

引用格式: CHEN Jiajun, LI Kaixiang, LI Renjian, et al. AtG-DeepLab V3+ Endoscopic Image Enhancement Algorithm Based on Self-attention Mechanism Optimization[J]. Acta Photonica Sinica, 2023, 52(8):0817001

陈家俊,李开祥,李仁剑,等. 基于自注意力机制优化的 AtG-DeepLab V3+内窥图像增强算法[J]. 光子学报, 2023, 52(8): 0817001

基于自注意力机制优化的 AtG-DeepLab V3+ 内窥图像增强算法

陈家俊¹, 李开祥¹, 李仁剑^{1,2}, 邵春蕾¹, 李贵叶¹, 陈玲玲¹

(1 深圳技术大学 健康与环境工程学院, 深圳 518118)

(2 北京航空航天大学, 电子信息工程学院, 北京 100191)

摘要: 基于自注意力机制和 DeepLab V3+ 网络联合构建了 AtG-DeepLab V3+ 开源算法进行内窥图像增强处理, 并采集测试靶和生物组织图像进行训练和测试。实现了内窥成像畸变和蜂窝状栅格结构的同时去除并能高清还原更多图像细节, 对比现有的内窥图像重建网络 U2 Net, Attention U-net 和 GARNN 等算法, 在峰值信噪比 (提升 66.4%, 51.9%, 154.6%)、结构相似度 (提升 55.6%, 45.9%, 231.5%) 等量化指标上均实现了较大幅度的提高。该算法为光学内窥图像处理提供了一个新的高效处理方案。

关键词: 内窥图像; 深度学习; 图像增强; 注意力门; 神经网络

中图分类号: TP391.41

文献标识码: A

doi: 10.3788/gzxb20235208.0817001

0 引言

光学内窥显微镜能适应弯曲、复杂的成像环境, 在医学临床诊断、工业检查以及环境监测等领域有着广泛的应用。在目前主流的光学内窥显微镜类型中, 光纤束式内窥镜有着重量轻、抗环境光干扰能力强以及改装潜力高 (如共聚焦、多光子内窥) 等优势, 日渐成为生物医学研究领域的重要研究工具。然而其传输图像受光纤导像束中光纤间隙限制, 有着明显的蜂窝状栅格效应。同时光纤束内窥系统通常搭配微透镜镜头使用以增加成像景深和收光效率, 而镜头引入的像差和畸变^[1]则会进一步降低成像质量, 往往难以获得理想的高分辨率、照明均匀的图像。因此, 消除光纤束内窥图像网格, 恢复图像质量对于内窥医学应用和诊断有着重要影响。传统的空域/频域滤波^[2]和插值算法^[3]可以应用以消除光纤束图像网格, 但图像质量并没有得到实质上的提高; 通过采集多幅图像计算中值图像或最大后验估计可以重建相对高分辨率光纤束图像^[4-5], 然而多次图像采集对实际应用条件提出了更高的要求。

近年来, 随着数据、算法和算力的飞速发展以及深度学习模型的进一步发展, 深度学习的图像识别、增强能力得到了显著的提升, 为相关学科研究更好地解析数据特征、开展决策优化等方面提供了强大的技术支撑。经过大量学习数据的训练, 深度学习网络可以对新图像进行前瞻性信息量的提升, 而卷积神经网络 (Convolutional Neural Network, CNN) 的发展使医学图像增强和识别取得了巨大突破。目前, 基于卷积神经网络进行内窥显微成像的计算机辅助诊断和辅助检测的研究正在迅速发展^[6-7]: 如利用神经网络识别分类功能的针对幽门螺杆菌^[8-9]、胃癌^[10]、结直肠息肉^[11]识别等已有报道。同时, CNN 的快速发展使得图像超分辨、图像重建有了更多的选择。卷积神经网络模型如 U-net^[12]、对抗生成网络^[13] (Generative Adversarial Network, GAN) 等开始应用于医学显微图像增强和重建中, 基于 U-net 模型的 CareNet 等^[14]实现了从低信

基金项目: 国家自然科学基金 (No. 52270008), 深圳市科技计划基础研究项目 (No. JCYJ20190813102005655), 深圳技术大学自制仪器项目 (No. 2020XZY003)

第一作者: 陈家俊, 2060411019@stumail.sztu.edu.cn

通讯作者: 陈玲玲, chenlingling@sztu.edu.cn

收稿日期: 2023-01-11; 录用日期: 2023-03-10

<http://www.photon.ac.cn>

噪荧光显微图像到高信噪比图像的映射。基于U-net改进的U2-Net^[15]、SRN-UNet^[16]等模型被应用于显著性目标检测任务如OCT^[17]、CT^[18]图像分割;引入自注意力机制的Attention U-net^[19]显著提高了U-net模型的特征敏感度和准确率^[20-22];DeepLab V3+^[23]使用了含有空洞卷积的空间金字塔池化模块(Atrous Spatial Pyramid Pooling, ASPP),适用于同尺寸图像的多类分割预测任务。在内窥图像方面,RAVID等^[24]报道了将深度神经网络(Deep Neural Networks, DNN)应用于光纤束内窥图像重建的首次尝试。该方法首先使用插值算法去除内窥光纤束蜂窝图案,然后训练DNN以提高分辨率。该模型的标签图像(Ground Truth, GT)由多幅重建算法生成,图像质量提升有限,且训练质量与配准算法强相关。JIANBO S等^[25-26]使用对抗生成网络(Generative Adversarial Restoration Neural Network, GARNN)和3D卷积网络分别实现了光纤束图像的重建,其数据集在经过精确的像素级配准(成像系统配准/VGGnet配准)后输入网络进行训练。尽管上述研究能提高图像质量,但在细节还原上仍存在一定挑战,更重要的是目前尚未有研究将深度学习引入以消除光纤束内窥图像的镜头畸变。

为了解决上述算法存在的不足,本文提出了基于自注意力机制和DeepLab V3+网络^[27]联合构建的AtG-DeepLab V3+深度学习模型以实现具有蜂窝结构和明显畸变的光纤束内窥图像重建,并选取最新光纤束内窥图像重建网络GARNN、经典图像重建网络U2-Net、引入自注意力机制的Attention U-net^[19]进行对比。本文在分辨率和畸变组合测试靶数据集下进行了验证实验,实验表明AtG-DeepLab V3+模型实现了光纤束内窥图像-高倍显微图像的直接映射,完成了光纤束内窥图像蜂窝网格结构和镜头畸变的同时去除,将分辨率从49.6 μm 提升至17.5 μm 。与U2-Net, Attention U-Net, GARNN相比,AtG-DeepLab V3+图像增强后在峰值信噪比(Peak Signal to Noise Ratio, PSNR)、结构相似度(Structural Similarity Index, SSIM)和标准均方根误差(Normalized Root Mean Square Error, NRMSE)指标提升均分别超过了51.0%, 45.0%和62.0%。在H&E染色的人类急性心肌梗死病理学组织切片数据集进行应用实验,AtG-DeepLab V3+将切片图像信噪比从8.73提升至15.28,分别高于对比网络66.4%, 51.9%, 154.6%。此外,基于AtG-DeepLab V3+模型的数据集不需要进行繁琐的精准图像配准预处理,大幅降低了图像预处理难度。我们相信本文构建的AtG-DeepLab V3+深度学习模型为基于光纤束的光学内窥图像增强提供了一个新的高速处理方案,尤其适用于大数据集的快速增强还原。

1 AtG-DeepLab V3+模型构建

1.1 DeepLab V3+模型

DeepLab V3+模型^[27]使用DeepLab V3+作为编码器结构,以Xception为主网络提取特征,整体采用编码-解码的网络架构,可以通过逐渐恢复空间信息来捕获更清晰的对象边界。编码器模块(Encoder)的主要功能是逐渐减少特征映射并捕获更高语义信息,使用步长为2的卷积代替池化,并采用空洞卷积,从而减少了分辨率的损失。解码器模块(Decoder)的主要功能是逐渐恢复空间信息,首先进行4倍双向上采样,然后与来自具有相同空间分辨率的网络主干的相应低层特征连接。深度卷积(为每个输入信道应用一个滤波器)和逐点卷积(组合信道之间深度卷积的输出)组合形成了减少参数量并扩大感受野的深度可分离卷积,能够通过以多个速率和多个有效视野探测传入特征或合并操作来编码上下文信息,从而产生更快、更强的编码-解码网络。

1.2 注意力机制

OKTAY O等^[19]提出了一种新的自注意力门(Attention Gate, AG)模型,可自动学习聚焦于不同形状和大小的目标结构,如图1虚线框所示,该模型以特征(x')和门控信号(g)作为输入,分别使用 1×1 卷积统一尺寸后逐项相加,得到叠加特征后激活,并使用双线性插值得到注意力系数的网格(α),用于识别显著图像区域,修剪特征响应,输入特征(x')与注意力系数(α)逐项相乘作为AG的输出,从较粗的尺度上收集的门控信号(g)所提供的激活和上下文信息来选择空间区域。注意力门实现了向前传播或向后传播过程中的神经激活,来自背景区域的梯度在反向过程中被向下加权,模型隐式学习抑制图像中的不相关区域,同时突出对特定任务有用的显著特征,可用于基于卷积网络的标准图像分析模型,提高模型的灵敏度和预测精度,适用于密集标签预测。

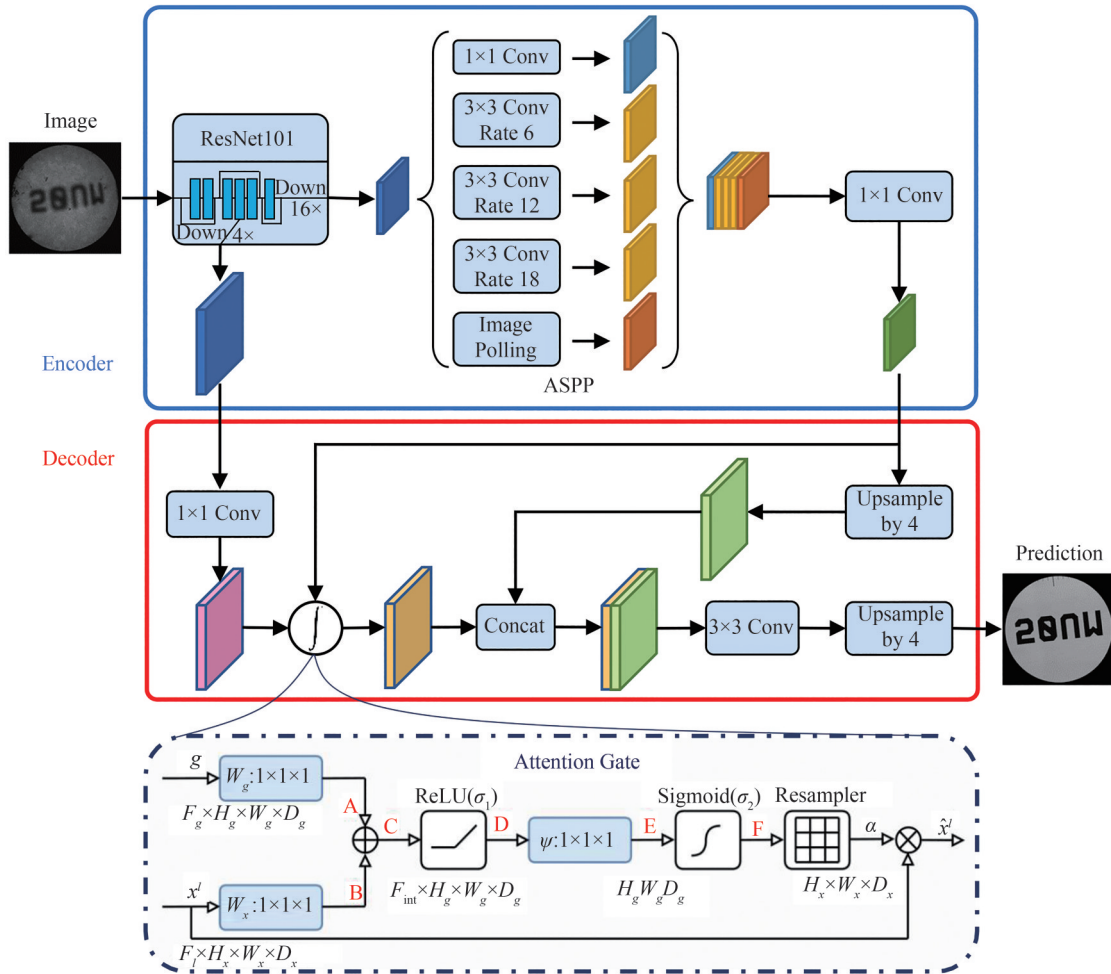


图1 基于自注意力机制和DeepLab V3+网络联合构建的AtG-DeepLab V3+模型结构图

Fig.1 AtG-DeepLab V3+ model structure based on self-attention mechanism and DeepLab V3+ network

1.3 基于自注意力机制和DeepLab V3+网络联合构建的AtG-DeepLab V3+模型

基于自注意力机制和DeepLab V3+网络联合构建的AtG-DeepLab V3+模型构建如图1所示,其使用ResNet101网络提取特征,改变原模型解码器部分下采样特征和上采样特征的简单拼接,将编码器部分ResNet101四倍下采样所得到的特征通道数由48扩大到128个,作为输入特征(x')。DeepLab V3+模型中的ASPP模块由一个 1×1 卷积、三个 3×3 卷积(rate={6, 12, 18})和全局平均池化层构成,其卷积和池化操作保证了其不变性,将提取到的不同类低层特征拼接并用一个 1×1 卷积减少通道数量作为高级抽象特征,AtG-DeepLab V3+模型将ASPP模块所提取到的高级抽象特征作为门控信号(g),将输入特征(x')和门控信号(g)输入注意力门模块(Attention Gate, AG),由门控信号(g)提供的激活和上下文信息计算注意力系数(α),注意力系数的网格重采样使用双线性插值完成。输入特征(x')用注意力系数(α)选择与给定任务相关的空间区域进行更新,实现对重要空域特征的筛选和提取,并将其与门控信号(g)四倍双线性上采样所得到的细化特征进行拼接,最后通过卷积层和双线性上采样得到最终输出。基于自注意力机制和DeepLab V3+网络联合构建的AtG-DeepLab V3+模型有效地结合了两者的特点,DeepLab V3+网络中的ASPP模块可以提取到图像不同尺度的高级抽象特征,自注意力门模块充分融合了图像下采样特征和高级抽象特征,可以消除内窥图像的蜂窝状栅格对图像分辨率的影响,使用注意力系数网格抑制了图像中的不相关区域,同时突出对特定任务有用的显著特征,增强了内窥图像对比度,使得AtG-DeepLab V3+模型在充分提取高级抽象特征的基础上进一步加强对重点空间区域及其特征的感知,在此基础上使用双线性插值得到细节丰富的增强图像,更加适用于光学内窥的图像增强相关研究。

2 光学内窥成像系统及其搭建

2.1 基于光纤束的光学内窥成像系统

基于光纤束的光学内窥成像系统如图2所示,由光源(Thorlabs, QTH10/M)、内窥光纤束(Fujikura, FIGH-30-850N, 30 000根纤芯, 外包层直径 $950\pm 50\ \mu\text{m}$)、物镜(Olympus, UPLFLN $10\times/0.3\ \text{NA}$)、无限远矫正套筒透镜(Tube lens, Thorlabs, TTL180-A)和成像探测器 sCMOS (Andor, Zyla5.5, $6.5\ \mu\text{m}/\text{像素}$, $2\ 560\times 2\ 160$)等部件组成。内窥光纤束远端微小透镜组(工作距离 $3\sim 10\ \text{mm}$, 视场角 80°)收集样本信号,经内窥光纤束传输至物镜焦点,物镜与无限远矫正套筒透镜组成 $4f$ 成像系统,对光纤束端面进行放大成像。光纤束式内窥镜由于其特殊的结构和检测环境,得到的图像背景灰度沿半径方向平均值逐渐减小,从而使整幅图像灰度严重不均,对比度下降,整体偏暗,从成像的分辨率来看,内窥镜成像的分辨率低,且带有蜂窝状栅格,使得图像中细节不够清晰。

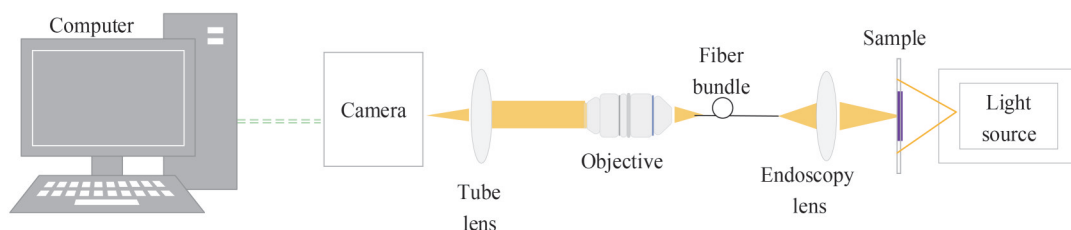


图2 基于光纤束的光学内窥成像系统图

Fig.2 Optical endoscopic imaging system diagram based on fiber bundle

2.2 畸变组合测试靶及生物样本数据集采集

分辨率和畸变组合测试靶(Thorlabs, R1L3S5P)包括NBS 1963A图案、扇形星(西门子星)图案、同心圆图案、网格图案、朗奇刻线等。NBS 1963A、扇形星和同心圆适于测量成像分辨率;网格图案适合测量成像系统引起的畸变;朗奇刻线适合评估分辨率、场曲和齐焦稳定性。生物组织样本为H&E染色的人类急性心肌梗死病理学组织切片(Carolina, #317120)。映射的每个数据集包括相同待测样本的光纤束内窥图像和不含前端内窥的高倍显微图像。

3 模型训练

3.1 数据集设置

本模型共进行两批次训练,首先进行测试靶图像的训练,模型验证后进行生物样本内窥图像的训练。预处理前单幅标签图像实际由多幅高分辨率的显微图像拼接而成($2\ 560\ \text{像素}\times 2\ 560\ \text{像素}$),拼接后图像单幅覆盖区域广,信息丰富度高。为使标签图像尺寸与输入内窥图像匹配,对标签图像进行下采样预处理($1\ 280\ \text{像素}\times 1\ 280\ \text{像素}$)。共采集到测试靶内窥图像及其标注图像35对,人体组织内窥图像及其标注图像20对,实际所使用的数据量(2.6 GB)达到了常见深度学习训练的数据量大小范围($0.5\sim 10\ \text{GB}$)^[21-24]。同时为防止训练过拟合,通过随机翻转、旋转进行数据增强,两批次训练图像数量分别扩充为280对和160对,分别取20对和8对用于模型测试。

3.2 实验环境及参数配置

实验仿真环境为Python3.9、Anaconda3,学习框架为PyTorch、OpenCV,开发工具为PyCharm-community-2022.1.1,硬件环境为深度学习GPU云服务器主机,基础架构为GPU驱动511.65、CUDA11.5,其中GPU使用1块TeslaV100-32,资源配置为10核40 GB内存和1卡32 GB显存。训练详细参数配置见表1,模型训练使用Adam优化器,其中Beta_1为一阶矩估计的指数衰减率,Beta_2为二阶矩估计的指数衰减率。

损失函数采用集成了Sigmoid层的二进制交叉熵损失,如式(1)、(2)所示。

$$\hat{y}' = \frac{1}{1 + e^{-\hat{y}}} \quad (1)$$

$$L(\hat{y}', y) = -[y \log(\hat{y}') + (1 - y) \log(1 - \hat{y}')] \quad (2)$$

式中, \hat{y} 代表网络的输出, \hat{y}' 代表经过 Sigmoid 层激活后的值, y 代表目标值,范围为 0 到 1。最终保存损失最小的模型。

表 1 模型参数配置
Table 1 Model parameter configuration

Parameter	Value
Learning rate	0.001
Beta_1	0.9
Beta_2	0.999
Momentum	0.9
Batch size	2
Decay	0
Epochs	200

4 实验结果分析

4.1 测试靶内窥成像训练结果分析

为了便于将改进后的模型与现有深度学习网络模型进行对比,本文选取了 U2-Net^[15]、Attention U-Net^[19]、GARNN 等^[26]网络进行对比,为了验证模型改进的必要性,对改进前后的模型进行消融实验。同时,本文引入峰值信噪比^[28](Peak Signal to Noise Ratio, PSNR)、结构相似性指数^[29](Structural Similarity, SSIM)和标准均方根误差(Normalized Root Mean Square Error, NRMSE)对增强后的信号图像进行定量分析。各模型增强后的测试靶图像如图 3 所示,其中 U2-Net 模型未能提取到图 3(c1)中的条纹,在图 3(c3)中提取到了部分同心圆周线并增强了对比度,在图 3(c4)放大了图像扭曲度;Attention U-Net 在图 3(d1)、(d3)中几乎丢失了全部细节信息,在图 3(d4)中加重了图案畸变;GARNN 模型在图 3(e1)、(e3)中未能增强细节信息与背景的对比度,在图 3(e2)、(e4)中细化并增强了部分细节特征,整体未能消除图案的畸变。DeepLab V3+模型和 AtG-DeepLab V3+模型在 4 种图案中均提取到了重要细节特征,并消除了图案的扭曲畸变,但从两者所恢复图像的局部细节放大对比可以看出 AtG-DeepLab V3+在图案分辨率和细节特征提取方面明显优于前者,在图 3(g1)(g2)中 AtG-DeepLab V3+所增强的图像中数字清晰、线条分明,分辨率高

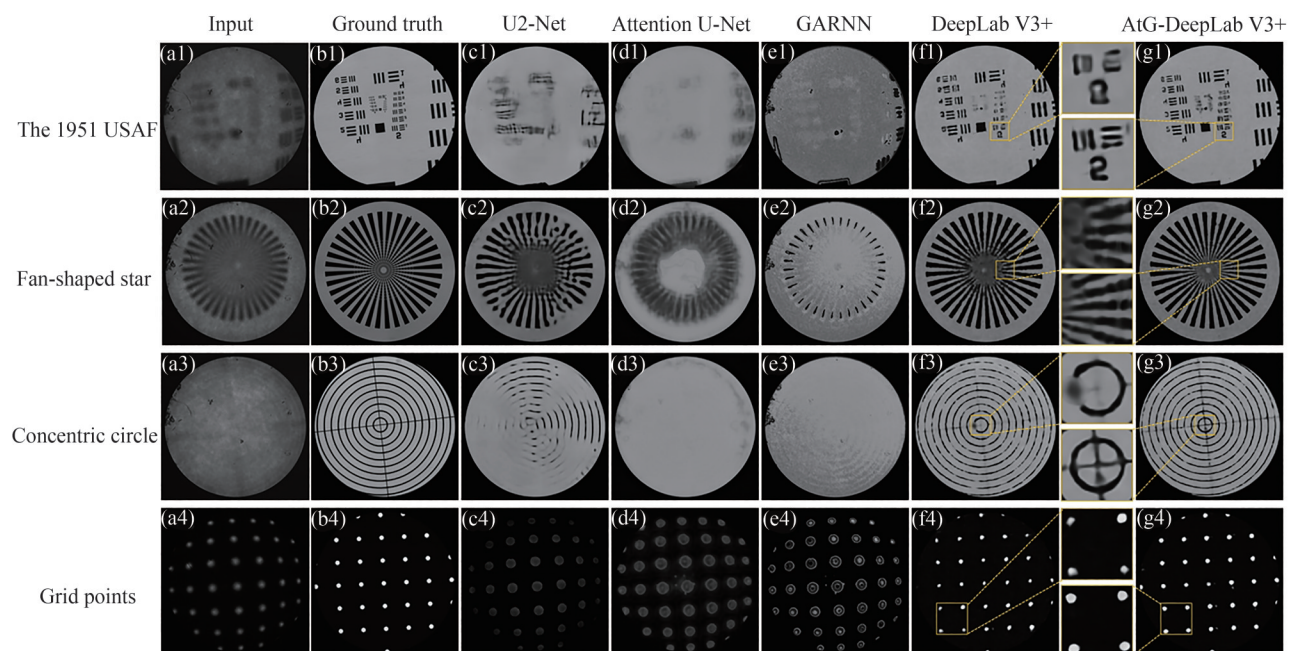


图 3 各模型增强测试靶图像效果对比

Fig.3 Comparison of test target images enhanced by each model

于 DeepLab V3+ 模型; 图 3(g3) 中 AtG-DeepLab V3+ 所增强的同心圆清晰均匀, 而 DeepLab V3+ 模型所增强的同心圆部分轮廓模糊, 且没有清晰地提取到十字架线条图案; 图 3(g4) 可以看出 AtG-DeepLab V3+ 模型所增强的网格点轮廓形状均近似圆形, 大小一致, 而 DeepLab V3+ 模型所增强的网格点轮廓形状各异, 大小不同。AtG-DeepLab V3+ 模型增强后的图像分辨率得到了显著提升, 如图 4 所示, 图 4(a) 为内窥采集分辨率板图案, 第 3 组 3 号元素能清晰分辨, 对应分辨率为 $49.61 \mu\text{m}$ 。图 4(b) 为 AtG-DeepLab V3+ 训练结果, 训练后最小能清晰分辨元素为第 4 组 6 号, 对应分辨率为 $17.54 \mu\text{m}$ 。

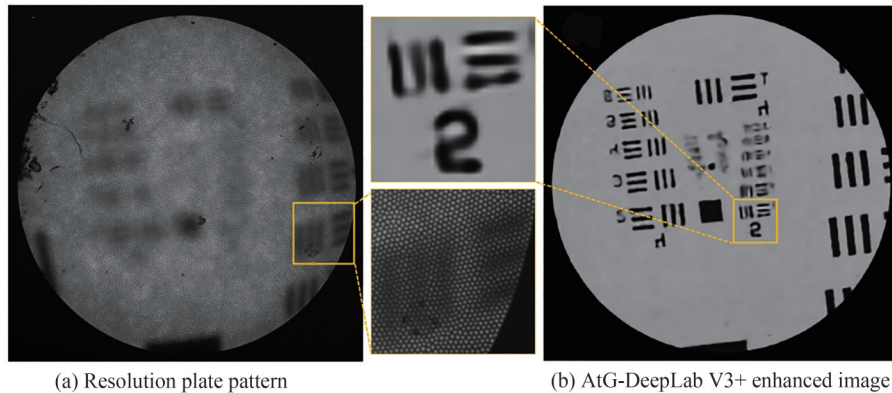


图 4 AtG-DeepLab V3+ 模型增强图像分辨率提升对比
Fig.4 Comparison of image resolution enhancement by AtG-DeepLab V3+ model

各模型增强图像的量化对比结果如图 5 所示, AtG-DeepLab V3+ 对内窥图像的增强效果在 PSNR (~ 27)、SSIM (~ 0.94) 和 NRMSE (~ 0.10) 三个评价方法的量化指标都优于其他模型, 分别高出 U2-Net 模型 54.2%、20.2% 和 68.8%, 高出 Attention U-Net 模型 66.3%、40.7% 和 71.5%, 高出 GARNN 模型 87.2%、52.4% 和 77.5%, 高出 DeepLab V3+ 模型 7.5%、2.1% 和 14.4%。值得注意的是, DeepLab V3+ 和 AtG-DeepLab V3+ 模型在图像增强后的量化指标上明显优于 U2-Net, Attention U-Net 和 GARNN 模型, 这与图 3 的整体图像恢复情况表现一致; 消融实验所得数据表明, 在 DeepLab V3+ 模型已经能较好地进行图像增强基础上, 引入自注意力机制的 AtG-DeepLab V3+ 模型仍在整体量化指标上进行了提升, 而其提升的部分重点来源于重要的细节特征还原 (如图 3 黄色框内放大部分的细节展示), 充分验证了增加自注意力模块的必要性。为验证模型稳定性, 对标准测试靶数据集进行 8 折内部交叉验证, 所得到的 PSNR, SSIM 和 NRMSE 指标分别为 27.83 ± 2.91 、 0.95 ± 0.03 和 0.19 ± 0.19 , 与测试集表现基本一致。

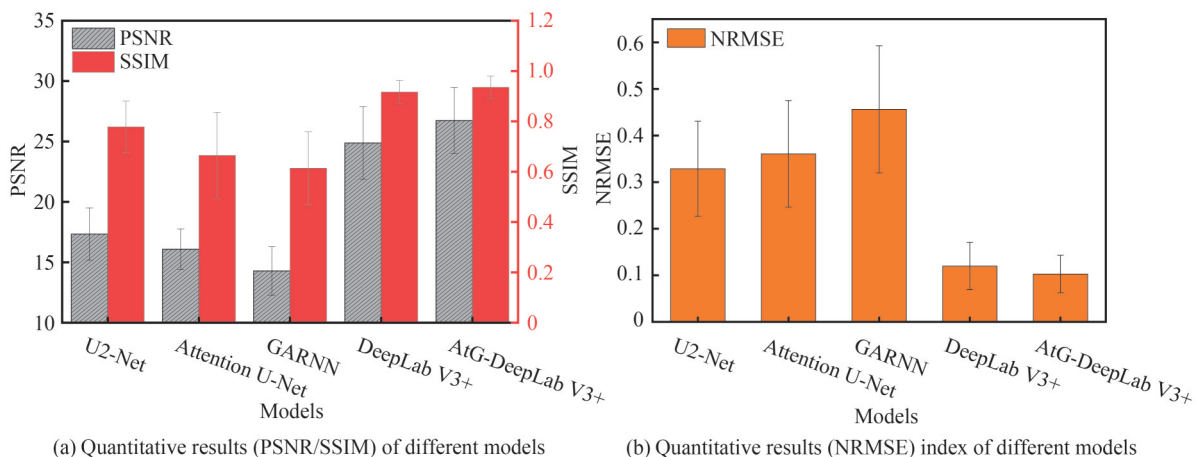


图 5 各模型增强测试靶图像评价指标对比
Fig.5 Comparison of evaluation indicators of test target images enhanced by each model

各模型的训练参数如表 2 所示, 各模型中 Attention U-Net 和 GARNN 训练完成后的单幅图像重建速度

最快,约为其余模型的40%,但 Attention U-Net 训练所占用的 GPU 内存最多,为 AtG-DeepLab V3+ 模型的221%,GARNN 网络训练时长最长,为 AtG-DeepLab V3+ 模型的185%。U2-Net 网络的训练时长和训练完成后的单幅图像重建速度与 AtG-DeepLab V3+ 接近,但其所占用的 GPU 内存也非常高,为 AtG-DeepLab V3+ 模型的219%。加入了自注意力机制优化后的 AtG-DeepLab V3+ 模型的训练参数与原始 DeepLab V3+ 模型较为接近,参数量增加了0.7%,训练时长增加了2.2%,GPU 占用量增加了3.2%,但训练完成后的单幅图像重建时间减少了4.2%。

表2 模型训练参数指标
Table 2 Model training parameter index

Model	GPU occupancy/MB	Training time/(h/200 epoch)	Parameter quantity/MB	Reconstruction rate/(s/piece ⁻¹)
U ² -Net	31 869	11	516	0.034 2
Attention U-Net	32 155	10.3	103	0.012 8
GARNN	19 752	17	22	0.013 7
DeepLab V3+	14 122	9	696	0.033 0
AtG-DeepLab V3+	14 570	9.2	701	0.031 6

综合表2与图3、4,可以发现 AtG-DeepLab V3+ 在训练时长和图像处理细节效果综合上存在优势。其在数据处理上的优势来源于使用了减少参数量并扩大感受野的深度可分离卷积。在图像处理效果上的优势主要来源于 DeepLab V3+ 模型的高级抽象特征提取和自注意力机制的重要空域特征选取这两个方面。DeepLab V3+ 模型的 ASPP 模块对于给定的输入以不同采样率的空洞卷积并行采样,以不同的比例捕捉图像的上下文,更好的提取不同层次的特征。通过进一步对自注意力机制的引入联合构建模型,有效地抑制了不相关区域的激活,更好地实现了对图像重要空域特征的筛选和提取,实现了内窥图像的高质量重建。

4.2 人体组织内窥成像训练结果分析

各模型增强后的生物组织图像如图6所示。其中 U2-Net 模型增强了组织与背景的对比度,在输入图像的基础上细化了部分特征,但相较于真实图像,未能实现准确完整的增强;Attention U-Net 和 GARNN 模型都仅增强了组织背景的对比度,但未能准确区分背景与组织,未能增强组织细节。DeepLab V3+ 模型和 AtG-DeepLab V3+ 模型均提取到了人体组织细节特征,并消除了图案的扭曲畸变,对内窥图像的增强效果

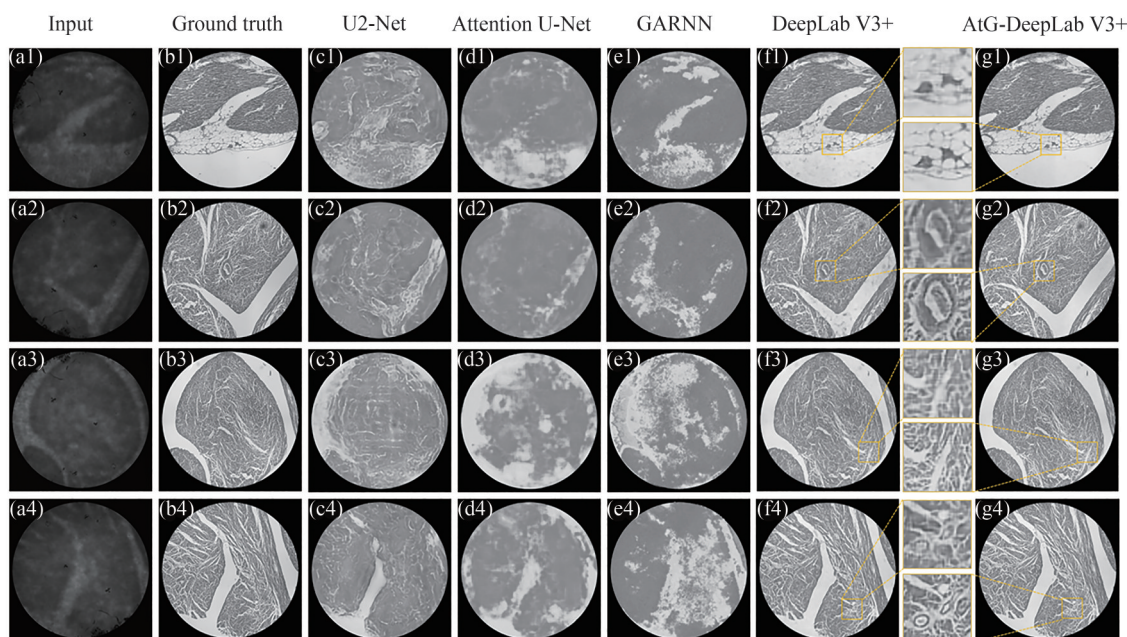


图6 各模型增强生物组织图像效果对比
Fig.6 Comparison of biological tissue images enhanced by each model

明显优于其他模型。如图6(f1)~(f4)、(g1)~(g4)所示,相较于DeepLab V3+模型,AtG-DeepLab V3+模型所恢复的人体组织纹理更清晰,细节更丰富,更加接近真实图像。

各模型增强图像的量化对比结果如图7所示,AtG-DeepLab V3+对内窥图像的增强效果在PSNR(~ 25)、SSIM(~ 0.78)和NRMSE(~ 0.10)三个评价方法的量化指标都优于其他模型,分别高出U2-Net模型66.4%、55.6%和71.6%,高出Attention U-Net模型51.9%、45.9%和62.4%,高出GARNN模型154.6%、231.5%和82.7%,高出DeepLab V3+模型13.0%、18.6%和28.3%。从评价指标上可以看出DeepLab V3+和AtG-DeepLab V3+模型增强后的图像在量化指标上明显优于U2-Net, Attention U-Net和GARNN模型,这与图6中各模型对内窥图像的增强效果表现一致;在DeepLab V3+模型已经能较好地恢复组织结构的基础上,引入自注意力机制的AtG-DeepLab V3+模型不仅在评价指标上有了进一步提升,而且实现了重要的组织细节特征还原(如图6黄色框内放大部分的细节展示),更具有生物医学内窥观测的实用意义。为验证模型稳定性,对生物组织样本数据集进行8折内部交叉验证,所得到的PSNR, SSIM和NRMSE指标分别为 24.79 ± 1.43 、 0.76 ± 0.05 和 0.10 ± 0.02 ,与测试集表现基本一致。

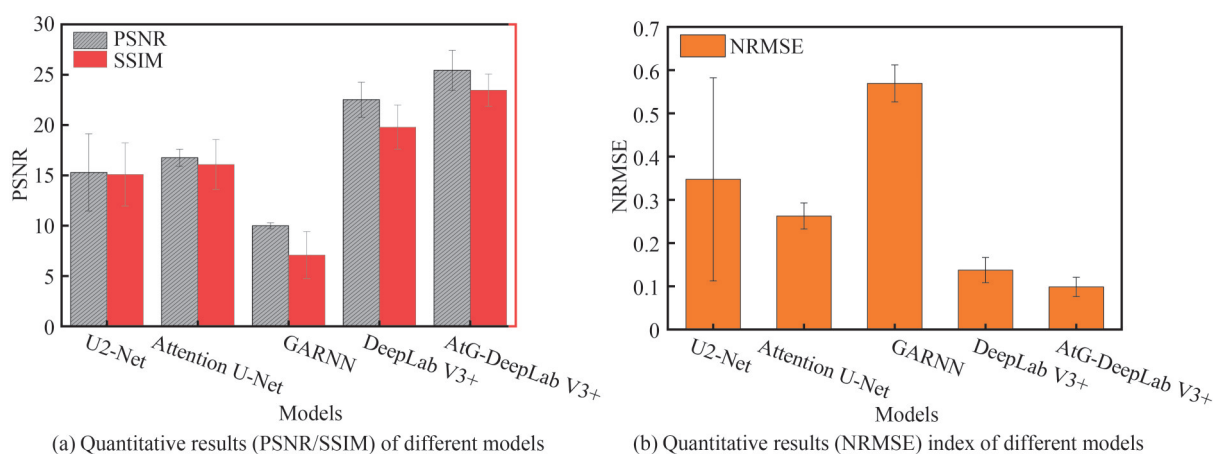


图7 各模型增强生物组织图像评价指标对比

Fig.7 Comparison of evaluation indicators of biological tissue image enhanced by each model

5 结论

本文针对光纤束内窥成像分辨率低、蜂窝图案严重、图像畸变严重等影响成像质量等问题,提出了基于注意力门控机制的AtG-DeepLab V3+卷积深度学习图像增强方案。通过在自主搭建的内窥-显微数据库上的应用显示,AtG-DeepLab V3+重建图像清晰地去除了光纤束端引入的蜂窝结构和图像畸变,将内窥图像分辨率从 $49.6 \mu\text{m}$ 提升至 $17.5 \mu\text{m}$ 。相比于最近发展的适用于光纤束蜂窝去除的GANRR网络,AtG-DeepLab V3+在成像测试靶图像重建明显更优,在PSNR和SSIM等量化指标上分别高出87.2%和52.4%。在对H&E染色的人类急性心肌梗死病理学组织成像中,AtG-DeepLab V3+能够从极低信噪比的原始内窥图像(PSNR=8.73)中重建高信噪比图像(PSNR=25.43),并精准还原出图像细节(SSIM由0.12提升至0.78)。同时,我们证明了引入注意力门控机制的重要性,AtG-DeepLab V3+在测试靶和组织成像重建应用中SSIM、PSNR、NRMSE均领先于U2-Net、Attention U-Net、GARNN和原始DeepLab V3+模型。本文构建的AtG-DeepLab V3+算法为基于光纤束的光学内窥图像增强提供了一个新的高速处理方案,引入的深度学习图像增强机制有望进一步集成于商用医学内窥,实现实时动态的内窥图像增强,促进内窥医疗的快速发展。与现有的针对成像系统的GARNN^[26]、3D CNN^[25]等图像增强模型相类似,本模型训练数据来源于单一成像系统,因此仅适用于与该系统具有相同结构特征的内窥图像增强任务,但我们相信通过多批次多类特征的内窥图像数据集进行大规模训练可以增强模型的泛化能力。此外,该模型还可通过级联最新的网络模型(如图神经网络)或者使用轻量化网络等方式进一步提高模型实现目标任务的效果和效率。

参考文献

- [1] YU Haiyang, ZHANG Huiying, LIU Simeng, et al. Research of fiber braking in the single fiber scanning endoscopic imaging system[J]. *Acta Photonica Sinica*, 2021, 50(5): 0511003.
于海洋,章慧英,刘思梦,等.单光纤扫描内窥成像制动技术研究[J].*光子学报*, 2021, 50(5): 0511003.
- [2] SHINDE A, MATHAM M V. Pixelate removal in an image fiber probe endoscope incorporating comb structure removal methods[J]. *Journal of Medical Imaging and Health Informatics*, 2014, 4(2): 203-211.
- [3] HAN J H, YOON S M. Depixelation of coherent fiber bundle endoscopy based on learning patterns of image prior[J]. *Optics Letters*, 2011, 36(16): 3212-3214.
- [4] CHEON G W, CHA J, KANG J U. Random transverse motion-induced spatial compounding for fiber bundle imaging[J]. *Optics Letters*, 2014, 39(15): 4368-4371.
- [5] SHAO J, LIAO W, LIANG R, et al. Resolution enhancement for fiber bundle imaging using maximum a posteriori estimation[J]. *Optics Letters*, 2018, 43(8): 1906-1909.
- [6] MIN J K, KWAK M S, CHA J M. Overview of deep learning in gastrointestinal endoscopy[J]. *Gut and Liver*, 2019, 13(4): 388.
- [7] SHARMA P, PANTE A, GROSS S A. Artificial intelligence in endoscopy[J]. *Gastrointestinal Endoscopy*, 2020, 91(4): 925-931.
- [8] ITOH T, KAWAHIRA H, NAKASHIMA H, et al. Deep learning analyzes *Helicobacter pylori* infection by upper gastrointestinal endoscopy images[J]. *Endoscopy International Open*, 2018, 6(2): E139-E144.
- [9] SHICHIJO S, NOMURA S, AOYAMA K, et al. Application of convolutional neural networks in the diagnosis of *Helicobacter pylori* infection based on endoscopic images[J]. *EBio Medicine*, 2017, 25: 106-111.
- [10] HIRASAWA T, AOYAMA K, TANIMOTO T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images[J]. *Gastric Cancer*, 2018, 21(4): 653-660.
- [11] BYRNE M F, CHAPADOS N, SOUDAN F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model[J]. *Gut*, 2019, 68(1): 94-100.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2015: 234-241.
- [13] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [14] WEIGERT M, SCHMIDT U, BOOTHE T, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy[J]. *Nature Methods*, 2018, 15(12): 1090-1097.
- [15] QIN X, ZHANG Z, HUANG C, et al. U²-Net: Going deeper with nested U-structure for salient object detection[J]. *Pattern Recognition*, 2020, 106: 107404.
- [16] TIAN Huijuan, ZHAI Jiahao, LIU Jianxin, et al. A low-quality iris image segmentation algorithm based on SRN-UNet[J]. *Acta Photonica Sinica*, 2022, 51(2): 0210006.
田会娟,翟佳豪,柳建新,等.基于SRN-UNet的低质量虹膜分割算法[J].*光子学报*, 2022, 51(2): 0210006.
- [17] KIM S H, KIM J, YANG S, et al. Automatic and quantitative measurement of alveolar bone level in OCT images using deep learning[J]. *Biomedical Optics Express*, 2022, 13(10): 5468-5482.
- [18] LIU T, LU Y, ZHANG Y, et al. A bone segmentation method based on Multi-scale features fuse U2Net and improved dice loss in CT image process[J]. *Biomedical Signal Processing and Control*, 2022, 77: 103813.
- [19] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[J]. *arXiv preprint arXiv:1804.03999*, 2018.
- [20] GUO C, SZEMENVEI M, YI Y, et al. Sa-unet: Spatial attention u-net for retinal vessel segmentation[C]. *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021: 1236-1242.
- [21] WANG B, QIU S, HE H. Dual encoding u-net for retinal vessel segmentation[C]. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2019: 84-92.
- [22] LIAN S, LUO Z, ZHONG Z, et al. Attention guided U-Net for accurate iris segmentation[J]. *Journal of Visual Communication and Image Representation*, 2018, 56: 296-304.
- [23] WANG J, LIU X. Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network[J]. *Computer Methods and Programs in Biomedicine*, 2021, 207: 106210.
- [24] RAVID, SZCZOTKA A B, PEREIRA S P, et al. Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy[J]. *Medical Image Analysis*, 2019, 53: 123-131.
- [25] SHAO J, ZHANG J, LIANG R, et al. Fiber bundle imaging resolution enhancement using deep learning[J]. *Optics Express*, 2019, 27(11): 15880-15890.

- [26] SHAO J, ZHANG J, HUANG X, et al. Fiber bundle image restoration using deep learning[J]. *Optics Letters*, 2019, 44(5): 1080-1083.
- [27] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 801-818.
- [28] HUYNH-THU Q, GHANBARI M. Scope of validity of PSNR in image/video quality assessment [J]. *Electronics Letters*, 2008, 44(13): 800-801.
- [29] 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.

AtG-DeepLab V3+ Endoscopic Image Enhancement Algorithm Based on Self-attention Mechanism Optimization

CHEN Jiajun¹, LI Kaixiang¹, LI Renjian^{1,2}, SHAO Chunlei¹, LI Guiye¹, CHEN Lingling¹

(1 *College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen 518118, China*)

(2 *College of Electronic Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China*)

Abstract: The optical endoscope imaging technology based on fiber bundle has the advantages of strong flexibility, no radiation and ease of integration. However, the structural characteristics of the fiber bundle inherently cause the honeycomb grid artefacts. In addition, the endoscope microlens introduces image distortion and low numerical aperture. As a result, there is a substantial decrease in the image quality for such endoscopic systems. In order to address this challenge, tremendous efforts have been made and a number of algorithms have been developed, such as spatial/frequency domain filtering, interpolation, etc., to successfully eliminate the image grid artefacts introduced by fiber bundle. Nevertheless, the image quality has not been substantially improved especially in terms of reduced spatial resolution and image distortion. In recent years, the image recognition and enhancement capabilities of Deep Learning (DL) have been significantly improved and thus, DL has been explored to apply for the reconstruction of endoscopic images, for example, using U2-Net, Attention U-net, and GARNN models. While the image quality could be improved to some extent, it remains challenging in the restoration of fine details. More importantly, until now, there has been no research on the utilization of DL to eliminate image distortion introduced by microlens in the endoscopic image. In order to address this challenge, we develop a new open source algorithm AtG-DeepLab V3+ by effectively integrating the self-attention mechanism and the DeepLab V3+ network for the enhancement processing of endoscopic images. This developed model adapts the coding-decoding network as a whole architecture and uses the ResNet101 network to extract features. Encoding context information by probing incoming features at multiple rates and multiple valid horizons or merging operations allows high-level abstract features to be extracted and clearer objects to be captured by progressively recovering spatial information. The self-attention gate integrated with the model decoder can effectively suppress the activation of irrelevant regions, better realize the screening and extraction of important spatial features of images, and achieve the distortion correction and high detail restoration of the endoscope image. In addition, we develop an optical endoscopy imaging system to collect test target images and biological tissue images for model training and testing. The training results of test target images demonstrate that the developed AtG-DeepLab V3+ model can effectively correct the intrapexial image distortion and remove the honeycomb grid structure artefacts. The results of local detail magnification and comparison illustrate that this model has a good performance in improving the image resolution limit and detail feature extraction compared with other current models (i.e., U2-Net, Attention U-net and GARNN). The training and testing results of biological tissue images exhibit that the fine detailed features of human tissue can be extracted by the AtG-DeepLab V3+ model, and the distortion of the pattern can be effectively eliminated. The

recovered images show a clearer tissue texture and richer fine details, which are highly similar to the real images. Compared with the existing U²-Net, Attention U-net and GARNN algorithms, the AtG-DeepLab V3+ model can achieve a much higher peak signal-to-noise ratio (increased by 66.4%, 51.9%, 154.6%), structural similarity (increased by 55.6%, 45.9%, 231.5%), demonstrating a significant improved enhancement effect of endoscopic image. We believe that this algorithm can provide a new efficient processing scheme for optical endoscopy image processing.

Key words: Endoscopic image; Deep learning; Image enhancement; Attention gate; Neural network

OCIS Codes: 170.2150; 200.4260; 100.3010