

专栏：〈水下图像处理〉

## 结合 Transformer 与生成对抗网络的水下图像增强算法

袁红春<sup>1</sup>, 张波<sup>1</sup>, 程心<sup>2</sup>

(1. 上海海洋大学 信息学院, 上海 201306; 2. 上海海洋大学 海洋生物资源与管理学院, 上海 201306)

**摘要:** 由于水下环境的多样性和光在水中受到的散射及选择性吸收作用, 采集到的水下图像通常会严重的质量退化问题, 如颜色偏差、清晰度低和亮度低等, 为解决以上问题, 本文提出了一种基于 Transformer 和生成对抗网络的水下图像增强算法。以生成对抗网络为基础架构, 结合编码解码结构、基于空间自注意力机制的全局特征建模 Transformer 模块和通道级多尺度特征融合 Transformer 模块构建了 TGAN (generative adversarial network with transformer) 网络增强模型, 重点关注水下图像衰减更严重的颜色通道和空间区域, 有效增强了图像细节并解决了颜色偏差问题。此外, 设计了一种结合 RGB 和 LAB 颜色空间的多项损失函数, 约束网络增强模型的对抗训练。实验结果表明, 与 CLAHE (contrast limited adaptive histogram equalization)、UDCP (underwater dark channel prior)、UWCNN (underwater based on convolutional neural network)、FUnIE-GAN (fast underwater image enhancement for improved visual perception) 等典型水下图像增强算法相比, 所提算法增强后的水下图像在清晰度、细节纹理和色彩表现等方面都有所提升, 客观评价指标如峰值信噪比、结构相似性和水下图像质量度量的平均值分别提升了 5.8%、1.8%和 3.6%, 有效地提升了水下图像的视觉感知效果。

**关键词:** 图像处理; 水下图像增强; Transformer; 生成对抗网络; 多项损失函数

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1001-8891(2024)09-0975-09

## Underwater Image Enhancement Algorithm Combining Transformer and Generative Adversarial Network

YUAN Hongchun<sup>1</sup>, ZHANG Bo<sup>1</sup>, CHENG Xin<sup>2</sup>

(1. School of Information, Shanghai Ocean University, Shanghai 201306, China;

2. School of Marine Living Resource Sciences and Management, Shanghai Ocean University, Shanghai 201306, China)

**Abstract:** Owing to the diversity of underwater environments and the scattering and selective absorption of light in water, acquired underwater images usually suffer from severe quality degradation problems, such as color deviation, low clarity, and low brightness. To solve these problems, an underwater image enhancement algorithm that combines a transformer and generative adversarial network is proposed. Based on the generative adversarial network, a generative adversarial network with transformer (TGAN) network enhancement model is constructed by combining the coding and decoding structure, global feature modeling transformer module based on the spatial self-attention mechanism, and channel-level multi-scale feature fusion transformer module. The model focuses on color and spatial channels with more serious underwater image attenuation. This effectively enhances the image details and solves the color-deviation problem. Additionally, a multinomial loss function, combining RGB and LAB color spaces, is designed to constrain the adversarial training of the network enhancement model. The experimental results demonstrate that when compared to typical underwater image enhancement algorithms, such as contrast-limited adaptive histogram equalization (CLAHE), underwater dark channel prior (UDCP), underwater based on convolutional neural network (UWCNN), and

收稿日期: 2023-04-09; 修订日期: 2023-05-17.

作者简介: 袁红春 (1971-), 男, 博士, 教授, 主要研究方向为智能信息处理。

通信作者: 张波 (1996-), 男, 硕士研究生, 主要研究方向为水下图像增强。E-mail: 806232525@qq.com。

基金项目: 国家自然科学基金 (41776142)。

fast underwater image enhancement for improved visual perception (FUnIE-GAN), the proposed algorithm can significantly improve the clarity, detail texture, and color performance of underwater images. Specifically, the average values of the objective evaluation metrics, including the peak signal-to-noise ratio, structural similarity index, and underwater image quality measure, improve by 5.8%, 1.8%, and 3.6%, respectively. The proposed algorithm effectively improves the visual perception of underwater images.

**Key words:** image processing, underwater image enhancement, Transformer, generative adversarial network, multinomial loss function

## 0 引言

建设海洋强国是实现中华民族伟大复兴的重大战略任务,有效实施利用海洋资源战略并掌握高效开发海洋资源技术的国家,必将在未来的经济和社会发展中占取先机。获取海洋信息是充分了解海洋的关键,作为海洋信息的载体,水下图像在海洋资源探索和水下作业任务方面发挥着重要作用。但是,由于水下成像环境的复杂和光在水中传播的散射效应<sup>[1]</sup>等因素,采集的水下图像通常会产生严重的退化问题,主要呈现为颜色偏差、对比度低和细节模糊等,严重限制了水下图像在海洋勘测、水下目标检测<sup>[2]</sup>和水下机器人<sup>[3]</sup>等实际场景中的应用。因此,亟须寻找一种有效的增强方法来解决水下图像的退化问题。

水下图像处理一般可以分为图像复原和图像增强<sup>[4]</sup>两大类。其中,基于非物理模型的增强方法主要是通过修改像素值来提高水下图像的视觉质量,忽略了水下图像的物理退化过程。例如,Hitam 等人<sup>[5]</sup>提出了一种直方图均衡算法用于处理水下图像,减少了噪声的产生同时提高了对比度。此外,Huang 等人<sup>[6]</sup>构建了一种基于自适应参数采集的全局直方图拉伸方法(relative global histogram stretching, RGHS),保留了图像细节的同时有效消除了噪声的影响。基于物理模型的复原方法主要是对成像过程进行建模,通过反演水下物理成像模型获得高质量的水下图像。Paulo 等人<sup>[7]</sup>利用暗通道先验(dark channel prior, DCP)原理,设计了一种估计水下环境传输的水下 DCP 算法,克服了 DCP 在水下成像应用中的局限性。此外,Peng 等人<sup>[8]</sup>设计了一种基于图像模糊和光吸收的水下场景深度估计方法(image blurriness and light absorption, IBLA),在不同色调和光照条件下都能很好地恢复水下图像。基于物理模型的复原方法考虑了造成水下图像退化的原因,使图像增强效果得到进一步提高,但该方法需要物理先验知识,对模型的参数估计存在一定的偏差,恢复结果易产生失真问题。基于数据驱动的增强方

法被认为是水下图像增强领域的深度学习技术,在水下图像增强方面表现出卓越的性能。例如,Li 等人<sup>[9]</sup>将水下成像模型与水下场景的光学特性相结合,提出了一种利用卷积神经网络的水下图像增强模型(UWCNN),能够直接构建清晰的水下图像。Islam 等人<sup>[3]</sup>基于条件生成对抗网络设计了一种实时增强模型(FUnIE-GAN),根据图像的全局内容、色彩表现和局部纹理评估图像质量。Salma 等人<sup>[10]</sup>构建了含有两个生成器的条件生成对抗网络算法(generative adversarial network with dual generator dewatering, DGD-cGAN),分别用于去除色偏和雾化。尽管深度学习算法在水下图像增强领域已经取得了不错的成果,但是未能考虑到水下图像在不同颜色通道和空间区域衰减不一致的问题,这限制了基于数据驱动方法性能的进一步提高。

近几年,Transformer<sup>[11]</sup>因其强大的特征提取能力在计算机视觉领域得到了越来越多的关注,且自注意力机制可以有效地对全局特征进行建模并融合多尺度特征,提升算法对图像细节的重建能力。针对现有方法所存在的问题,本文提出了一种结合 Transformer 与生成对抗网络(generative adversarial network, GAN)的水下图像增强算法(TGAN)。该算法以生成对抗网络为基础框架,生成网络中引入了基于空间自注意力机制的全局特征建模 Transformer 模块(global feature modeling transformer, GFMT)和多尺度特征融合 Transformer 模块(multiscale feature fusion transformer, MSFFT),加强了算法对衰减更严重的颜色通道和空间区域的关注,可有效解决现有方法增强后的图像存在的局部色偏、清晰度低、亮度低等问题。

## 1 相关理论

### 1.1 生成对抗网络

GAN 包括生成网络和判别网络两个部分,通过对抗训练在生成网络和判别网络之间建立一个数学函数。生成网络采用参数化概率生成模型,接收随机噪声并通过逆变换采样,有效地拟合输入训练集

数据的实际分布情况，进而生成符合输入训练集数据特点的近似信息。判别网络是一种分类模型，对于输入的两种数据，其数据处理流程是将真实的训练数据记为真标签，生成网络生成的虚假数据记为假标签，通过计算损失值，对真实数据的判别趋近于真，对生成数据的判别趋近于假，并反馈给生成网络，不断训练模型，更新网络的参数。GAN 的目标函数可以用式(1)表示：

$$\min_G \max_D V(D, G) = E_y [\log D(y)] + E_x [\log (1 - D(G(x)))] \quad (1)$$

式中： $x$  表示输入生成网络的随机噪声； $y$  表示真实的训练数据； $G(x)$ 表示生成网络生成的虚假数据； $E$  表示数学期望；判别网络  $D$  旨在最大化  $V(D, G)$ ；生成网络  $G$  旨在最小化  $V(D, G)$ 。通过这种对抗性的学

习方式，GAN 可以生成高质量的、与真实数据相似的虚假数据。GAN 的整体网络结构如图 1 所示。

## 1.2 Transformer 模型

Transformer 是一种基于自注意力机制（self-attention mechanism）的深度学习模型，由编码器和解码器组成。编码器用于将输入序列编码成一个高维的特征向量，而解码器则使用编码器的输出来生成目标序列。Transformer 的核心算法是大量的多头自注意力模块，能够直接提取数据样本的整体特征，从而更好地处理长序列和对序列的全局信息进行建模。同时，Transformer 也使用了残差连接和层标准化技术来加速训练，并且可以通过并行计算来进一步提高效率。单层 Transformer 模型的结构如图 2 所示。

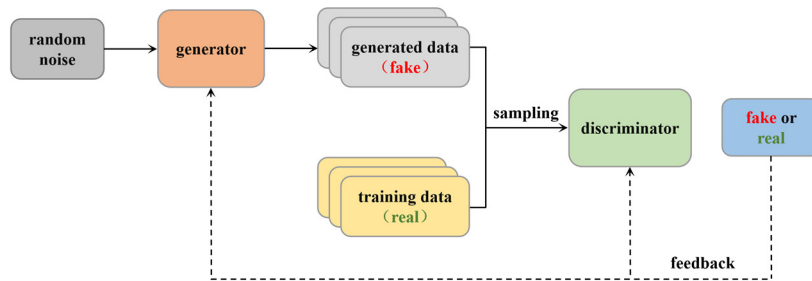


图 1 生成对抗网络结构

Fig.1 Generative adversarial network structure

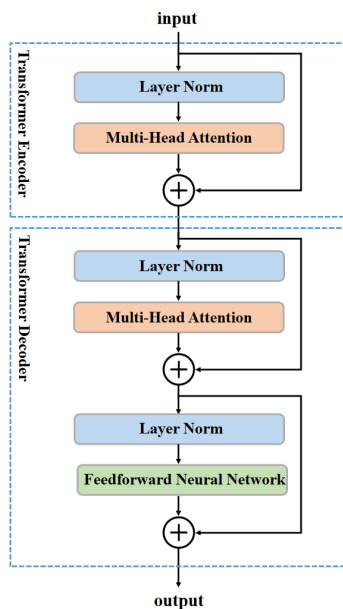


图 2 单层 Transformer 模型

Fig.2 A single-layer Transformer model

## 2 本文算法

### 2.1 TGAN 网络增强模型

本文提出的水下图像增强网络模型 TGAN 由生成网络和判别网络两部分组成。生成网络的目的是

学习清晰的水下图像的概率分布，并且通过反向传播算法训练生成网络中的权重，最终生成质量更好的水下图像，其由 4 部分组成，分别是编码器、GFMT 模块、MSFFT 模块和解码器。编码器由连续的卷积层和下采样操作组成，它负责提取输入图像的特征。解码器由连续的反卷积层和上采样操作组成，它负责将编码器提取的特征图映射回原始图像的尺寸，并还原更精细的图像细节。对于输入生成网络中的原始图像，首先经过 5 次下采样操作，然后作为 GFMT 模块的输入，再经过 GFMT 模块特征重映射后，输出直接发送到第一个上采样模块。同时 4 个下采样模块的输出作为 MSFFT 模块的输入，解码器中 4 个不同尺度的上采样模块将接收来自 MSFFT 模块的 4 个输出。

判别网络旨在区分出生成网络输出的图像和参考图像，其结构类似于 PatchGAN<sup>[12]</sup>网络，包含 5 个下采样模块，输出为  $16 \times 16 \times 1$  大小的补丁块，该输出表示两个输入图像之间的平均差异程度。当某个对应区域存在较大的差异时，输出中对应该位置的像素值会越接近于 0，反之则接近于 1，每个补丁像素点对应于输入图像的一个局部感受野。这种方法的优点在于，不仅减少了参数数量和计算量，还能够对图像进行局部细节增强。TGAN 的网络结构及数据处理流程如图 3 所示。

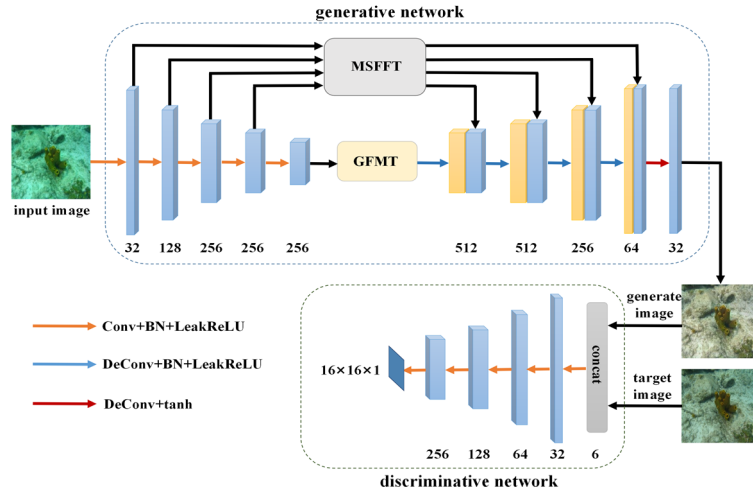


图3 TGAN网络结构

Fig.3 TGAN network structure

## 2.2 GFMT 全局特征建模模块

针对水下图像在不同空间区域退化不均匀的问题,本文设计了一种基于空间自注意力机制的全局特征建模 Transformer 模块 (GFMT) 来替换生成网络中原有的瓶颈层 (bottleneck) [13]。GFMT 模块能够准确地建模水下图像的全局特征,加强网络对衰减较为严重的空间区域的关注,从而使增强后的水下图像细节更清晰,内容分布更均匀。GFMT 模块的详细结构如图4所示。

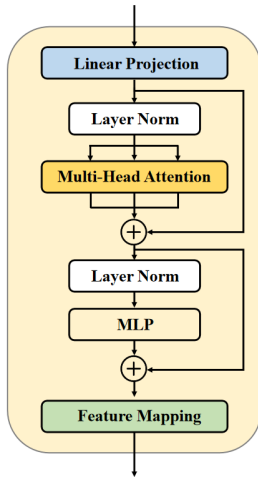


图4 GFMT 模块结构

Fig.4 GFMT module structure

GFMT 模块的数据处理流程如下,对于输入 GFMT 的特征图  $F_{in}$ ,首先使用线性投影将特征图拉伸成特征序列  $S_{in}$ ,可表示为:

$$S_{in} = W * F_{in} \quad (2)$$

式中:  $W$  表示权重矩阵;  $W * F_{in}$  表示线性投影操作。

然后将特征序列  $S_{in}$  输入到 Transformer 层中,

Transformer 层包含多头注意力模块 (Multi-Head Attention, MHA) 和前馈网络 (Feed Forward Networks, FFN), FFN 包括一个标准化层和一个全连接层 (多层感知机, Multilayer Perceptron, MLP), Transformer 层的输出可由下式计算:

$$S' = MHA(LN(S_{in})) + S_{in} \quad (3)$$

$$S = FFN(LN(S')) + S' \quad (4)$$

式中:  $LN$  (layer normalization) 表示层标准化;  $S$  表示 Transformer 层的输出序列。经过特征重映射后 GFMT 模块的输出与输入大小保持一致。

## 2.3 MSFFT 特征提取模块

为解决水下图像在不同颜色通道衰减不一致的问题 [14], 本文设计了一种多尺度特征融合 Transformer 模块 (MSFFT), MSFFT 模块利用通道自注意力机制对编码器输出的特征进行通道级多尺度特征融合,并将融合结果传递给解码器,能够加强网络对衰减更严重的颜色通道的关注,从而实现色彩均匀的水下图像增强,有效解决水下图像存在的颜色偏差问题。MSFFT 模块的详细结构如图5所示。

MSFFT 模块的输入是不同尺度的特征图  $F_i (i=1,2,3,4)$ , 其整体的数据处理流程是首先使用线性投影操作将特征图拉伸成特征序列  $S_i (i=1,2,3,4)$ , 然后计算得到6个矩阵  $Q_i (i=1,2,3,4)$ 、 $K$  和  $V$ :

$$Q_i = S_i W_Q, K = S W_K, V = S W_V \quad (5)$$

式中:  $W_Q (i=1,2,3,4)$ ,  $W_K$  和  $W_V$  表示可学习的权重矩阵;  $S$  是由  $S_i (i=1,2,3,4)$  在通道维度上叠加生成的。

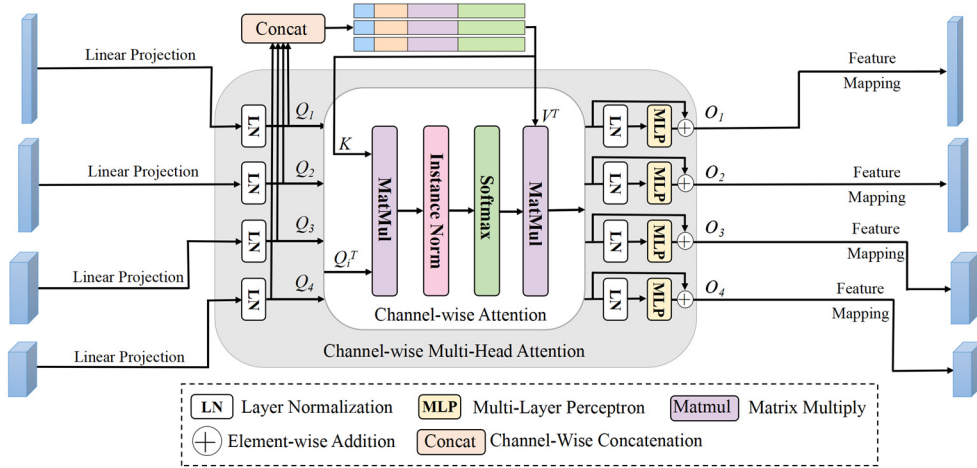


图 5 MSFFT 模块的详细结构

Fig.5 Detailed structure of MSFFT module

计算得到的 6 个矩阵  $Q_i (i=1,2,3,4)$ 、 $K$  和  $V$  经过通道注意力输出  $CA_i$ ，可表示为：

$$CA_i = \text{SoftMax} \left( \text{IN} \left( \frac{Q_i^T K}{\sqrt{C}} \right) \right) V^T \quad (6)$$

式中：IN (instance normalization) 表示实例归一化操作<sup>[15]</sup>； $C$  表示 4 个尺度上通道数之和； $Q_i^T$  和  $V^T$  表示  $Q_i$  和  $V$  的转置矩阵。这个注意力操作是沿着通道轴而不是经典的补丁轴<sup>[16]</sup>执行，引导网络关注图像质量下降更严重的颜色通道。

$CA_i$  经过前馈网络传播得到输出  $Q_i (i=1,2,3,4)$ ，也就是通道级多头注意力模块 (CMHA) 的输出可以表示为：

$$Q_i = CA_i + \text{MLP}(\text{LN}(CA_i)) \quad (7)$$

式中：MLP 代表多层感知机，公式(7)的操作需要依次执行 4 次。

最后，对 4 个特征序列  $Q_i (i=1,2,3,4)$  执行特征映射，将它们重组为特征图作为解码器上采样模块的输入。

## 2.4 损失函数

为了利用 LAB 颜色空间更宽的色域表示范围以及对亮度和对比度进行更准确的描述，本文设计了一种结合 RGB 和 LAB 颜色空间的双颜色空间损失函数来训练增强网络模型。首先，将 RGB 颜色空间的图像转换到 LAB 颜色空间，可表示为：

$$L^{G(x)}, A^{G(x)}, B^{G(x)} = \text{RGB2LAB}(G(x)) \quad (8)$$

$$L^y, A^y, B^y = \text{RGB2LAB}(G(y)) \quad (9)$$

式中： $x$ 、 $y$  和  $G(x)$  分别表示原始的水下图像、原始图像对应的参考图像和生成网络输出的图像。

LAB 颜色空间中的损失函数可表示为公式(10)：

$$\text{Loss}_{\text{LAB}}(G(x), y) = E_{x,y} \left[ \left( L^y - L^{G(x)} \right)^2 - \sum_{i=1}^n Q(A_i^y) \log(Q(A_i^{G(x)})) - \sum_{i=1}^n Q(B_i^y) \log(Q(B_i^{G(x)})) \right] \quad (10)$$

式中： $Q$  表示量化操作，用于将连续值映射到离散值； $E$  表示数学期望； $L$ 、 $A$ 、 $B$  分别表示 LAB 颜色空间中图像的亮度分量、从绿色到红色的分量和从蓝色到黄色的分量。

生成网络的 3 个损失函数分别为 RGB 颜色空间下的  $L_2$  损失函数（表示为  $\text{Loss}_{\text{RGB}}$ ），色彩感知损失  $\text{Loss}_{\text{per}}$ <sup>[17]</sup>和  $\text{Loss}_{\text{LAB}}$ ，可表示为：

$$\text{Loss}_{\text{RGB}} = E_{x,y} \left[ \|y - G(x)\|_1 \right] \quad (11)$$

$$\text{Loss}_{\text{per}} = \sqrt{\frac{(512 + r_{\text{mean}})r^2}{256} + 4g^2 + \frac{(767 - r_{\text{mean}})b^2}{256}} \quad (12)$$

式中： $r$ 、 $g$ 、 $b$  分别表示生成图像与参考图像在红、绿、蓝颜色通道上的差值； $r_{\text{mean}}$  表示生成图像与参考图像红色通道的平均值。

此外，引入标准的 GAN 损失函数来表示生成图像和参考图像之间的差异，公式如下：

$$L_{\text{GAN}}(G, D) = E_y[\log D(y)] + E_x[\log(1 - D(G(x)))] \quad (13)$$

式中： $D$  表示判别网络，旨在最大化  $L_{\text{GAN}}(G, D)$ ，以准确地区分生成的图像与参考图像； $G$  表示生成网络，作用是将生成图像和参考图像之间的差异最小化。

整体的损失函数可表示为：

$$L = \arg \min_G \max_D L_{\text{GAN}}(G, D) + \alpha \text{Loss}_{\text{LAB}}(G(x), y) + \beta \text{Loss}_{\text{RGB}}(G(x), y) + \gamma \text{Loss}_{\text{per}}(G(x), y) \quad (14)$$

式中： $\alpha$ 、 $\beta$ 、 $\gamma$  为超参数，进行大量实验对比后分别



设置为 0.2, 0.2, 0.6。

### 3 实验结果与分析

#### 3.1 实验数据

现有的水下图像数据集或多或少存在着图像数量、水下场景少,甚至不是真实的水下场景等缺点,限制了基于数据驱动的水下图像增强方法的性能。因此本文构建了包含 4900 张图像的数据集 UITD,分为训练集和测试集两个部分。训练集包含 2400 对场景丰富的水下图像和对应的参考图像,其中 1000 对来自 LSUI 数据集<sup>[18]</sup>,800 对来自 UIEB 数据集<sup>[19]</sup>,600 对来自 EUVP 数据集<sup>[3]</sup>。测试集由两部分组成,第一部分 Test-1 是从 EUVP 数据集中挑选的 60 张有参考的水下图像,第二部分 Test-2 是从 UIEB 数据集中挑选的 40 张无参考的水下图像。

#### 3.2 实验设置

本实验在 Ubuntu 操作系统下,通过 PyTorch 深度学习框架实现。硬件配置如下:AMD 5950X CPU,64 GB 运行内存,NVIDIA GeForce RTX3090 (24 GB) GPU。

训练时,batchsize 设置为 6,所有输入图像调整到固定大小( $256 \times 256 \times 3$ ),像素值归一化到 $[0, 1]$ 区间。使用 Adam 优化算法对模型进行 800 个 epoch 训练,优化算法参数 $\beta_1$ 设置为 0.9, $\beta_2$ 设置为 0.999,对于前 500 个 epoch 和后 300 个 epoch,网络初始学习率分别设置为 0.0005 和 0.0002,学习率每 40 个 epoch 下降 20%。

#### 3.3 实验分析

本文进行了消融实验和对比实验,并在独立测试集上对实验结果进行了定性和定量的对比分析。对于带有参考图像的测试数据集,本文使用峰值信噪比(peak signal-to-noise ratio, PSNR)<sup>[20]</sup>和结构相似性(structural similarity, SSIM)<sup>[21]</sup>指标进行全参考评估,这两个指标反映了待评价图像与参考图像的接近程度,其中 PSNR 值越高代表图像内容越接近,SSIM 值越高代表结构和纹理越相似。对于无参考测试数据集中的图像,本文使用了无参考评价指标 UCIQE (underwater color image quality evaluation)<sup>[22]</sup>,UIQM (underwater image quality measure)<sup>[23]</sup>和 NIQE (natural image quality evaluator)<sup>[24]</sup>。UCIQE 利用色度、饱和度和对比度的线性组合进行定量评估,分别量化色偏、低清晰度和低对比度。UIQM 是水下图

像色度量、水下图像清晰度度量和水下图像对比度度量的线性组合。NIQE 利用在自然图像中观察到的信息对图像质量进行评估。其中,UIQM 和 UCIQE 值越高则图像质量越好,NIQE 值越小则图像越自然。

##### 3.3.1 消融实验

为了验证本文模型的 GFMT 模块,MSFFT 模块和 Loss<sub>LAB</sub> 损失项 3 个组件对于增强效果的影响,本文在 Test-1 数据集上进行了消融实验。对比的模型包括:不含 3 个组件的基准模型(benchmark model, BL),仅包含 GFMT 模块的模型(BL+GFMT),仅包含 MSFFT 模块的模型(BL+MSFFT),仅包含 Loss<sub>LAB</sub> 损失项的模型(BL+Loss<sub>LAB</sub>)以及完整模型(TGAN)。除对比项外,其余各方面均保持一致。不同模型在测试集 Test-1 上的定性对比如图 6 所示。

观察图 6 可知,BL+GFMT 增强结果比 BL 具有更少的噪声和伪影,这得益于 GFMT 模块能够关注水下图像衰减更严重的空间区域,但水下图像的颜色偏差问题仍然存在;由于 MSFFT 模块能够将不同尺度的特征融合之后进行增强,BL+MSFFT 的增强结果有效提高了水下图像的对比度和饱和度,但是图像的亮度方面表现不佳;BL+Loss<sub>LAB</sub> 的增强结果整体颜色更接近参考图像。虽然 BL+GFMT 和 BL+MSFFT 的增强结果分布均匀,但整体颜色不够准确,完整模型的增强结果具有最好的视觉质量。由此可知,所研究的 3 个组件在增强过程中都有其特定的功能性,它们的集成可以提高网络的整体性能。

同时,为客观评价各组件的功能性,对 Test-1 数据集进行定量分析,得到的结果如表 1。完整模型在 Test-1 数据集上取得了最好的定量性能,增强后的图像最接近参考图像,这证明了结合 GFMT 模块,MSFFT 模块和 Loss<sub>LAB</sub> 损失项 3 个组件对于增强效果的有效性。

##### 3.3.2 对比实验

为了进一步评估本文算法对水下图像增强的有效性,分别在 Test-1 和 Test-2 数据集上将本文算法与现有的 7 个经典的水下图像增强方法进行定性和定量的对比分析,对比方法包括基于非物理模型的增强方法(CLAHE<sup>[5]</sup>,RGHS<sup>[6]</sup>)、基于物理模型的复原方法(UDCP<sup>[7]</sup>,IBLA<sup>[8]</sup>)和基于数据驱动的增强方法(UWCNN<sup>[9]</sup>,FUnIE-GAN<sup>[10]</sup>,DGD-cGAN<sup>[11]</sup>)。

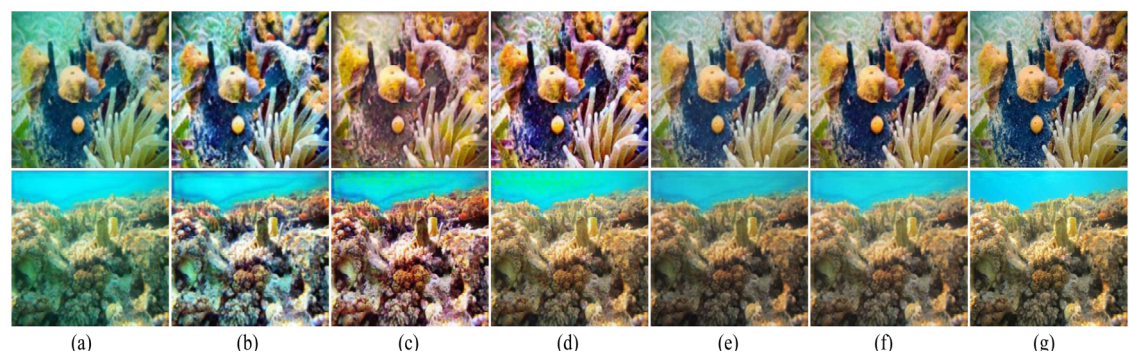


图 6 消融实验定性对比(a)水下图像; (b)BL; (c)BL+GFMT; (d)BL+MSFFT; (e)BL+LOSSLAB; (f)TGAN; (g)参考图像

Fig.6 Qualitative comparison of ablation experiments. (a)Underwater images; (b)BL; (c)BL+GFMT; (d)BL+MSFFT; (e)BL+LOSSLAB; (f)TGAN; (g)Reference images

表 1 在测试集 Test-1 上的消融实验结果

Table 1 Experimental results of ablation study on Test-1

Models	PSNR	SSIM
BL	19.2556	0.7014
BL+GFMT	21.6849	0.7635
BL+MSFFT	22.3719	0.7813
BL+LOSSLAB	21.4161	0.7281
TGAN	24.0546	0.8257

不同方法在测试集 Test-1 上的定性对比结果如图 7 所示，基于非物理模型的 CLAHE 方法增强过的图像会出现曝光现象，与参考图像在颜色和细节上存在一定差距。RGHS 方法处理深绿色图像时效果较差。基于物理模型的 UDCP 方法会加重图像的色偏现象，尤其是绿色和红色偏差。IBLA 方法能解决一定程度上的色偏问题，但整体增强效果欠佳。基于数据驱动的 UWCNN 方法处理后的图像呈现出黄色颜色偏差，且亮度偏暗。FUnIE-GAN 方法对深绿色图像增强效果较差，且图像的细节方面还有提升空

间。DGD-cGAN 方法增强后的图像亮度会得到一定提升，但绿色偏差问题依然存在。相比之下，本文所提算法在颜色校正、提升亮度、增强对比度和饱和度等方面均展示出极具竞争力的性能，并且有效地增强了图像细节，在部分图片上取得了比参考图像更好的视觉效果。

为客观验证本文算法的性能，采用 PSNR 和 SSIM 两种参考图像质量评价指标，将本文算法与上述其他方法做定量对比分析。两种指标的均值如表 2 所示。由表 2 可知，本文算法在 PSNR 和 SSIM 评价指标上的平均值为 24.0546 和 0.8257，均优于其他对比方法，分别比第二名提升了 5.8%和 1.8%，这证明了 GFMT 和 MSFFT 模块可以在捕捉全局特征的同时，有效加强对衰减严重的颜色通道和空间区域的关注，从而使本文算法在图像细节增强、校正色偏、提高对比度和提升亮度方面都有表现出很好的效果。

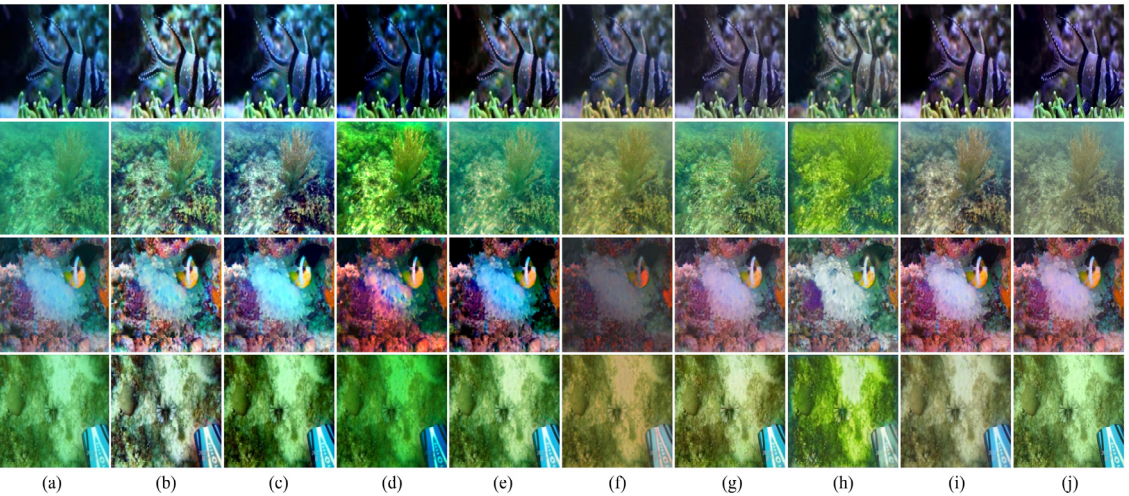


图 7 不同方法在测试集 Test-1 上的定性对比(a)水下图像; (b)CLAHE; (c)RGHS; (d)UDCP; (e)IBLA; (f)UWCNN; (g)FUnIE-GAN; (h)DGD-cGAN; (i)本文方法; (j)参考图像

Fig.7 Qualitative comparison of different methods on Test-1. (a)Underwater images; (b)CLAHE; (c)RGHS; (d)UDCP; (e)IBLA; (f)UWCNN; (g)FUnIE-GAN; (h)DGD-cGAN; (i)Our method; (j)Reference images



表 2 不同方法在测试集 Test-1 上的定量对比

Table 2 Quantitative comparison of different methods on Test-1

Methods	PSNR	SSIM
CLAHE	18.4342	0.7653
RGHS	18.2053	0.7672
UDCP	14.0555	0.5650
IBLA	19.9222	0.7487
UWCNN	18.1209	0.7420
FUnIE-GAN	22.7413	0.8112
DGD-cGAN	17.3954	0.6955
TGAN	24.0546	0.8257

为了更加全面和准确地评估本文算法的性能，在无参考测试集 Test-2 上，同样将本文算法与上述其他方法进行定性和定量对比分析。定性对比分析结

果如图 8。观察图 8 可以发现，CLAHE 方法不能完全消除色偏问题。RGHS 方法虽提升了一定亮度，但在颜色校正方面效果不佳。基于物理模型的 UDCP 和 IBLA 方法复原后的图像加重了色偏且亮度过低。UWCNN 方法增强后的图像颜色偏黄，且亮度偏低。FUnIE-GAN 方法处理后的图像偏暗，细节纹理不够突出。DGD-cGAN 方法处理后的图像整体颜色偏绿，增强效果较差。与上述方法相比，本文算法在对比度和清晰度提升，细节增强以及校正颜色偏差等方面展示出最令人满意的效果，增强后的图像纹理更加突出且视觉感受更好。

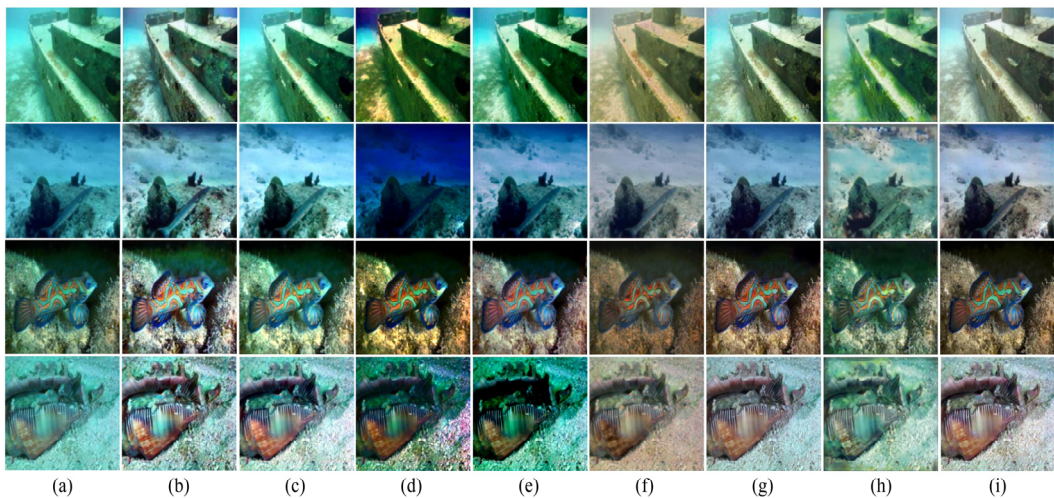


图 8 不同方法在测试集 Test-2 上的定性对比(a)水下图像; (b)CLAHE; (c)RGHS; (d)UDCP; (e)IBLA; (f)UWCNN; (g)FUnIE-GAN; (h)DGD-cGAN; (i)本文方法

Fig.8 Qualitative comparison of different methods on Test-2. (a)Underwater images; (b)CLAHE; (c)RGHS; (d)UDCP; (e)IBLA; (f)UWCNN; (g)FUnIE-GAN; (h)DGD-cGAN; (i)Our method

为客观地评价各方法在无参考测试集 Test-2 上的性能，本文选取 3 个非参考图像质量评价指标 UCIQE、UIQM 和 NIQE 对增强结果进行定量对比分析。表 3 展示了不同方法在测试集 Test-2 上的评价指标平均得分。从表 3 中可以看出，本文方法在 UCIQE 和 UIQM 评价指标上取得了最优的结果，NIQE 评价指标仅大于 IBLA 方法，这进一步说明了本文方法在图像细节增强、自然表现和视觉感知提升等方面的表现更为突出。

表 3 不同方法在测试集 Test-2 的定量对比

Table 3 Quantitative comparison of different methods on Test-2

Methods	UCIQE	UIQM	NIQE
CLAHE	0.4516	3.1570	6.5814
RGHS	0.4673	2.4674	6.4705
UDCP	0.4216	2.0992	5.7852
IBLA	0.4731	2.3331	5.7619
UWCNN	0.3508	3.0378	6.7935
FUnIE-GAN	0.4314	3.0997	6.2796
DGD-cGAN	0.3689	3.1810	7.2689
TGAN	0.4846	3.2963	5.7743

## 4 结论

水下图像增强是完成各种水下作业的重要支撑技术。针对水下图像存在的对比度低、清晰度差和色彩失真等退化问题，本文提出了一种基于 Transformer 和生成对抗网络的水下图像增强算法（TGAN）。生成网络中集成的 Transformer 模块可以解决现有方法中没有考虑到水下图像在不同颜色通道和空间区域衰减不一致的问题，其中 GFMT 模块能够准确获取水下图像的全局特征，加强网络对图像衰减较为严重区域的关注，从而使算法增强后的图像细节更加清晰。MSFFT 模块可以捕获多尺度特征，聚焦退化严重的颜色通道，实现色彩均匀的水下图像增强。此外，结合 RGB 和 LAB 颜色空间特征设计的多项损失函数提高了输出图像的亮度和对比度。实验表明，本文算法在评价指标 PSNR、SSIM、UCIQE 和 UIQM 上均取得了最优结果，在水



下图增强方面能够有效地校正色偏, 均匀地提高清晰度, 同时提升图像的整体视觉感知效果。但基于 Transformer 网络的训练通常需要较大的数据集, 未来的工作将尝试对模型进行改进, 以提升在小数据集上模型的训练效果。

### 参考文献:

- [1] YANG M, HU J T, LI C Y, et al. An in-depth survey of underwater image enhancement and restoration[J]. *IEEE Access*, 2019, 7: 123638-123657.
- [2] ANWAR S, LI C Y. Diving deeper into underwater image enhancement: a survey[J]. *Signal Processing: Image Communication*, 2020, 89: 115978.
- [3] Islam M J, XIA Y, Sattar J. Fast underwater image enhancement for improved visual perception[J]. *IEEE Robotics and Automation Letters*, 2020, 5: 3227-3234.
- [4] 晋玮佩, 郭继昌, 祁清. 基于条件生成对抗网络的水下图像增强[J]. *激光与光电子学进展*, 2020, 57(14): 141002.  
JIN W P, GUO J C, QI Q. Underwater image enhancement based on conditional generative adversarial network[J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141002.
- [5] Hitam M S, Awalludin E A, Yusoff W N J H W, et al. Mixture contrast limited adaptive histogram equalization for underwater image enhancement[C]//*International Conference on Computer Applications Technology (ICCAT)*, 2013: 1-5.
- [6] HUANG D M, WANG Y, SONG W, et al. Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition[C]//*24th International Conference on MultiMedia Modeling (MMM)*, 2018(10704): 453-465.
- [7] Drews Paulo, Nascimento E, Moraes F, et al. Transmission estimation in underwater single images[C]//*IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2013: 825-830.
- [8] 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.
- [9] LI C Y, Anwar S, Porikli F. Underwater scene prior inspired deep underwater image and video enhancement[J]. *Pattern Recognition*, 2020, 98: 107038-107038.
- [10] Gonzalez Sabbagh S, Robles Kelly A, Gao S. DGD-cGAN: a dual generator for image dewatering and restoration[J]. arXiv preprint arXiv:2211.10026, 2022.
- [11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 1049-5258.
- [12] LI C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks[C]//*Computer Vision-ECCV*, 2016: 702-716.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 7780459.
- [14] WANG H, CAO P, WANG J, et al. Utransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(3): 2441-2449.
- [15] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization[J]. arXiv preprint arXiv:1607.08022, 2016.
- [16] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth  $16 \times 16$  words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [17] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//*Computer Vision-ECCV*, 2016: 694-711.
- [18] PENG L T, ZHU C L, BIAN L H. U-shape transformer for underwater image enhancement[C]//*Computer Vision-ECCV*, 2023: 290-307.
- [19] 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.
- [20] Korhonen J, You J. Peak signal-to-noise ratio revisited: Is simple beautiful[C]//*Fourth International Workshop on Quality of Multimedia Experience*, 2012: 37-38.
- [21] Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//*20th International Conference on Pattern Recognition*, 2010: 2366-2369.
- [22] YANG M, Sowmya A. An underwater color image quality evaluation metric[J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 6062-6071.
- [23] 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.
- [24] Mittal A, Soundararajan R, Bovik A. Making a "Completely Blind" image quality analyzer[J]. *IEEE Signal Processing Letters*, 2013, 20(3): 209-212.