



基于物理先验的深度特征融合水下图像复原

张心祎, 谭耀, 邢向磊

引用本文:

张心, 谭耀, 邢向磊. 基于物理先验的深度特征融合水下图像复原[J]. *智能系统学报*, 2023, 18(6): 1185–1196.

ZHANG Xinyi, TAN Yao, XING Xianglei. Deep feature fusion for underwater-image restoration based on physical priors[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(6): 1185–1196.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202304038>

您可能感兴趣的其他文章

样本仿真结合迁移学习的声呐图像水雷检测

Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning
智能系统学报. 2021, 16(2): 385–392 <https://dx.doi.org/10.11992/tis.202101030>

多感知兴趣区域特征融合的图像识别方法

Image recognition method based on multi-perceptual interest region feature fusion
智能系统学报. 2021, 16(2): 263–270 <https://dx.doi.org/10.11992/tis.201906032>

利用残差密集网络的运动模糊复原方法

Image restoration with residual dense network
智能系统学报. 2021, 16(3): 442–448 <https://dx.doi.org/10.11992/tis.201912002>

基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion
智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects
智能系统学报. 2020, 15(3): 560–567 <https://dx.doi.org/10.11992/tis.201904020>

基于生成对抗网络的机载遥感图像超分辨率重建

Super-resolution reconstruction of airborne remote sensing images based on the generative adversarial networks
智能系统学报. 2020, 15(1): 74–83 <https://dx.doi.org/10.11992/tis.202002002>

DOI: 10.11992/tis.202304038

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20230807.1532.002>

基于物理先验的深度特征融合水下图像复原

张心祎, 谭耀, 邢向磊

(哈尔滨工程大学智能科学与工程学院, 黑龙江哈尔滨 151001)

摘要: 由于水下环境的浮游生物悬浮杂质及不同光谱吸收率等干扰因素, 水下图像往往会出现图像模糊、颜色失真和光照不均等退化问题。本文提出联合水下物理成像规律与数据驱动深度学习方法的水下图像重建模型。利用深度神经网络推断物理成像模型中的可学习参数, 通过调制卷积和物理先验知识分别生成基于数据驱动的复原特征图和基于物理先验的复原特征图, 引入混合注意力机制的深层特征级融合, 重建最终的复原图像。实验结果表明该方法可以在减少噪声、提高对比度的同时, 恢复图像的细节, 提高水下图像的可视化质量和目标检测精度, 增强水下学习模型的鲁棒性和泛化能力。

关键词: 深度学习; 水下图像恢复; 神经网络; 信息分离; 编码器; 解码器; 特征提取; 图像融合

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 1673-4785(2023)06-1185-12

中文引用格式: 张心祎, 谭耀, 邢向磊. 基于物理先验的深度特征融合水下图像复原 [J]. 智能系统学报, 2023, 18(6): 1185-1196.

英文引用格式: ZHANG Xinyi, TAN Yao, XING Xianglei. Deep feature fusion for underwater-image restoration based on physical priors[J]. CAAI transactions on intelligent systems, 2023, 18(6): 1185-1196.

Deep feature fusion for underwater-image restoration based on physical priors

ZHANG Xinyi, TAN Yao, XING Xianglei

(College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: Due to interference factors such as suspended impurities of plankton and varying spectral absorption rates in an underwater environment, underwater images often suffer from degradation issues such as image blur, color distortion, and uneven illumination. This paper proposes an underwater-image reconstruction model that combines physical imaging principles with data-driven deep-learning methods. Using a deep neural network to infer the learnable parameters in the physical imaging model, the model generates data-driven restoration feature maps and physically informed restoration feature maps through modulated convolution and prior physical knowledge, respectively. Deep feature fusion with a mixed-attention mechanism is introduced to reconstruct the final image. Experimental results showed that this method can reduce noise, improve contrast, and restore image details, enhancing the visual quality and target detection accuracy of underwater images and increasing the robustness and generalizability of the underwater learning model.

Keywords: deep learning; underwater-image restoration; neural networks; information separation; encoder; decoder; feature extraction; image fusion

随着社会的发展, 人类对海洋资源的探索更加深入, 我们越来越需要清晰的水下图像来进行水下生物, 水下操作系统的研究。然而, 获取的水下图像往往降质十分严重, 其中最主要的因素是水对于不同波长的光吸收率不同, 红色光波长最长吸收最为严重, 而蓝色光波长最短吸收最少, 因此水下图像总是呈现出蓝绿色调, 同时伴随着色偏和色彩不鲜明等问题。同时由于水中光

线传播距离短, 受到水分子、悬浮颗粒、藻类、微生物等的散射和吸收作用, 光线的强度、频率和颜色均发生变化, 从而导致水下图像总是模糊, 充满了噪点。水下图像复原的方法可以被粗略的分为传统方法和基于深度学习的方法。早期人们尝试去通过调整像素值来提高视觉质量, 例如动态像素的延展^[1]、像素点的自适应分配^[2]以及图像融合^[3]。尽管这些方法可以在一定程度上提高视觉质量, 但它们忽略了水下成像机制, 因此往往会导致过度增强或增强不足的问题, 并且会导致伪影的产生。例如在一些水下场景中^[4]处理的

收稿日期: 2023-04-18. 网络出版日期: 2023-08-07.

基金项目: 国家自然科学基金项目 (62076078, 61703119).

通信作者: 邢向磊. E-mail: xingxl@hrbeu.edu.cn.

效果并不总是很好。如今被广泛使用的是水下成像的物理模型,物理模型通过先验知识来评估水下成像模型的参数,这些基于先验知识的方法包括红通道先验^[5]、水下暗通道先验^[6]、最小信息先验^[7]等。尽管这些物理模型取得了不错的复原效果,但是其计算时间较长,同时其对水下成像公式中参数评估的准确性十分依赖于水下成像的类型,而且模型中的先验并不总是成立的,因此这类方法的泛化能力较差。神经网络模型的发展为水下图像处理提供了新的思路,随着 Swin-Transformer^[8]、ConvNext^[9]等网络架构的提出,神经网络处理视觉任务获得了进一步的提升。但是与水下图像相对应的目标图像难以获取,现有的数据集中数据数量相比于其他任务如分类,语义分割等也十分稀少,这也限制了数据驱动模型的应用,为了解决数据集问题,文献[10-11]使用了生成对抗网络和水下图像生成公式来合成水下图像和清晰图像用于有监督学习。为了避免使用成对的训练数据,UcycleGAN^[12]使用了一个弱监督的水下图像修复网。

水下图像质量低主要体现在色偏和图像模糊两个方面,对于图像模糊的处理往往需要关注图像内物体的边缘细节,这是局部性的、像素级别的处理,而造成色偏的主要原因是水对于光的均匀吸收造成水下图像整体偏绿或偏蓝,这往往是全局性的与物体所处的位置无关。

考虑到以通用暗通道先验 (generalization of the dark channel prior, GDCP)^[13]为代表的传统方法中物理先验知识的成功应用,本文提出了基于物理先验的特征融合水下图像恢复网络。首先考虑到水下图像的成像特点,将水下图像中局部空间信息和全局色彩信息进行分离,提取出保留像素空间位置关系的低通道数特征向量和宽、高均为 1 且只包含全局性色彩信息的高通道数特征向量。并且在空间信息提取模块中,通过空间注意力机制使模型专注于水下图像中某些局部信息,提高恢复图像中物体边缘细节的表现,减少模糊。在色彩信息提取模块中,考虑到不同成像条件对水下图像色彩的影响,通过引入通道注意力机制使模型专注于不同通道上色彩、对比度、亮度等全局信息的恢复,减少生成图像中的色偏,使生成图像尽可能地接近于陆地图像 (Groud-Truth)。提取出不同信息后,在解码器中逐步生成恢复图像,通过调制操作联合不同通道上的全局信息,使不同层级的解码器将不同尺度的色彩信息逐步融合进空间信息中。最后,在物理先验模块中,将水下光学成像模型 IMF(image model func-

tion) 嵌入到最终的生成器中,通过引入注意力机制,使网络在处理不同的水下图像时关注不同的成像机制,以综合利用神经网络和传统方法的优势。

1 相关知识

1.1 水下图像生成

Jaffe-McGlamery 模型^[11,14]是一种用于描述水下光学传输的经典模型,其假设水体是均质的,并考虑了散射、吸收和反射等因素对光线的影响,如图 1 所示。

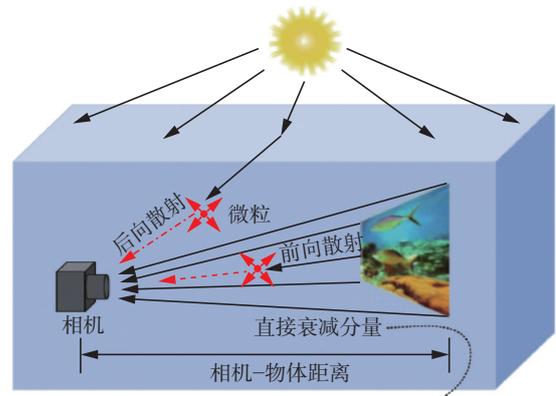


图 1 水下图像成像过程

Fig. 1 Underwater image imaging process

该模型可抽象表达为

$$E_T(x, y) = E_d(x, y) + E_f(x, y) + E_b(x, y) \quad (1)$$

式中将水下获得的信号分为 E_d 直接照射、 E_f 前向散射和 E_b 后向散射 3 个部分。 (x, y) 表示图像中像素的坐标; $E_T(x, y)$ 代表相机捕捉到的总光能量; $E_d(x, y)$ 代表直接照射部分是指光线从物体表面反射并在未经散射的情况下到达相机的部分; $E_f(x, y)$ 代表前向散射,是物体表面反射的光在到达相机这段距离上发生的散射; $E_b(x, y)$ 代表后向散射部分,是指自然光到达物体表面之前被反射进相机的部分。

在水下环境中,光线很容易被吸收和散射,导致光线强度衰减快,前向散射能够传递的距离有限,因此前向散射所占比例较小。同时在水中水分子和悬浮颗粒的影响下,散射光线更容易向后方散射,这会进一步导致前向散射的减少,从而扩大后向散射所占比例,因此为了便于计算,我们常常忽略 IMF 中的前向散射部分,而使用简化后的 Jaffe-McGlamery 模型,该模型如下表示:

$$I_c(x, y) = J_c(x, y)t_c(x, y) + A_c(1 - t_c(x, y)) \quad (2)$$

式中: (x, y) 表示像素点在图像中的坐标, c 代表红绿蓝 (r、g、b) 通道, $I_c(x, y)$ 表示相机直接拍摄的光强度, $t_c(x, y)$ 表示在水中介质传输率, $J_c(x, y)t_c(x, y)$

表示场景能量在水的吸收作用后剩余的部分, $J_c(x, y)$ 表示未衰减的图像, A_c 表示全局均匀环境光。

在水中, $t_c(x, y)$ 也被称为介质传输率, 其值受到水下衰减系数和物体与相机间距离的影响, 因此 $t_c(x, y)$ 可以通过以下关系式表示:

$$t_c(x, y) = e^{-\beta_c d(x)} \quad (3)$$

式中: β_c 为不同通道的水下衰减系数, 其由光的波长决定; $d(x)$ 表示目标的深度信息, 即目标物体到相机的距离。

使用介质传输率和全局均匀环境光, 可以得到未衰减图像的计算公式:

$$J_c(x, y) = \frac{I_c(x, y) - A_c}{\max(t_c(x, y), t_0)} + A_c \quad (4)$$

由于使用简化后的 Jaffe-McGlamery 模型会导致一些信息的丢失, 同时为了保存一些深度信息, 因此在实际应用中, 设置 $t_0 = 0.1$ 作为下边界, 同时裁剪 $J_c(x, y)$ 在 13~255 之间。

1.2 暗通道先验算法

单图像复原的通用暗通道先验 GDCP 算法^[13]提出了一种通用的基于物理先验的水下图像恢复方法, 其通过水下图像的局部光强度估算介质传输率 \tilde{t}_c , 通过全局光强度最高的点估计全局均匀背景光 A_c , 最后通过式 (4) 得到恢复的水下图像。其具体表达式如下:

$$\tilde{t}_{\text{rgb}}(x) = 1 - \min_{y \in \Omega(x)} \left\{ \min_{c \in \{r, g, b\}} \frac{I_c(y)}{A_c} \right\} \quad (5)$$

该公式通过点 x 某一邻域内 r 、 g 、 b 三通道中光强的最小值估算不同通道的介质传输率。其中 $y \in \Omega(x)$ 代表 y 是始于 x 邻域内的一个点。该公式的物理含义为点 x 处的介质传输率是以 x 为中心的某一区域内, r 、 g 、 b 三通道中, 像素值最小的点与均匀背景光 A_c 的比值。为了计算 A_c , 需要先计算暗通道先验的光强图 $I_{\text{Dcp}}^{\text{rgb}}(x)$, 其数学表达式如下:

$$I_{\text{Dcp}}^{\text{rgb}}(x) = \min_{y \in \Omega(x)} \{ \min_{c \in \{r, g, b\}} I_c(y) \} \quad (6)$$

该公式的物理含义为点 x 处的暗通道先验光强是以 x 为中心的某一邻域内, r 、 g 、 b 三通道中, 像素值最小的点所对应的光强。

A_c 的数学表达式为

$$A_c = I_c(\arg \max_{x \in P^{0.1\%}} \sum_{c \in \{r, g, b\}} I_c(x)) \quad (7)$$

其中 $P^{0.1\%}$ 是 $I_{\text{Dcp}}^{\text{rgb}}(x)$ 中前 0.1% 最大像素值点所构成的位置集。

该公式的物理含义为在 $I_{\text{Dcp}}^{\text{rgb}}(x)$ 中前 0.1% 最大像素值点所构成的位置集中, 对 r 、 g 、 b 三通道光强求和并找出其最大值, 该值所对应的点处的光强即为均匀背景光。

最后将计算得到的 $\tilde{t}_{\text{rgb}}(x)$ 、 A_c 代入式 (7) 得到处理后得水下图像。

尽管 GDCP 算法取得了不错的恢复效果, 但是由于物理参数很难估计准确, 因此限制了该算法的应用。

2 基于物理先验的深度特征融合网络

本文网络的整体流程如图 2 所示, 受水下图像光学成像公式 (式 (1)) 和 GDCP 算法^[13] 中利用背景光和色彩信息计算传输率的启发, 本文提出通过数据驱动的深度学习方法, 自适应地学习式 (4) 中的全局背景光 A_c 和介质传输率图 \hat{t}_c 。在解码器中充分利用局部空间信息和全局色彩信息估算传输率特征图 \hat{t}_d , 再通过物理先验模块 (physical priors, PPAM) 中的调制卷积和物理先验知识分别生成基于数据的恢复图像 I_g 和基于物理先验的恢复图像 I_p , 最后通过 PPAM 模块中的混合注意力机制, 将 I_g 与 I_p 进行深度特征级融合, 并重建为最终的恢复图像 I_{rec} 。

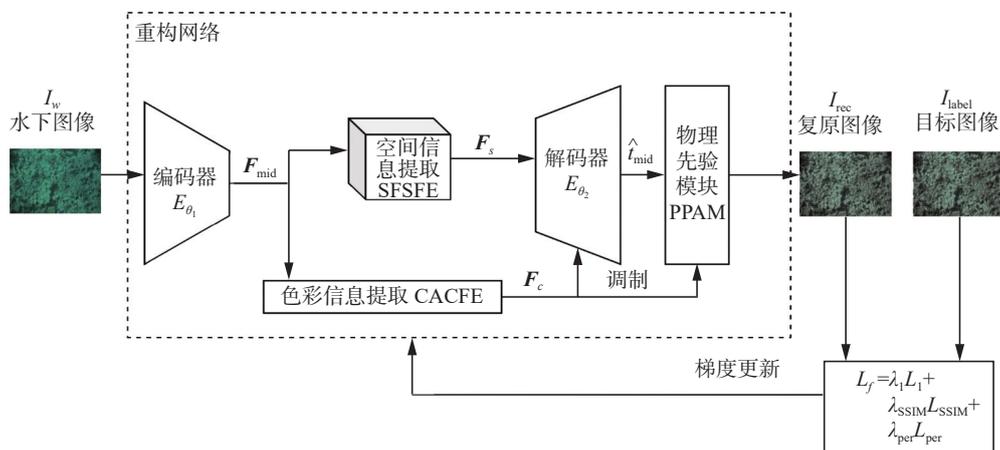


图 2 基于物理先验的深度特征融合水下图像恢复网络的整体流程

Fig. 2 Overall flow of underwater image restoration network based on physical prior deep feature fusion

恢复图像 I_{rec} 的具体数学表达式如下:

$$\begin{cases} I_{rec} = PPAM(D_{\theta_2}(F_s, F_c), F_c) \\ F_s = SFSFE(F_{mid}) \\ F_c = CACFE(F_{mid}) \\ F_{mid} = E_{\theta_1}(I_w) \end{cases} \quad (8)$$

式中: PPAM、SFSFE、CACFE 分别代表物理先验模块, 空间注意力空间信息提取 (spatial attention spatial feature extraction, SASFE) 模块, 通道注意力色彩信息提取 (channel attention color information extraction, CACFE) 模块; F_{mid} 是中间特征向量; F_s 是空间特征向量; F_c 是色彩特征向量; D_{θ_2} 是以 θ_2 为参数得解码器; E_{θ_1} 是以 θ_1 为参数的编码器; I_w 是水下图像; 具体实现将在后文介绍。

由图 2 可知, 本文网络的整体流程由 5 个部分组成, 分别为编码器、解码器、SFSFE、CACFE、PPAM。

网络的输入为归一化至 $[-1, 1]$ 范围内的水下图像 $I_w \in \mathbf{R}^{3 \times h \times w}$, 其中 h 和 w 分别代表水下图像的高和宽, 本文中均取 256。将 I_w 送入编码器编码, 编码器输出中间特征向量 $F_{mid} = E_{\theta_1}(I_w)$, $F_{mid} \in \mathbf{R}^{c \times h_{mid} \times w_{mid}}$ 。其中 h_{mid} 和 w_{mid} 分别代表了 F_{mid} 的高和宽, 本文中均为 16, c 代表通道数。然后 F_{mid} 经过 SFSFE 和 CACFE 分别得到空间特征向量 $F_s \in \mathbf{R}^{c \times h_s \times w_s}$ 和色彩特征向量 $F_c \in \mathbf{R}^c$; 最后在解码器中通过调制操作将 F_s 与 F_c 融合, 输出介质传输率向量 $\hat{t}_c = D_{\theta_2}(F_s, M(F_c)) \in \mathbf{R}^{3 \times h \times w}$, M 是调制操作。下面对网络进行详细介绍。

2.1 编码器

图 3 详细的展示了编码器结构, 其由 1 个卷积模块和 4 个残差模块组成。

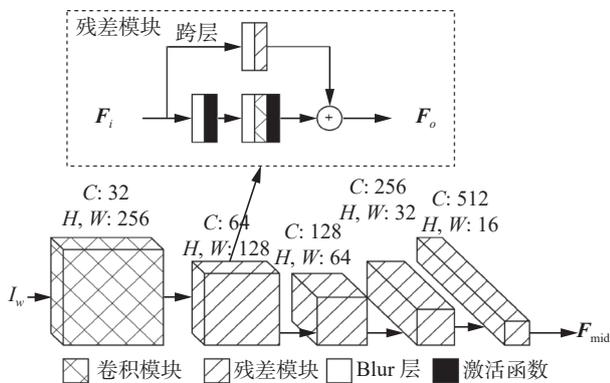


图 3 编码器结构

Fig. 3 Encoder structure

首先水下图像 I_w 由一个卷积操作编码为一个水下特征向量 $F_w = \text{Conv}(I_w) \in \mathbf{R}^{32 \times 256 \times 256}$, 得到的水下特征向量送入后续的残差模块, 特征向量通道数由 32 逐渐增长至 512, 其增长因子是 2, 在通道

数增长的同时, 在每一个残差模块中, 都对特征向量进行下采样, 逐步提高神经网络的感受野, 其下采样因子也为 2, 经过四个残差结构后, 可以得到中间特征向量 $F_{mid} = E_{\theta_1}(I_w) \in \mathbf{R}^{512 \times 16 \times 16}$ 。

本文的编码器和解码器由残差模块作为基本单元组成, 具体结构如图 3 中所示, 每一个残差模块由 3 个卷积层 2 个激活层组成, 使用残差模块的目的是提高数据的保真度并且可以避免梯度消失的问题^[15], 在每一个残差模块中, 卷积层具有相同数量的滤波器。残差模块中所有的卷积层卷积核大小都为 3, 其步长均为 1。残差模块的具体数学表达式如下:

$$F_o = B(F_i * k^{\text{skip}} + b^{\text{skip}}) + R(B(R(F_i * k^1 + b^1)) * k^2 + b^2) \quad (9)$$

式中: F_o 代表输出特征, F_i 代表输入特征, $*$ 代表卷积操作, k^1 , k^2 , k^{skip} 分别代表第 1 层、第 2 层和跨层的卷积核参数, b^1 , b^2 , b^{skip} 分别代表第 1 层、第 2 层和跨层的卷积偏置值, $R(\cdot)$ 代表 Relu 激活函数, $B(\cdot)$ 代表 Blur 下采样操作。

在残差模块中使用了 Blur^[16] 进行下采样, 这是因为使用 Blur 进行下采样可以降低图像中的高频信号, 减少噪声和细节, 从而避免神经网络学习到不必要的细节信息。此外, 使用 Blur 进行下采样可以使图像保持平滑的特性, 避免出现锯齿和马赛克等问题。最后, 使用 Blur 进行下采样可以避免出现深度网络中的梯度消失问题, 使得神经网络更加稳定和易于训练。

2.2 空间注意力空间特征提取模块

SFSFE 由 2 个空间注意力模块所组成, 空间注意力模块如图 4 所示。

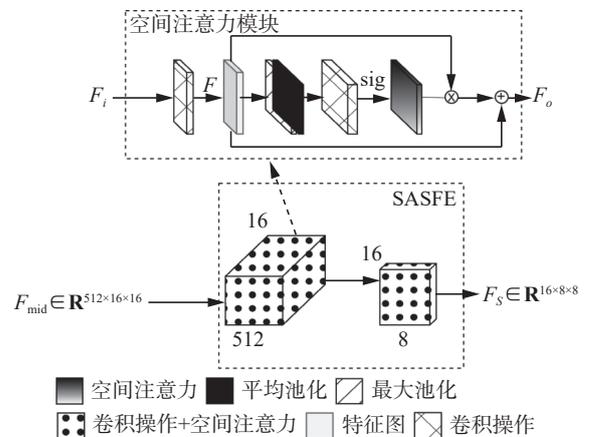


图 4 空间注意力空间特征提取模块

Fig. 4 Spatial attention spatial feature extraction module

空间注意力模块输出的计算公式如下:

$$F_o = (\sigma(f^{7 \times 7}(\text{cat}[F_{Avg}, F_{Max}])) \otimes F) \oplus F \quad (10)$$

式中: F_o 为输出向量, $F = \text{Conv}(F_i)$ 代表输入特征经

卷积操作后输出的特征图, $f^{7 \times 7}$ 是卷积核大小为 7 的卷积操作是图 5 中的第 2 个卷积操作, $F_{\text{Avg}} = \text{Avgpool}(F)$ 、 $F_{\text{Max}} = \text{Maxpool}(F)$ 分别代表在通道方向上的平均池化和最大池化后得到的特征图, $F_{\text{cat}} = \text{cat}[F_{\text{Avg}}, F_{\text{Max}}]$, sig 是 sigmoid 激活函数, \oplus 和 \otimes 分别代表了像素级别的加法和乘法。

第 1 个空间注意力模块不改变特征向量的通道数, 在第 2 个注意力模块中, 在通道方向上进行压缩, 最终输出的空间信息特征向量 $F_s \in \mathbf{R}^{8 \times 16 \times 16}$ 。

2.3 通道注意力色彩信息提取模块

不同于 GDCP^[13] 直接使用局部光信息与全局均匀背景光估算介质传输率图 $\hat{t}_c \in \mathbf{R}^{3 \times 256 \times 256}$, 本文设计了 CACFE 模块以使用神经网络提取水下图像的全局色彩信息, 以便在后续的模块中计算全局均匀背景光 A_c 。

考虑到每个通道空间所包含的信息意义不同, 因此提取色彩信息时每个通道应该具有不同的贡献。使用通道注意力模块提取水下图像的色彩信息。通道注意力模块如图 5 所示。

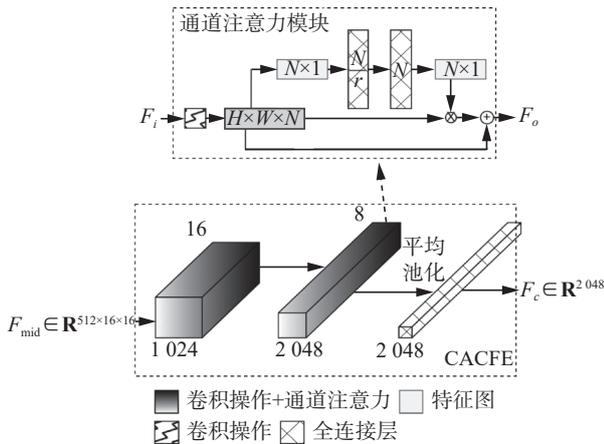


图 5 通道注意力色彩信息提取模块

Fig. 5 Channel attention color information extraction module

其输入特征是 $F_i \in \mathbf{R}^{N \times H \times W}$ 的空间向量, 其中 F_i 是输入特征图, H 和 W 分别代表特征图的高和宽。首先使用全局平均池化, 从而得到一个信息描述符 $Z \in \mathbf{R}^{N \times 1}$, 这是一个嵌入式的逐通道特征响应全局分布。第 k 个 z 可以表述为

$$z_k = \frac{1}{H \times W} \sum_i^H \sum_j^W F_k(i, j) \quad (11)$$

其中 $k \in [1, N]$, 为了更好地利用通道之间的相互依赖性, 使用了自门控机制来生成每个通道调制的权重集合。

$$s = \sigma(W_2 * (\delta(W_1 * z))) \quad (12)$$

其中: σ 代表 sigmoid 激活函数, δ 代表 Relu 激活函数, $*$ 代表卷积操作, W_1 和 W_2 分代表了 2 个卷积层

的权重, 其输出通道分别为 $\frac{N}{r}$ 和 N , 其中 r 等于 16, 其目的是为了减少运算量, 加快程序的运行速度。同时, 为了避免梯度消失和保持原有特征的特性, 以残差的形式处理通道注意力的权重。

$$F_o = \text{Conv}(F_i \oplus F_i \otimes S) \quad (13)$$

其中 $F_o \in \mathbf{R}^{c \times H/2 \times W/2}$ 是输出特征向量, 其高宽通过卷积操作变为原来的 1/2, \oplus 和 \otimes 分别代表了像素级别的加法和乘法。

色彩信息提取模块由两个通道注意力模块^[17] 依次串联组成, 考虑到色彩信息与空间信息无关, 因此中间特征向量每通过一个通道注意力模块其宽高都会缩减为原来的 1/2, 在最后一个全连接层前使用全局平均池化替代下采样以消除空间信息, 最终输出色彩信息特征向量 $F_c \in \mathbf{R}^{2048}$ 。

2.4 解码器

受 GDCP^[13] 算法中利用局部信息和全局色彩信息计算传输率图的启发, 在解码器中将 F_c 与 F_s 相结合计算 \hat{t}_c , 以便在 PPAM 模块中利用式 (4) 生成一张基于物理先验的恢复图像。

如图 6 所示, 本文的解码器共由 9 个残差模块组成, 前 4 个残差模块只进行通道方向上的扩展并不进行上采样操作, 受文献 [18] 的启发, 本文使用调制操作, 将色彩信息逐步融入进特征向量中。

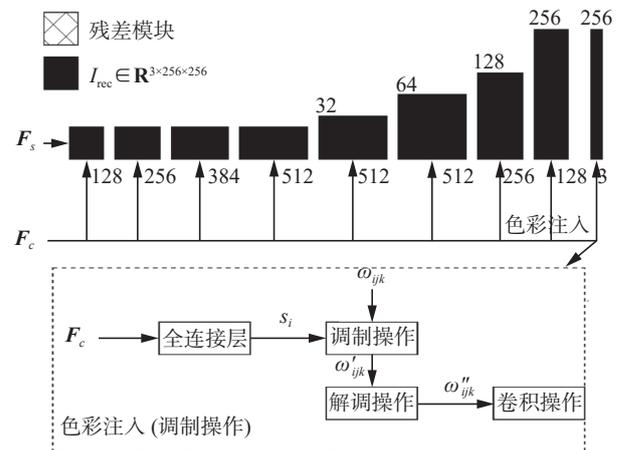


图 6 解码器结构

Fig. 6 Decoder structure

简单来说调制操作是将学习到的调制向量映射到特定层的平均值和方差上, 调制操作的数学表达式如下:

$$w'_{ijk} = s_i \cdot w_{ijk} \quad (14)$$

其中: w 和 w' 分别代表原始权重和调制权重, “ \cdot ” 代表元素相乘, s_i 是通过全连接层所学到的与第 i 个输入特征图相对应的比例。本文中为色彩信息, j 和 k 分别为特征图和卷积核的空间下标。经过

调制和卷积操作后, 输出向量的标准差为

$$\sigma_j = \sqrt{\sum_{i,k} w'_{ijk}{}^2} \quad (15)$$

为了将输出特征图恢复为单位标准差, 本文需要解调操作, 即将上述标准差再次嵌入到卷积权重中:

$$w''_{ijk} = \frac{w'_{ijk}}{\sqrt{\sum_{i,k} w'_{ijk}{}^2 + \varepsilon}} \quad (16)$$

ε 的作用是防止分母为 0。

后 5 个残差模块在通道方向上不断的压缩, 同时依旧使用 Blur 对图像进行上采样, 在最后的

残差模块中, 生成传输率特征图 $\hat{t}_c \in \mathbf{R}^{3 \times 256 \times 256}$ 。

2.5 基于混合注意力机制的物理先验模块

为了更好地利用已有的物理先验知识和深度神经网络的非线性学习能力, 本文设计了物理先验模块。

如图 7 所示, 在物理先验模块中, 首先利用 \hat{t}_c 与 F_c 通过调制操作生成一张基于神经网络的恢复图像, 称之为生成特征向量 F_g ; 同时利用 \hat{t}_c 与 F_c 通过全连接层计算 1.1 中的式 (4) 中的参数并生成一张基于物理先验的恢复图像, 称之为物理特征向量 F_p 。将生成特征向量和物理特征向量沿通道堆叠后, 利用混合注意力机制生成的权重图, 经过卷积操作后得到最终恢复的图像。

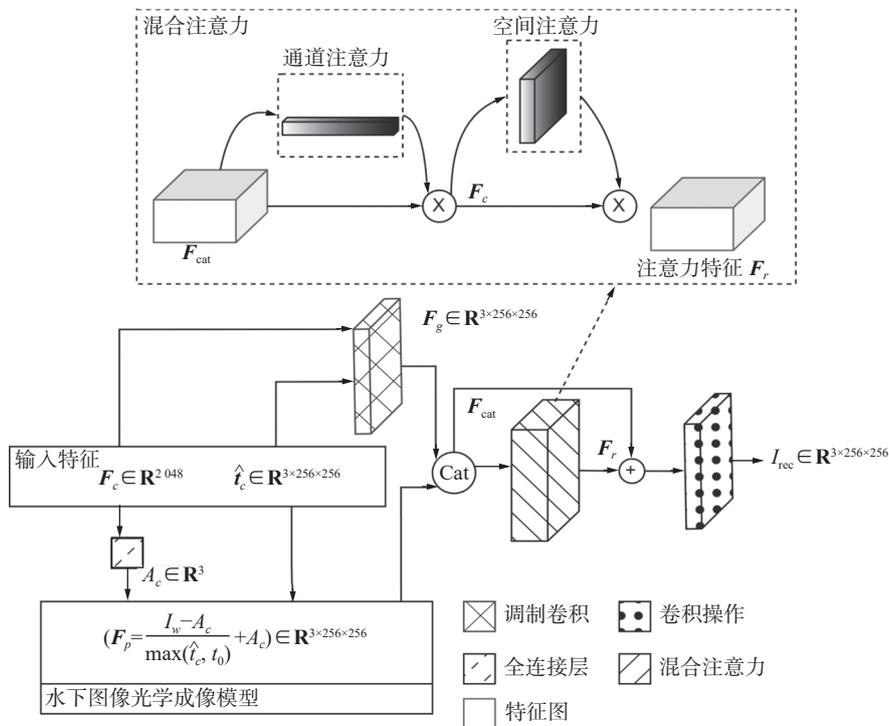


图 7 物理先验模块

Fig. 7 Physical prior module

其中 F_p 具体表达式如下:

$$F_p = \frac{I_w - A_c}{\max(\hat{t}_c, t_0)} + A_c \quad (17)$$

其中: $t_0=0.1$, $A_c = \text{mlp}(F_c) \in \mathbf{R}^3$ 是通过 mlp 全连接层计算出的均匀背景光。

混合注意力流程如图 7 中所示, 其具体数学表达式如下:

$$F_r = M_s(M_c(F_{\text{cat}})) \quad (18)$$

其中 M_s 和 M_c 分别代表空间注意力和通道注意力, 具体实现见 2.2、2.3 节。PPAM 的数学表达式如下:

$$I_{\text{rec}} = \tanh(\text{Conv}(F_r) \oplus F_{\text{cat}}) \quad (19)$$

其中 I_{rec} 为恢复图像, $F_{\text{cat}} = \text{concat}[F_p, F_g]$ 代表是将物理特征向量和生成特征向量沿通道方向堆叠后

的堆叠特征图, Conv、 \oplus 分别为卷积操作、元素加法。

2.6 损失函数

为了训练本文网络, 本文使用了 3 种损失函数的组合, 分别为 L_1 损失、SSIM^[19] 损失、感知路径损失^[20]。

L_1 损失, 也被称为平均绝对误差 (mean absolute error, MAE), 是将每个样本的预测值与真实值之差的绝对值求和后再求平均值。

$$L_1 = \frac{1}{n} \sum_{i=1}^n |I_{\text{label}} - I_{\text{rec}}| \quad (20)$$

式中: I_{label} 和 I_{rec} 分别代表了目标图像和复原图像在 i 处的像素值, n 为整张图像的像素点个数。

相比于 L_2 损失, L_1 损失对离群值更稳健, 即数据集中存在的一些噪声点或异常值对损失的影响较小。并且 L_1 损失更倾向于保留细节, 在训练过程中易于优化, 训练稳定, 因此 L_1 损失更适合于重构图像的任务^[21]。

同时, 为了增强生成图像对于局部结构和细节的表现能力, 提高原始图像与生成图像在亮度, 对比度和结构上的相似性, 本文也使用结构相似度 SSIM^[19] 作为损失函数的一部分。

$$S = \frac{(2\mu_{\text{rec}}\mu_{\text{label}} + C_1)(2\sigma_{\text{relabel}} + C_2)}{(\mu_{\text{rec}}^2 + \mu_{\text{label}}^2 + C_1)(\sigma_{\text{rec}}^2 + \sigma_{\text{label}}^2 + C_2)} \quad (21)$$

式中: S 代表结构相似度, μ_{rec} 和 μ_{label} 分别代表复原图像和目标图像的均值, σ_{rec} 和 σ_{label} 分别表示复原图像和目标图像的标准差, σ_{relabel} 表示两幅图像的协方差, C_1 和 C_2 是两个常数, 用于避免分母为 0 的情况。

$$L_{\text{SSIM}} = 1 - S \quad (22)$$

像素级别的差异有时并不一定能够反映人眼对图像质量的感知差异, 因此引入了感知路径损失, 其通过将图像转换为特征向量表示, 并比较它们在特征空间中的相似度, 可以更好地捕捉人眼对图像质量的感知, 文献^[22]的实验结果表明感知路径损失可以提高生成图像对于细节的表现能力和网络风格转换的能力。

本文中的感知路径损失是基于在 ImageNet 数据集上预训练的 VGG-19 网络来计算的, 其具体表达式为

$$L_{\text{per}} = \sum_{m=1}^H \sum_{n=1}^W |\varphi_j(I_{\text{rec}})(m, n) - \varphi_j(I_{\text{label}})(m, n)| \quad (23)$$

式中: φ_j 表示 VGG-19 中的第 j 个卷积层, H 、 W 分

别代表图像的高和宽, m 、 n 分别代表对应的像素点, 本文使用特征提取层的输出之间的差值来衡量重构图像与真实图像之间的差值。

最终的整体损失函数为

$$L_f = \lambda_1 L_1 + \lambda_{\text{SSIM}} L_{\text{SSIM}} + \lambda_{\text{per}} L_{\text{per}} \quad (24)$$

式中: λ 用于平衡不同损失之间的范围, 最终设置 $\lambda_1 = 1$, $\lambda_{\text{SSIM}} = 1.1$, $\lambda_{\text{per}} = 0.1$ 。

3 实验结果与分析

本文使用 Adam 优化器来训练网络, 设置其学习率为 0.0002, β_1 为 0.9, β_2 为 0.999。设置学习率每 10 个 epoch 缩小到原来的 1/5, 最终学习率固定在 0.000001, 设定 batch size 为 8, 将图片随机裁剪为 256, 随机使用水平翻转, 垂直翻转等数据增强手段。

3.1 数据集

由于水下环境的复杂性和特殊性导致获取真实水下图像数据并进行标注是十分困难和昂贵的, 因此现有的真实水下图像数据集中图像数量有限, 成像条件单一。综上所述, 本文使用文献^[23–24]中的合成数据集进行模型预训练, 并在 Heron Island Coral Reef 数据集 (HICRD)^[25] 上进行迁移训练, 以下对数据集进行详细介绍。

3.1.1 合成数据集

如表 1 所示, 文献^[26]中给出了不同成像条件下的衰减系数, 其中 I、IA、IB、II 和 III 是开阔水域, 1、3、5、7 和 9 是沿海水域, 1 代表最干净, 9 代表最浑浊。

表 1 不同成像条件下的衰减系数

Table 1 Attenuation coefficient under different imaging conditions

颜色	I	IA	IB	II	III	1	3	5	7	9
蓝	0.982	0.975	0.968	0.940	0.890	0.875	0.80	0.670	0.50	0.290
绿	0.961	0.955	0.950	0.925	0.885	0.885	0.820	0.730	0.610	0.460
红	0.805	0.804	0.830	0.800	0.750	0.750	0.710	0.670	0.620	0.550

本文使用表 1 中的衰减系数, IMF 图像公式和 NYU-V2RGB-D^[27] 数据集, 按照文献^[23–24]中所使用的方法合成了一个水下图像数据集。其中 NYU-V2RGB-D 数据集作为清晰图像的来源。首先随机生成深度系数 $d(x)$, 其中 x 代表位置坐标, 然后使用生成的深度系数与数据集中的深度图像元素相乘, 得到合成水下图像所需的深度图。最后我们生成背景光强度 A_c , 得到所有参数后送入 1.1 节中的水下图像生成公式 (式 (2)) 中, 得到合成的水下图像。

为每种水质类型生成 5 张合成图像, 这样每

张清晰图像都会有 50 张与之对应的水下图像。

3.1.2 真实数据集

Heron Island Coral Reef 数据集 (HICRD)^[25] 包含了来自 8 个不同场地的原始水下图像, 每个场地都有详细的元数据, 包括水参数 (漫反射衰减)、最大潜水深度和相机型号等信息。其中 6 个场地有详细的衰减系数。根据原始图像的深度信息和物体与相机之间的距离, 具有相似深度、恒定距离和良好视觉质量的图像被标记为高质量。文献^[25]中精确计算了未衰减的原始图像, 并手工剔除了一些效果不尽如人意的图像, 最终得到了

一个包含 6003 张低质量水下图像、3673 张高质量水下图像和 2000 张未衰减的原始图像的数据集。

3.2 合成数据集输出结果

如图 8 所示, 本文使用 UWCNN^[23] 和 UIE-DAL^[24] 作为对比对象, UWCNN 使用合成的数据集进行训练, 针对不同的水质环境, 训练不同模型, 每次使用时需先根据水质类型选取对应的模型参数, 由于类型 I、IA、IB 的衰减系数十分接近, 为了简化训练过程, UWCNN 合并了类型 I、IA、IB 因此共 8 个模型; UIE-DAL 使用了对抗网络的思想, 使用编码器获取剔除了水质环境因素的特征向量, 使用该特征向量直接恢复目标图

像, 是一种通用模型。从对比图可知, UWCNN 所恢复的图像在细节方面有所不足, 但是其色彩表现稍好; UIE-DAL 会在图像中添加伪影, 使生成的图像偏红; 本文模型在所有的水质类型中均可以做到对水下图像的恢复, 特别是在类型 5、7、9 这样的极端环境中, 其恢复效果要远好于 UWCNN 和 UIE-DAL, 但是在这样的极端环境当中, 本文生成的图像与目标图像仍有较大差距, 恢复图像中色彩信息表现不足。这是因为本文采用了空间信息和色彩信息相分离的生成结构, 在极端环境下其色彩信息很少, 因此恢复的图像总是缺乏色彩的。



图 8 合成数据集重构效果

Fig. 8 Reconstruction effect of composite dataset

通过对比类型 7、II、III, 可以发现经本文模型处理后的图像显著地减少了黄色和蓝色色偏。

相比于 UWCNN^[23] 和 UIE-DAL^[24], 本文模型

所生成的图像中伪影和色偏显著减少, 并且本文的模型可以很好的保留图像中的细节部分, 如在类型 1、3 中仍可看清床上的物品细节。

3.3 真实数据集输出结果

如图 9 所示, 使用 HICRD 来测试本文的模型, 图 9(b) 是由 UIE-DAL 的预训练模型生成, 图 9(c) 是直接使用本文模型在合成数据集上训练的权重, 图 9(d) 是本文模型在 HICRD 上进行迁移训练的结果。通过图 9 可以看出, 本文模型在真实数据集上的泛化性显然好于 UIE-DAL, 尽管使用

预训练模型时, 大部分情况下, 输出图像与目标图像有明显差异, 输出图像的亮度显然高于目标图像, 但这有可能是目标图像的不完全恢复所造成的, 因为在第 2 行中, 显然使用合成数据集训练的模型的输出结果更好, 即更倾向于 GroudTruth。这也从侧面证实了使用合成数据集进行训练的可靠性。

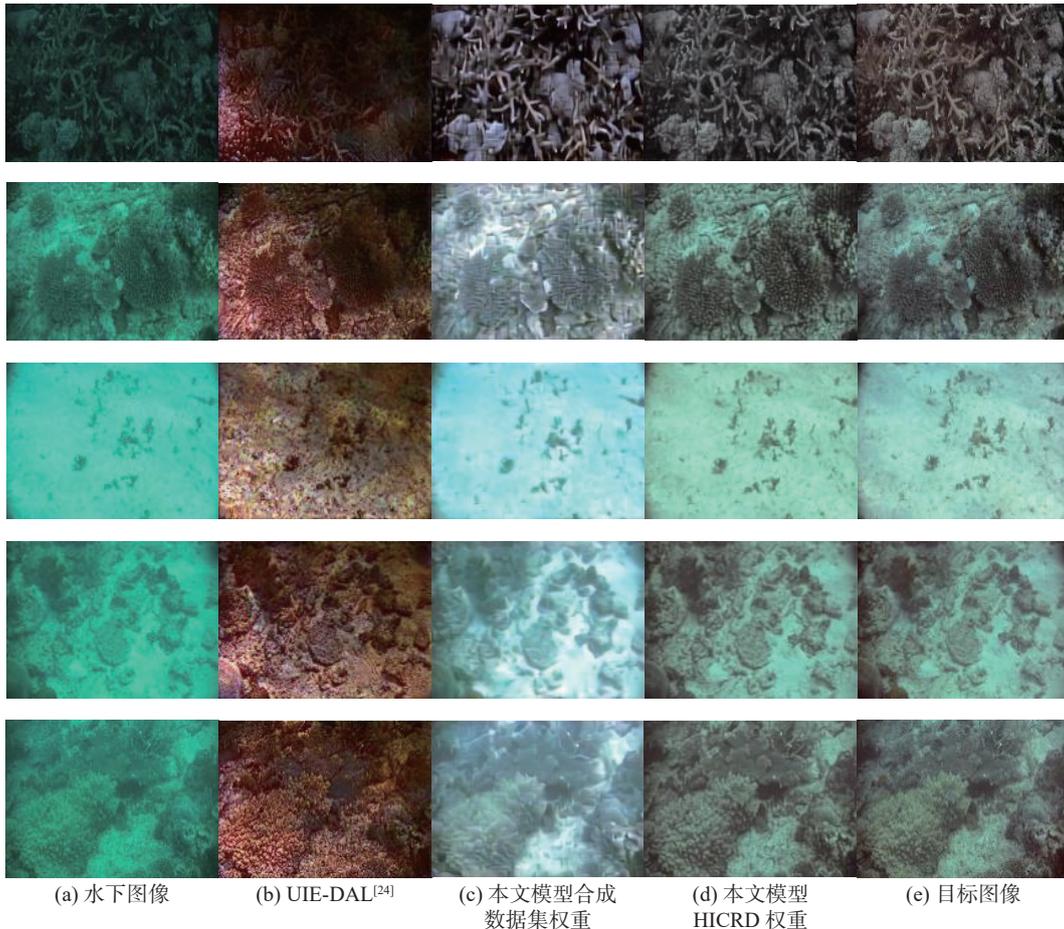


图 9 HICRD 重构效果
Fig. 9 Reconstruction effect of HICRD

通过生成图像与原图像之间的对比, 发现生成图像的绿色色偏明显减少。不论使用何种模型, 恢复图像的纹理细节表现显然好于水下图像, 这证明了本文模型对于水下图像具有明显的增强效果, 具有良好的泛化性, 同时由于本文模型具有通用性, 当在对应数据集上进行训练时, 可以快速地逼近目标图像。

3.4 定量分析指标

使用 SSIM 和 PSNR 进行不同模型的定量对比。SSIM 是一个评估图像质量的指标, 它考虑图像的结构信息和亮度信息, 更能反映出人眼对图像细节和结构的敏感度, 更接近于人类主观感受, 其值范围从 0 到 1, 值越大, 表示两幅图像越

相似, 具体公式见式 (21)。

PSNR 是最常用的评估图像质量的指标之一, 表示生成图像与目标图像之间的差异大小, 值越大, 表示图像保真度、图像去噪的效果越好。

3.5 定量分析

表 2 给出了本文模型与当前主流模型在合成数据集上训练后的定量表现。从表 2 可以看出, 随着成像条件变差, 模型恢复效果也逐渐变差, 在大多数的成像条件下, 本文模型可以取得超越主流模型的性能表现, 往往具有更高的 SSIM 值和 PSNR 值, 在类型 1 成像条件下, 尽管没有达到最大的 SSIM 值, 但是其与 UWCNN 模型的差距非常小。

表 2 不同模型的定量分析
Table 2 Quantitative analysis of different models

指标	类型	RED ^[5]	UDCP ^[28]	ODM ^[7]	UIBLA ^[29]	UWCNN ^[23]	UIE-DAL ^[24]	本文
SSIM	1	0.7406	0.7629	0.7240	0.6957	0.8558	0.9126	0.9484
	3	0.6639	0.6614	0.6765	0.5765	0.7951	0.9126	0.9218
	5	0.5934	0.4269	0.6441	0.4748	0.7266	0.8972	0.8817
	7	0.5089	0.2628	0.5632	0.3052	0.6070	0.7734	0.7850
	9	0.3192	0.1624	0.4178	0.2202	0.4920	0.6232	0.7122
	I	0.8816	0.8264	0.8172	0.7449	0.9376	0.9129	0.9361
	II	0.8837	0.8387	0.8251	0.8017	0.9236	0.9164	0.9637
	III	0.7911	0.7587	0.7546	0.7655	0.8795	0.9164	0.9460
	PSNR	1	15.596	15.757	16.085	15.079	21.790	26.4488
3		12.789	14.474	14.282	13.442	20.251	26.4488	26.3053
5		11.123	10.862	14.123	12.611	17.517	23.6697	23.2338
7		9.991	9.467	12.266	10.753	14.219	20.5793	21.5283
9		11.620	9.317	9.302	10.090	13.232	17.6551	20.0403
I		19.545	18.816	18.095	17.488	25.927	23.1015	28.0644
II		20.791	17.204	17.610	18.064	24.817	26.1602	28.6587
III		16.690	14.924	16.710	17.100	22.633	26.1602	26.9307

3.6 高级任务表现

水下图像处理是为了更好地执行后续的高级视觉任务,例如显著性目标检测。显著性目标检测(saliency object detection)是指在一张图片中自动识别出最显著、最引人注目的区域,即显著性区域,通常被认为是人眼在图片中第一时间注意到的区域。

图 10 给出了使用预训练模型 BASNet^[30] 在水下图像和复原图像上进行显著性目标检测的结果。从左至右分别为水下图像、复原图像、水下图像显著目标检测结果、复原图像显著目标检测结果。



图 10 重构图像显著目标检测表现

Fig. 10 Performance of significant target detection in reconstructed images

通过分析图 10 可以发现,复原图像显著目标检测效果远好于水下图像显著目标检测,例如在第 1 行中,使用复原图像可以更好地检测出水下蛙人的腿部和脚部细节;在第 2 行中,使用复原图像可以更好地检测出水下雕像的轮廓,并且可以检测出雕像手中的一些细节;在第 3 行中,原图像无法检测出的右下角石块在复原图像中可以被检测出;在第 4 行中,与原始水下图像相比复原图像可以检测出更多的人形雕塑。

4 结束语

本文提出一种将物理先验知识和深度神经网络相结合的水下图像复原模型。受暗通道先验中使用 RGB 色彩信息计算传输率的启发,该模型分别提取水下图像的空间信息和色彩信息,再通过调制操作将色彩信息与空间信息融合以推断水下传输率,最后通过混合注意力机制将基于水下光学成像模型和调制卷积的复原特征图融合,生成最终的恢复图像。实验的结果表明,使用平均绝对误差损失,结构相似度损失和感知路径损失可以有效训练本文模型。本文模型在可以极大地改善水下图像的成像效果,减少图像中的模糊、色偏等问题,提高目标检测的检测效果。由于采用了信息分离的结构,本文模型具有良好的泛化性,适用于多种水下成像条件,特别是在极端的水下环境当中,本文模型仍可以恢复图像的部分细节。

参考文献:

[1] IQBAL K, ODETAYO M, JAMES A, et al. Enhancing

- the low quality images using unsupervised colour correction method[C]//2010 IEEE International Conference on Systems, Man and Cybernetics. Istanbul: IEEE, 2010: 1703–1709.
- [2] ABDUL GHANI A S, MAT ISA N A. Underwater image quality enhancement through integrated color model with Rayleigh distribution[J]. *Applied soft computing*, 2015, 27: 219–230.
- [3] 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.
- [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. Providence: IEEE, 2012: 81–88.
- [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] 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.
- [7] LI Chongyi, GUO Jichang, CONG Runmin, 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] LIU Ze, LIN Yutong, CAO Yue, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2022: 9992–10002.
- [9] LIU Zhuang, MAO Hanzhi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11966–11976.
- [10] LI Jie, 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.
- [11] 李微, 毕晓君. 改进 U-Net 网络的水下图像增强 [J]. *应用科技*, 2021, 48(3): 34–40.
LI Wei, BI Xiaojun. Underwater image enhancement based on improved U-Net model[J]. *Applied science and technology*, 2021, 48(3): 34–40.
- [12] LI Chongyi, GUO Jichang, GUO Chunle. Emerging from water: underwater image color correction based on weakly supervised color transfer[J]. *IEEE signal processing letters*, 2018, 25(3): 323–327.
- [13] PENG Y T, CAO Keming, COSMAN P C. Generalization of the dark channel prior for single image restoration[J]. *IEEE transactions on image processing*, 2018, 27(6): 2856–2868.
- [14] JAFFE J S. Computer modeling and the design of optimal underwater imaging systems[J]. *IEEE journal of oceanic engineering*, 1990, 15(2): 101–111.
- [15] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132–7141.
- [16] DONG Chao, LOY C C, HE Kaiming, et al. Image super-resolution using deep convolutional networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(2): 295–307.
- [17] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11531–11539.
- [18] KARRAS T, LAINE S, AITTALA M, et al. Analyzing and improving the image quality of StyleGAN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 8107–8116.
- [19] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE transactions on image processing: a publication of the IEEE signal processing society*, 2004, 13(4): 600–612.
- [20] KARRAS T, AILA T LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation[EB/OL]. (2017–10–27)[2023–04–18]. <https://arxiv.org/abs/1710.10196>.
- [21] ARORA S, COHEN N, HAZAN E. On the optimization of deep networks: implicit acceleration by overparameterization[EB/OL]. (2018–02–19)[2023–04–18]. <https://arxiv.org/abs/1802.06509>.
- [22] JUSTIN J, ALEXANDRE A, LI Feifei. Perceptual losses for real-time style transfer and super-resolution[C]//Computer Vision–ECCV 2016. Amsterdam: Springer International Publishing, 2016: 694–711.
- [23] LI Chongyi, ANWAR S, PORIKLI F. Underwater scene prior inspired deep underwater image and video enhancement[J]. *Pattern recognition*, 2020, 98: 107038.
- [24] UPLAVIKAR P, WU Zhenyu, WANG Zhangyang. All-in-one underwater image enhancement using domain-adversarial learning[EB/OL]. (2019–05–30)[2023–04–18]. <https://arxiv.org/abs/1905.13342>.

- [25] HAN Junlin, 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] JERLOV N G. *Marine optics*[M]. [S. l.]: Elsevier, 1976.
- [27] NATHAN S, DEREK H, PUSHMEET K, et al. Indoor segmentation and support inference from RGBD images[C]// *Computer Vision—ECCV 2012*. Florence: Springer International Publishing, 2012: 746–760.
- [28] LI Chongyi, GUO Jichang, CHEN Sanji, et al. Underwater image restoration based on minimum information loss principle and optical properties of underwater imaging[C]// *2016 IEEE International Conference on Image Processing*. Phoenix: IEEE, 2016: 1993–1997.
- [29] PENG Yantsung, PAMELA C C. Underwater image restoration based on image blurriness and light absorption[J]. *IEEE transactions on image processing*, 2017, 26(4): 1579–1594.
- [30] QIN Xuebin, ZHANG Zichen, HUANG Chenyang, et al. BASNet: boundary-aware salient object detection[C]// *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2020: 7471–7481.

作者简介:



张心祎, 硕士研究生, 主要研究方向为深度学习与水下图像增强。



谭耀, 硕士研究生, 主要研究方向为深度学习与水下图像增强。



邢向磊, 教授, 博士生导师, 主要研究方向为计算机视觉, 模式识别与机器学习。以第一完成人获黑龙江省高等学校科学技术奖(自然科学类)一等奖, 《智能系统学报》优秀论文奖。

2023 首届全国人工智能应用场景创新挑战赛总决赛暨 全国人工智能应用场景创新峰会 2023 1st China's Innovation Challenge on Artificial Intelligence Application Scene Finals (CICAS2023) & National Artificial Intelligence Application Scenario Innovation Summit

中国人工智能学会、科技部新一代人工智能发展研究中心联合东莞市人民政府, 将于 2023 年 12 月 28—30 日在东莞松山湖科学城举办“‘场景驱动·数智强国’——2023 首届全国人工智能应用场景创新挑战赛总决赛暨全国人工智能应用场景创新峰会”。

本届大赛活动采用“1+1+8+6+N”模式, 即 1 场展览展示、1 场开幕式暨主论坛、8 场总决赛项目路演、6 场专题论坛和 N 场特色活动(项目招商对接会、未来科创场景+宴、走读松山湖等), 搭建全栈人工智能场景应用落地、创新创业、招商引资、成果转化的产业化合作交流平台。

时间地点:

举办时间: 2023 年 12 月 28—30 日

活动地点: 东莞松山湖科学城(松山湖凯悦酒店、松山湖科学城展览馆)

联系方式:

全国人工智能应用场景创新挑战赛组委会秘书处

联系人: 王老师(15726613955)、杨老师(15295651069)

邮 箱: zwh@cicas.cn