

基于内容与风格双域改善的水下图像增强方法

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 浙江杭州 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, 然而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, 尤其对于困难样本很难做出有效的增强. 针对以上问题, 本文提出了一种基于内容与风格双域改善的水下图像增强方法. 首先, 设计了一个简单有效的框架将水下图像解耦为内容特征与风格编码并重建. 对于内容域, 设计了双域自适应高频调整模块增强原始水下图像内容特征中的细节并抑制噪声; 对于风格域, 提出了建立风格库并在推理过程中匹配最优风格来实现风格的变换. 最后, 通过构建内容一致性损失和重建损失以监督模型完成水下图像的增强与重建. 实验结果表明, 在水下图像增强领域公开的 UIEB 数据集、EUVP 数据集和 RUIE 数据集上, 本文方法的主观视觉效果和客观评价指标(PSNR、SSIM、UIQM 和 UCIQE)均优于已有方法, 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.1089.202*.论文编号

Underwater Image Enhancement Method with Content and Style Dual-domain Improvement

Liu Chunxiao*, Cao Yue, Wang Chenghua, 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. However, existing underwater image enhancement methods struggle to deal with the complex realities of underwater environments, especially in effectively enhancing challenging samples. To address these issues, we propose a novel method that leverages dual-domain improvements in content and style. Initially, a straightforward and effective framework is designed to decouple underwater images into content features and style codes, followed by reconstruction. For the content domain, a dual-domain adaptive high-frequency adjustment block is designed to enhance detail and suppress noise. For the style domain, we establish a style library that enables style transformation by matching the optimal style during inference. Finally, the model completes the enhancement and reconstruction of underwater images by utilizing a content consistency loss

收稿日期: 2024-02-**; 修回日期: 20**-**-**. 基金项目: 国家自然科学基金(61976188); 浙江省自然科学基金(LY24F020004); 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项目(2023R408035, 2023R408072). 刘春晓(1979—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 论文通信作者, 主要研究方向为计算机视觉与计算机图形学、机器学习与智能系统、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与深度学习; 王成骅(2002—), 男, 本科生, 主要研究方向为图像处理与深度学习; 周子翔(2002—), 男, 本科生, 主要研究方向为图像处理与深度学习.

and a reconstruction loss for supervision. Experimental results on publicly available underwater image datasets such as UIEB, EUVP, and RUIE show that our method outperforms existing methods in both subjective visual effects and objective evaluation metrics (PSNR, SSIM, UIQM, and UCIQE). Especially, 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愈发频繁,这些水下作业均需要稳定清晰的水下画面图像作为辅助分析的手段才得以顺利开展.然而,由于水下环境的特殊性,要想获取高质量的水下图像通常面临着多重挑战,如水体对不同色光产生不同程度的吸收和散射导致颜色失真、水体浑浊导致对比度降低和视觉信息的损失等.因此,研发能够克服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两类:基于图像处理的传统方法和基于深度学习的现代方法,其中前者又可以分为基于模型的方法和模型无关的方法.基于模型的方法借助人先验知识来估计水下成像模型^[1]中的未知参数;模型无关的方法则依赖于设计适当的图像增强技术.但是,它们都需要调整复杂的参数以适应不同的水下环境,且增强效果有限.相对来说,基于深度学习的方法往往具备更好的性能,然而它们也面临着限制与挑战.其中最显著的是,大部分基于深度学习的技术依赖于成对的训练数据.换言之,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)^[2]通过同时使用多个已有水下图像增强方法并人工挑选最优结果的方式构建而成.尽管 UIEB 数据集中包含丰富的水下图像,但其较小的数据量(仅 890 对原始图像和参考图像)以及部分质量不佳的参考图像限制了已有深度学习方法的性能表现.尽管部分研究如水下场景启发的方法^[3]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收效甚微.

为了解决上述问题,本文提出了内容与风格双域改善的水下图像增强方法,主要创新如下:

(1) 提出了基于内容与风格解耦重建的水下图像增强新框架.对解耦得到的内容特征和风格编码分别进行改善,最终通过重建得到增强结果;

(2) 提出了双域自适应高频调整模块,通过自适应的提取和调整高频信息,同时学习水下图像与参考图像之间内容特征的互相映射;

(3) 提出了新的推理方式.通过统计最优匹配次数建立风格库.在推理过程中,内容特征分别与风格库中的所有风格融合并从中挑选最优结果.

实验证明,本文方法在 UIEB、EUVP 和 RUIE 等多个数据集上均有良好的增强结果,表现出了优异的泛化性能,并且在困难样本增强方面远远优于其他方法.

1 相关工作

1.1 基于图像处理的传统方法

传统的水下图像增强方法可以分为基于模型的方法和模型无关的方法. Berman 等^[4]提出了一种用于水下图像颜色恢复的方法,该方法主要利用霾线(haze-lines)技术来校正水下图像的色偏. Galdran 等^[5]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摄影中常见的丢失. Peng 等^[6]扩展了暗通道先验(Dark Channel Prior, DCP)^[7]方法,使其不再局限于去雾.它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. Hitam 等^[8]结合了

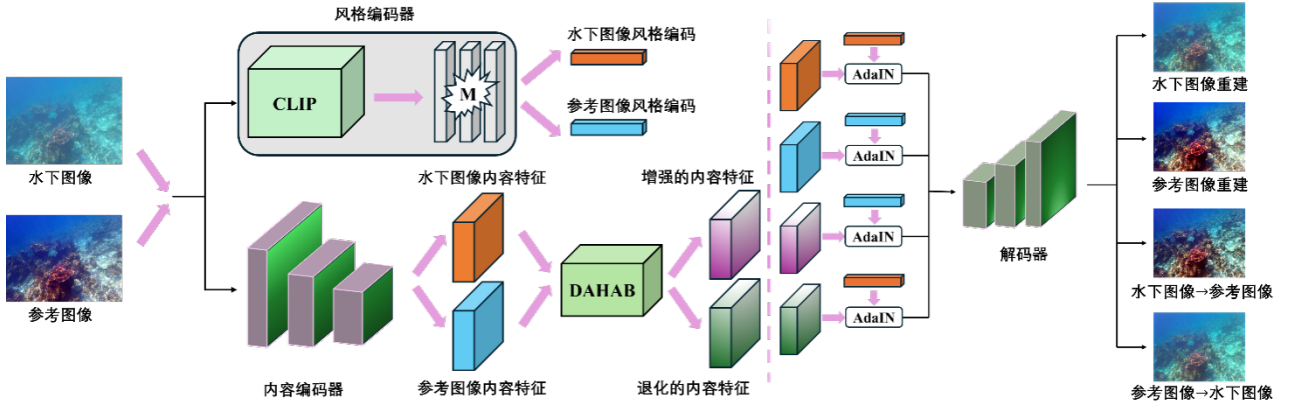


图1 本文方法的整体框架

对比度受限的自适应直方图均衡化 (contrast limited adaptive histogram equalization, CLAHE) 技术和混合模式处理. Zhang 等^[9]对传统的 Retinex 理论^[10]的进行了扩展, 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. 此外, 还有 Fusion^[11]和 MMLE^[12]等方法也取得了一定的成功.

1.2 基于深度学习的现代方法

近年来, 越来越多的研究者尝试将深度学习应用于水下图像增强, 如 MLFcGAN^[13] 和 DehazeNet^[14]等. Li 等^[2]构建了一个具有 950 幅水下图像的增强基准数据集, 并在此基础上提出了 Water-Net 模型. Islam 等^[15]提出了基于条件生成对抗网络的实时水下图像增强模型, 称为 FGAN, 该方法通过全局内容、颜色、局部纹理和风格信息来评估感知图像质量, 并以此建立目标函数监督对抗训练. Chen 等^[16]提出了一种内容和风格分离从而进行水下图像增强的方法, 称为 UIESS. 该方法将风格分为真实、合成和干净三个域, 并尝试将风格从真实域映射到干净域, 以此实现水下图像增强. Huang 等^[17]提出了一种基于均值教师的半监督水下图像恢复框架 (Semi-supervised Underwater Image Restoration, Semi-UIR), 该方法通过半监督框架缓解了水下图像成对数据短缺的问题, 并建立可靠库 (Reliable Bank) 进一步确保方法的可靠性. Fu 等^[18]提出了一种通过关注水类型不敏感表示来增强水下图像的方法, 称为 SCNet, 该方法在 UNet 架构的每个尺度上进行空间和通道的归一化,

以减少不同水类型对增强过程的影响. Li 等^[19]提出了第一个基于比较学习框架的水下图像增强方法 CLUIE, 它在增强过程中对比使用多个增强参考, 并提出了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.

2 本文方法

本文整体框架如图 1 所示, 大致可分为左右两个部分. 左部实现了水下图像风格与内容的解耦, 右部则是将不同的内容和风格重新融合以实现图像重建, 并在此过程中对水下图像进行增强.

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, 即:

$$C_{raw}, C_{ref} = E^C (raw, ref) \quad (1)$$

$$S_{raw}, S_{ref} = E^S (raw, ref) \quad (2)$$

其中, $\{raw, ref\}$ 分别表示水下图像和对应的参考图像, $\{E^C, E^S\}$ 分别表示内容编码器和风格编码器, $\{C_{raw}, C_{ref}\}$ 与 $\{S_{raw}, S_{ref}\}$ 分别表示内容特征和风格编码. 内容特征包含了该图像所有的空间结构信息, 而风格编码则与结构信息无关, 它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.

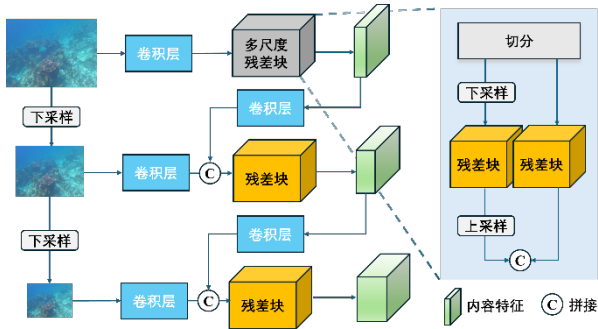


图 2 内容编码器

风格编码器由 CLIP 图像编码器^[20]与风格网络 M 组成, 其中风格网络 M 是一系列线性层和 ReLU 激活函数. 在整个训练过程中, CLIP 图像编码器参数固定. CLIP 图像编码器具有极佳的零样本学习 (Zero-shot) 能力, 在面对各类未接触过的复杂水下图像时, 依然能保持强大的特征提取与编码能力, 能够有效提取和区分不同图像的深层特征. 这不但有助于风格网络 M 获取图像对应的风格编码, 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. 将输入图像进行两次下采样, 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 (Muti-ResBlock)^[21]中. 多尺度残差块具体结构如图 2 右部所示. 首先将输入的特征在通道方向上切分为两部分, 前 16 个通道的特征分别经过下采样、残差块、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 (Concat) 得到第一个尺度的内容特征. 由于多尺度残差块具备多个尺度的视野, 因此能够更好的提取内容特征, 并且还能一定程度上减少网络的参数. 对于另外两个尺度, 首先将对应分辨率的图像输入卷积层使其通道数扩展到 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. 综上, 内容编码器将得到三个尺度的内容特征, 在提取到足够高级的特征同时, 保证细节不会丢失.

2.2 双域自适应高频调整模块

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. 然而, 水下图像到参考图像的增强过程, 使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, 本文提出双域自适应高频调整模块 (Dual-domain Adaptive High-frequency Adjust Block, DAHAB) 对它们的内容特征进行调整.

DAHAB 将学习两种内容特征的映射. 首先, DAHAB 需要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 $F_{enhance} : C_{raw} \rightarrow C_{ref}$. 其次, DAHAB 还需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 $F_{degrade} : C_{ref} \rightarrow C_{raw}$. 对于两种不同的映射分别设置一个 DAHAB. 记调整后的内容特征分别为 C_{raw}^{en} 和 C_{ref}^{deg} .

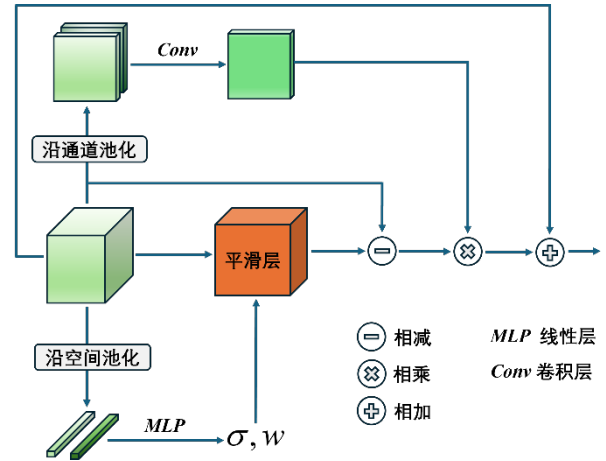


图 3 双域自适应高频调整模块

DAHAB 具体结构如图 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, 其数学表示如下:

$$\sigma, w = MLP(AvgPool_s(F) + MaxPool_s(F)) \quad (3)$$

其中, MLP 是线性层, $AvgPool_s$ 和 $MaxPool_s$ 分别表示沿空间方向的平均池化和最大池化. F 是输入的内容特征. 预测的参数均为三元组, 即 $\sigma = \{\sigma_i\}, w = \{w_i\}, i = 1, 2, 3$, 且 $\sigma_i > 0, \sum w_i = 1$. 这些参数将用于初始化平滑层, 以自适应地筛选高频信息. 平滑层对于输入特征的每一个通道, 都会设置核大小为 3、5、7 三种高斯核, 它们的方差则由 $\sigma = \{\sigma_i\}$ 决定. 当三种高斯核得到三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} . 随后, 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, 即:

$$\tilde{F} = \text{Conv}([\text{AvgPool}_C(F), \text{MaxPool}_C(F)]) \quad (4)$$

其中, Conv 是核大小为 3×3 的卷积层, $[\cdot, \cdot]$ 代表特征拼接操作, AvgPool_C 和 MaxPool_C 分别表示沿通道方向的平均池化和最大池化. $\tilde{F} \in R^{H \times W \times 1}$ 代表了不同空间位置需要关注的程度. DAHAB 最终的输出结果可表示为:

$$F_{out} = F + \tilde{F} \times (F - F_{smooth}) \quad (5)$$

其中, $F - F_{smooth}$ 代表了高频信息, 通过与 \tilde{F} 相乘再与 F 相加, DAHAB 自适应的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, \tilde{F} 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 $F_{enhance}$ 与 $F_{enhance} : C_{raw} \rightarrow C_{ref}$ 两种映射关系.

2.3 水下图像增强

如图 1 左部所示, 在获得内容特征以及风格编码后, 模型将重新融合它们并重建对应的图像. 具体来说共有四个分支, 即:

$$\overline{raw} = D(A(C_{raw}, S_{raw})) \quad (6)$$

$$\overline{ref} = D(A(C_{ref}, S_{ref})) \quad (7)$$

$$\widetilde{ref} = D(A(C_{raw}^{en}, S_{ref})) \quad (8)$$

$$\widetilde{raw} = D(A(C_{ref}^{deg}, S_{raw})) \quad (9)$$

其中, D 是与内容编码器结构相对应的多尺度解码器, A 是 Xun. 等人提出的自适应实例归一化 (Adaptive Instance Normalization, AdaIN)^[22]. AdaIN 能够有效的融合风格与内容, 其数学表示如下:

$$\text{AdaIN}(x, a, b) = a \frac{x - \mu(x)}{\sigma(x)} + b \quad (10)$$

其中, $x, \mu(x), \sigma(x)$ 分别表示内容特征及其均值与方差, a 和 b 是风格编码输入线性层后动态产生的拉伸和偏置系数. 四个分支共同训练并使用同

一个解码器, 它们每一个都是必要的, 去除任意一个都会使模型失衡从而导致性能下降.

在公式(8)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. 然而需要注意的是, 在实际推理过程中, 并没有成对的参考图像提供风格. 本文于此提出两种不同的参考风格获取方式: (1)训练一个风格映射网络, 将水下图像风格域映射到参考图像风格域; (2)在多个视觉效果良好的非成对水下图像中匹配合适的参考风格. 由于研究^{[16][23]}已对推理方式 1 有详细的介绍, 故不在此赘述. 本文重点讨论推理方式 2, 并将在实验部分验证其有效性.

具备良好视觉效果的水下图像往往具备一定的共性, 例如较好的清晰度、饱和度以及光照条件. 那么, 能否利用这些图像的风格作为参考风格实现水下图像增强? 为了验证该猜想, 本文在 UIEB^[2]基准数据集上进行了实验. UIEB 共有 890 幅参考图像, 对每一幅参考图像计算两个常用的水下图像无参考指标: 水下图像质量测量 (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] (值越大图像质量越好), 筛选出两个指标均大于均值的参考图像, 以此初步获取具备良好视觉效果的水下图像, 共计 242 幅. 随后, 对于每一幅水下图像, 使用这 242 幅参考图像的风格生成不同的增强结果, 然后加权它们的 UIQM 与 UCIQE 计算评估总分, 并根据评估总分对所有结果排序, 确定每一幅水下图像匹配到的最佳参考图像.

经统计, 242 幅参考图像中, 共有 77 幅被匹配为最佳. 然而, 匹配次数最多的前 16 幅参考图像, 已经占据了 98% 最佳匹配结果, 当数量扩大到 32 幅时, 其占比已经达到了 99.6%. 以上数据说明, 选取 K 幅参考图像作为风格库用于增强非成对的水下图像是可行的. 在推理过程中, 直接读取预先编码的 K 幅参考图像风格, 将输入的水下图像内容特征复制 K 份后与之融合, 批量生成 K 个增强结果, 并计算它们的评估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. 并且相比于推理方式 1, 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

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 L1 距离, 即:

$$L_{con} = \alpha \sum_{i=1}^3 \|C_{raw}, C_{ref}\| \quad (11)$$

其中 α 是权重参数. 一个适当的 α 是必要的, 过大的 α 会导致内容编码器忽略两种内容特征中的高频差异, 过小的 α 则会导致内容编码器不能有效的去除风格.

重建损失: 为了使内容特征融合风格编码后能重建回对应的图像, 本文对每个分支均使用 L1 损失进行约束, 如下:

$$L_{rec} = \sum_{i=1}^3 \|\overline{raw}, Sc(raw)\| + \sum_{i=1}^3 \|\overline{ref}, Sc(ref)\| + \sum_{i=1}^3 \|\widetilde{raw}, Sc(raw)\| + \sum_{i=1}^3 \|\widetilde{ref}, Sc(ref)\| \quad (12)$$

其中, $Sc(\cdot)$ 表示缩放(Scale)到对应尺度.

3 实验结果与分析

3.1 实验设置

本节将分别在 UIEB^[2]、EUVP^[12]和 RUIE^[26]三个数据集上进行实验, 并与七种最先进的水下增强方法进行比较, 包括两种传统方法(GDCP^[6]、fusion^[11])和五种基于深度学习的方法(FGAN^[15]、UIESS^[16]、Semi-UIR^[17]、SCNet^[18]和 CLUIE^[19]). 将 UIEB 数据集中 890 对图像进行随机分割, 取 800 对用训练/验证模型, 记作 UIEB-train. 剩余的 90 对图像使用推理方式 1 进行测试, 记作 UIEB-test. 此外, UIEB 数据集中包含 UIEB-challenge 子集, 该

子集中含有 60 张无参考的复杂水下退化图像. 本节将在 UIEB-challenge、EUVP、RUIE 三个无参考数据集上使用推理方式 2 进行测试, 以证明该方法的有效性.

本文方法实现基于 Pytorch 框架, 并在 NVIDIA RTX 3090 GPU 上进行. 采用 Adam 优化器, 学习率为 0.0001, 通过余弦退火^[27]逐渐降低到 $1e-6$, 批量处理大小为 8. 在整个训练过程中, 输入的图像都会被缩放为 256×256 , 并应用 224×224 的随机裁剪和概率为 0.5 的水平/垂直翻转作为数据增强.

在推理方式 2 中, 风格库大小 K 设为 16, 并根据 1:1 加权 UIQM 与 UCIQE 计算评估总分. 在内容一致性损失中, α 设为 0.1.

3.2 主观对比

本节从 UIEB-test、UIEB-challenge、EUVP 和 RUIE 中分别选取了几幅具备常见退化的水下图像, 如图 4、5、6、7 中 a 列所示, 分别存在有雾、偏色、光照不佳等问题.

对比发现, GDCP 并不能有效的对水下图像进行增强, 它的增强结果中同时出现了亮度较暗或亮度饱和度过高两种极端情况, 如图 4a 第一行第三行所示. Fusion 能够一定程度上增强图像细节, 然而它的增强结果整体较为暗淡, 视觉效果不佳. FGAN 过于强调算法的速度, 导致无法应对各类严重退化, 尤其是输入图像具有较强的色偏时, 完全无法对其进行颜色校正, 如图 4e 第二行, 图 5d 第一行第二行所示. 此外, FGAN 在部分图像中还会产生明显的菱形伪影, 如图 4e 第一行、图 6d 第二行所示. UIESS 同样尝试将水下图像解耦为内容域和风格域并进行增强, 然而它并没有在内容域做出任何改善, 仅尝试在风格域做单一的增强映射. 这导致它不能很好的保持细节, 且泛化性能有限, 仅在部分图像上实现了有效的增强. Semi-UIR 在多数图像上取得了较好的增强结果, 但光照不佳的水下图像对它而言仍是个挑战, 如图 5f 第三行所示. 此外, 相比本文方法, 它不能较好的去雾以恢复细节, 如图 7 第一行所示, 本文方法产生的

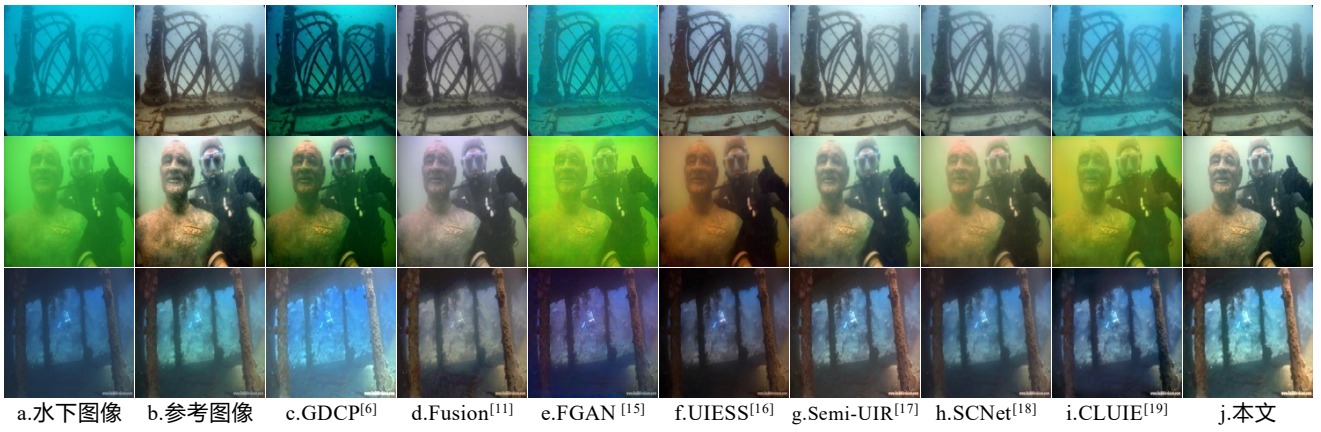


图 4 UIEB-test 数据集上 8 种方法视觉效果对比

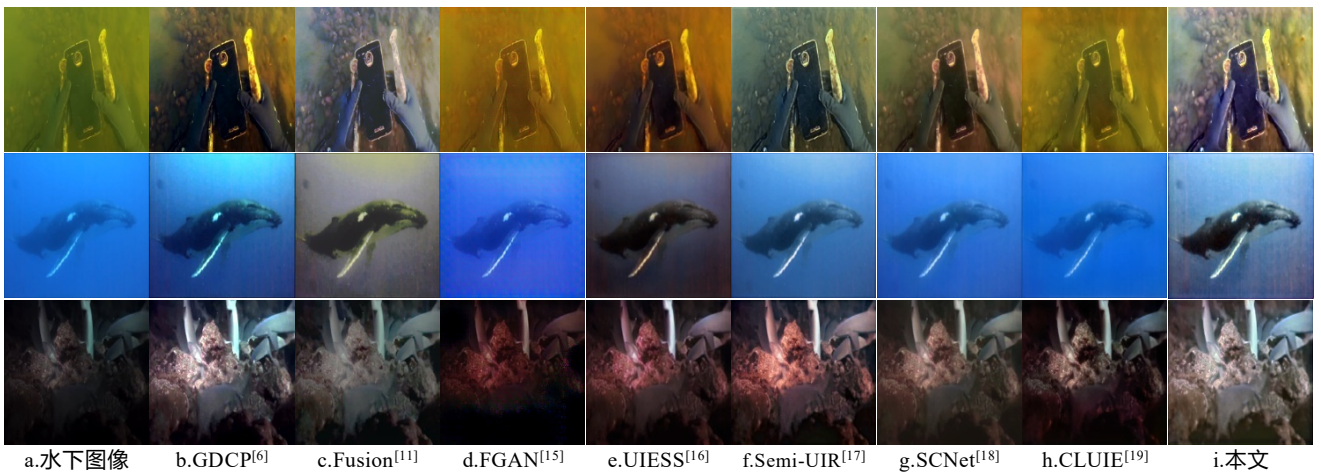


图 5 UIEB-challenge 数据集上 8 中方法视觉效果对比

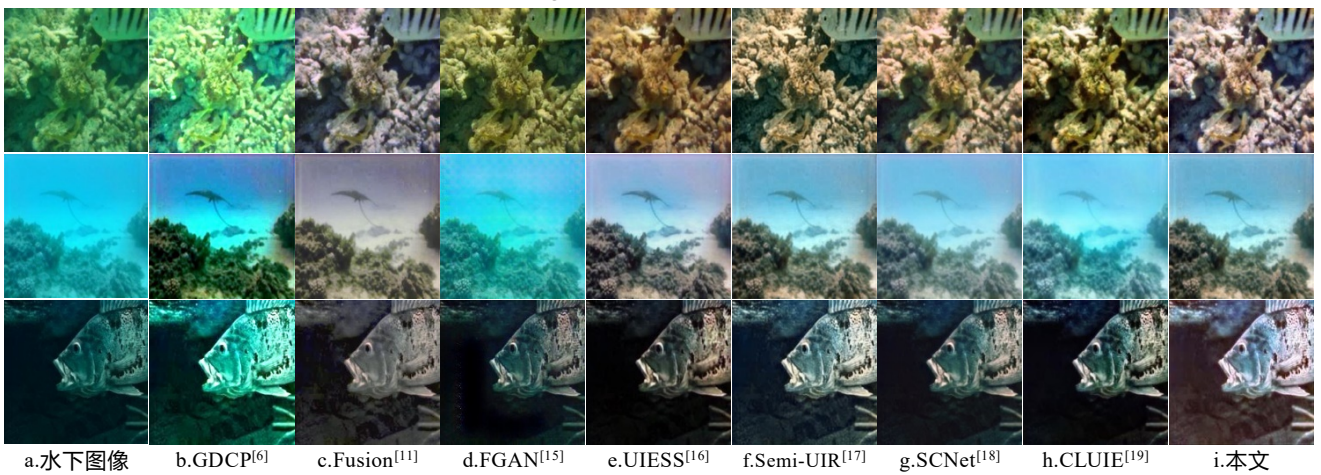


图 6 EUVP 数据集上 8 种方法视觉效果对比

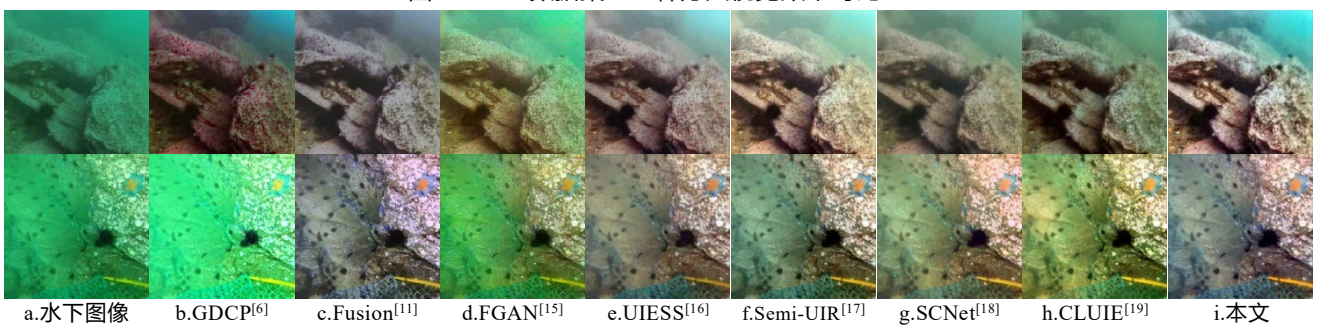


图 7 RUIE 数据集上 8 种方法视觉效果对比

结果其石头纹理更加清晰立体,且上方背景两块石头的边界也更加明显. SCNet 能够在一定程度上校正图像的色偏,但当色偏过于严重时依然会有残留,且在去雾方面仍待改进. CLUIE 在面对部分图像时有较好的增强结果,但同样无法克服严重退化,如图 4i 第二行、图 5h 第二行所示,无法达到增强的效果. 相比之下,本文方法在面对各类水下图像时,均表现出优秀的增强效果,尤其在一些困难样本上给人留下了深刻的印象.

3.3 客观对比

在客观对比部分,本文选择了峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 和结构相似性指数 (Structural Similarity Index Measure, SSIM)^[28] 作为指标以评价各个方法在 UIEB-test 上的结果. 对于 UIEB-challenge、EUVP 和 RUIE 等无参考数据集,本文选择了 UIQM^[24] 和 UCIQE^[25] 进行评价.

有参考指标对比如表 1 所示. 可以发现本方法在 PSNR 上取得了次优的结果,相较于第一名 Semi-UIR 仅落后 0.13,相较于第三名 SCNet 则高出 1.58. 可以明显看出在 PSNR 方面本文方法和 Semi-UIR 大大优于其他方法. 在 SSIM 方面,本文方法取得最优结果,相较于第二名 Semi-UIR 领先了 0.09,相较于第三名 SCNet 领先了 0.039. 需要注意的是,尽管本文方法在 PSNR 上略逊于

Semi-UIR, 但正如研究^{[29][30]}所言, SSIM 作为衡量标准相比于 PSNR 更为准确客观,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 一样以非归一化绝对散度作为度量.

表 1 UIEB-test 数据集上八种方法的 PSNR 和 SSIM 对比

方法	PSNR	SSIM
GDCP ^[6]	13.60	0.724
fusion ^[11]	18.32	0.852
FGAN ^[15]	18.08	0.761
UIESS ^[16]	19.02	0.809
Semi-UIR ^[17]	24.31	<u>0.915</u>
SCNet ^[18]	22.60	0.885
CLUIE ^[19]	21.02	0.883
本文	<u>24.18</u>	0.924

注. 粗体表示最优结果, 划线表示次优结果.

无参考指标如表 2 所示. 本文方法在六个结果中取得了四个最优两个次优的优良成绩. 尽管在 UIEB-challenge 和 EUVP 两个数据集上,本文方法的 UCIQE 分别低于 UIESS 和 GDCP, 但图 4、5、6、7 均说明本文方法的增强结果在视觉上优于它们. 此外,本文方法在 UCIQE 指标与它们相近的同时, UIQM 指标远高于它们,同样体现了本文方法的优越性. 综上所述,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.

表 2 UIEB-challenge、EUVP、RUIE 三个数据集上八种方法的 UIQM 和 UCIQE 对比

方法	UIEB-challenge		EUVP		RUIE	
	UIQM	UCIQE	UIQM	UCIQE	UIQM	UCIQE
输入	2.14	0.357	2.62	0.421	2.96	0.32
GDCP ^[6]	2.094	<u>0.450</u>	2.19	0.486	2.67	<u>0.431</u>
fusion ^[11]	2.549	0.379	<u>3.20</u>	0.441	<u>3.34</u>	0.379
FGAN ^[15]	<u>3.02</u>	0.409	3.02	0.445	3.21	0.362
UIESS ^[16]	2.65	0.461	3.04	0.458	3.17	0.419
Semi-UIR ^[17]	2.85	0.430	2.95	0.441	3.28	0.399
SCNet ^[18]	2.74	0.379	3.14	0.444	3.31	0.355
CLUIE ^[19]	2.71	0.412	2.93	0.454	3.10	0.370
本文	3.16	<u>0.450</u>	3.24	<u>0.462</u>	3.37	0.450

注. 粗体表示最优结果, 划线表示次优结果.

3.4 消融实验

为了验证本文方法各个环节的有效性,进行了消融实验,如下所示: (1)移除风格编码器中

CLIP 图像编码器; (2)移除 DAHAB; (3)移除内容一致性损失; (4)移除参考图像 → 水下图像分支以及公式(12)中对应的部分. 其中,对比实验(1)中,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. 此外, 为了证明 DAHAB 的优越性, 增加实验(5), 使用广泛应用的卷积块注意力模块 (Convolutional Block Attention Module, CBAM)^[31] 替换 DAHAB. 结果如表 3 所示.

实验结果表明, 本文方法中的每一个部分都是必要的, 去除任意环节均会导致指标下降. 此外, 根据图 8 中展示的视觉效果来看, 本文方法的增强结果相比其他对比实验的增强结果细节更加清晰, 颜色更加自然.

表 3 消融实验指标对比

方法	UIEB-test		UIEB-challenge		EUVP		RUIE	
	PSNR	SSIM	UIQM	UCIQE	UIQM	UCIQE	UIQM	UCIQE
(1)	22.69	0.900	2.73	0.420	2.95	0.454	3.23	0.391
(2)	23.26	0.889	2.82	0.413	3.05	0.456	3.27	0.420
(3)	23.74	0.894	2.94	0.434	3.12	0.441	3.30	0.436
(4)	23.45	0.896	2.79	0.421	3.14	0.448	3.27	0.446
(5)	23.70	0.910	3.02	0.436	3.16	0.453	3.30	0.443
本文	24.18	0.924	3.16	0.450	3.24	0.462	3.37	0.450

注. 粗体表示最优结果.

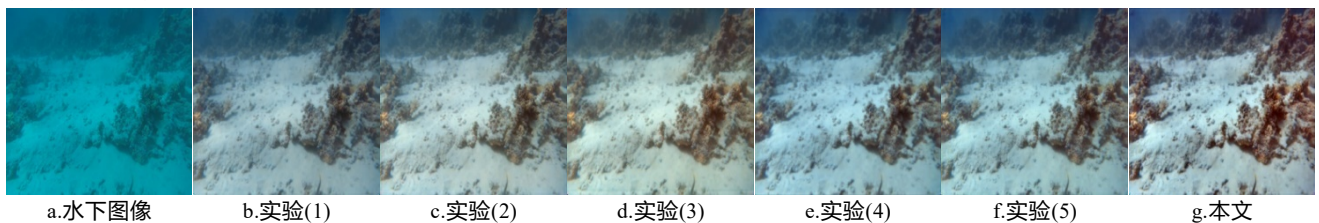


图 8 消融实验视觉效果对比

UIQM 与 UCIQE 的权重. 本节还将讨论在推理方式 2 中, 以不同权重加权 UIQM 与 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与 UCIQE 更大的权重. 然而, 实验发现这会导致增强结果出现偏色、过饱和等问题, 如图 8 所示. 图 8a 按 1:1 加权, 图 8b 按 1:5 加权. 可以明显看出图 8b 的增强结果偏向红色. 相对而言, 图 8a 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 与 UCIQE.

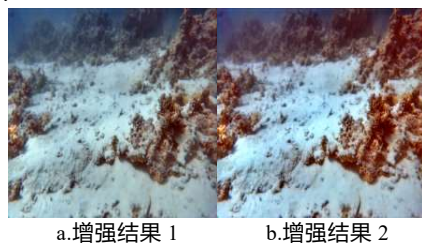


图 9 不同 UIQM 与 UCIQE 权重下增强结果对比

4 结 语

本文提出了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. 对于风格域, 本文提出了 DAHAB 对内容特征的高频信息进行调整; 对于风格域, 本文通过建立风格库并使用内容特征匹配最优风格编码的方式进行增强. 相比于其他方法, 本文方法在 UIEB、EUVP 和 RUIE 三个数据集上取得了更好的指标与视觉效果, 并且在困难样本的增强上有较大优势.

参考文献(References):

- [1] Schechner Y Y, Karpel N. Clear underwater vision[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2004: 536-543
- [2] 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, 2019, 29: 4376-4389
- [3] Li C Y, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement[J]. Pattern Recognition, 2020, 98: 107038-107049

- [4] 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*, 2020, 43(8): 2822-2837
- [5] 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
- [6] Peng Y T, Cao K, Cosman P C. Generalization of the dark channel prior for single image restoration[J]. *IEEE Transactions on Image Processing*, 2018, 27(6): 2856-2868
- [7] He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 33(12): 2341-2353
- [8] Hitam M S, Awalludin E A, Yusof 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*. Piscataway: IEEE, 2013: 1-5
- [9] Zhang S, Wang T, Dong J, et al. Underwater image enhancement via extended multi-scale retinex[J]. *Neurocomputing*, 2017, 245: 1-9
- [10] Land E H. The retinex theory of color vision[J]. *Scientific American*, 1977, 237(6): 108-129
- [11] Ancuti C O, Ancuti C, De Vleeschouwer C, et al. Color balance and fusion for underwater image enhancement[J]. *IEEE Transactions on Image Processing*, 2017, 27(1): 379-393
- [12] Zhang W, Zhuang P, 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
- [13] Liu X, Gao Z, Chen B M. MLFcGAN: Multilevel feature fusion-based conditional GAN for underwater image color correction[J]. *IEEE Geoscience and Remote Sensing Letters*, 2019, 17(19): 1488-1492
- [14] Cai B, Xu X, Jia K, et al. DehazeNet: An end-to-end system for single image haze removal[J]. *IEEE Transactions on Image Processing*, 2016, 25(11): 5187-5198
- [15] Islam M J, Xia Y, Sattar J. Fast underwater image enhancement for improved visual perception[J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 3227-3234
- [16] Chen Y W, Pei S C. Domain Adaptation for Underwater Image Enhancement via Content and Style Separation[J]. *IEEE Access*, 2022, 10: 90523-90534
- [17] Huang S, Wang K, Liu H, et al. Contrastive semi-supervised learning for underwater image restoration via reliable bank[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2023: 18145-18155
- [18] 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
- [19] Li K, 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
- [20] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//*Proceedings of Machine Learning Research*. New York: PMLR, 2021: 8748-8763
- [21] Cui Y, Ren W, Cao X, et al. Focal network for image restoration[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2023: 13001-13011
- [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: 1501-1510
- [23] Jiang Q, Zhang Y, Bao F, et al. Two-step domain adaptation for underwater image enhancement[J]. *Pattern Recognition*, 2022, 122: 108324-108338
- [24] 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
- [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, 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 International Conference on Learning Representations*. Toulon: OpenReview.net, 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] Setiadi D R I M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography[J]. *Multimedia Tools and Applications*, 2021, 80(6): 8423-8444
- [30] 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
- [31] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//*Proceedings of the European Conference on Computer Vision*. Heidelberg: Springer, 2018: 3-19

作者姓名 (按论文署名顺序填写)	单位	工作邮箱 (域名为工作单位的邮箱)	特殊情况说明 (包括单位邮箱涉密不可公开、非员工无单位邮箱等情况请说明)
刘春晓	浙江工商大学计算机科学与技术学院	cxliu@mail.zjgsu.edu.cn	电话号码 15968806010
曹越	浙江工商大学计算机科学与技术学院	1805090223@pop.zjgsu.edu.cn	电话号码 13216959963
王成骅	浙江工商大学计算机科学与技术学院	2012190218@pop.zjgsu.edu.cn	电话号码 15336540831
周子翔	浙江工商大学计算机科学与技术学院	2011080223@pop.zjgsu.edu.cn	电话号码 15967196978