[13]和

表 1　4 个数据集上 8 种方法视觉效果对比

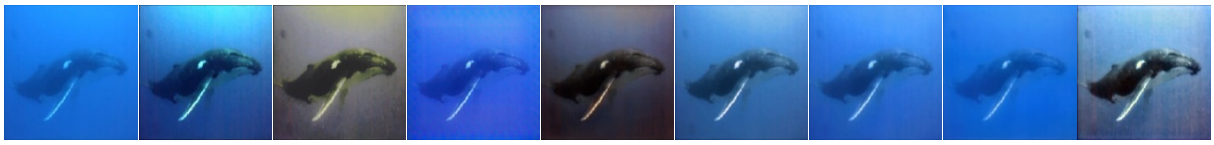
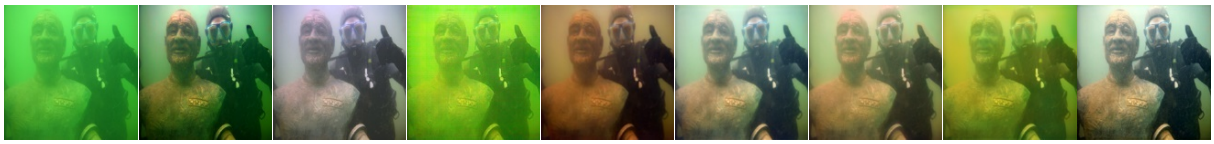
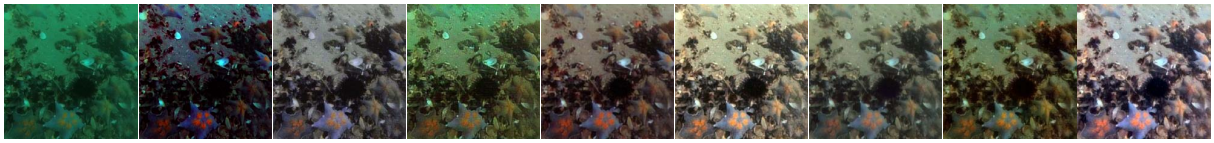
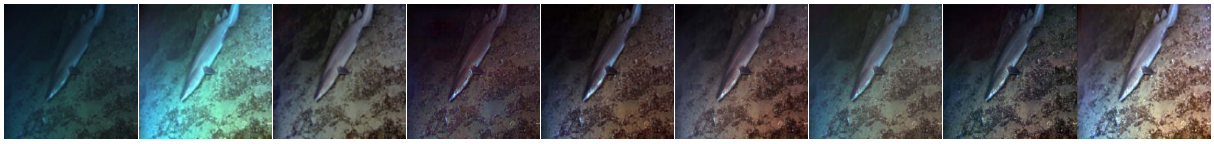
图编号	水下图像	GDCP[8]	Fusion[10]	FGAN[12]	UIESS[13]	Semi-UIR[14]	SCNet[15]	CLUIE[16]	本文
①									
②									
③									
④									
⑤									
⑥									
⑦									
⑧									
⑨									
⑩									

表 2　UIEB-test 数据集上 8 种方法的 PSNR 和 SSIM 对比

方法	PSNR/dB	SSIM	方法	PSNR/dB	SSIM
GDCP[8]	13.60	0.724	Semi-UIR[14]	**24.31**	<u>0.915</u>
Fusion[10]	18.32	0.852	SCNet[15]	22.60	0.885
FGAN[12]	18.08	0.761	CLUIE[16]	21.02	0.883
UIESS[13]	19.02	0.809	本文	<u>24.18</u>	**0.924**

注：粗体表示最优值，下划线表示次优值。

表 3　3 个数据集上 8 种方法的 UIQM、UCIQE 和 MUSIQ 对比

方法	UIEB-challenge			EUVP			RUIE		
	UIQM	UCIQE	MUSIQ	UIQM	UCIQE	MUSIQ	UIQM	UCIQE	MUSIQ
输入	2.14	0.357	38.53	2.62	0.421	35.17	2.96	0.320	38.20
GDCP[8]	2.09	0.450	39.97	2.19	**0.486**	36.43	2.67	0.431	**40.32**
Fusion[10]	2.55	0.379	40.09	3.20	0.441	33.45	3.34	0.379	36.56
FGAN[12]	3.02	0.409	37.72	3.02	0.445	37.13	3.21	0.362	36.76
UIESS[13]	2.65	**0.461**	33.42	3.04	0.458	33.38	3.17	0.419	31.79
Semi-UIR[14]	2.85	0.430	39.35	2.95	0.441	35.80	3.28	0.399	34.31
SCNet[15]	2.74	0.379	37.81	3.14	0.444	34.60	3.31	0.355	33.79
CLUIE[16]	2.71	0.412	38.55	2.93	0.454	34.33	3.10	0.370	35.01
本文	**3.16**	0.450	**41.90**	**3.24**	0.462	**38.21**	**3.37**	**0.450**	38.57

注：粗体表示最优值，下划线表示次优值。

GDCP[8]。实验结果表明，本文方法增强后的水下图像整体效果更自然、视觉效果更佳，且面对各类复杂水下退化时展现出了强大的泛化性能。

3.4　消融实验

通过消融实验验证本文方法各个组件的有效性，方案如下。（1）方案 1：移除风格编码器中 CLIP 图像编码器，使用卷积层和自适应平均池化替代 CLIP 进行特征提取和浓缩；（2）方案 2：移除 DAHAB 中的 M_{degrade} 映射，由于该方案 DAHAB 不生成 $C_{\text{ref}}^{\text{deg}}$，因此不交叉重建 $I_{\text{raw}}^{\text{cross}}$；（3）方案 3：移除内容一致性损失 L_{content}；（4）方案 4：移除整个 DAHAB，直接基于提取到的 C_{raw} 交叉重建 $I_{\text{ref}}^{\text{cross}}$。实验结果如表 4 所示。可以看出，本文方法中每个组件都是必要的，去除任意组件均会导致性能下降。

表 4　消融实验结果对比

方案	CLIP 编码器	M_{degrade}	L_{content}	DAHAB	CBAM[32]	UIEB-test	
						PSNR/dB	SSIM
1		✓	✓	✓		22.69	0.900
2	✓		✓	✓		23.45	0.896
3	✓	✓		✓		23.74	0.894
4	✓	✓	✓			23.26	0.889
5	✓	✓	✓		✓	23.70	0.910
本文	✓	✓	✓	✓		**24.18**	**0.924**

注：粗体表示最优值。

（1）各组件作用分析。CLIP 图像编码器具备强大的零样本学习能力，能够帮助模型适应各类复杂的水下风格，将其替换为简单的卷积层和池化层会导致风格编码器的性能大幅下降。DAHAB 中，M_{degrade} 看似不直接参与水下图像增强，然而其为 DAHAB 提供了另一个视角，以关注内容特征中关键的高频信息，去除该部件会导致 DAHAB 的性能下降。内容一致性损失 L_{content} 是内容编码器能够正确地提取去风格化的内容特征的重要保证，其利用水下图像和参考图像内容相似而风格不同的特性，对内容编码器提取的成对内容特征进行监督，去除该损失则会导致内容编码器无法正确地区分内容与风格。DAHAB 通过调整内容特征的高频信息，解决水下退化导致的细节模糊、缺失等问题，移除 DAHAB 则会影响增强结果的细节与清晰度。

（2）DAHAB 的优越性。为了证明 DAHAB 的优越性，增加方案 5，即使用广泛应用的卷积块注意力模块（convolutional block attention module，CBAM）[32]替换 DAHAB。通过指标对比可知，DAHAB 优于 CBAM。CBAM 虽然考虑了通道注意力与空间注意力，却未区分低频与高频信息；DAHAB 则专注于代表细节的高频信息，因此能取得更好的结果。

5 种消融方案的视觉效果对比如图 4 所示。可以看出，相较于对比方法的增强结果，本文方法的增强结果的细节更加清晰，颜色更加自然。



a. 水下图像　b. 方案 1　c. 方案 2　d. 方案 3　e. 方案 4　f. 方案 5　g. 本文

图 4　消融实验视觉效果对比

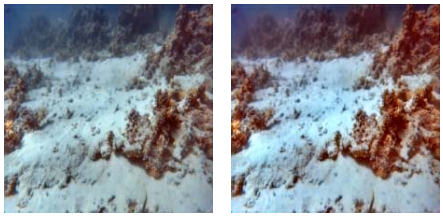
（3）内容一致性损失的权重。本文对不同权重下内容一致性损失对模型性能的影响进行实验，结果如表 5 所示。可以看出，选择一个适当的α是必要的，如果α过大，会导致内容编码器忽略 2 种内容特征中的高频差异；否则会导致内容编码器不能有效地去除风格。

表 5　在 UIEB-test 数据集上不同α的指标对比

α	PSNR/dB	SSIM
1.00	<u>24.01</u>	<u>0.915</u>
0.10	**24.18**	**0.924**
0.01	23.89	0.902

注：粗体表示最优值，下划线表示次优值。

（4）UIQM 与 UCIQE 的权重。在第 2 种推理方式中，以不同权重加权 UIQM 和 UCIQE 计算评估总分对增强结果的影响。直觉上来说，UIQM 与 UCIQE 处于不同的量级，在计算评估总分时理应给予 UCIQE 更大的权重。然而，通过实验发现，这会导致增强结果出现偏色、过饱和等问题，如图 5 所示。其中，图 5a 按 1∶1 加权，图 5b 按 1∶5 加权。可以看出，图 5b 的增强结果偏向红色；相对而言，图 5a 的增强结果更为自然，视觉效果更佳。经过多次实验比较，本文最终按照 1∶1 加权 UIQM 和 UCIQE。



a. 增强结果 1　　b. 增强结果 2

图 5　不同 UIQM 和 UCIQE 权重下增强结果对比

4　结　语

本文提出一种基于内容与风格双域改善的水下图像增强方法。与现有的方法不同，本文首先建立了一个风格与内容解耦重建框架，并在解耦与重建的过程中分别处理内容域与风格域；对于内容域，提出了 DAHAB 对内容特征的高频信息进行调整；对于风格域，通过建立风格库并使用内容特征匹配最优风格特征的方式进行增强。在 UIEB、EUVP 和 RUIE 这 3 个数据集上，与 7 种方法进行对比实验的结果表明，本文方法取得了更好的指标与视觉效果。

尽管在匹配最佳风格时使用并行的方式保证效率，但不可避免地增加了所需的显存，导致本文方法在处理分辨率较大的图像时可能需要进一步精简风格库中风格特征的数量，因此性能会略微降低。如何更加高效地匹配内容特征与风格特征，将是未来工作的重点。

参考文献（References）：

[1] 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

[2] Li C Y, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement[J]. Pattern Recognition, 2020, 98: Article No.107038

[3] 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

[4] 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

[5] Hitam M S, Awalludin E A, Yussof W N J H W, *et al*. Mixture contrast limited adaptive histogram equalization for underwater image enhancement[C] //Proceedings of the International Conference on Computer Applications Technology. Los Alamitos: IEEE Computer Society Press, 2013: 1-5

[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] Land E H. The Retinex theory of color vision[J]. Scientific American, 1977, 237(6): 108-129

[8] Peng Y T, Cao K M, Cosman P C. Generalization of the dark channel prior for single image restoration[J]. IEEE Transactions on Image Processing, 2018, 27(6): 2856-2868

[9] Zhang S, Wang T, Dong J Y, *et al*. Underwater image enhancement via extended multi-scale Retinex[J]. Neurocomputing, 2017, 245: 1-9

[10] 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

[11] Zhang W D, Zhuang P X, Sun H H, *et al*. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement[J]. IEEE Transactions on Image Processing, 2022, 31: 3997-4010

[12] 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

[13] Chen Y W, Pei S C. Domain adaptation for underwater image enhancement via content and style separation[J]. IEEE Access, 2022, 10: 90523-90534

[14] Huang S R, Wang K Y, Liu H, *et al*. Contrastive semi-supervised learning for underwater image restoration via reliable bank[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE

Computer Society Press, 2023: 18145-18155

[15] Fu Z Q, Lin X P, Wang W, *et al*. Underwater image enhancement via learning water type desensitized representations[C] //Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Los Alamitos: IEEE Computer Society Press, 2022: 2764-2768

[16] Li K Q, Wu L, Qi Q, *et al*. Beyond single reference for training: underwater image enhancement via comparative learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(6): 2561-2576

[17] Isola P, Zhu J Y, Zhou T H, *et al*. Image-to-image translation with conditional adversarial networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5967-5976

[18] Yang S, Zhou Y F, Liu Z W, *et al*. Rerender a video: zero-shot text-guided video-to-video translation[C] //Proceedings of the SIGGRAPH Asia Conference Papers. New York: ACM Press, 2023: 1-11

[19] Zheng C X, Liu B Z, Xu X M, *et al*. Learning an interpretable stylized subspace for 3D-aware animatable artforms[OL]. [2024-02-28]. https://ieeexplore.ieee.org/document/10430412

[20] Radford A, Kim J W, Hallacy C, *et al*. Learning transferable visual models from natural language supervision[C] //Proceedings of the 38th International Conference on Machine Learning. New York: ICML, 2021: 8748-8763

[21] Cui Y N, Ren W Q, Cao X C, *et al*. Focal network for image restoration[C] //Proceedings of the IEEE/CVF Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 12955-12965

[22] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1501-1519

[23] Jiang Q, Zhang Y F, Bao F X, *et al*. Two-step domain adaptation for underwater image enhancement[J]. Pattern Recognition, 2022, 122: Article No.108324

[24] 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

[25] Yang M, Sowmya A. An underwater color image quality evaluation metric[J]. IEEE Transactions on Image Processing, 2015, 24(12): 6062-6071

[26] Liu R S, Fan X, Zhu M, *et al*. Real-world underwater enhancement: challenges, benchmarks, and solutions under natural light[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(12): 4861-4875

[27] Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts[C] //Proceedings of the 5th International Conference on Learning Representations. Washington DC: ICLR, 2017: 1-16

[28] 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

[29] Ke J J, Wang Q F, Wang Y L, *et al*. MUSIQ: multi-scale image quality transformer[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 5128-5137

[30] Setiadi D R I M. PSNR vs SSIM: imperceptibility quality assessment for image steganography[J]. Multimedia Tools and Applications, 2021, 80(6): 8423-8444

[31] Sara U, Akter M, Uddin M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. Journal of Computer and Communications, 2019, 7(3): 8-18

[32] Woo S, Park J, Lee J Y, *et al*. CBAM: convolutional block attention module[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 3-19