

融合Transformer和卷积的结直肠息肉分割算法

刘宏滨, 顾德

江南大学物联网工程学院, 江苏 无锡 214122

【摘要】结直肠息肉大小不一、形态各异,特别是边界模糊导致难以准确定位,小尺寸息肉容易漏检,使得息肉分割存在较大的挑战。针对以上问题,提出一种融合Transformer和卷积的息肉分割算法。首先使用Transformer提取图像全局特征,保证网络全局建模的能力,提高息肉主体区域和模糊边界的定位能力。然后,引入卷积加强网络对息肉细节的处理能力,细化边界分割效果,提高小尺寸息肉捕获能力。最后,将Transformer和卷积提取的特征进行深度融合,实现特征互补。将该算法在CVC-ClinicDB和Kvasir-SEG数据集上进行实验,其相似性系数分别为95.4%和93.2%,平均交并比分别为91.3%和88.6%。进一步在CVC-ColonDB、CVC-T和ETIS数据集上测试泛化能力,其相似性系数分别为81.3%、90.9%和80.1%。结果表明,所提算法在息肉分割的准确度上有较大的提升。

【关键词】息肉分割;特征融合;Transformer;卷积

【中图分类号】R318;TP183

【文献标志码】A

【文章编号】1005-202X(2024)03-0316-07

Colorectal polyp segmentation algorithm integrating Transformer and convolution

LIU Hongbin, GU De

School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China

Abstract: In response to the challenges of varied sizes and diverse shapes of colorectal polyps, especially with blurred boundaries that often complicates localization and smaller polyps being particularly prone to oversight, a colorectal polyp segmentation algorithm integrating Transformer and convolution is proposed. Transformer is employed to extract global features from images for ensuring the network's capability for global modeling and improving the localization capability for both main polyp regions and vague boundaries. Subsequently, convolution is introduced to augment the network's ability to process polyp details, refining boundary segmentation and enhancing the capture capability for small-sized polyps. Finally, a deep fusion of the features extracted by Transformer and convolution is carried out to realize feature complementarity. The experimental evaluation using CVC-ClinicDB and Kvasir-SEG datasets show that the algorithm has similarity coefficients of 95.4% and 93.2%, and mean intersection over union of 91.3% and 88.6%, respectively. Further tests on the generalization capability of the algorithm are conducted on CVC-ColonDB, CVC-T, and ETIS datasets, in which similarity coefficients of 81.3%, 90.9% and 80.1% are obtained. The results indicate a notable improvement in the accuracy of polyp segmentation achieved by the proposed algorithm.

Keywords: polyp segmentation; feature fusion; Transformer; convolution

前言

结直肠癌是一种常见的癌症,其死亡率排在所有癌症的前三位^[1]。结直肠息肉是结直肠内部生长

的肿瘤或增生物,具有潜在的恶性转化风险,这使得结直肠息肉成为结直肠癌的主要风险因素。在临床上,结肠镜检查可以有效提供结直肠息肉的位置和状态信息,帮助完成对息肉的分割处理。然而,息肉的形态多样,大小不一,颜色和纹理与周围的正常组织相似,边缘不清晰,导致息肉分割的准确率不高。自动息肉分割技术不仅可以提高诊断的准确性和可靠性,而且可以节省医疗资源和人力成本,提高工作效率,具有重要的现实意义^[2]。

自动息肉分割技术主要有两类,一种是基于手工特征的传统方法,另一种是基于深度学习的方法。

【收稿日期】2023-12-20

【基金项目】江苏省自然科学基金(BK20231036, BK20180594)

【作者简介】刘宏滨, 硕士, 研究方向:深度学习、医学图像处理, E-mail: 2469607141@qq.com

【通信作者】顾德, 博士, 副教授, 研究方向:无线传感器网络拓扑识别和基于图像的生物医学信息识别, E-mail: gude@jiangnan.edu.cn

传统方法所依赖的手工特征通常包括颜色、纹理和形状特征,在特定情况下传统方法可以获得比较好的分割效果^[3]。但是由于手工特征的表征能力差,当息肉形状复杂时分割的准确度很难保证^[4]。随着深度学习的快速发展,基于深度学习的息肉分割方法越来越多地得到应用。

一种基于深度学习的息肉分割方法是卷积神经网络方法,卷积神经网络因其具有局部感知和平移不变等特性,在语义分割任务中得到了广泛应用并取得显著成果。其中,Long等^[5]提出使用反卷积操作来获得与原始输入图像相同尺寸的输出,在语义分割领域取得突破性的进展。Ronneberger等^[6]提出U型网络,在生物医学图像分割中成功应用。Zhou等^[7]为更好地融合和利用不同深度的特征,在U-Net模型的基础上引入嵌套和密集的捷径连接,提高整个息肉区域的分割效果^[6]。Fan等^[8]提出并行反向注意力模块,实现更精准的息肉定位。尽管以卷积为基础进行改进的网络在息肉的细节分割上都取得了不错的效果^[9-10],但是因为卷积神经网络受感受野的限制,使得模型在关注上下文信息时仍然存在难题。

另一种基于深度学习的息肉分割方法是Transformer方法,近年来Transformer在自然语言处理领域大放异彩,因此出现了将Transformer用于计算机视觉的模型^[11-12]。许多基于视觉Transformer的网络在息肉分割任务中取得了不错的效果^[13-14]。Xie等^[15]将分层Transformer作为编码器,有效实现Transformer在息肉分割中的应用。Duc等^[16]提出可以在多尺度特征上对全局语义信息进行建模的网络,在息肉分割上取得很好的效果。然而,基于Transformer的网络为发挥全局建模的能力对全局特征图进行处理,导致全局特征的均一化趋势,使其在预测图像细节方面的能力受到限制^[17-18]。

针对上述工作中存在的一些问题,本文提出一种Transformer和卷积并行处理的双分支网络结构,Transformer分支重点提取全局特征,加强对息肉主体区域的定位能力,卷积分支补充细节信息,细化边缘分割。设计深浅特征桥接(Deep and Shallow Feature Bridging, DSFB)模块来提高对浅层特征的抓取能力,设计特征后处理(Feature Post-Processing, FPP)模块对深层和浅层特征进行特征处理,设计特征整合输出(Feature Integration Output, FIO)模块高效地实现了双分支特征融合。

1 本文方法

本文提出一种双分支网络,网络整体结构如图1所示,该网络主要包括3部分,即Transformer分支

(TB)、卷积分支(CB)和双分支特征整合输出模块。TB和CB并行独立下采样提取特征,其中,TB重点捕获全局上下文信息,CB关注局部特征,使得模型能够同时理解图像的局部细节和全局结构,从而高效地进行息肉分割。

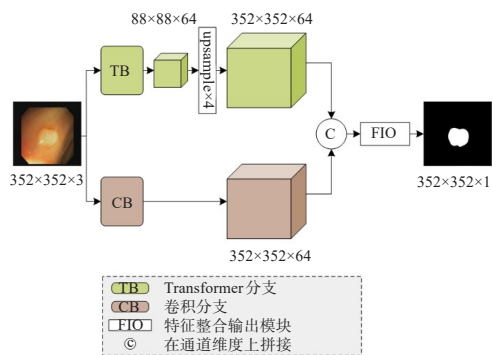


图1 本文网络整体结构

Figure 1 Overall structure of the proposed network

1.1 Transformer分支

Transformer分支的整体结构如图2所示,该分支的结构设计受到U-Net中U型框架的影响^[6],使用了编码器-解码器的结构。Transformer分支中,使用PVTv2作为编码器提取息肉图像的全局特征^[19],PVTv2的Stage结构如图3所示。PVTv2以视觉Transformer为基础,通过堆叠4个阶段的模块来实现。输入图像通过补丁嵌入将原图分割成4x4大小的补丁,然后将数据展平得到可以输入Transformer编码器的格式。首个阶段将图像处理成原图尺寸的四分之一,而后每个阶段的特征图尺寸都将减半,该下采样操作可以得到4个不同尺寸的特征图。多尺度特征能够捕获多层次信息,增强对不同尺寸息肉的识别和分割能力^[20]。Transformer编码器的前向过程通过深度卷积和零填充实现位置信息的学习,代替了固定位置编码。这样PVTv2编码器编码过程中得到的特征既可以保留全局上下文信息又可以实现多尺度特征提取。

PVTv2编码器得到的不同尺度特征包含浅层的高分辨率特征和深层的低分辨率特征,高分辨率特征图对应着更多息肉细节信息,低分辨率特征图对应着更多空间语义信息^[21]。息肉的边界和纹理在浅层特征中得到更好的表达,而息肉的类别和形态在深层特征中得到更好的识别。融合这两种特征能够帮助模型同时捕获细节和语义信息,解决模糊边界难以准确定位的问题,因此本文设计了多尺度特征融合解码器MFFD,MFFD包含4个DSFB模块和3个FPP模块,结构上为逐级连接,MFFD的结构图见图2,DSFB、FPP和所用残差模块的结构如图4所示。

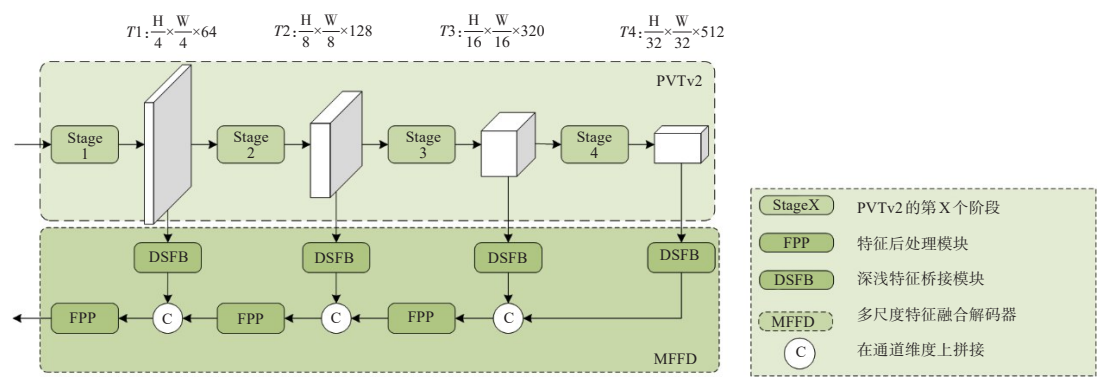


图 2 Transformer 分支结构

Figure 2 Structure of the Transformer branch

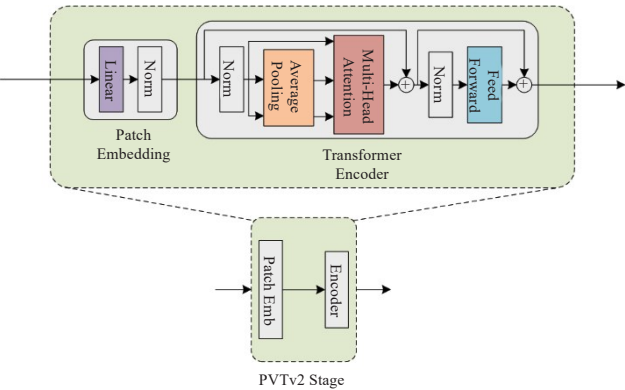


图 3 PVTv2 网络结构

Figure 3 PVTv2 network architecture

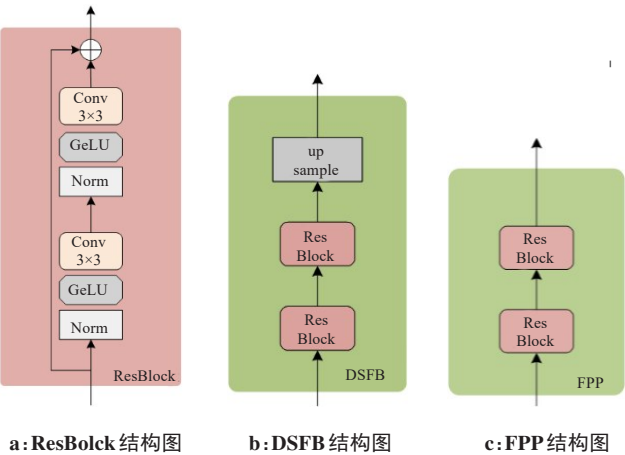


图 4 Transformer 分支中的模块

Figure 4 Modules in the Transformer branch

DSFB 模块将 PVTv2 编码器得到的 4 个不同级别的特征图进行过滤,此操作旨在滤除息肉背景和周围组织区域的冗余信息,保留更重要的空间语义信息。4 个不同级别特征图通过两次残差模块过滤后统一将尺寸上采样到 88×88 以便于后续融合。残差模块的设计受到 ResNet^[22]和 Transformer^[11]模块的启发,使用层标准化提高训练的稳定性 and 收敛速度,同时 GeLU 激活

函数能够引入非线性,帮助模型学习更复杂的特征表示。卷积操作能够在保留空间信息的同时学习局部特征。残差连接可以帮助网络学习恒等映射,保留原始特征信息。总的来说,DSFB 模块用来丰富多尺度特征表示,提取关键信息,统一特征图的尺寸。FPP 模块堆叠了两次残差模块,将多个级别的特征图自上而下逐级拼接融合,细化网络特征。

1.2 卷积分支

卷积分支的设计如图 5 所示,整体框架设计仍然遵循 U 型架构^[6],卷积分支的编码器部分通过 4 次下采样获得 4 个不同尺寸特征图,每次下采样都将特征图的尺寸减半。解码器实现逐级融合,提高特征图的分辨率直到达到原图尺寸。高分辨率特征图能够保留更多的细节信息,包括息肉边缘、纹理和形状等,能够提供更精细的空间信息,并且可以提供更多的像素信息,有助于改善小尺寸息肉的识别和分割,降低漏检率。因此卷积分支为了能够保留高分辨率特征图,首次下采样仅将特征图尺寸减半,即卷积分支获得的特征图尺寸是 Transformer 分支特征图的两倍。

为了提高编码器的特征提取能力,本文按照 ConvNeXt 的网络设计了卷积模块(Conv Block)^[23],如图 6 所示。卷积模块通过 7×7 的深度卷积扩大感受野,通过层标准化来防止过拟合。使用两次 1×1 的卷积操作来调整通道数,模块中使用 GeLU 激活函数来引入非线性,使用了通道缩放来帮助网络捕捉更多的特征信息以改进网络的性能^[24]。模块最后引入残差连接来避免梯度消失,加快网络的收敛速度。解码器部分逐级实现多尺度特征融合,帮助网络在上采样中捕获不同尺度的特征信息,以适应不同大小的息肉分割。逐级融合后的上采样操作逐步恢复图像分辨率,同时保留深层的语义信息和浅层的细节信息。卷积分支的设计提升了网络在像素级别细节信息上预测分割的能力。

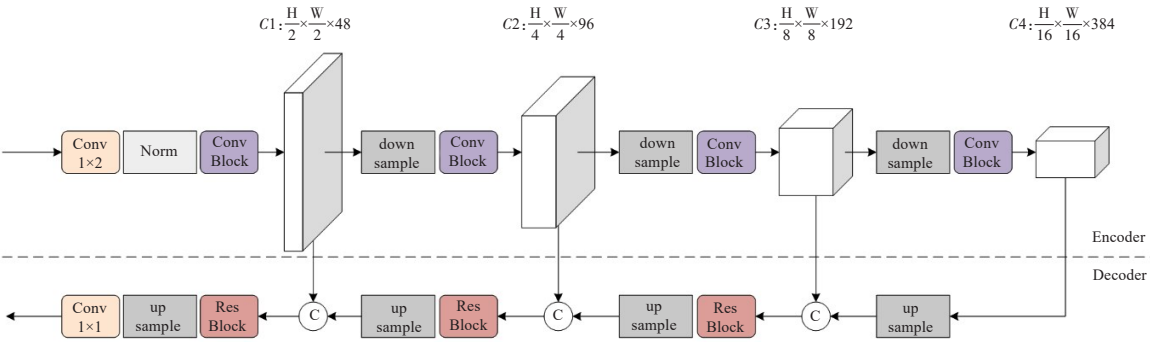


图5 卷积分支
Figure 5 Convolution branch

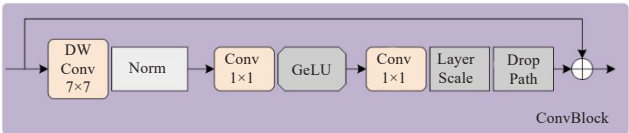


图6 卷积模块
Figure 6 Convolution block

1.3 双分支特征整合输出

FIO 模块用于融合卷积分支和 Transformer 分支的输出。该模块包括两个残差模块和一个 1×1 卷积层,最后通过 Sigmoid 激活函数输出融合结果。FIO 模块结构如图 7 所示。通过残差模块,FIO 能够优化整合的特征,使整合后的特征能够保留重要信息,并且有助于网络的训练和收敛。FIO 模块提供一个简单而有效的方式来融合和优化特征,相比于复杂的融合结构,FIO 模块能够在保持网络性能的同时,简化网络结构,为双分支网络提供一个强有力的特征表示。

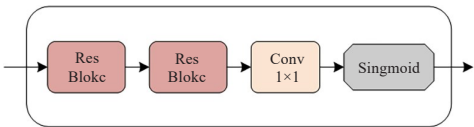


图7 特征整合输出模块
Figure 7 Feature integration output module

2 实验结果与分析

2.1 数据集及评估指标

为了验证本文算法的学习能力和泛化能力,使用以下 5 个公开的结直肠息肉数据集进行训练和测试:CVC-ClinicDB^[25]、Kvasir-SEG^[26]、CVC-ColonDB^[27]、CVC-T^[28]和 ETIS^[29]。其中 CVC-ClinicDB 由医学图像计算与计算机辅助国际会议(MICCAI)息肉分割

挑战赛于 2015 年发布,Kvasir-SEG 由挪威奥斯陆大学医院内窥镜专家采集并标注,CVC-ColonDB 是由都柏林城市大学机器视觉组提供的息肉数据集,CVC-T 是一个常用息肉分割测试集,ETIS 数据集由 MICCAI 息肉分割挑战赛于 2017 年发布。数据集的图像数量和图像尺寸信息如表 1 所示。

表 1 数据集详细信息
Table 1 Details of the datasets

| 数据集 | 图像数量 | 图像尺寸 |
|--------------|-------|---------------------|
| Kvasir-SEG | 1 000 | 720×576~1 920×1 072 |
| CVC-ClinicDB | 612 | 384×288 |
| CVC-ColonDB | 380 | 574×500 |
| ETIS | 196 | 1 225×966 |
| CVC-T | 60 | 574×500 |

为了全面评估结直肠息肉分割模型的性能,本文采用几个在息肉分割任务中非常流行的指标进行定量评估,包括平均相似性系数(mDice)、平均交并比(mIoU)、精准率(Precision)和召回率(Recall):

$$mDice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{1}$$

$$mIoU = \frac{TP}{TP + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

其中,TP指的是被模型正确分割为息肉的样本数;FP指的是被模型错误分割为息肉的样本数;FN指的是被模型错误分割为背景的样本数。mDice、mIoU、精准率和召回率数值越高表示分割效果越好。

2.2 实验细节

本文实验使用深度学习框架为Pytorch 1.12.1,使用V100 GPU来加速。为了保证算法对比的客观性,本文的实验设置保持一致,按照PraNet的实验将输入图像尺寸都调整到352×352^[8],使用了高斯模糊、水平和垂直翻转、仿射变换等数据增强方法。使用AdamW优化器,初始学习率为0.000 1,迭代次数设置为200,每次迭代的批处理量为8。学习能力实验中分别将数据集Kvasir-SEG和CVC-ClinicDB以8:1:1的比例随机划分为训练集、验证集和测试集,随机数为41。泛化能力实验中将Kvasir-SEG和CVC-ClinicDB的训练集和验证集分别组合作为合并训练集和合并验证集,其中每个数据集的图像文件顺序不变,保持原始数据集的结构和内容完整性。实验使用了二值交叉熵损失(Binary Cross Entropy, BCE)与平均相似性系数损失的和作为模型的损失:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (5)$$

$$L_{total} = L_{BCE} + L_{Dice} \quad (6)$$

其中, N 表示样本数量, y_i 表示第*i*个样本的实际标签, \hat{y}_i 表示第*i*个样本的预测值。

2.3 学习能力验证

评估学习能力可以验证模型在息肉分割时的有效性。通过实验评估本文算法在数据集Kvasir-SEG和CVC-ClinicDB上的学习能力,与目前经典的5个模型进行比较,包括U-Net^[6]、UNet++^[7]、PraNet^[8]、UACANet^[30]、ColonFormer^[16],相应的实验结果如表2所示。

从表2的实验结果可以看出,本文算法在CVC-ClinicDB数据集上的4个指标均高于其他算法。具体来说,相比于基线网络U-Net, mDice、mIoU、精准率、召回率4项指标分别提升3.9%、4.8%、1.7%和

表2 不同算法在CVC-ClinicDB和Kvasir-SEG数据集上的结果

Table 2 Results of different algorithms on CVC-ClinicDB and Kvasir-SEG datasets

| 数据集 | 算法 | mDice | mIoU | 精准率 | 召回率 |
|--------------|-------------|-------|-------|-------|-------|
| CVC-ClinicDB | U-Net | 0.915 | 0.865 | 0.938 | 0.918 |
| | UNet++ | 0.845 | 0.756 | 0.832 | 0.892 |
| | PraNet | 0.907 | 0.858 | 0.913 | 0.923 |
| | UACANet | 0.910 | 0.865 | 0.911 | 0.917 |
| | ColonFormer | 0.942 | 0.897 | 0.949 | 0.951 |
| | 本文算法 | 0.954 | 0.913 | 0.955 | 0.954 |
| Kvasir-SEG | U-net | 0.863 | 0.818 | 0.890 | 0.909 |
| | Unet++ | 0.748 | 0.631 | 0.887 | 0.687 |
| | PraNet | 0.908 | 0.856 | 0.935 | 0.903 |
| | UACANet | 0.901 | 0.856 | 0.938 | 0.889 |
| | ColonFormer | 0.927 | 0.877 | 0.936 | 0.923 |
| | 本文算法 | 0.932 | 0.886 | 0.941 | 0.929 |

3.6%;在Kvasir-SEG数据集上的4项指标分别为93.2%、88.6%、94.1%和92.9%,也均高于其他对比算法。实验结果表明本文算法在学习能力表现上相比于其他网络提升比较明显,能更精准地实现息肉分割。

图8展示了不同算法的分割结果可视化。图8a表示原始图像,图8b表示标签图,图8c~h分别表示U-Net、UNet++、PraNet、UACANet、ColonFormer及本文方法的分割结果。

2.4 泛化能力分析

评估息肉分割模型的泛化能力是非常重要的,在实际的临床环境中,病人的肠道状况和息肉的类型、形态、大小等可能存在很大的多样性,有些情况

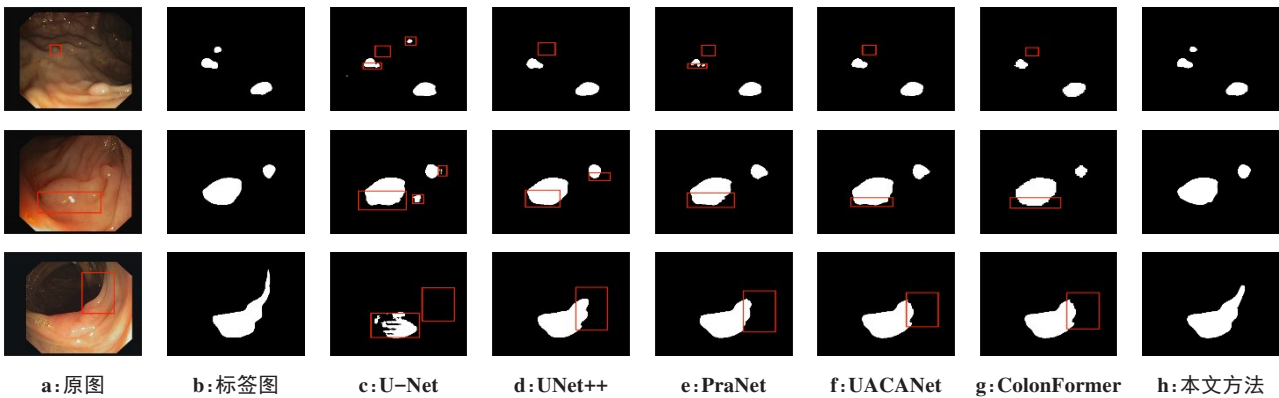


图8 不同算法的分割结果可视化

Figure 8 Visualization of segmentation results obtained by different algorithms

第1~3行展示了数据集中3张不同图像在不同算法下的分割效果,其中第1行图像选自CVC-ClinicDB,第2、3行图像选自Kvasir-SEG

在训练数据集中可能并未出现,因此需要确保模型具有良好的泛化能力能够适应各种变化。数据集 Kvasir-SEG 和 CVC-ClinicDB 有不同分辨率和不同背景复杂度的息肉图像,包含各种类型的息肉图像信息,因此本文实验在二者的合并数据集上进行训练,在 CVC-ColonDB、ETIS 和 CVC-T 上做泛化能力测试。相应的实验结果如表 3 所示。

表 3 不同算法在 CVC-ColonDB、ETIS 和 CVC-T 数据集上的结果
Table 3 Results of various algorithms on CVC-ColonDB, ETIS, and CVC-T datasets

| 数据集 | 算法 | mDice | mIoU | 精准率 | 召回率 |
|-------------|-------------|-------|-------|-------|-------|
| CVC-ColonDB | U-Net | 0.512 | 0.444 | 0.542 | 0.539 |
| | UNet++ | 0.483 | 0.410 | 0.526 | 0.517 |
| | PraNet | 0.712 | 0.640 | 0.753 | 0.739 |
| | UACANet | 0.783 | 0.704 | 0.825 | 0.812 |
| | ColonFormer | 0.811 | 0.730 | 0.767 | 0.764 |
| | 本文算法 | 0.813 | 0.731 | 0.845 | 0.824 |
| ETIS | U-Net | 0.398 | 0.335 | 0.455 | 0.414 |
| | UNet++ | 0.401 | 0.344 | 0.461 | 0.447 |
| | PraNet | 0.628 | 0.567 | 0.631 | 0.665 |
| | UACANet | 0.766 | 0.689 | 0.708 | 0.728 |
| | ColonFormer | 0.789 | 0.711 | 0.733 | 0.768 |
| | 本文算法 | 0.801 | 0.733 | 0.800 | 0.833 |
| CVC-T | U-Net | 0.710 | 0.627 | 0.684 | 0.791 |
| | UNet++ | 0.707 | 0.624 | 0.675 | 0.800 |
| | PraNet | 0.851 | 0.797 | 0.826 | 0.903 |
| | UACANet | 0.910 | 0.849 | 0.868 | 0.942 |
| | ColonFormer | 0.894 | 0.826 | 0.815 | 0.895 |
| | 本文算法 | 0.909 | 0.842 | 0.876 | 0.957 |

从表 3 数据可以看出本文算法在 CVC-ColonDB 和 ETIS 上提升较大,取得最优结果,mDice 和 mIoU 在 CVC-ColonDB 上分别达到 81.3% 和 73.1%,与 U-Net 相比分别提升 30.1% 和 28.7%,在 ETIS 上达到 80.1% 和 73.3%,与 U-Net 相比分别提升 40.3% 和 39.8%。因为 CVC-T 数据集较小,所以本文算法在该数据集上的 mDice 和 mIoU 没有达到最高,略低于 UACANet,但是精准率和召回率仍高于 UACANet。

2.5 消融实验

为了评估本文模型每个模块的作用,本文通过一系列消融实验来定量评估每个模块的有效性。在 Kvasir-SEG 和 CVC-ClinicDB 数据集上进行实验,实验结果如表 4 所示。

表 4 消融实验结果
Table 4 Ablation experiment results

| 数据集 | 消融设置 | mDice | mIoU | 精准率 | 召回率 |
|---------------|----------|-------|-------|-------|-------|
| CVC- ClinicDB | w/o FIO | 0.942 | 0.899 | 0.946 | 0.950 |
| | w/o FPP | 0.941 | 0.900 | 0.939 | 0.951 |
| | w/o DSFB | 0.947 | 0.902 | 0.943 | 0.947 |
| | w/o CB | 0.941 | 0.896 | 0.936 | 0.956 |
| | 本文模型 | 0.954 | 0.913 | 0.955 | 0.954 |
| Kvasir-SEG | w/o FIO | 0.901 | 0.845 | 0.924 | 0.910 |
| | w/o FPP | 0.903 | 0.845 | 0.936 | 0.899 |
| | w/o DSFB | 0.902 | 0.846 | 0.950 | 0.890 |
| | w/o CB | 0.892 | 0.839 | 0.934 | 0.886 |
| | 本文模型 | 0.932 | 0.886 | 0.941 | 0.929 |

本文将 PVTv2 作为消融实验的基线,并从本文模型中移除或者更换模块进行具体实验,比较移除或者更换前后的结果来评估模块的有效性。标准版本表示为“本文模型(PVTv2+DSFB+FPP+FIO+CB)”。其中,CB 表示整个卷积分支。

将 DSFB 模块移除,训练了“without DSFB(w/o DSFB)”版本。与本文模型相比较,移除 DSFB 模块的模型在两个数据集上的性能表现都出现了大幅的下降。尤其在 Kvasir-SEG 上的 mDice 指标下降明显,下降幅度为 3%。表明了 DSFB 能够有效感知整体区域,减弱背景噪声的影响,尤其对背景复杂度高的图像效果更加明显。

用两次卷积核大小为 3×3,步长为 1 的卷积操作代替特征后处理模块 FPP,版本表示为“w/o FPP”。与标准模型相比,该版本在两个数据集上的性能也出现下降。在 CVC-ClinicDB 和 Kvasir-SEG 上的 mDice 指标分别下降 1.3% 和 2.9%。可以验证 FPP 模块在特征图拼接后实现了对不同尺寸特征图的有效处理。

移除特征整合输出模块 FIO,训练版本“w/o FIO”。与标准模型相比,在 CVC-ClinicDB 和 Kvasir-SEG 上的 mDice 指标分别下降 1.2% 和 3.1%。可以验证 FIO 模块对两个独立下采样分支的特征整合起到显著效果。

移除卷积分支 CB,训练版本“w/o CB”,结果显示移除 CB 后模型的各项性能指标下降幅度最大,在 CVC-ClinicDB 和 Kvasir-SEG 数据集上的 mDice 分别下降 1.3% 和 4%。这表明卷积分支对于完整模型的重要作用,卷积分支的存在强化了模型对于图像细节的感知能力,为模型的细化分割起到补充作用,为更精确的预测提供了关键支持。

3 结 论

针对结直肠息肉图像中息肉大小存在差异、形状不规则、分布不均匀、边界模糊等复杂特征,提出了一种融合Transformer分支和卷积分支的息肉分割算法。其中,Transformer分支和卷积分支独立并行计算,Transformer分支的编码器发挥全局建模优势整合上下文信息,并且通过多尺度特征融合解码器将浅层特征和深层特征进行信息交互;卷积分支借助U型架构补充图像的局部特征和纹理,增强息肉边界和小尺寸息肉分割能力;特征整合输出模块将全局特征和局部特征充分融合,整体提高息肉分割准确度。为了验证上述方法的可行性,本文在5个息肉数据集上进行实验来证明。实验结果表明,与目前公开的其他息肉分割模型相比,本文提出的方法表现出了一定的优势,模型的学习能力和泛化能力都得到了很大程度的提升。

【参考文献】

- [1] Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023[J]. CA Cancer J Clin, 2023, 73(1): 17-48.
- [2] 朱世祺,徐昶,周鑫,等.基于DeepLab V3+深度神经网络的结直肠息肉内镜图像分割[J].中国医学物理学杂志, 2023, 40(8): 944-949. Zhu SQ, Xu C, Zhou X, et al. Colorectal polyp segmentation in endoscopic images using DeepLab V3+ [J]. Chinese Journal of Medical Physics, 2023, 40(8): 944-949.
- [3] Yue GH, Li SY, Cong RM, et al. Attention-guided pyramid context network for polyp segmentation in colonoscopy images[J]. IEEE Trans Instrum Meas, 2023, 72: 1-13.
- [4] Liu GQ, Jiang Y, Liu D, et al. A coarse-to-fine segmentation frame for polyp segmentation via deep and classification features[J]. Expert Syst Appl, 2023, 214: 118975.
- [5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 3431-3440.
- [6] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [7] Zhou ZW, Rahman Siddiquee MM, Tajbakhsh N, et al. UNet++: a nested U-Net architecture for medical image segmentation[C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer International Publishing, 2018: 3-11.
- [8] Fan DP, Ji GP, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer International Publishing, 2020: 263-273.
- [9] Srivastava A, Jha D, Chanda S, et al. MSRF-net: a multi-scale residual fusion network for biomedical image segmentation[J]. IEEE J Biomed Health Inform, 2022, 26(5): 2252-2263.
- [10] Jin QC, Hou HY, Zhang GX, et al. FEGNet: a feedback enhancement gate network for automatic polyp segmentation[J]. IEEE J Biomed Health Inform, 2023, 27(7): 3420-3430.
- [11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Adv Neural Inf Process Syst. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [12] Liu Z, Lin YT, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 9992-10002.
- [13] Sanderson E, Matuszewski BJ. FCN-transformer feature fusion for polyp segmentation[C]//Annual Conference on Medical Image Understanding and Analysis. Cham: Springer International Publishing, 2022: 892-907.
- [14] Wang JF, Huang QM, Tang FL, et al. Stepwise feature fusion: local guides global[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2022. Cham: Springer Nature Switzerland, 2022: 110-120.
- [15] Xie EZ, Wang WH, Yu ZD, et al. SegFormer: simple and efficient design for semantic segmentation with transformers [C]//35th Conference on Neural Information Processing Systems (NeurIPS 2021). 2021: 12077-12090.
- [16] Duc NT, Oanh NT, Thuy NT, et al. ColonFormer: an efficient transformer based method for colon polyp segmentation[J]. IEEE Access, 2022, 10: 80575-80586.
- [17] Park KB, Lee JY. SwinE-Net: hybrid deep learning approach to novel polyp segmentation using convolutional neural network and Swin transformer[J]. J Comput Des Eng, 2022, 9(2): 616-632.
- [18] Wang WH, Xie EZ, Li X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 548-558.
- [19] Wang WH, Xie EZ, Li X, et al. PVT v2: improved baselines with pyramid vision transformer [J]. Comput Vis Media, 2022, 8(3): 415-424.
- [20] Zhang YD, Liu HY, Hu Q. Transfuse: fusing transformers and CNNs for medical image segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 14-24.
- [21] Wei J, Hu YW, Zhang RM, et al. Shallow attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 699-708.
- [22] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [23] Liu Z, Mao HZ, Wu CY, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 11966-11976.
- [24] Touvron H, Cord M, Sablayrolles A, et al. Going deeper with image transformers [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 32-42.
- [25] Bernal J, Sánchez FJ, Fernández-Esparrach G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians[J]. Comput Med Imaging Graph, 2015, 43: 99-111.
- [26] Jha D, Smedsrud PH, Riegler MA, et al. Kvasir-SEG: a segmented polyp dataset [C]//MultiMedia Modeling. Cham: Springer International Publishing, 2020: 451-462.
- [27] Tajbakhsh N, Gurudu SR, Liang JM. Automated polyp detection in colonoscopy videos using shape and context information[J]. IEEE Trans Med Imaging, 2016, 35(2): 630-644.
- [28] Vázquez D, Bernal J, Sánchez FJ, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. J Healthc Eng, 2017, 2017: 4037190.
- [29] Silva J, Histace A, Romain O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer[J]. Int J Comput Assist Radiol Surg, 2014, 9(2): 283-293.
- [30] Kim T, Lee H, Kim D. UACANet: uncertainty augmented context attention for polyp segmentation[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery, 2021: 2167-2175.

(编辑:薛泽玲)