doi：10. 12068/j. issn. 1005－3026. 2024. 12. 004

基于Transformer的多尺度水下图像增强网络

杨爱萍，方思捷，邵明福，张腾飞
（天津大学 电气自动化与信息工程学院，天津 300072）

摘 要：基于CNN（convolutional neural network）的水下图像增强方法容易忽略全局特征，导致复原图像出现颜色失真、对比度下降等现象，影响全局视觉感知效果．因此，提出一种基于Transformer的多尺度水下图像增强网络．针对全局特征缺失问题，融入水下图像先验设计位置编码模块，构建适用于水下场景的Swin Transformer模块，并通过自注意力机制针对性地提取图像全局特征，提升全局感知性能；针对局部细节模糊现象，设计CNN模块关注水下图像纹理、边缘等局部特征，改善细节感知效果；构建转移融合模块，将Swin Transformer的全局注意力转移到卷积特征上，达成全局和局部特征的高效融合与利用．实验结果表明，所提方法在EUVP子集上的PSNR值最高可达23. 47 dB，可有效增强全局视觉感知能力，显著改善图像视觉质量．

关 键 词：水下图像增强；深度学习；Transformer；卷积神经网络；转移融合
中图分类号：TP 391 **文献标志码**：A **文章编号**：1005-3026(2024)12-1696-10

Transformer-based Multi-scale Underwater Image Enhancement Network

YANG Ai-ping, *FANG Si-jie*, *SHAO Ming-fu*, *ZHANG Teng-fei*
(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China. Corresponding author: YANG Ai-ping, E-mail: yangaiping@tju.edu.cn)

Abstract: CNN (convolutional neural network)-based underwater image enhancement methods neglect global visual perception, leading to color distortion and contrast degradation. A Transformer-based multi-scale underwater image enhancement network (MTransNet) is proposed. To address the problem of lacking global visual perception, a position encoding module is designed based on underwater image priors and a Swin Transformer module which is applicable to underwater scenes is constructed. Furthermore, self-attention mechanism is built to improve global perception performance. As for the detail blurring that exists in current methods, a CNN module is developed to capture local features such as textures or edges, to improve local perception performance. The transfer fusion module is built to transfer global attention of Swin Transformer to local convolutional feature, achieving full fusion and utilization of global feature and local feature. The PSNR value on subsets of EUVP can reach up to 23. 47 dB, which demonstrates the method can significantly enhance global visual perception and increase image visual quality.

Key words: underwater image enhancement; deep learning; Transformer; convolutional neural network; transfer fusion

在复杂的水下环境中，由于可见光受到水体吸收和散射等影响，拍摄的水下图像往往存在对比度低、细节模糊及颜色失真等问题，直接影响水下目标识别与跟踪[1-2]等视觉任务，研究水下图像增强具有重要的现实意义．

水下图像增强可分为传统方法和基于深度学习的方法．传统方法可分为非物理模型方法和基于物理模型的方法．非物理模型方法[3-5]主要通

收稿日期：2023-06-06
基金项目：国家自然科学基金资助项目（62071323）．
作者简介：杨爱萍（1977-），女，山东聊城人，天津大学副教授，博士生导师．

过调整像素值或像素分布改善水下图像色偏和对比度.由于未考虑水下图像退化机制,该类方法存在颜色失真、对比度受限、图像锐化程度过高等现象.基于物理模型的方法根据水下图像先验信息或深度信息估计模型的透射率分布图,然后通过反演水下成像物理模型[6-7]得到复原图像.该类方法增强图像效果显著,但由于水下深度信息估计不准确以及先验信息存在场景受限等问题,复原图像仍存在色偏、伪影等现象.

近年来,基于深度学习的水下图像增强方法成为领域研究热点.Wang等[8]首次提出一种基于卷积神经网络(convolutional neural network,CNN)的端到端水下图像增强网络(underwater image enhancement network,UIE-Net),通过联合训练颜色校正子网络和去雾子网络,实现图像增强,并利用数据增强策略,实现数据集扩充,提升模型泛化能力.为进一步增强网络实时性,Islam等[9]提出以条件生成对抗网络为基础的改善视觉质量的快速水下图像增强模型(fast underwater image enhancement for improved visual perception based on conditional GAN,FUnIE-GAN),并根据全局内容、颜色、局部纹理和样式信息构建感知损失评估图像质量;该方法能快速处理水下图像,但是复原图像细节损失过多.Li等[10]提出一种基于门控融合的水下网络(Water-Net)并构建第一个水下图像增强基准数据集(underwater image enhancement benchmark,UIEB),其通过网络生成的置信度图融合预处理图像,有效改善色偏、增强对比度;由于该方法采用全卷积形式,导致特征提取不充分,细节信息损失较多.为校正水下图像色偏,Li等[11]提出一种基于介质透射率分布图引导的多颜色空间融合的水下图像增强模型(underwater image enhancement via medium transmission-guided multi-color space embedding,Ucolor),基于注意力机制,通过多颜色空间编码与反透射率分布图引导的解码网络,实现水下图像增强.为进一步增强融合效果,减少光线水下传播衰减损失,Qi等[12]提出一种基于相关特征匹配和联合学习的水下图像协同增强网络(underwater image co-enhancement network,UICoE-Net),通过融合不同水下图像提供的互补信息,在编-解码器结构中引入相关特征匹配单元,实现2个分支之间的相互关联,提高视觉质量;但该方法直接学习局部映射关系,忽视全局信息,引入颜色校正的不确定性,导致局部增强效果不佳.

基于卷积神经网络的深度学习方法虽然可取得比传统方法更优的性能,但由于卷积运算重点关注局部而忽视全局特性,导致上述方法复原图像全局视觉感知效果较差.Transformer[13]和ConvNeXt[14]可以一定程度上弥补传统卷积运算的缺陷.ConvNeXt是一种基于CNN的模型,能够有效地捕捉局部信息和空间结构.Fan等[15]提出了一种轻量级注意力引导的ConvNeXt网络(lightweight attention guided ConvNeXt network,LACN)用于低照度图像增强,该模型可有效实现图像增强,并较好地保留色彩信息,实现噪声抑制,但ConvNeXt网络泛化能力较弱,对于复杂场景图像处理能力较弱,且对全局信息关注不足.基于Transformer的特征提取机制通过图片序列的上下文信息关注全局特征,可充分发挥全局注意力特性,有效提升特征提取能力.为了解决计算机视觉任务中实例规模大、分辨率高等问题,Liu等[16]提出Swin Transformer(Transformer using shifted windows)模型,利用移位窗口方案进行特征表示,这种分层架构可以在各种比例下进行建模,使得Swin Transformer可以获得多尺度特征信息,并广泛兼容各类计算机视觉任务.

受上述文献启发,本文提出基于Transformer的多尺度水下图像增强网络(Transformer-based multi-scale underwater image enhancement network,MTransNet),实现对全局和局部特征的综合利用.通过融入水下图像先验信息,设计位置编码模块,建立Swin Transformer模块中图像块相对位置关系.构建CNN模块和Swin Transformer模块,通过卷积运算和自注意力机制针对性地提取图像的局部信息和全局特征.设计转移融合模块,将Swin Transformer的全局注意力转移到卷积特征,实现全局和局部特征的高效利用.

1　基于Transformer的多尺度水下图像增强网络

基于Transformer的多尺度水下图像增强网络(MTransNet)采用编-解码器网络结构,网络整体框架如图1所示.首先,输入图像通过位置编码模块,将生成特征送入Swin Transformer模块,利用Swin Transformer模块和CNN模块分别提取水下图像对应的全局特征和局部特征.为了增强局

部特征的全局建模能力,利用转移融合模块将Swin Transformer提取的全局注意力转移到CNN模块的局部特征中.

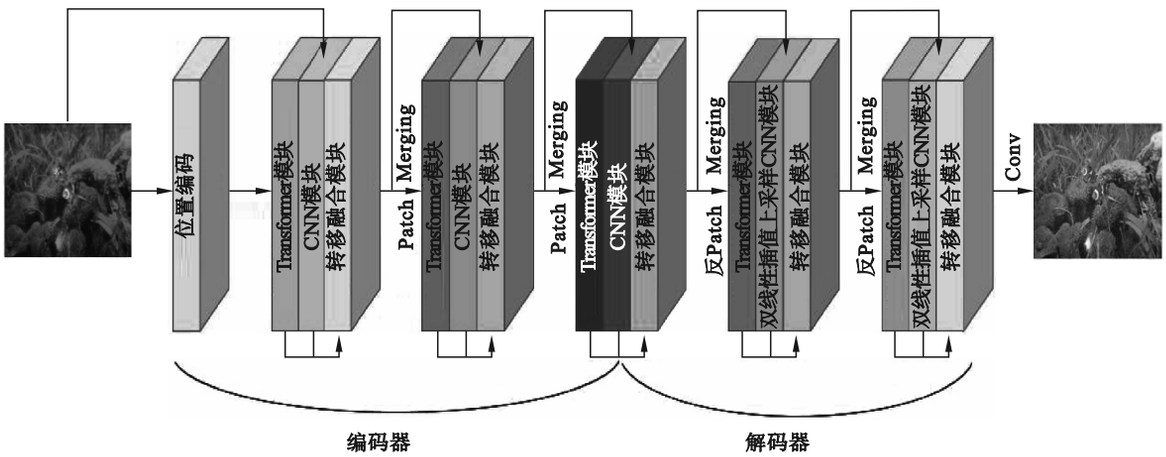

图1 网络结构

Fig. 1 Network structure

CNN模块的前2个卷积层与ReLU激活函数相连,最后1个卷积层使用最大池化操作进行下采样.Swin Transformer通过Patch Merging操作改变图像的尺寸大小和通道数,确保Transformer特征与CNN特征相匹配.转移融合模块可以将Transformer全局注意力特性转移至CNN细节特征,而不改变输入特征的大小和通道数.融合特征分别进入下一级CNN模块和Swin Transformer模块,以提取不同尺度的局部特征和全局特征,并通过转移融合模块得到融合特征.最终送入第三级编码器模块,获得编码器输出特征.

解码器部分与编码器部分网络结构保持一致,编码器输出特征进入CNN模块和Swin Transformer模块,通过双线性插值和反Patch Merging操作,实现上采样,再通过转移融合模块得到融合特征,送入下一级解码器模块,最终对融合特征进行解码,获得复原图像.

具体地,对于大小为$H\times W\times 3$的输入特征x,通过位置编码模块和CNN模块后,输出特征维度大小为$H/2\times W/2\times 96$. Swin Transformer模块不改变输入特征的维度大小,可保证Transformer和CNN输出特征的维度一致性,转移融合模块同样不改变融合特征的大小.之后,大小为$H/2\times W/2\times 96$的融合特征x_1被送入第二级编码器模块,融合特征x_1经过Patch Merging下采样操作,输出特征维度大小为$H/4\times W/4\times 192$,二级CNN模块输出特征维度大小为$H/4\times W/4\times 192$,融合特征x_2大小保持不变;第三级编码器模块输入特征为x_2,输出融合特征x_3维度大小为$H/8\times W/8\times 384$. 解码器模块输入特征y_1大小为$H/8\times W/8\times 384$,Swin Transformer模块通过反Patch Merging操作,输出大小为$H/4\times W/4\times 192$的Transformer特征,CNN模块通过双线性插值上采样得到大小为$H/4\times W/4\times 192$的CNN特征,转移融合模块不改变融合特征的维度大小;融合特征y_2进入下一级解码器模块,得到大小为$H/2\times W/2\times 96$的融合特征y_3;最后,通过卷积层得到大小为$H\times W\times 3$的复原图像I.具体过程可表示为

$$x_i=F_{\text{merge}}(f_{\text{CNN}}(x),f_{\text{Trans}}(f_{\text{position}}(x))), \tag{1}$$

$$x_{i+1}=F_{\text{merge}}(f_{\text{CNN}}(x_i),f_{\text{Trans}}(f_{\text{patch}}(x_i))), \tag{2}$$

$$y_1=x_3=F_{\text{merge}}(f_{\text{CNN}}(x_{i+1}),f_{\text{Trans}}(f_{\text{patch}}(x_{i+1}))), \tag{3}$$

$$y_{j+1}=F_{\text{merge}}(f'_{\text{CNN}}(y_j),f_{\text{Trans}}(f'_{\text{patch}}(y_j))), \tag{4}$$

$$I=\text{Conv}(y_{j+1}). \tag{5}$$

其中:$i,j=1,2$;x表示输入特征;x_i,y_j分别为编码器、解码器模块输出的i,j尺度的融合特征;I表示MTransNet网络复原后的图像;$f_{\text{position}},f_{\text{Trans}},f_{\text{CNN}},f_{\text{patch}}$分别表示位置编码模块、Swin Transformer模块、CNN模块和Patch Merging操作对应的非线性映射关系;$f'_{\text{patch}},f'_{\text{CNN}}$分别为反Patch Merging操作和双线性插值上采样CNN模块对应的非线性映射关系;F_{merge}表示转移融合模块的映射关系.

1.1 位置编码模块

位置嵌入对于保留图像的空间信息至关重要,然而基于Transformer的网络结构常以水平或垂直顺序嵌入,忽略水下图像不同空间区域衰减程度的差异.受Guo等[17]启发,本文设计适合水下图像的Swin Transformer位置编码模块.因此,

MTransNet提出一种新的水下图像位置嵌入方法,引入与水下图像衰减相关的先验信息,获取水下图像深度图$D(x)$,并在位置编码中嵌入水下图像深度信息,具体可表示为

$$T(x)=\max_{c\in\{\mathrm{r,g,b}\},y\in\Omega(x)}\frac{A^c-J^c(y)}{\max(A^c,1-A^c)},\tag{6}$$

$$D(x)=-\frac{\ln T(x)}{\beta}.\tag{7}$$

其中:J为输入图像;x,y表示像素点;$\Omega(x)$为以x为中心的局部区域;c表示图像RGB三通道之一;β为水下图像总的介质衰减系数;A为环境光;$T(x)$为透射率分布图.

位置编码模块网络结构如图2所示.分块大小设置为2×2,输入图像经过分块操作和线性嵌入层后,输出向量维度由$H\times W\times C$变为$H/2\times W/2\times C$,其中通道C设置为96.位置编码信息分为空间位置信息和水下深度信息,采用三角函数编码:

$$\mathrm{PE}(\mathrm{pos}^D,2i)=\sin(\mathrm{pos}^D/10\,000^{2i/P_{\mathrm{total}}}),\tag{8}$$

$$\mathrm{PE}(\mathrm{pos}^D,2i+1)=\cos(\mathrm{pos}^D/10\,000^{2i/P_{\mathrm{total}}}).\tag{9}$$

其中:PE代表位置嵌入信息;pos^D代表位置维度(垂直方向、水平方向、空间深度);每一个位置维度P_{total}设置为32,总的位置维度对应编码层96个维度通道;i为位置索引.位置信息向量与线性嵌入层输出向量相加,得到编码器输入特征向量.

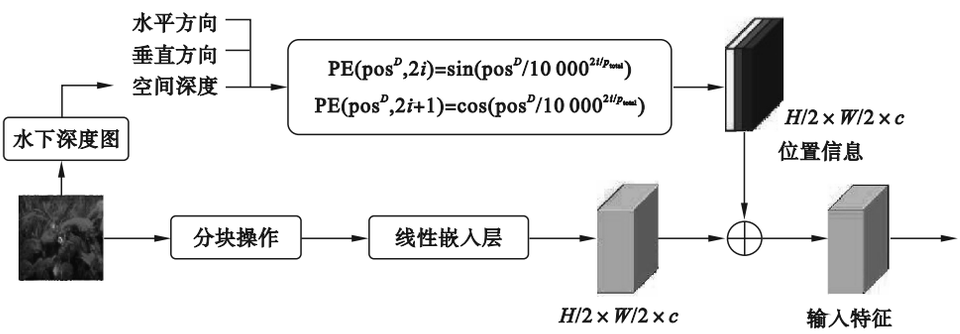

图2　位置编码模块

Fig. 2　Position encoding module

1.2　Swin Transformer模块

Swin Transformer是一种基于移动窗口的层级式Transformer模型,可以作为视觉领域通用骨干网络.CNN模块通过卷积和下采样操作增大每个卷积核的感受野,实现多尺度变换.Swin Transformer针对图片、视频等存在尺度变化的对象,设计Patch Merging操作,通过融合相邻的patch增大感受野,建立层级式结构,实现多尺度特征提取.Swin Transformer模块通过分块操作,在每一个窗口内作多头自注意力,利用多维投影模拟卷积神经网络多输出通道的效果,同时移动窗口实现全局建模,如图3所示.Swin Transformer模块由2个Transformer基本模块组成,第1个Transformer模块在单个窗口内实现多头自注意力,第2个Transformer模块通过向右下角循环移动窗口实现重叠区域内的多头自注意力,达到局部和全局特征的统一,其前馈神经网络层、残差连接和层归一化处理则保持不变.MTransNet采用三级Swin Transformer结构,每一级注意力头的数量分别为3,6和9.

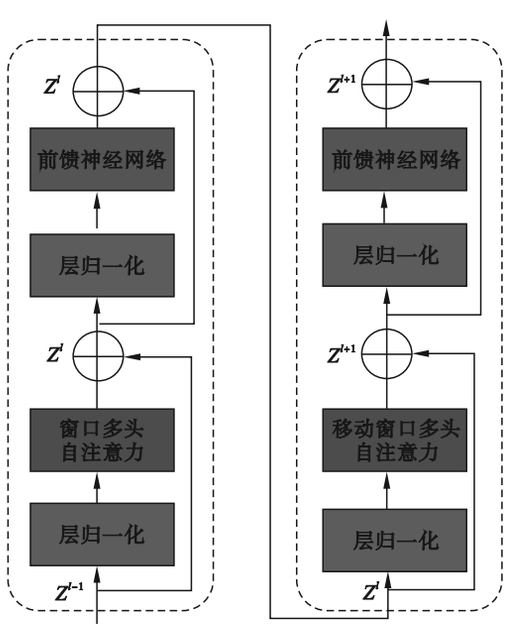

图3　Swin Transformer模块

Fig. 3　Swin Transformer module

1.3　转移融合模块

基于自注意力机制的Transformer特征可获取大范围全局信息,但细节纹理较为粗糙.而CNN关注局部像素之间的关系,可获取清晰的细

节信息和局部特征.受风格迁移[18]和图像超分辨率[19]等启发,本文提出一种转移融合模块,在不破坏CNN细节特征的前提下,将Swin Transformer的全局注意力转移到CNN特征上,实现全局特征和局部特征有机结合,具体结构如图4所示.该过程可表示为

$$F_{\mathrm{m}}^{s}=G_{\gamma}^{s}(F_{\mathrm{t}}^{s})\otimes \mathrm{IN}(F_{\mathrm{c}}^{s})\oplus G_{\beta}^{s}(F_{\mathrm{t}}^{s}). \tag{10}$$

其中:F_{m}^{s}代表融合特征;F_{c}^{s}代表CNN特征;F_{t}^{s}代表Swin Transformer特征;IN代表归一化操作;$s\in\{1,2,3\}$表示特征融合的不同尺度.$G_{\gamma}(g),G_{\beta}(g)$均由两个卷积层组成,代表不同的映射关系.$\otimes,\oplus$分别表示逐像素相乘、逐像素相加.经过$\gamma$映射层的Swin Transformer特征$G_{\gamma}(F_{\mathrm{t}}^{s})$与实例归一化操作后的卷积特征$F_{\mathrm{c}}^{s}$逐像素相乘,与经过$\beta$映射层的Swin Transformer特征$G_{\beta}(F_{\mathrm{t}}^{s})$逐像素相加,得到融合特征$F_{\mathrm{m}}^{s}$.

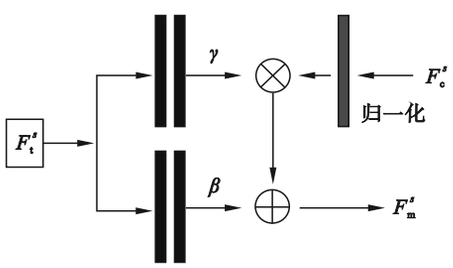

图4 转移融合模块

Fig. 4 Transfer fusion module

1.4 损失函数

本文采用L_2损失作为网络的总损失函数,通过缩小复原图像和参考图像之间的距离,增强水下图像质量,改善视觉感知效果.

L_2损失函数:通过逐点计算复原图像和参考图像像素之间的差异来约束图像信息损失,可表示为

$$L_2=\frac{1}{HW}\sum_{m=1}^{H}\sum_{n=1}^{W}(\hat{I}(m,n)-I(m,n))^2. \tag{11}$$

其中:m,n分别为水平和垂直方向上像素点的索引;H,W表示图像水平和垂直方的像素点的个数;$\hat{I}$为水下参考图像;I为MTransNet生成的复原图像.

2 实验与结果分析

2.1 实验设置

1)实验环境配置:Intel(R) Xeon(R) i7-12700 k处理器,5.0 GHz主频和NVIDIA RTX3090 GPU,编程语言和模型搭建平台为Python和PyCharm;训练框架基于Pytorch,采用Adam优化器;动量衰减指数设置为:$\beta_1=0.9,\beta_2=0.99$,初始学习率设为0.000 1,批处理大小(batch size)设为8.训练共进行100次迭代,耗时约8 h.

2)数据集设置:为了增强MTransNet的泛化能力和鲁棒性,选取增强水下视觉感知(enhancing underwater visual perception, EUVP)数据集[9]中ImageNet子集的2 992对水下图像、Dark子集的1 008对水下图像、UIEB数据集[10]的800对水下图像组成训练集,共包含4 800对原始水下图像和清晰的参考图像.选取ImageNet训练集中剩余的708对水下图像、Dark训练集中202对水下图像作为Test-910测试集;选取UIEB训练集中剩余的90对水下图像作为Test-R90测试集;选取RUIE测试数据集[20]中的水下图像质量子集UIQS,将其按照水下图像质量评价指标(underwater color image quality evaluation, UCIQE)得分由高到低划分为A,B,C,D,E 5类,从A,B,D,E类中各选取一幅图片,C类选取2幅图片,构造Test-UIQS测试集,检测MTransNet与对比方法的色彩校正能力.MTransNet和对比方法均使用相同的数据集进行测试,对比方法均使用原论文作者开源代码,保证实验结果真实有效.

3)对比方法:从主观视觉效果和客观评价指标两方面与先进的水下图像增强算法进行对比,包括基于扩散的方法(fusion-based method)[21],基于Retinex的方法(Retinex-based)[22],最小色彩损失和局部自适应对比度增强方法(minimal color loss and locally adaptive contrast enhancement method for underwater image enhancement, MMLE)[23]等;基于CNN的Water-Net[10],Ucolor[11];基于GAN的FUnIE-GAN[9],水下生成对抗网络(underwater GAN, UGAN)[24].

4)评价指标:对于Test-910和Test-R90测试集,使用峰值信噪比(peak signal-to-noise ratio, PSNR)和结构相似性(structural similarity, SSIM)指标进行评估.对于Test-UIQS测试集,采用无参考评价指标UCIQE和水下图像质量指标(underwater image quality measures, UIQM)来衡量各方法性能.为避免UCIQE和UIQM指标存在某种程度上的不公平,同样引入用户研究(user study, US)对主观视觉感知进行评分,邀请20位志愿者对增强图像的感知质量进行独立评分,评分等级由1至5(1为最差,5为最好).志愿者通过

判断增强图像是否存在颜色偏差、伪影,增强结果是否具有良好的对比度、主观感受以及是否清晰自然来进行主观评分.

2.2　主观评价

从不同的测试数据集中选取具有不同特点的水下图像进行对比实验.首先在Test-910测试集上进行比较,如图5所示.传统方法中fusion-based方法取得最好的视觉效果,但对于弱光条件下的水下图像增强效果不佳(图5b第6幅图像);Retinex-based方法颜色校正不合理,亮度、对比度较低,整体视觉效果较差;MMLE方法增强图像对比度过高,出现过饱和和颜色失真(图5d第5幅图像).深度学习方法中,UGAN方法整体色偏去除不彻底(图5f中第3,4幅图像);Water-Net方法能够增强视觉效果,但整体色调偏暗、亮度偏低(图5g中第2,3,4幅图像);Ucolor方法增强图像整体亮度有明显的提升,视觉效果有一定改善,但存在颜色校正不彻底问题(图5h中第1,2,3,6幅图像). FUnIE-GAN和本文方法MTransNet均取得较好的视觉效果.FUnIE-GAN方法颜色校正合理,有效去除水下图像色偏.MTransNet方法在第1,2,5,6幅图像取得最佳效果,显著增强图像清晰度和细节信息,有效改善亮度、对比度,获得最佳的主观视觉感受.

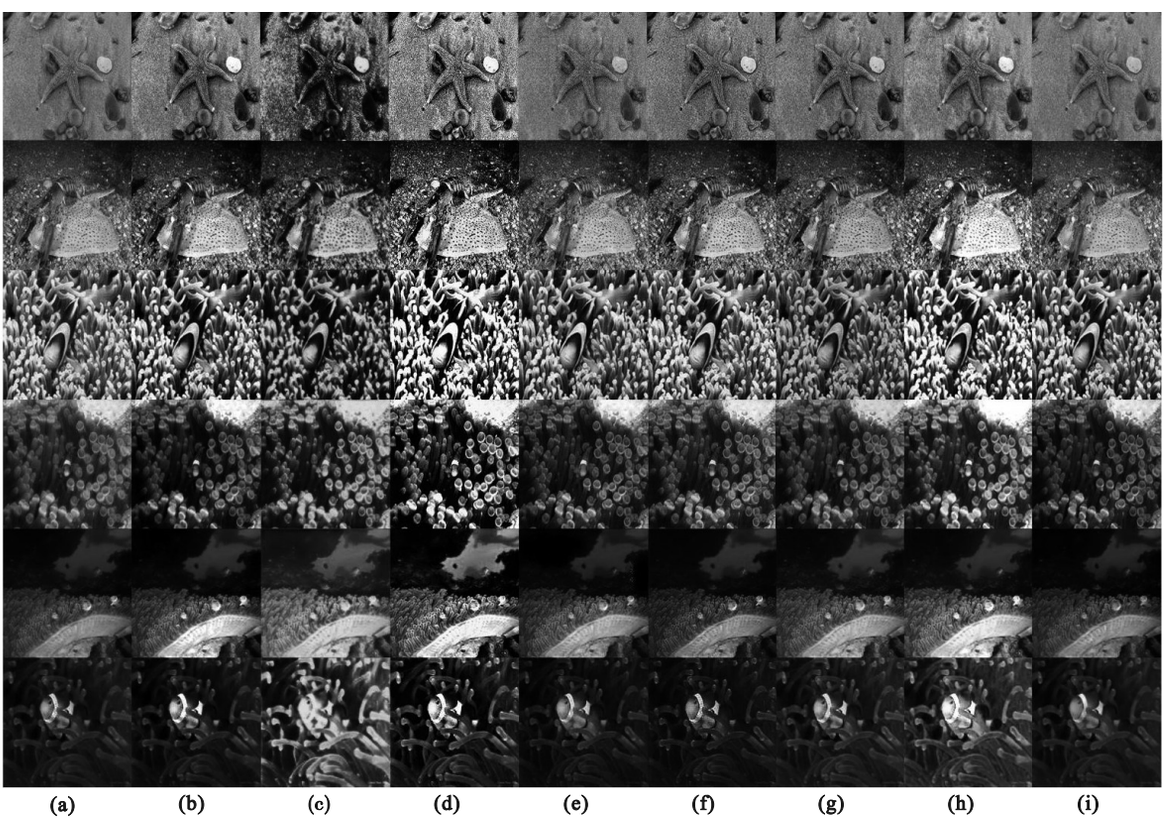

图5　Test-910数据集对比结果

Fig. 5　Comparison results of Test-910 dataset

(a)—原始图像;(b)—fusion-based;(c)—Retinex-based;(d)—MMLE;(e)—FUnIE-GAN;(f)—U-GAN;(g)—Water-Net;(h)—Ucolor;(i)—MTransNet.

各方法在Test-R90测试集上的实验结果如图6所示.由于UIEB数据集原始水下图像较为清晰,因此各增强方法均取得不错的视觉效果.第1幅图像MMLE方法复原图像中彩带的颜色发生改变,Ucolor方法复原图像地面颜色失真,且图像清晰度不如MTransNet方法.第2,3幅图像fusion-based方法、Ucolor方法、MTransNet方法视觉效果较好,但fusion-based方法复原得到的图像亮度偏低,Ucolor方法复原图像远景区域出现过曝光,视觉效果不理想.第4幅图像,fusion-based方法和Water-Net方法可有效去除色偏,但整体亮度较低;Retinex-based方法虽然有效增强图像亮度,但背景区域出现伪影和不真实特征;FUnIE-GAN,UGAN和Ucolor方法未能有效去

除色偏，恢复图像颜色不真实；MMLE，MTransNet方法均取得较好的视觉效果.第5，6幅图像MMLE，MTransNet方法取得相近的视觉效果，fusion-based方法和Retinex-based方法第5幅图像地面颜色失真；Water-Net方法亮度偏低；Ucolor方法与MTransNet方法相比，恢复图像细节模糊.综上所述，Test-R90测试集上MTransNet方法颜色校正自然、图像清晰、亮度较高，取得最佳视觉效果.

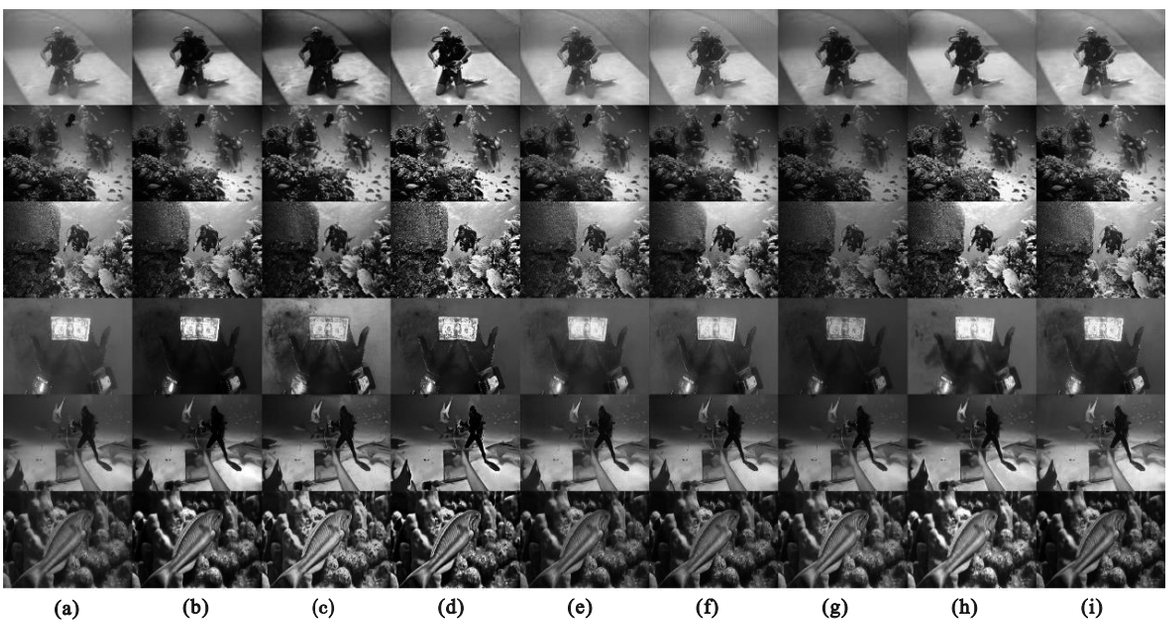

图6 Test-R90数据集对比结果

Fig. 6 Comparison results of Test-R90 dataset

(a)—原始图像；(b)—fusion-based；(c)—Retinex-based；(d)—MMLE；(e)—FUnIE-GAN；(f)—U-GAN；(g)—Water-Net；(h)—Ucolor；(i)—MTransNet.

Test-UIQS测试集上的主观对比结果如图7所示.由于该测试集没有参考图像，深度学习方法无法有针对性地学习数据集中水下图像的特点，而传统方法并不依赖于参考图像，所以该数据集中传统方法整体视觉效果略优于深度学习方法.但传统方法受到水下图像先验和假设条件限制，对不满足先验条件的场景复原效果欠佳.如fusion-based方法(图7b第3，4幅图像)和Retinex-based方法(图7c第3，4，5幅图像)得到的图像均存在色偏.MMLE方法假设条件符合大多数水下环境，但综合考虑Test-R90，Test-910数据集实验结果，该方法存在对比度过度增强现象，增强背景水体时出现色偏问题.基于深度学习的方法中，FUnIE-GAN和UGAN方法增强效果不明显，Water-Net方法第2，3幅图像存在色偏现象，而MTransNet方法复原图像清晰度高，整体视觉质量理想.

2.3 客观评价

表1为各方法在不同数据集上的客观评价指标，其中加粗、下划线分别代表最高值与次高值.Test-910和Test-R90数据集使用有参考评价指标PSNR和SSIM，Test-UIQS测试集使用无参考评价指标UCIQE和UIQM以及用户研究US衡量MTransNet与对比方法的性能.

对于有参考评价指标，FUnIE-GAN方法在Test-910测试集上PSNR指标取得次高值、SSIM指标取得最高值，图5中主观视觉效果突出；在Test-R90测试集上PSNR和SSIM指标较低，图6中主观视觉质量较差，说明该方法泛化性能差.MMLE方法在传统方法中整体视觉效果表现不错，但存在对比度过度增强和色偏现象，PSNR指标和SSIM指标在对比方法中处于中等水平.Water-Net和Ucolor方法在Test-910测试集上评价指标较低，在Test-R90测试集高于传统方法，与主观视觉效果一致.本文方法MTransNet在Test-910和Test-R90数据集上PSNR指标取得最高值，SSIM指标取得次高值，说明MTransNet方法在对比方法中综合表现最好，可有效改善图像视觉质量，泛化性能较好.

对无参考评价指标，传统方法MMLE取得最

高的用户研究分数(US),MTransNet在深度学习方法中取得最高的感知分数. Retinex-based方法和MMLE方法虽然取得最高的UCIQE和UIQM指标,但是主观衡量上并没有取得最好的视觉感知效果.

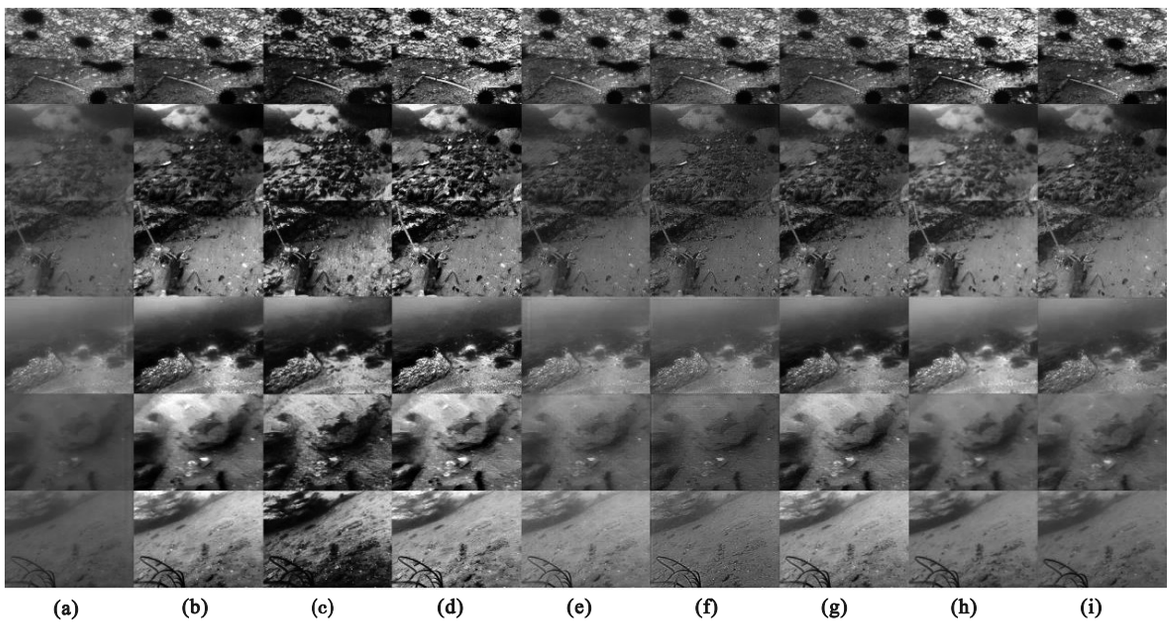

图7　Test-UIQS数据集对比结果

Fig. 7　Comparison results of Test-UIQS dataset

(a)—原始图像;(b)—fusion-based;(c)—Retinex-based;(d)—MMLE;(e)—FUnIE-GAN;(f)—U-GAN;(g)—Water-Net;(h)—Ucolor;(i)—MTransNet.

表1　对比方法客观评价结果

Table 1　Objective evaluation results of comparative methods

方法	Test-910		Test-R90		Test-UIQS		
	PSNR/dB	SSIM	PSNR/dB	SSIM	US	UCIQE	UIQM
Input	17. 24	0. 79	16. 31	0. 79	1. 00	0. 48	0. 84
fusion-based[21]	21. 83	0. 83	17. 83	0. 82	2. 80	0. 62	1. 22
Retinex-based[22]	20. 06	0. 82	18. 09	0. 83	2. 60	**0. 71**	1. 14
MMLE[23]	22. 34	**0. 85**	18. 44	0. 85	**3. 00**	0. 64	**1. 34**
FUnIE-GAN[9]	23. 06	**0. 85**	17. 64	0. 79	1. 60	0. 55	0. 81
UGAN[24]	20. 76	0. 80	17. 32	0. 77	1. 40	0. 51	0. 77
Water-Net[10]	21. 41	0. 83	18. 90	**0. 87**	2. 40	0. 56	0. 97
Ucolor[11]	20. 52	0. 81	18. 41	0. 85	2. 10	0. 53	0. 88
MTransNet	**23. 47**	0. 84	**19. 52**	0. 86	2. 50	0. 62	1. 21

综上所述,MTransNet方法主客观综合性能表现最佳,可显著实现水下图像增强,实现视觉质量提升,并具有较强的泛化能力,可广泛应用于各类复杂水下环境.

2.4　消融实验

设计消融实验分别验证MTransNet所设计的Swin Transformer模块、CNN模块、位置编码模块的有效性及合理性,具体情况如下:

1)无Transformer表示移除Swin Transformer模块的消融实验;

2)无CNN表示移除CNN模块的消融实验;

3)无PE(position encoding)表示移除位置编码模块的消融实验.

如表2所示,无论是移除Transformer模块、

CNN模块,还是位置编码模块,PSNR和SSIM客观评价指标均有明显下降.如图8所示,从主观视觉上看,无Transformer缺少全局注意力特征优势,图像整体视觉效果较差.无CNN由于卷积运算局部特征增强优势的缺失,图像细节和边缘保留不足.无PE移除了位置编码模块导致图像空间信息保留不充分,因此图像增强存在不均匀的现象.消融实验充分证明融合Transformer与CNN特征以及位置编码模块的优势,该方法可实现全局特征和局部特征的联合优化,增强水下图像的视觉质量.

表2 消融实验客观评价结果

Table 2 Objective evaluation results of ablation experiment

评价指标	PSNR/dB	SSIM
原始图像	16.31	0.79
MTransNet	**19.52**	**0.86**
无Transformer	17.50	0.79
无CNN	17.70	0.81
无PE	18.21	0.82

图8 消融实验

Fig. 8 Ablation experiment

(a)—原始图像;(b)—MTransNet;(c)—无Transformer;(d)—无CNN;(e)—无PE.

3 结 语

本文提出一种基于Transformer的多尺度水下图像增强网络(MTransNet).网络整体采用对称的编-解码网络结构,通过深度图引导设计位置编码模块,确保水下图像序列位置的相对稳定.Swin Transformer模块通过自注意力机制增强图像全局特征,弥补卷积运算过于关注局部像素的缺陷.利用转移融合模块将Transformer的全局注意力特性转移到CNN特征,实现局部细节和全局特征的高效统一,有效改善图像的整体视觉质量.

参考文献:

[1] Jiang Q P, Gu Y S, Li C Y, et al. Underwater image enhancement quality evaluation: benchmark dataset and objective metric [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(9): 5959-5974.

[2] Raveendran S, Patil M D, Birajdar G K. Underwater image enhancement: a comprehensive review, recent trends, challenges and applications [J]. *Artificial Intelligence Review*, 2021, 54(7): 5413-5467.

[3] Hummel R. Image enhancement by histogram transformation [J]. *Computer Graphics and Image Processing*, 1977, 6(2): 184-195.

[4] Reza A M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement [J]. *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, 2004, 38(1): 35-44.

[5] Hitam M S, Yussof W N J H W, Awalludin E A, et al. Mixture contrast limited adaptive histogram equalization for underwater image enhancement [C]//2013 International Conference on Computer Applications Technology (ICCAT). Sousse, 2013: 1-5.

[6] McGlamery B L. A computer model for underwater camera systems [J]. *Proceedings of the SPIE*, 1980, 208: 221-231.

[7] Jaffe J S. Computer modeling and the design of optimal underwater imaging systems [J]. *IEEE Journal of Oceanic Engineering*, 1990, 15(2): 101-111.

[8] Wang Y, Zhang J, Cao Y, et al. A deep CNN method for underwater image enhancement [C]//2017 IEEE International Conference on Image Processing (ICIP). Beijing, 2017: 1382-1386.

[9] 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.

[10] 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.

[11] Li C Y, Anwar S, Hou J H, et al. Underwater image enhancement via medium transmission-guided multi-color space embedding [J]. *IEEE Transactions on Image Processing*, 2021, 30: 4985-5000.

[12] Qi Q, Zhang Y C, Tian F, et al. Underwater image co-enhancement with correlation feature matching and joint learning [J]. *IEEE Transactions on Circuits and Systems for*

Video Technology，2022，32(3)：1133−1147.

[13] Vaswani A，Shazeer N，Parmar N，et al. Attention is all you need [J]. *Advances in Neural Information Processing Systems*，2017，30：6000−6010.

[14] Liu Z，Mao H Z，Wu C Y，et al. A ConvNet for the 2020s[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition.New Orleans，2022：11976−11986.

[15] Fan S J，Liang W，Ding D R，et al. LACN：a lightweight attention-guided ConvNeXt network for low-light image enhancement [J]. *Engineering Applications of Artificial Intelligence*，2023，117：105632.

[16] Liu Z，Lin Y T，Cao Y，et al.Swin Transformer：hierarchical vision Transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision.Montreal，2021：9992−10002.

[17] Guo C L，Yan Q X，Anwar S，et al. Image dehazing Transformer with transmission-aware 3D position embedding [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition.New Orleans，2022：5802−5810.

[18] Jiang L M，Zhang C X，Huang M Y，et al. TSIT：a simple and versatile framework for image-to-image translation[C]// Computer Vision−ECCV 2020：16th European Conference Proceedings，Part III 16.Glasgow，2020：206−222.

[19] Wang X T，Yu K，Dong C，et al.Recovering realistic texture in image super-resolution by deep spatial feature transform [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Salt Lake City，2018：606−615.

[20] 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.

[21] 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.

[22] Fu X Y，Zhuang P X，Huang Y，et al. A retinex-based enhancing approach for single underwater image[C]//2014 IEEE International Conference on Image Processing (ICIP). Paris，2014：4572−4576.

[23] 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.

[24] Fabbri C，Islam M J，Sattar J.Enhancing underwater imagery using generative adversarial networks [C]//2018 IEEE International Conference on Robotics and Automation (ICRA).Brisbane，2018：7159−7165.