

DOI:10.19651/j.cnki.emt.2416275

融合局部和全局特征的息肉分割模型^{*}张攀峰^{1,2} 杨贺¹ 神显豪^{2,3} 程小辉^{1,2} 杜慧⁴

(1. 桂林理工大学计算机科学与工程学院 桂林 541006; 2. 广西嵌入式技术与智能系统重点实验室 桂林 541006;
3. 桂林理工大学物理与电子信息工程学院 桂林 541006; 4. 桂林理工大学公共管理学院 桂林 541006)

摘要: 针对现有模型在息肉分割中存在复杂区域分割困难、边缘细节信息丢失、泛化能力不足等问题,提出一种融合局部和全局特征的息肉分割模型。以卷积神经网络和 Transformer 作为并行编码器,使模型可以兼顾多种尺度的局部细节特征和全局语义特征;在跳跃连接处构建注意力增强模块和多尺度残差模块,前者强化模型对重要信息的关注度,后者高效探索目标区域并准确预测其边界,同时促进不同层次特征之间的交互;在解码阶段采用基于残差的逐步上采样特征融合方式汇聚各阶段特征,进一步增强模型的感知能力,丰富息肉特征;最后使用高效预测头促进浅层特征的融合,输出分割结果。该模型在多个对比实验中表现最优,同次优模型相比,在 Kvasir、CVC-ClinicDB 数据集上,mDice 平均提升了 1.21%;mIoU 平均提升了 1.82%;在 CVC-ColonDB、ETIS 数据集上,mDice 平均提升了 2.67%,mIoU 平均提升了 2.83%。实验结果表明,相比于现有主流模型,该模型具有较优的分割精度和泛化性能。

关键词: 息肉分割;Transformer;卷积神经网络;注意力增强模块;多尺度残差模块;特征融合

中图分类号: TP391;TN911 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Polyp segmentation model based on fusion of local and global features

Zhang Panfeng^{1,2} Yang He¹ Shen Xianhao^{2,3} Cheng Xiaohui^{1,2} Du Hui⁴

(1. College of Computer Science and Engineering, Guilin University of Technology, Guilin 541006, China;

2. Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin 541006, China;

3. College of Physics and Electronic Information Engineering, Guilin University of Technology, Guilin 541006, China;

4. College of Public Administration, Guilin University of Technology, Guilin 541006, China)

Abstract: To solve the problems of difficulty in segmentation of complex areas, loss of edge details, and insufficient generalization ability in polyp segmentation by existing models. This paper proposed a polyp segmentation model based on fusion of local and global features. Convolutional neural network and Transformer are used as parallel encoders to make the model take into account both the local detail features and global semantic features of multiple scales. The attention enhancement block and the multi-scale residual block are constructed at the jump junction. The former enhances the model's focus on important information, while the latter efficiently explores the target regions and accurately predicts their boundaries, while promoting the interaction between different levels of features. The residual-based stepwise upsampling feature fusion method is used in the decoding stage to gather the features of each stage, which further enhanced the perception ability of the model and enriched the polyp features. Finally, the efficient prediction head is used to promote the fusion of shallow features and output the segmentation results. The model performs best in the comparative experiments. Compared with the sub-optimal model, on the Kvasir and CVC-ClinicDB datasets, it achieved an average mDice improvement of 1.21% and an average mIoU improvement of 1.82%; on the CVC-ColonDB and ETIS datasets, it achieved an average mDice improvement of 2.67% and an average mIoU improvement of 2.83%. The experimental results show that the proposed model has better segmentation accuracy and generalization performance than the existing mainstream models.

Keywords: polyp segmentation; Transformer; convolutional neural network; attention enhancement block; multi-scale residual block; feature fusion

0 引言

结直肠癌是全球癌症死亡的主要原因^[1],具有高病发

率、高死亡率。大多数结直肠癌是由恶性病变的肠粘膜息肉引起的。因此,通过早期检测和切除息肉对于预防结直肠癌非常重要。结肠镜检查对结直肠癌的检测和诊断非常

收稿日期:2024-06-20

* 基金项目:国家自然科学基金(62362018,62362017,61862019)、广西重点研发计划(桂科 AB23075116,桂科 AB24010338)项目资助

有效,可以实时提供患者体内息肉的具体状态,以便其得到及时的治疗^[2]。虽然结肠镜检查是结直肠癌检测的黄金标准^[3-4],但是由于息肉大小不一、形态复杂、对比度低,且与粘膜之间的边界不清晰,使得检测效果很大程度上依赖医生的技术水平和临床经验,导致在结肠镜检查过程中存在较高的误检率^[5]。更为重要的是,有关调查表明,息肉的检出概率每提升1个百分点,患结直肠癌的概率就会降低约3个百分点^[6]。因此,设计一种能够在早期自动精准检测息肉的方法对预防结直肠癌具有非常重要的意义。

目前,息肉分割已经成为医学图像领域重点关注的研究问题^[7]。随着深度学习技术在该领域的快速发展,一些基于卷积神经网络的方法应用在息肉分割方面已取得了不小的进步。Zhou等^[8]借助于大量的跳跃连接结构,构建了UNet++模型,减少了相邻层特征之间的语义差异,ResUNet++^[9]利用残差和注意力机制构建深层网络,展现出超越U-Net的息肉分割效果,但上述模型在分割中存在着边界缺失问题;PraNet^[2]利用反向注意力机制,探索息肉内部与边界的关系,改善了边界分割效果,但由于缺少对多尺度特征的提取,该模型对于小目标息肉感知能力不佳;在此基础上,Lou等^[10]利用轴向注意力和多尺度特征提取模块构建出CaraNet模型,提高了对小目标息肉的分割效率,但相邻层特征难以充分交互融合。此外,上述方法主要以卷积神经网络为主干,受卷积操作有限的局部感受野的影响,这类方法能有效感知图像的局部细节特征,但对息肉全局特征的探索缺乏效率,而息肉在大小、纹理上具有多变性,使得基于卷积的方法难以有效的从全局方位建模息肉图像,从而在分割过程中容易出现难以定位息肉、误分割息肉的现象。

最近,许多基于Transformer的网络被应用到视觉任务中,取得了不错的效果^[11-14]。Zheng等^[15]使用Transformer作为网络的编码器,在自然图像分割任务上获得了最优的结果;许建等^[16]利用Transformer编码器和通道特征增强,提升了对小尺寸息肉的分割效果;Wang等^[17]利用金字塔视觉Transformer提取息肉图像的全局信息,并使用渐进式解码器来强调局部特征,在息肉分割任务上取得了最优表现,但对于息肉边缘的模糊性缺乏关注;在此基础上,Yue等^[18]采用相同的主干网络,引入边界生成模块和边界细化模块,Wu等^[19]以Swin Transformer网络为主干,引入多尺度特征提取机制和反向注意力机制,二者均提升了息肉边缘的分割精度。Transformer网络拥有覆盖全局的高效感受野,能有效捕获图像中的全局依赖关系,但在提取图像局部细节方面由于缺乏归纳偏差而常常表现不佳,这种不足在医学图像分割中尤为明显^[20],使得以Transformer网络为主干的方法对于小目标息肉和复杂的局部息肉区域分割效果不佳。

为了解决上述问题,提高模型分割精度、泛化性能以及边缘分割的准确性,本文基于经典卷积网络和Transformer

网络(以PraNet^[2]和SSFormer^[17]为主),提出一种融合局部特征和全局特征的息肉分割模型,主要包括以下工作:

1)为了兼顾息肉的全局特征和局部特征,利用卷积神经网络和Transformer作为并行主干,分别从细节特征和全局特征的角度出发提取息肉的特征,使模型能够全面地建模多种尺寸大小的息肉,提高泛化性能;

2)为了强化模型对息肉特征和息肉边界的感知能力,在跳跃连接处创建注意力增强模块(attention enhancement block, AEB)和多尺度残差模块(multi-scale residual block, MSRB)。AEB通过对图像的全局和局部的建模,多角度强化关键区域,提升模型对重要特征的关注度,减少由于息肉的多变性而导致的信息丢失;MSRB具有多种感受野,可以进行多层次的特征提取和特征交互,使模型高效分割各种复杂的息肉区域,准确预测目标边界;

3)为了增强相邻阶段特征之间的相关性,减少特征融合不当而造成的语义差异和信息损失,提出基于残差的逐步上采样特征融合(stepwise upsampling feature fusion based on residual, SUFF)的方式来充分融合各阶段特征信息,提升模型识别各种目标种类的能力;

4)为了防止边界等重要细节特征丢失,在解码器末尾使用高效预测头(prediction head, PH),能有效将浅层特征融合到即将输出的结果中,完善细节信息,产生分割结果。

1 网络模型结构

本文的模型框架如图1所示,首先,利用局部特征编码器和全局特征编码器分别提取多尺度局部细节特征和全局语义特征;其次,利用空间注意力增强模块(spatial attention enhancement block, SAEB)和通道注意力增强模块(channel attention enhancement block, CAEB)分别强化局部和全局编码器提取的特征,并将输出合并,合并后的特征将被送入MSRB中,MSRB利用多种感受野探索目标区域及其边界,促进局部和全局信息的交互;然后,使用SUFF解码方式融合各阶段特征,对于浅层特征,因为融合时机较晚,在预测头中继续进行融合,最终输出息肉分割结果。

1.1 编码器

为了克服卷积神经网络和Transformer网络作为主干网络在提取图像特征中存在的局限性,本文将两种主干网络结合,采用并行的编码器架构,从而实现优势互补。主干网络的选取参考基础网络PraNet和SSFormer的编码器。

1)局部特征编码器

由图1可知,局部特征编码器采用Res2Net101^[21]来逐层获取图像的多尺度局部细节特征,主要分为5个阶段,head阶段由3个卷积层和一个最大池化层顺序组成,对图像进行初步处理;剩余阶段均由多个Res2Net block构成($\times i$ 表示*i*个)。Res2Net block的结构如图2(a)所示,核心在于:(1)使用残差连接操作将输入与输出结合,避免已

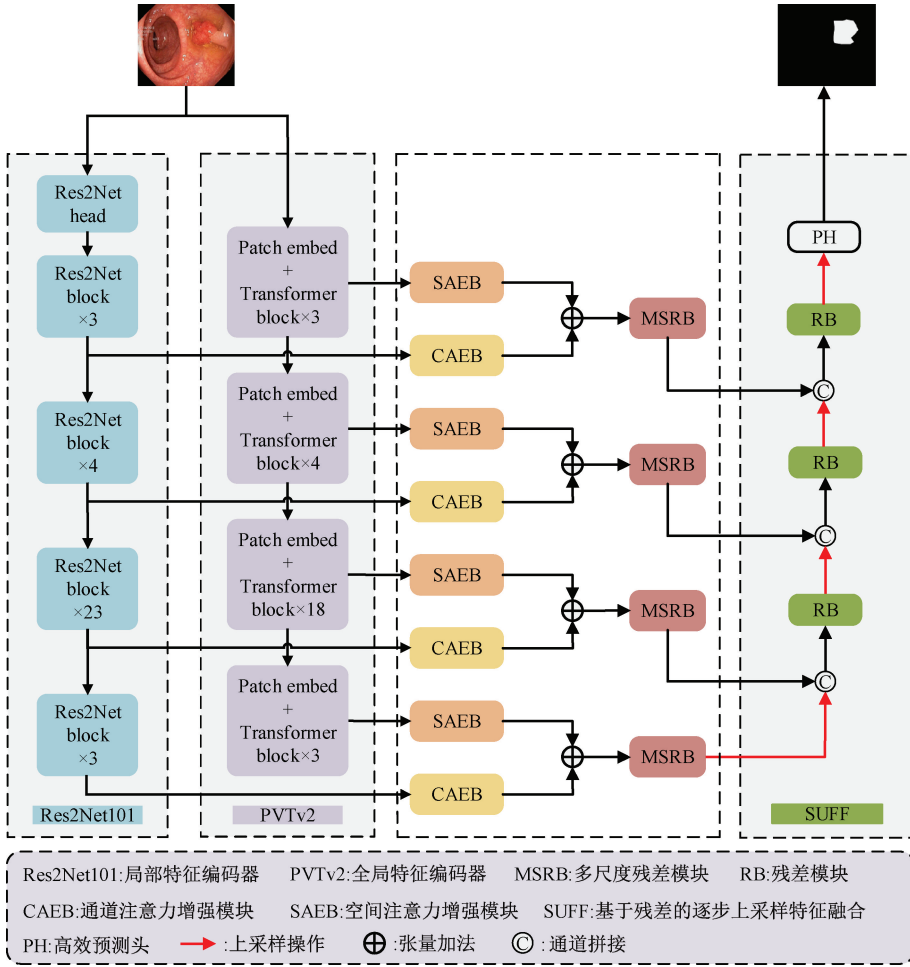


图 1 网络总体结构

Fig. 1 Overall structure of network

提取的特征信息的丢失,保证网络的性能不会下降。
 (2)多尺度信息的提取,对在通道维度拆分后的特征图的处理过程如下:

$$\mathbf{b}_i = \begin{cases} \mathbf{a}_i & i = 1 \\ \text{Conv}(\mathbf{a}_i) & i = 2 \\ \text{Conv}(\mathbf{b}_{i-1} + \mathbf{a}_i) & i \geq 3 \end{cases} \quad (1)$$

该操作中右侧卷积可以处理左侧卷积的输出结果,扩大了感受野,使得 Res2Net block 拥有多种感受野组合,擅长捕获多尺度特征,有利于局部细节特征的提取。二者结合,可以使局部特征编码器的性能不会因为层次的加深而降低,学习过程更加稳定,加快收敛速度。

2)全局特征编码器

如图 1 所示,全局编码器利用金字塔视觉 Transformer v2(PVTv2)^[22]来获取图像的全局语义特征,主要采用 4 个处理阶段来逐步建模图像全局,提取多尺度语义特征。每个阶段包含一个嵌入层(Patch embed)和多个 Transformer block($\times i$ 表示 i 个),嵌入层通过卷积和展平操作将图像分块、降维和拉直,降维操作使得 PVTv2

在每个阶段可以将特征图下采样来灵活地改变其空间尺寸,这与传统 Transformer 网络^[23-24]保持特征图空间尺度大小不变的做法不同,使网络能够从息肉图像提取多尺度特征信息;Transformer block 负责处理嵌入层的输出,结构如图 2(b)所示,该模块主要包含空间缩减注意力层(spatial reduction attention, SRA)和卷积前馈层(convolutional feed forward, CFF)。CFF 借助于深度可分离卷积进行位置编码的零填充,实现对位置信息的建模;SRA 的核心是空间缩减自注意力机制,具体信息处理过程如下:

$$SR(x) = \text{Norm}(\text{Reshape}(x, R_i)W^S) \quad (2)$$

$$\text{Attention}(q, k, v) = \text{Softmax}\left(\frac{qk^T}{\sqrt{d_{\text{head}}}}\right)v \quad (3)$$

式中:SR 为空间缩减操作;Norm 为组归一化;Reshape 为尺寸重塑操作;输入序列 x 的尺寸: $(H_i W_i) \times C_i$; H_i 、 W_i 、 C_i 、 R_i 分别为第 i 阶段特征图的高、宽、通道数以及 SR 操作缩减的倍数; $W^S \in \mathbf{R}^{(R_i^2 C_i) \times C_i}$ 为线性映射矩阵;Attention 为自注意力机制; q 、 k 、 v 为特征图序列。式(2)通过重塑

操作将输入序列 x 的尺寸重塑为 $(H_i W_i / R_i^2) \times R_i^2 C_i$; 再通过与线性映射矩阵相互作用, 将尺寸变为 $(H_i W / R_i^2) \times C_i$, 相比于原序列, 输出序列尺寸缩减为原来的 $1/R_i^2$ 。式(2)的输出同时又是式(3)自注意力计算中 k, v 的输入, 这种尺寸缩减操作很大程度上节省了自注意力机制对输入序列进行运算时的计算开销, 从而可以高效地建模图像的全局依赖关系。

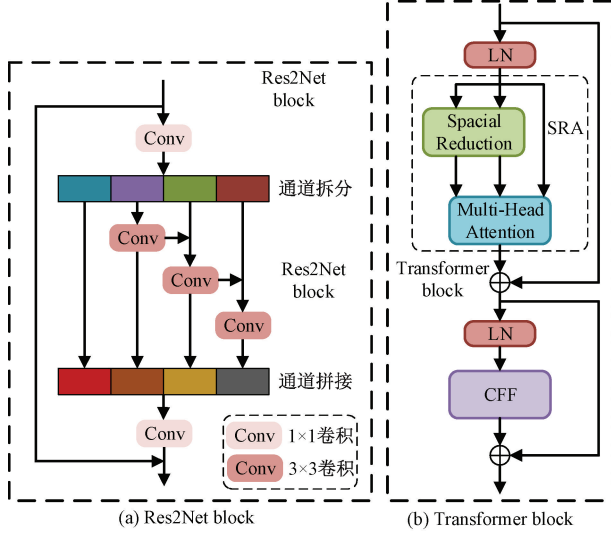


图2 Res2Net block 和 Transformer block 的结构

Fig. 2 The structure of the Res2Net block and the Transformer block

1.2 注意力增强模块

由于息肉所处环境的复杂性, 使得编码器提取的特征中含有影响分割精度的噪声因素, 为了提高模型对于息肉特征的敏感度, 本文在模型的跳跃连接处创建了注意力增强模块, 对编码器提取的特征进行强化处理, 由于两个编码器所提取特征的侧重点不同, 本文有针对性的创建了两种增强模块, 由图1可知, SAEB用于强化局部特征编码器提取的特征; CAEB用于强化全局特征编码器提取的特征。

1) 空间注意力增强模块(SAEB)

SAEB的结构如图3(b)所示, 不同于传统的空间注意力机制^[25], SAEB的任务是: (1)赋予目标区域较大的权重, 以增强局部细节并抑制不相关因素; (2)通过残差连接进一步强化目标区域的重要性, 预防关键信息的丢失。具体的信息处理过程可以表示为:

$$F_2 = \text{Conv1}(\text{Conv1}(F_1)) \quad (4)$$

$$F_3 = S(f^{7 \times 7}(\text{Concatenation}(f^{\text{Maxpool}}(F_2), f^{\text{Avgpool}}(F_2)))) \quad (5)$$

$$F_4 = F_3 \otimes F_2 \oplus F_1 \quad (6)$$

式中: F_i 为特征图; $f^{7 \times 7}$ 为 7×7 的卷积; Concatenation 为通道拼接; f^{Maxpool} 和 f^{Avgpool} 分别为在通道方向的最大池化和平均池化; S 为 Simoid 激活函数。首先, 连续的卷积操作作用于增强局部区域内相邻像素间的交互, 主要用于防止较小的息肉组织区域的注意力分散, 随后式(5)中的池化

和卷积操作可以将注意力集中到单一的空间层面上, 激活操作作用于压缩特征值, 得到特征图中每个区域的权重, 最后, 在式(6)中, 乘法操作将学习到的权重分配给对应区域的像素点, 越重要的区域分配的权重值越高; 加法操作将原始特征引入, 其主要作用在于: (1)进一步增加主要目标区域重要性; (2)由于息肉具有多变性, 使得部分复杂息肉区域在注意力学习过程中容易被忽略, 原始特征的引入可以弥补该部分特征, 供后续模块识别, 减少信息损失。

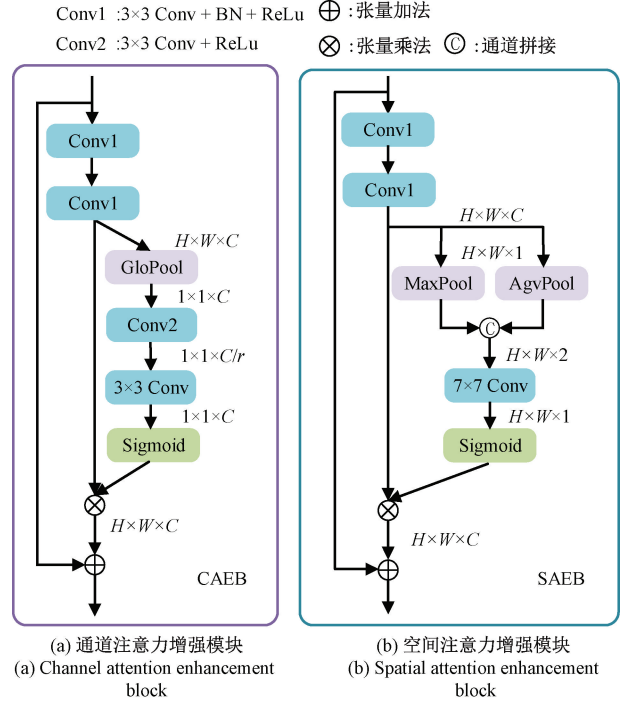


图3 通道注意力增强模块和空间注意力增强模块的结构

Fig. 3 The structure of the channel attention enhancement block and spatial attention enhancement block

2) 通道注意力增强模块(CAEB)

CAEB针对全局特征编码器, 处理具有全局意义的特征信息, 其结构如图3(a)所示, 主要任务是: (1)使网络的注意力集中在包含重要语义特征的通道, 而忽略无关通道; (2)使用残差连接进一步强化关键通道, 同时预防信息丢失。具体的信息处理过程可以表示为:

$$F_2 = \text{Conv1}(\text{Conv1}(F_1)) \quad (7)$$

$$F_3 = S(f^{3 \times 3}(\text{Conv2}(f^{\text{GloPool}}(F_2)))) \quad (8)$$

$$F_4 = F_3 \otimes F_2 \oplus F_1 \quad (9)$$

式中: $f^{3 \times 3}$ 为 3×3 的卷积; f^{GloPool} 为全局平均池化; S 为 Simoid 激活函数。同 SAEB, 式(7)中连续的卷积可以防止注意力分散, 式(8)中的全局平均池化可以将特征图的尺寸压缩, 使注意力集中于通道层面, 卷积和激活操作通过压缩和扩展特征图尺寸来学习各个通道特征的重要性, 建模通道间的相互依赖性; 式(9)中, 乘法操作将学习到的通道权重分配给每个通道, 让网络更加关注重要的通道特

征,抑制无用的特征,残差连接的加法操作将初始输入的特征再次引入,强化关键通道,同时预防注意力学习过程中由于息肉的多变性而导致的部分重要通道被忽视的问题。

1.3 多尺度残差模块(MSRB)

与 PraNet 及其改进网络^[2,10,19]的反向注意力机制不同,MSRB 不仅可以探索息肉的边界信息,而且能够弥补多尺度特征信息的缺失,结构如图 4 所示,MSRB 由 3 个并行分支构成,Conv1 由 3×3 Conv、BN、ReLU 组成,左侧分支由两个串联卷积层组成;中间分支由四个串联卷积层和两个残差连接组成,其本身可以被视为一个多尺度特征提取器;右侧分支为残差连接。左、中分支的最大感受野范围分别为 5×5 、 9×9 ,使得 MSRB 能够同时提取到多种不同尺度的息肉特征信息,提高对目标区域的探索效率;同时,感受野的多样性,能够实现对复杂区域进行多范围的探索,可以较容易地发现息肉区域和周围粘膜之间细微的不同,有利于在复杂的环境中准确预测息肉边界,引导模型准确识别目标;右侧分支的残差连接,不仅能够防止梯度消失,还可以补充在特征提取过程中未能捕捉到的和丢失的信息,有助于形成更多的息肉细节。

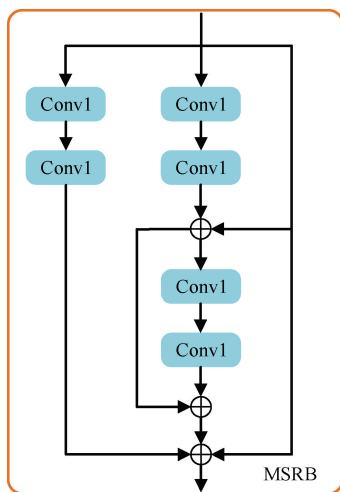


图 4 多尺度残差模块的结构

Fig. 4 The structure of Multi-scale residual block

由图 1 可知,MSRB 处理的特征图是由 SAEB 和 CAEB 的输出合并得到的,包含局部细节特征和全局语义特征,MSRB 利用并行分支在不同尺度范围内交互两种不同层次的特征,使模型能够充分利用两种特征之间的互补性,有利于提升模型的感知性能,以应对息肉的多变性。

1.4 基于残差的逐步上采样特征融合(SUFF)

SSFormer 网络在特征融合阶段直接将各阶段的特征直接上采样到 88×88 后进行聚合,对于较低分辨率的特征图来说,上采样倍数过多容易造成特征丢失,导致各阶段特征之间较弱的相关性,在进行特征融合极易产生语义

差异,造成信息损失,影响模型性能。为了使各阶段特征的融合尽可能平滑、减少语义差异,本文在解码阶段使用 SUFF 方式融合各阶段特征。由图 1 可知,SUFF 采用自下而上的方式,先将末尾阶段的语义特征进行两倍上采样,然后与相邻阶段同尺寸特征图拼接,通过残差块促进特征融合,输出的特征图继续重复上述过程。具体过程如式(10)所示。

$$y_i = \text{RB}(\text{Concatenation}(x_i, \text{Up}(y_{i+1}))) \quad (10)$$

式中: y_i 为融合后的第 i 层特征图, x_i 为融合前的第 i 层特征图,其中 $i \in [1, 4]$, $y_4 = x_4$;RB 为残差模块;Up 为两倍上采样;Concatenation 为通道拼接。由于相邻层特征之间具有较强的相关性,易于融合,并且残差的引入可以弥补上采样和融合过程中的信息损失,所以这种渐进式的特征融合方式能够有效促进深层特征与浅层特征的融合,逐步丰富特征类别,增加模型所能感知的息肉样式。

1.5 预测头(PH)

不同于 FCN-Transformer^[26]仅使用 PH 进行特征图的预测分割,本文利用其特殊的结构进行浅层特征融合,并取得了较好的效果。如图 5(a)所示,PH 由两个残差模块和一个卷积构成,残差模块结构如图 5(b)所示。残差模块负责将网络浅层的细粒度特征与已经融合好的较深层的粗粒度特征融合,卷积模块负责输出预测分割图。如图 1 所示,由于浅层特征进入 SUFF 模块的时机较晚,可能会出现与深层特征融合不充分的问题,因为浅层特征包含大量的细节信息,融合不当会导致包括边界在内的诸多细节特征的丢失,这也是被许多模型所忽视的^[8,17]。因此,本文将浅层特征输入预测头中继续进行特征融合;残差块的使用,可以在融合过程中持续补充输入特征,防止融合不当造成信息损失,信息融合完毕后,预测头通过卷积操作产生息肉分割结果。

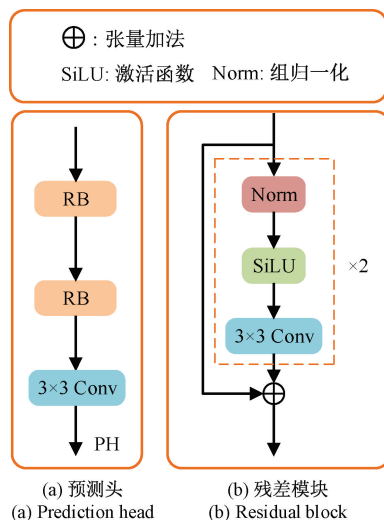


图 5 预测头、残差模块的结构

Fig. 5 The structure of prediction head, residual block

2 实 验

2.1 实验数据及数据增强

本文使用公开的息肉数据集 Kvasir^[27]、CVC-ClinicDB^[28]、CVC-ColonDB^[29]和 ETIS^[30]来验证模型的分割性能。根据 PraNet 等模型^[2,31-32]的划分方法,训练数据集构成如下:从 Kvasir 和 CVC-ClinicDB 数据集中随机选择 900 张和 550 张息肉图像。测试数据集则由 Kvasir 和 CVC-ClinicDB 中剩余的图像以及 CVC-ColonDB 和 ETIS 中的全部图像构成,共 798 张。由于数据集的图像分辨率各不相同,为方便训练和测试,按照 PraNet 等模型^[2,17,20]的设置将所有数据集的图像分辨率调整为 352×352 。在加载训练集图像及其对应的标签时,除了将图像进行标准化之外,还随机使用下列数据增强操作来扩大训练集:高斯模糊、色彩变换、水平和垂直翻转、仿射变换。其中高斯模糊和颜色变换的操作仅应用于图像,其余增强操作均一致地应用于图像及其对应的标签。

2.2 实验细节

本文的模型和实验环境基于 pytorch 框架,操作系统为 Windows 10,显卡为 NVIDIA GeForce RTX 3090。模型训练采用 Adam 优化器,初始学习率为 1×10^{-4} ;批处理大小为 8,训练轮次 epoch 为 120。采用加权 IoU 损失 L_{wIoU} 和加权二值交叉熵损失 L_{wbce} 之和作为损失函数。为了提高息肉边界分割的准确性,本文遵循 PolypPVT 等模型^[31,33]中对难分辨像素的处理方法,在损失函数中应用式(11)赋予边界像素比较高的权重,具体计算过程为:

$$WF = 1 + 5 \times |f^{Avgpool}(mask) - mask| \tag{11}$$

式中: $f^{Avgpool}$ 为尺寸为 31×31 的平均池化, $mask$ 为图像的标签,输出结果 WF 为各像素点权重的权重图。

2.3 评价指标

为了准确地评估模型分割性能,本文采用医学图像分割中高度认可的评价指标:平均 Dice 相似性系数($mDice$)、平均交并比($mIoU$)、精确率($Precision$)、召回率($Recall$)。各指标计算式分别为:

$$mDice = \frac{1}{K} \sum_{i=1}^K \frac{2|X \cap Y|}{|X| + |Y|} \tag{12}$$

$$mIoU = \frac{1}{K} \sum_{i=1}^K \frac{TP}{TP + FP + FN} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

式中: X 为模型输出的结果, Y 为 X 对应的标签, K 为类别数量, TP 为息肉区域中被正确预测为息肉的像素个数, FP 为非息肉区域中被错误预测为息肉的像素个数, FN 为息肉区域中被错误预测为非息肉的像素个数。

2.4 实验结果分析

1) 对比实验

为验证本文模型的性能,将其与目前最优的一些模型进行对比:U-Net++ (2018)^[8]、ResUNet++ (2021)^[9]、PraNet (2020)^[2]、DCRNet (2022)^[34]、SSFormer (2022)^[17]、CaraNet (2023)^[10]、BRNet (2024)^[18]、MSRAformer (2022)^[19]、PolypPVT (2023)^[31]。U-Net++ 和 ResUNet++ 是息肉分割的基准模型,PraNet 和 SSFormer 是本文所参照的基础模型,DCRNet 和 CaraNet 是基于 PraNet 改进提出的,DCRNet 采用了更先进的主干网络和并行注意力模块;CareNet 基于 PraNet 网络的框架,引入了轴向注意力机制和多尺度特征提取模块;PolyPVT 和 BRNet 采用与 SSFormer 相同的主干网络,在此基础上,PolyPVT 构建了多层次特征提取和跨层特征融合模块;BRNet 构建了边界生成和边界细化模块;MSRAformer 是两种基础模型的结合,编码器采用 Transformer 网络,解码器基于 PraNet 网络,构建了多尺度特征提取模块,改进了反向注意力机制。上述模型中,除基准模型外,其余均是目前公认的在同种类型模型中处于最新、最优水平的模型,采用这些与本文模型联系密切的模型做对比,能够最大程度的体现其先进性。所有模型设置完全相同的实验细节,实验结果如表 1、2 所示,最优值采用加粗表示,分割可视化结果如图 6 所示。

表 1 不同模型在 Kvasir、CVC-ClinicDB 上的分割结果

Table 1 Segmentation results of different models on Kvasir and CVC-ClinicDB

模型	Kvasir				CVC-ClinicDB			
	<i>mDice</i>	<i>mIoU</i>	<i>Precision</i>	<i>Recall</i>	<i>mDice</i>	<i>mIoU</i>	<i>Precision</i>	<i>Recall</i>
U-Net++	0.849 1	0.765 7	0.864 3	0.887 3	0.858 7	0.785 2	0.895 0	0.853 5
ResUNet++	0.795 4	0.708 3	0.870 1	0.806 0	0.765 1	0.684 4	0.868 0	0.755 9
PraNet	0.896 2	0.833 7	0.925 9	0.897 4	0.875 1	0.804 8	0.903 0	0.881 5
DCRNet	0.881 5	0.815 4	0.878 9	0.923 3	0.882 2	0.824 4	0.886 0	0.899 0
SSFormer	0.916 5	0.860 5	0.923 0	0.930 8	0.922 1	0.863 9	0.918 1	0.932 0
CaraNet	0.918 0	0.865 9	0.935 7	0.921 5	0.937 0	0.888 7	0.934 0	0.947 8
MSRAformer	0.920 7	0.866 8	0.904 6	0.954 9	0.930 2	0.879 2	0.925 7	0.945 9
PolypPVT	0.919 1	0.868 6	0.912 1	0.937 5	0.938 0	0.890 1	0.935 2	0.948 3
BRNet	0.920 5	0.871 0	0.935 2	0.929 1	0.931 2	0.883 0	0.952 5	0.924 3
本文	0.931 0	0.887 9	0.944 2	0.938 9	0.950 3	0.907 2	0.950 4	0.952 5

表 2 不同模型在 CVC-ColonDB、ETIS 上的分割结果
Table 2 Segmentation results of different models on CVC-ColonDB and ETIS

模型	CVC-ColonDB				ETIS			
	<i>mDice</i>	<i>mIoU</i>	<i>Precision</i>	<i>Recall</i>	<i>mDice</i>	<i>mIoU</i>	<i>Precision</i>	<i>Recall</i>
U-Net++	0.707 2	0.618 8	0.744 5	0.753 2	0.615 4	0.528 3	0.625 5	0.732 3
ResUNet++	0.636 7	0.550 6	0.701 5	0.657 5	0.348 4	0.292 1	0.386 9	0.372 1
PraNet	0.752 5	0.672 0	0.752 5	0.809 4	0.665 1	0.581 1	0.634 9	0.761 2
DCRNet	0.761 0	0.678 5	0.740 4	0.869 2	0.599 0	0.523 3	0.552 0	0.791 8
SSFormer	0.797 2	0.716 3	0.826 6	0.815 3	0.760 5	0.668 5	0.703 7	0.897 7
CaraNet	0.764 8	0.697 5	0.783 6	0.799 2	0.766 4	0.678 5	0.717 1	0.890 9
MSRAformer	0.772 4	0.697 6	0.777 0	0.827 7	0.706 5	0.633 7	0.660 5	0.816 1
PolypPVT	0.795 2	0.722 9	0.808 1	0.840 0	0.753 5	0.666 0	0.685 4	0.927 5
BRNet	0.792 2	0.722 2	0.816 6	0.816 5	0.741 6	0.666 1	0.706 1	0.842 5
本文	0.804 8	0.731 9	0.813 7	0.842 9	0.779 7	0.700 6	0.717 1	0.935 6

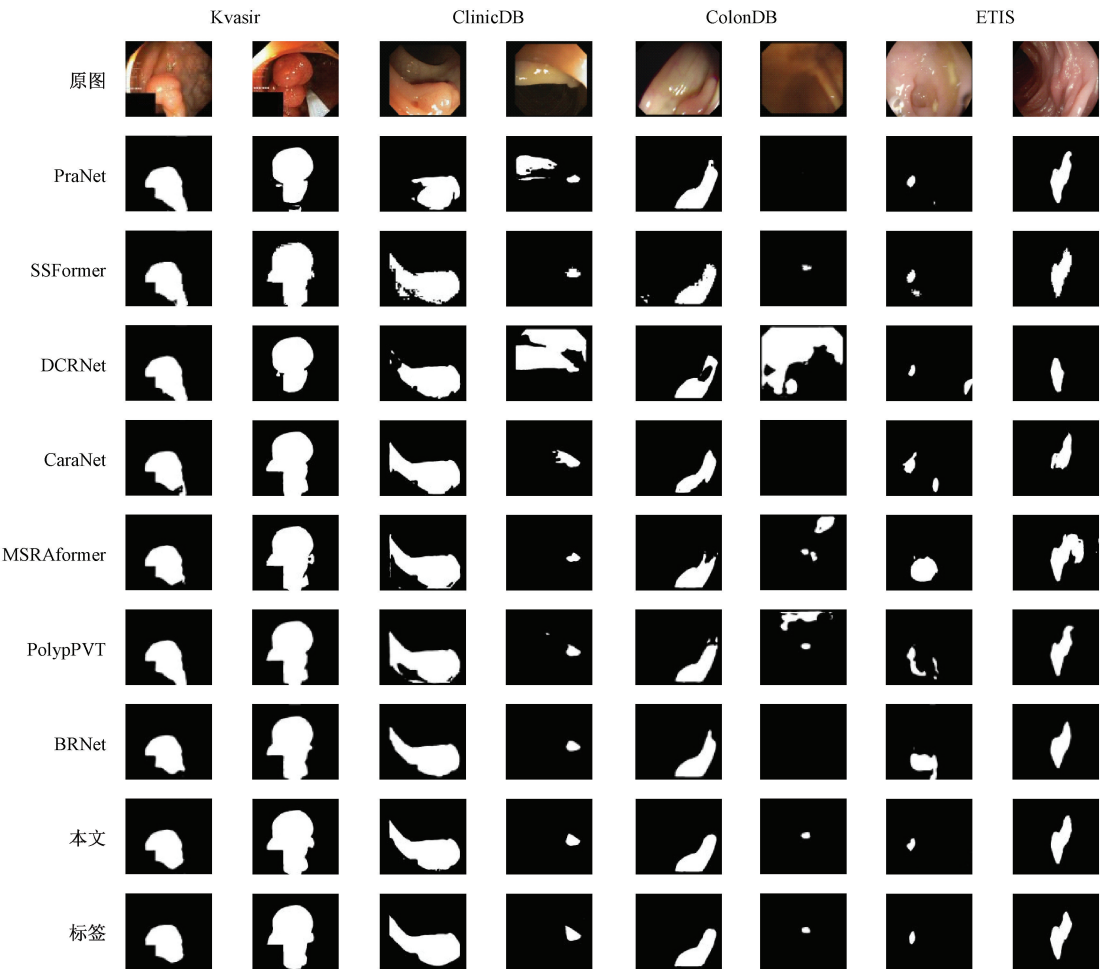


图 6 可视化分割结果
Fig. 6 Visual segmentation results

表 1 展示了各模型在 Kvasir 和 CVC-ClinicDB 数据集上的息肉分割性能,从中可以得知,本文模型的 *mDice* 分别为 93.10%、95.03%,*mIoU* 分别为 88.79%、90.72%,*Precision* 分别为 94.42%、95.04%,3 项指标均获得了最

优值,与次优算法 PolypPVT 相比, $mDice$ 分别提升了 1.19%、1.23%, $mIoU$ 分别提升了 1.93%、1.71%, $Precision$ 分别提升了 3.21%、1.52%。综合对比各项指标可知,本文模型总体性能要优于其他模型,具有先进的学习能力。

表 2 给出了各模型在数据集 CVC-ColonDB 和 ETIS 上的泛化性能测试结果。在 CVC-ColonDB 数据集上,本文模型的 4 项指标分别为 80.84%、73.19%、81.37%、84.29,与 PraNet 相比,各指标分别提升了 5.23%、5.99%、6.12%、3.35%;在具有较多小目标息肉的数据集 ETIS 上,本文模型的各项指标均为最优值,相比于性能较为领先的 CaraNet, $mDice$ 、 $mIoU$ 、 $Recall$ 分别提升了 1.33%、2.21%、4.47%。综合对比可知,本文模型的总体性能仍为最优,具有较强的泛化能力。此外,在 ETIS 上的分割效果表明本文模型能有效提升复杂的小目标息肉的分割精度。

各模型的可视化分割结果如图 6 所示。从中可以发现以下问题:对于边界轮廓较为复杂的息肉,如第 2、3、8 列,许多模型存在漏检、边缘分割粗糙的问题;对于小目标息肉,如第 4、6、7 列,部分模型难以准确定位其位置;对于与息肉相似的非息肉区域,如第 1、4、7、8 列,部分模型存在一定的误检率;对于息肉内部的一些重要区域,部分模型不能有效感知,导致该部分区域被忽略,如第 3、5 列。相比之下,本文模型能够较为精准的感知并定位息肉,抑制无关信息的干扰,实现接近真实情况的分割效果,而且息肉边缘的分割较为清晰、准确。综上所述,本文模型能够很好地弥补现有模型对各种复杂息肉和息肉边缘分割困难的不足。

2)消融实验

为探索本文模型中各模块对于整体分割性能的影响,本文在 Kvasir、CVC-ClinicDB 数据集上进行了消融实验。M1 表示仅使用局部特征编码器和全局特征编码器,跳跃连接结构仅采用卷积,除此之外不添加任何模块;M2 表示在 M1 的基础上给添加了注意力增强模块 SAEB 和 CAEB;M3 是在 M2 的基础上添加多尺度残差模块 MSRB;M4 是在 M3 的基础上添加了基于残差的逐步上采样特征融合解码方式 SUFF;M5 即本文网络,在 M4 的基础上添加了 PH 预测头模块。

消融实验结果如表 3 所示,其中最优值采用加粗表示。从表中可以看出,M2 的总体性能要优于 M1, $mDice$ 分别提升了 0.22、0.10 个百分点, $mIoU$ 分别提升了 0.36、0.26 个百分点,说明了 SAEB 和 CAEB 能有效提升模型对重要特征信息的关注度,但性能提升并不明显,原因在于 SAEB 和 CAEB 提取的信息不能得到很好的融合;M3 相比于 M2,性能提升较为明显, $mIoU$ 分别提升了 0.65、0.50 个百分点,在 Kvasir 上 $Precision$ 提升了 1.61 个百分点,说明 MSRB 可以有效增强特征之间的交互,捕获多尺度特征,识别包括边界在内的更多复杂的特征信息;M4 相比于 M3, $mDice$ 分别提升了 0.49、0.73 个百分点, $mIoU$ 分别提升了 0.47、1.08 个百分点,提升较为明显,说明 SUFF 可以充分融合其上游模块提取的丰富的特征信息,增强模型的特征感知能力。M5 即本文模型,综合性能最优,与 M4 相比, $mDice$ 分别提升了 0.24、0.07 个百分点, $mIoU$ 分别提升了 0.35、0.08 个百分点,该部分的提升主要针对浅层细节特征,说明预测头能够有效的将浅层特征融合到特征图中。

表 3 在 Kvasir、CVC-ClinicDB 上的消融实验结果
Table 3 Ablation experiments results on Kvasir and CVC-ClinicDB

模型	Kvasir				CVC-ClinicDB			
	$mDice$	$mIoU$	$Precision$	$Recall$	$mDice$	$mIoU$	$Precision$	$Recall$
M1	0.918 6	0.869 6	0.925 8	0.938 3	0.937 1	0.888 0	0.938 7	0.941 5
M2	0.922 0	0.873 2	0.931 3	0.937 4	0.938 1	0.890 6	0.942 0	0.941 3
M3	0.924 5	0.879 7	0.947 4	0.921 4	0.942 3	0.895 6	0.942 3	0.946 7
M4	0.929 4	0.884 4	0.923 0	0.958 2	0.949 6	0.906 4	0.946 5	0.956 1
M5	0.931 0	0.887 9	0.944 2	0.938 9	0.950 3	0.907 2	0.950 4	0.952 5

3 结 论

针对现有模型在息肉分割中存在的复杂区域误分割、漏分割、边缘信息缺失、泛化能力不足的问题,提出一种融合局部和全局特征的息肉分割模型,利用并行的局部和全局特征编码器,以兼顾图像的局部细节特征和高级语义特征;设计注意力增强模块和多尺度残差模块,前者强化模型对关键区域的注意力,同时防止信息的丢失,后者具有

多种感受野,实现对息肉区域、边界高效的探索、多层次特征信息的交互;解码处设计基于残差的逐步上采样特征融合方式,逐步融合各阶段特征,避免信息丢失,丰富模型识别的特征种类;最后利用高效预测头促进浅层细节特征的融合,产生最佳分割结果。

实验结果表明,该模型的分割精度和泛化性能优于许多主流模型,能出色地分割小目标息肉、复杂形态的息肉,较为精准地定位、分割边界区域,辅助医生进行息肉诊断,

降低误检率,为临床应用提供重要参考价值。此外,该模型还可以扩展应用到其他医学图像分割领域,特别是对形态和纹理多样化且边界模糊不清的图像有较好的分割效果。

参考文献

- [1] SIEGEL R L, MILLER K D, FUCHS H E, et al. Cancer statistics [J]. CA: A Cancer Journal for Clinicians, 2022,72(1): 7-33.
- [2] FAN D P, JI G P, ZHOU T, et al. Pranet: Parallel reverse attention network for polyp segmentation[C]. Medical Image Computing and Computer-Assisted Intervention, 2020: 263-273.
- [3] SHAUKAT A, KAHN C J, BURKE C A, et al. ACG clinical guidelines: Colorectal cancer screening 2021 [J]. The American Journal of Gastroenterology, 2021, 116(3): 458-479.
- [4] 杨昆,孙宇锋,汪世伟,等. YOLOF-CBAM: 一种新的结直肠息肉实时分类与检测方法[J]. 电子测量技术, 2023,46(16):138-147.
YANG K, SUN Y F, WANG SH W, et al. YOLOF-CBAM: New real-time classification and detection method for colorectal polyps [J]. Electronic Measurement Technology, 2023,46(16):138-147.
- [5] PUYAL J G B, BHATIA K K, BRANDAO P, et al. Endoscopic polyp segmentation using a hybrid 2D/3D CNN[C]. Medical Image Computing and Computer Assisted Intervention, 2020: 295-305.
- [6] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv: 2010.11929, 2020.
- [7] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: A Cancer Journal for Clinicians, 2021,71(3):209-249.
- [8] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS 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, 2018: 3-11.
- [9] JHA D, SMEDSRUD P H, RIEGLER M A, et al. ResuNet++: An advanced architecture for medical image segmentation [C]. 2019 IEEE International Symposium on Multimedia, 2019: 225-2255.
- [10] LOU A, GUAN SH Y, LOEW M H. CaraNet: Context axial reverse attention network for segmentation of small medical objects [J]. Medical Imaging, 2023,10(1): 014005.
- [11] 金宇锋,陶重彝. 基于 Transformer 的融合信息增强 3D 目标检测算法 [J]. 仪器仪表学报, 2023, 44(12): 297-306.
JIN Y F, TAO CH B. Fusion information enhanced method based on transformer for 3D object detection [J]. Chinese Journal of Scientific Instrument, 2023, 44(12):297-306.
- [12] 刘肇隆,范馨月. 基于全尺度跳跃连接的 TransUNet 医学图像分割网络 [J]. 国外电子测量技术, 2023, 42(11):42-48.
LIU ZH L, FAN X Y. TransUNet medical image segmentation network based on full-scale jump connectivity [J]. Foreign Electronic Measurement Technology, 2023,42(11):42-48.
- [13] 刘铁,段勇. 融合 CNN 和 Transformer 的机器人室内场景识别 [J]. 电子测量与仪器学报, 2023, 37(5): 223-229.
LIU T, DUAN Y. Robot indoor scene recognition based on fusion of CNN and Transformer [J]. Journal of Electronic Measurement and Instrumentation, 2023,37(5):223-229.
- [14] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021,43(10):3349-3364.
- [15] ZHENG S X, LU J CH, ZHAO H SH, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:6877-6886.
- [16] 许建,王若涵. 一种基于局部通道注意力的可信息肉分割方法 [J]. 激光与光电子学进展, 2024: 1-17. DOI: 10.3788/LOP241160.
XU J, WANG R H. A trustworthy polyp segmentation method based on local channel attention [J]. Laser & Optoelectronics Progress, 2024: 1-17. DOI: 10.3788/LOP241160.
- [17] WANG J F, HUANG Q M, TANG F L, et al. Stepwise feature fusion: Local guides global [C]. Medical Image Computing and Computer Assisted Intervention, 2022:110-120.
- [18] YUE G H, LI Y Y, JIANG W CH, et al. Boundary refinement network for colorectal polyp segmentation in colonoscopy images [J]. IEEE Signal Processing Letters, 2024, 31:954-958.
- [19] WU C, LONG CH, LI SH J, et al. MSRAformer: Multiscale spatial reverse attention network for polyp segmentation [J]. Computers in Biology and Medicine,

- 2022, 151: 106274.
- [20] ZHANG Y D, LIU H Y, HU Q. Transfuse: Fusing transformers and CNNs for medical image segmentation [C]. Medical Image Computing and Computer Assisted Intervention, 2021: 14-24.
- [21] GAO SH H, CHENG M M, ZHAO K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.
- [22] WANG W H, XIE EN Z, LI X, et al. PVT v2: Improved baselines with pyramid vision transformer [J]. Computational Visual Media, 2022, 8(3): 415-424.
- [23] RANFTL R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction [C]. IEEE/CVF International Conference on Computer Vision, 2021: 12179-12188.
- [24] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention [C]. 38th International Conference on Machine Learning, 2021, 139: 10347-10357.
- [25] GUO CH L, SZEMENYEI M, YI Y G, et al. SA-UNet: Spatial attention U-Net for retinal vessel segmentation [C]. 25th International Conference on Pattern Recognition (ICPR), 2021: 1236-1242.
- [26] SANDERSON E, MATUSZEWSKI B J. FCN-Transformer feature fusion for polyp segmentation [C]. Medical Image Understanding and Analysis, 2022: 892-907.
- [27] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented polyp dataset [C]. Multimedia Modeling, 2020: 451-462.
- [28] BERNAL J, SÁNCHEZ F J, FERNÁNDEZESPARRACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians [J]. Computerized Medical Imaging and Graph, 2015, 43: 99-111.
- [29] TAJBAKSH N, GURUDU S R, LIANG J M. Automated polyp detection in colonoscopy videos using shape and context information [J]. IEEE Transactions on Medical Imaging, 2015, 35(2): 630-644.
- [30] SILVA J, HISTACE A, ROMAIN O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer [J]. International Journal of Computer Assisted Radiology and Surgery, 2014, 9(2): 283-293.
- [31] DONG B, WANG W, FAN D P, et al. Polyp-PVT: Polyp segmentation with pyramid vision transformers [J]. CAAI Artificial Intelligence Research, 2023, 2: 9150015.
- [32] HUANG X D, ZHUO L, ZHANG H, et al. Polyp segmentation network with hybrid channel-spatial attention and pyramid global context guided feature fusion [J]. Computerized Medical Imaging and Graphics, 2022, 98: 102072.
- [33] KIM T, LEE H, KIM D. Ucanet: Uncertainty augmented context attention for polyp segmentation [C]. 29th ACM International Conference on Multimedia, 2021: 2167-2175.
- [34] YIN Z J, LIANG K M, MA ZH Y, et al. Duplex contextual relation network for polyp segmentation [C]. 19th International Symposium on Biomedical Imaging, 2022: 1-5.

作者简介

张攀峰, 博士, 副教授, 硕士生导师, CCF 会员 (98351M), 主要研究方向为信息存储、信息安全、人工智能。

E-mail: panf_zhang@glut.edu.cn

杨贺, 硕士研究生, 主要研究方向为计算机视觉、医学图像分析。

E-mail: yhe20222022@163.com

杜慧 (通信作者), 硕士, 副教授, 主要研究方向为信息安全管理与保护。

E-mail: 2018060@glut.edu.cn