文章编号:1674-2974(2024)06-0040-12 DOI:10.16339/j.cnki.hdxbzkb.2024265

PRNet:渐进式消减不确定区域的息肉分割网络

何东之[1†],肖杏梅[1],李韫昱[1],薛永乐[1],李雲奇[2]
(1. 北京工业大学 信息学部,北京 100124;
2. 中国人民解放军第一医学中心,北京 100039)

摘 要:由于息肉图像的自动分割病灶区域大小不一和边界模糊,从而导致分割精度较低.针对这两个问题,本文提出先定位后逐步精细的渐进式消减网络(Progressive Reduction Network,PRNet).该网络采用Res2Net提取病灶区域特征,利用多尺度跨级融合模块将注意融合机制与跨级特征结合,有效应对病灶区域多尺度问题,提升定位准确度.在自上而下恢复图像分辨率的过程中,引入不确定区域处理模块和多尺度上下文感知模块.前者通过设定递减的阈值逐步挖掘息肉边缘信息,增强边缘细节特征的识别能力;后者则进一步探索病灶区域周围潜在的上下文语义,提升模型的整体表征能力.此外,本算法还设计了一个简单的特征过滤模块,用于筛选编码器特征中的有效信息.在Kvasir-SEG、CVC-Clinic和ETIS数据集上的实验结果表明,所提算法的Dice系数分别达到了92.09%、93.05%和74.19%,优于现有的息肉分割算法,展示出了较好的鲁棒性和泛化性.

关键词:医学图像处理;息肉分割;结肠镜;不确定区域;多尺度
中图分类号:TP311 **文献标志码**:A

PRNet: Progressive Reduction Network for Polyp Segmentation in Uncertain Areas

HE Dongzhi[1†],XIAO Xingmei[1],LI Yunyu[1],XUE Yongle[1],LI Yunqi[2]
(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
2. The First Medical Center of Chinese People's Liberation Army (PLA) General Hospital, Beijing 100039, China)

Abstract: Automatic segmentation of polyp images usually results in low segmentation accuracy due to the various sizes of lesion regions and blurry boundaries. Based on these two perspectives, a novel Progressive Reduction Network (PRNet) is proposed, which first locates polyps and then gradually refines their boundaries. The network utilizes Res2Net to extract features from the lesion region and leverages the multi-scale cross-level fusion module to improve localization accuracy. By combining the attention fusion mechanism with cross-level features in this module, the network can effectively solve the issue of multi-scale lesion areas. Furthermore, PRNet combines an uncertain region processing module and a multi-scale context-aware module when restoring image resolution from top to bottom. The former gradually mines polyp edge information by setting decreasing thresholds to enhance the

* **收稿日期**:2023-06-07
基金项目:国家自然科学基金资助项目(62005045),National Natural Science Foundation of China(62005045);教育部产学合作协同育人项目(220702278135421),University-Industry Collaborative Education Program, Ministry of Education, China(220702278135421)
作者简介:何东之(1970—),男,山东临沂人,北京工业大学副教授,博士,硕士生导师
† 通信联系人,E-mail: victor@bjut.edu.cn

recognition of edge detail features, while the latter, to improve the overall representation capability of the model, further explores the inherent potential contextual semantics of lesion regions. In addition, a simple feature filtering module is designed in this algorithm to filter the valid information in the encoder features. Experimental results on the Kvasir-SEG, CVC-Clinic, and ETIS datasets show that the Dice coefficients of the algorithm reach 92.09%, 93.05%, and 74.19%, respectively. Compared with other existing polyp segmentation algorithms, PRNet outperforms them and demonstrates its superior robustness and generalization.

Key words: medical image segmentation; polyp segmentation; colonoscopy; uncertain area; multi-scale

根据世界卫生组织国际癌症研究机构发布的2020年全球癌症(GLOBOCAN)统计报告[1],结直肠癌在世界范围内位居第三,死亡人数位居第二.它一般开始于被称作息肉的小型非肿瘤细胞群.随着时间推移,一些息肉可能会发展为恶性肿瘤,即结直肠癌.因此,定期筛查、分割和切除结直肠息肉是非常重要的.目前,结肠镜检查是发现结直肠息肉的常用方法,通过结肠内镜检查,能够直观地观察到肿瘤的外观以及结肠黏膜的病变情况.但在临床实践中,因为医师的主观判断和操作技巧等原因,四分之一的息肉无法被准确鉴别[2].即便息肉能够被准确鉴别,手动分割结直肠息肉仍是一个耗时的操作[3].因此,设计一种高效、准确的计算机辅助结直肠息肉分割算法对于减轻医护人员工作,减少医疗差错具有重要意义.

随着深度学习技术在医学图像领域的发展,许多深度卷积网络被应用于结肠镜图像息肉分割中并取得了较好的效果.完全卷积网络[4]是语义分割中的经典网络,其变体[5]成功应用在了息肉分割任务中.Jha等提出的ResUNet++[6]利用U-Net[7]的U形结构将残差模块[8]、挤压激励机制[9]和空洞空间金字塔池化模块[10]有效结合,在息肉分割数据集上实现了较高的分割精度.虽然上述算法取得了不错的效果,但由于结肠镜图像中的病灶区域具有形状大小各异、与周围正常黏膜组织对比度低的特点,因此精确分割息肉仍是一项具有挑战性的任务.

一些研究鉴于以上难点,提出了相应的解决方法.PraNet[11]利用部分解码器融合高级语义特征生成全局特征图,并借助反向注意力[12]通过逐步消除背景中的非目标区域逐步挖掘更精确的边界线索.ACSNet[13]利用边界信息概率值在0.5左右这一特点来提取边界,约束编码器特征并在解码路径中构建ASM模块融合高级语义特征、低级细节特征和全局上下文特征,减小层级特征之间的语义鸿沟.CCBANet[14]相比ACSNet对特征信息的提取更为细致,它在每个编码阶段之后,进行全局和局部信息的提取和整合,并将前景信息、背景信息和边缘信息三者融合以达到更好的分割性能.UACANet[15]则从不确定区域的角度出发,将前景、背景和不确定区域三部分结合共同挖掘息肉的边界信息.通过加入对不确定区域的探索,有效提升了息肉分割的精度.基于此思路,Guo等[16]通过使不确定区域被突出显示为更多的前景物体同时强行抑制背景中的噪声信息来逐步挖掘边界区域,并进一步提出特征增强结构来应对多尺度病灶区域的问题.

虽然上述算法能够在一定程度上解决息肉分割过程中的难题,但模型在设计过程中仍存在以下几个问题:对于多尺度特征的提取和层级特征之间的融合不够充分,造成一些小息肉的漏分割和较大息肉的误分割;(Convolutional Neural Networks,CNN)编码器利用卷积层提取的多级特征中,存在的大量冗余和干扰信息会影响后续的特征建模;对病灶中不确定区域的探索不足,使得在息肉边界处的分割精度不够理想.

针对上述问题,本文提出一种渐进式消减不确定区域的息肉分割模型PRNet,其能够更准确地定位病灶区域和更精确地分割息肉边缘.为适应结肠镜图像中不同大小的病灶区域,模型将跨级特征与多尺度注意融合机制[17](Multi-Scale Channel Attention, MSCA)结合获得全局特征,同时并行使用最大池化和平均池化操作获取注意力权重来对编码特征进行过滤和重定向.随后,将增强注意力和特定阈值结合共同提取息肉边缘,并将这一边缘特征作为引导,进一步探索不确定区域的全局距离依赖关系.最后,应用多尺度上下文感知模块(Multi-scale Contextual Awareness

Module, MCAM)挖掘潜在的病灶特征,完善对息肉区域的建模.在多个息肉分割数据集上进行实验验证,结果表明所提模型优于现有的息肉分割算法,并设置消融实验验证了所提模块的有效性.

1 方法

生物学研究[18]中指出,捕食的过程可以分为三个阶段,即探测、识别和捕获.受这一启发,在进行息肉分割时,同样需要首先确定息肉的大体位置,再逐步精细息肉区域的边缘.因此,本文提出一个包含多尺度跨级融合模块(Multi-scale Cross-level Fusion Module, MCFM)、特征过滤模块(Feature Filter Module, FFM)、不确定区域处理模块(Uncertain area Processing Module, UPM)和 MCAM 四个功能模块的渐进式消减网络 PRNet,总体结构如图 1 所示.首先将结肠镜息肉图像作为 Res2Net[19]编码网络的输入,逐层卷积提取,得到各级特征 f_1、f_2、f_3、f_4 和 f_5,并通过 MCFM 提取多种尺度病灶区域的特征,生成初始预测图.其次,将经过过滤模块的特征和预测图以及设定的阈值作为 UPM 的输入,挖掘病灶区域中的难分割部分并建立空间上下文的长距离依赖关系.最后,MCAM 进一步探索病灶区域周围潜在的语义特征,完善特征建模.

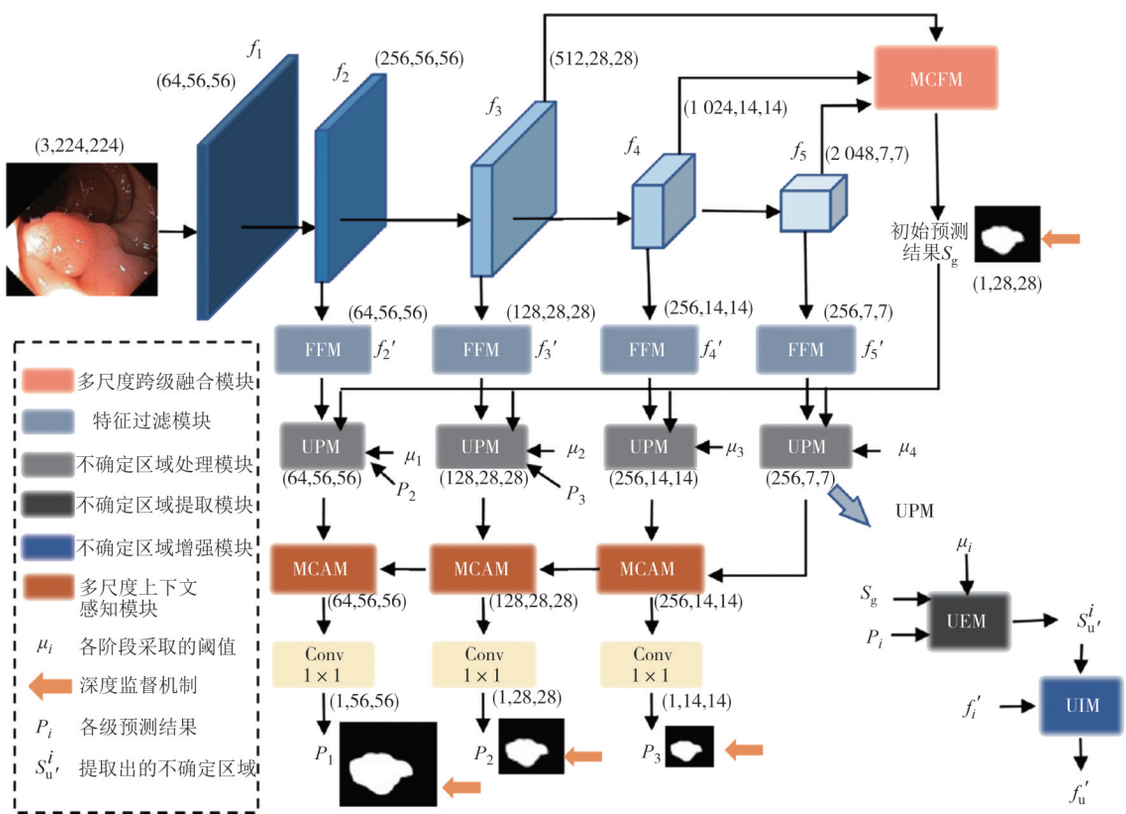


图 1 PRNet 整体架构

Fig.1 The overall structure of PRNet

1.1 多尺度跨级融合模块

经编码器提取到的各级特征中,低层特征由于其与输入图像相距较近,提取到的特征往往包含较多的细节信息,且分辨率较高,需要较多的计算资源;而高层特征则包含更强的抽象语义信息,可为网络提供更准确的分类和判断信息[20].因此,MCFM 只使用高层编码特征 f_3、f_4 和 f_5.图 2 展示了 MCFM 的详细结构.首先将高层特征送入感受野(Receptive Field, RF)结构中捕获多尺度病灶区域特征,得到 f_r_3、f_r_4 和 f_r_5.由于结肠镜图像中息肉的尺寸大小会有很大的差异,单一感受野往往不能同时捕捉多个尺寸的息肉.为了应对这一问题,本算法将 MSCA 和部分解码器(Partial Decoder, PD)结合实现跨级融合,以有效挖掘多尺度信息.对 f_r_3 具体展开进行描述,f_r_5 和 f_r_4 分别上采样 4 倍和 2 倍以匹配 f_r_3 大小,之后将三者在通道维度上拼接并送入 MSCA 中.由于 MSCA 不会改变输入特征的维度,所以之后加入 1×1 卷积块,重新将通道数变为 32.上述过程用公

式表示为:

$$f_f_3 = \mathrm{Conv}\left(\mathrm{MSCA}\left(\mathrm{C}\left(\mathrm{up}\left(f_r_4\right), \mathrm{up}\left(\mathrm{up}\left(f_r_5\right)\right), f_r_3\right)\right)\right) \tag{1}$$

式中:up 表示将图像的分辨率上采样 2 倍;C 表示通道维度上的合并拼接操作;Conv 表示经过 MSCA 之后的卷积恢复通道数操作.f_f_4 同样通过类似的计算得到,f_f_5 即为 f_r_5.最后融合特征 $f_f_j(j \in \{3,4,5\})$ 被送入 PD,获得初始分割图 S_g.PD 整合了三个层次的特征,其中两个是经过 MSCA 之后的特征 f_f_4 和 f_f_3,另一个是 f_5 经过 RF 之后的特征 f_f_5.因此,初始分割图 S_g 可以通过以下方式计算:

$$S_g = \mathrm{PD}\left(f_f_5, f_f_3, f_f_4\right) \tag{2}$$

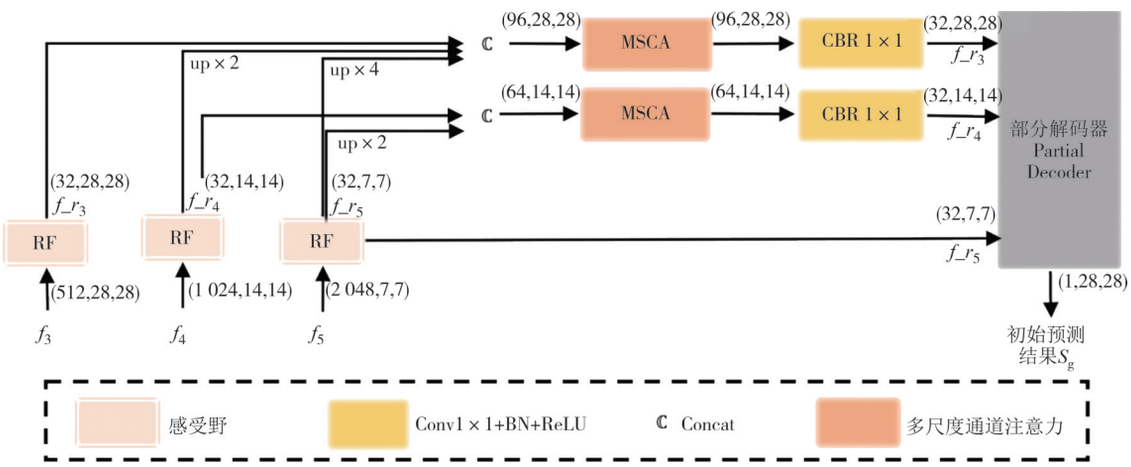


图 2　MCFM 的详细结构

Fig.2　The detailed stucture of MCFM

MSCA 的详细结构如图 3 所示.MSCA 将多尺度与通道注意相结合,是一种基于全局信息和局部信息双利用的自我注意.第一个分支采用全局平均池化来提取全局上下文,它可以保留相对较大息肉的全局位置.第二个分支则保持原始特征的大小来获得局部信息,这可以防止忽略较小的息肉对象.

不同层级特征对于息肉分割任务来说有不同的贡献,因此融合多层级特征能够获得更加全面的特征表示.通过上述的跨级融合,既保留了不同尺寸息肉的特征,又减小了不同层级特征之间的语义差异[21].

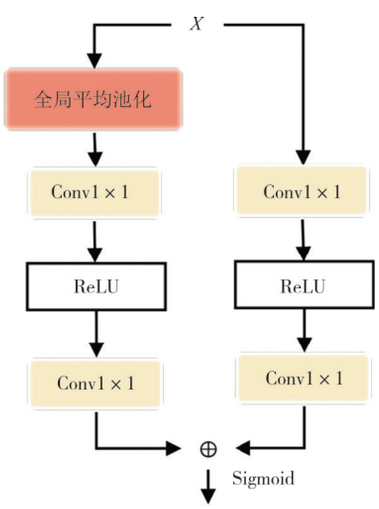


图 3　MSCA 的详细结构图

Fig.3　The detailed structure of MSCA

1.2　特征过滤模块

骨干网络提取到的多尺度特征存在大量冗余信息,并且提取到特征对于各通道的重要程度是相同的,这会导致一些干扰信息影响后续的特征建模.所以,本算法提出 FFM 解决上述问题,模块细节如图 4 所示.首先,对来自编码器的特征 $f_i(i \in \{2,3,4,5\})$ 分别进行全局平均池化(GAP)和全局最大池化(GMP)操作.一般来说,平均池化能够保留更多的背景信息,而最大池化能够保留更多的纹理细节信息.所以,本算法通过并行操作,最大限度地保留丰富的特征空间信息,减少细节丢失.之后使用共享的一维卷积[22]和 Sigmoid 函数归一化注意力权重,得到相对应的通道权重 wei^i,公式表示为:

$$\mathrm{wei}^i = \sigma\left(\mathrm{Conv1d}\left(\mathrm{GAP}\left(f_i\right)\right) \oplus \mathrm{Conv1d}\left(\mathrm{GMP}\left(f_i\right)\right)\right) \tag{3}$$

式中:σ 表示 Sigmoid 函数;conv1d 表示一维卷积;GAP 和 GMP 分别表示全局平均池化和全局最大池化.相比于一般的通道注意力,FFM 提取注意力的方式有以下两点不同:一是并行使用全局最大池化和

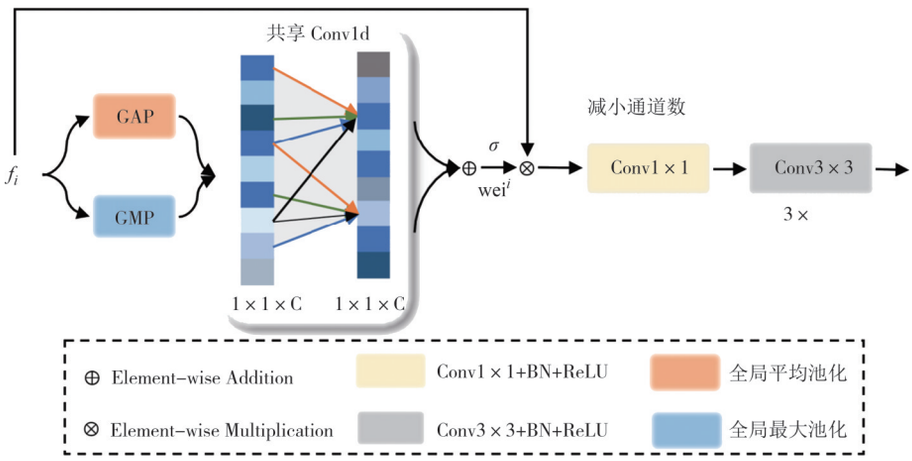


图4 FFM结构

Fig.4 The structure of FFM

全局平均池化两种操作,能够提取到前景和背景两方面的细节信息;二是使用共享一维卷积代替全连接层,能够较好地探索相邻通道间的依赖关系,且参数设计不受限于特定输入特征的大小.

最后,权重 wei^i 与输入特征 f_i 相乘,完成加权,并用1个1×1卷积和3个3×3卷积减小通道数,实现特征重标定.重标定后的特征记为 f_i',用公式表示为:

$$f_i' = \mathrm{Convs}\left(\mathrm{Conv}\left(f_i \otimes \mathrm{wei}^i\right)\right) \tag{4}$$

式中:Convs表示多个3×3卷积;Conv表示1×1卷积.

1.3 不确定区域处理模块

对于结肠镜图像中病灶区域边缘模糊的问题,一些研究试图借助边界信息或反向注意力解决.借助边界信息进行引导的网络[23-24],通常会引入额外的边缘数据,这对于较为庞大的数据集来说是一项耗时的工作.而使用反向注意力虽然可以获得边界线索,但不可避免地会引入背景中的噪声信息,网络性能也会因此受到限制.受文献[16]的启发,本文从不确定区域的角度出发,从中挖掘边界信息来应对息肉边缘对比度低的问题.

UPM由两部分组成,第一部分是不确定区域提取模块(Uncertain Area Extraction Module, UEM),其目的是尽可能提取出清晰准确的不确定区域,以便从中挖掘病灶边缘;第二部分是不确定区域增强模块(Uncertain Area Improvement Module, UIM),目的是增强模型对不确定区域的建模能力.

1.3.1 不确定区域提取模块

在初始预测结果 S_g 中,通常由于息肉边缘和周围组织的形态相似,使得 S_g 边缘部分的分割精度较低.对此,本算法引入增强注意力并使其与上一阶段的预测图 P_i 相结合(第三、四阶段的UEM除外),以提升 S_g 的有效性.增强后的预测图记为 S_e:

$$S_e = f_{\max}\left(g\left(S_g, \sigma, \lambda\right), \mathrm{Sigmoid}\left(S_g\right)\right) \oplus P_i \tag{5}$$

式中:$g(\cdot)$是一个经典的滤波器函数,标准偏差 σ=32,核大小 λ=4,之后进行归一化操作;$f_{\max}(.)$为最大值函数,用于突出 S_g 区域;⊕ 表示按元素相加操作.本算法将 S_e 看作两部分,一部分是由那些明确属于前景构成的确定性区域,其他则为不确定区域.具体来说,将大于 μ_i 的像素赋值为0,即将明确属于前景的像素区域不予考虑,剩余部分即为不确定区域,对不确定区域利用表达式(7)增强.具体过程如图5(a)所示,公式表示如下:

$$S_u^i = F\left(S_e\right) = \begin{cases} 0, & S_e \geqslant \mu_i \\ S_e, & S_e < \mu_i \end{cases} \tag{6}$$

$$S_{u'}^i = 1 - \frac{\mathrm{abs}\left(S_u^i - 0.5\right)}{0.5} \tag{7}$$

3.3节中对 μ_i 的设置进行了对比实验,本算法 μ_i 采用{0.75,0.75,0.7,0.7}.图5(b)是使用文献[13]中的方法提取的不确定区域,对比图5(a)和图5(b)发现,本文提出的方法提取出更加清晰的息肉边界的同时增强了其中的不明显区域,为之后的上下文信息探索提供更好的基础.

1.3.2 不确定区域增强模块

通过之前的工作UEM和FFM,获得了息肉图像中不易判别的区域 $S_{u'}^i$ 和有价值的编码特征 f_i'.二者共同作为UIM的输入,其中 $S_{u'}^i$ 作为引导特征,用以探索 f_i' 中较难分割的区域.具体如图6所示,使 $S_{u'}^i$ 和 f_i' 按元素相乘并接上3×3卷积和跳跃连接,从而增强网络对分割过程中不确定区域的建模能力.之后为了进一步建立病灶完整区域的长距离依赖,并考虑到模型大小,使用两次十字交叉注意力[25]来获得像

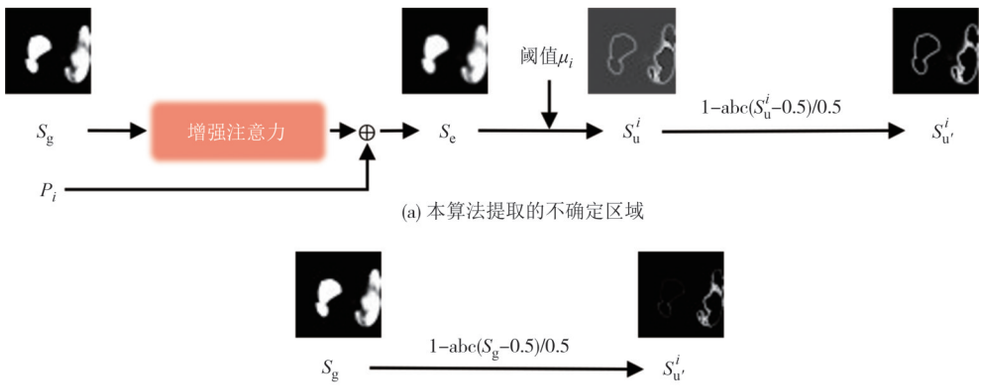


图 5　本算法提取的不确定区域(a)与文献[13]中提取的不确定区域(b)的比较

Fig.5　Comparison between the uncertain areas extracted using the algorithm in this paper and the reference[13]

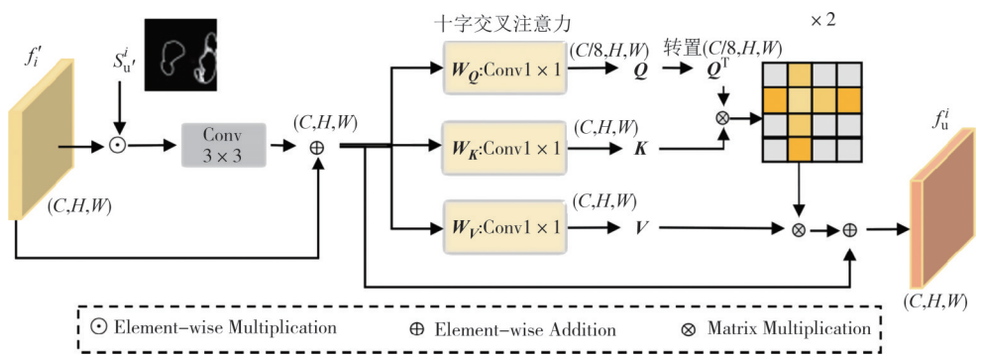


图 6　UIM 的详细结构

Fig.6　The detailed structure of UIM

素之间的全局关系,最终得到 UIM 的输出f_u^i,这一过程用公式表示为:

$$f_u^i = \mathrm{CCA}\left(\mathrm{CCA}\left(f_e^i \oplus \mathrm{Conv}\left(S_{u'}^i \otimes f_i'\right)\right)\right) \tag{8}$$

式中:CCA 表示十字交叉注意力;Conv 表示 3×3 卷积.

UEM 和 UIM 的紧密配合,增强了模型对病灶区域中难分割部分的建模能力,并通过自上而下设置的递减阈值,逐步缩小了不确定区域的范围,进而挖掘出息肉的边界信息.

1.4　多尺度上下文感知模块

为缓解各层级特征之间的语义差异并继续挖掘病灶区域周围隐藏的上下文信息,提出 MCAM.本文在文献[26]对于多尺度融合探索的基础上,以本级 UPM 的输出和上一级 UPM 或 MCAM 的输出作为该模块的输入,结构如图 7 所示.下面以f_u^4和f_u^5举例进行说明.首先使f_u^5上采样到f_u^4大小,并使二者相乘,以深层特征指导浅层特征语义表达,同时使用残差连接保留本级特征信息.之后将重新建模后的f_u^4与f_u^5拼接、卷积,得到初始聚合特征f_a.

在挖掘目标区域潜在上下文语义信息时,文献[27]中提出使用 1×1 卷积缩小通道数,并通过级联 4 个分支进行不同感受野的特征搜索.考虑到使用 1×1 卷积进行直接降维,会不可避免地丢失一些信息.因此本算法将通道数分为 4 组$\{f_a^1, f_a^2, f_a^3, f_a^4\}$,每一组采用大小不同的空洞速率进行学习.并考虑到每组之间的通道语义相关性,使用密集连接的方式融合多个尺度的信息,从而得到多尺度特征$\{f_a^{1'}, f_a^{2'}, f_a^{3'}, f_a^{4'}\}$,用公式表述为:

$$f_a^{j'} = \begin{cases} \mathrm{Conv}^r\left(f_a^{j+1} \oplus f_a^j\right), & j = 1 \\ \mathrm{Conv}^r\left(\left(\sum_{m=1}^{j-1} f_a^{m'}\right) \oplus f_a^{j+1} \oplus f_a^j\right), & j \in \{2, 3\} \\ \mathrm{Conv}^r\left(\left(\sum_{m=1}^{j-1} f_a^{m'}\right) \oplus f_a^j\right), & j = 4 \end{cases} \tag{9}$$

式中:Conv^r表示卷积核为 3×3,空洞速率为 r 的空洞卷积,$r \in \{1, 2, 3, 4\}$.最后将分支特征$f_a^{j'}$拼接得到f_a',

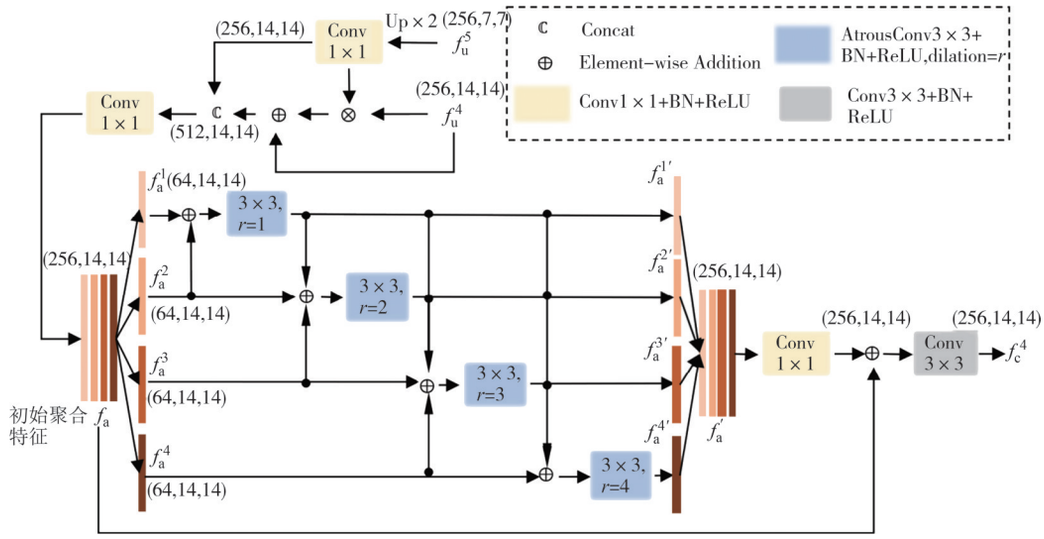


图 7 MCAM 结构

Fig.7 The structure of MCAM

并经过 1×1 卷积、残差连接和 3×3 卷积,得到 MCAM 的最终输出.公式表示为:

$$f_c^4 = \mathrm{Conv3}\left(f_a \oplus \mathrm{Conv1}\left(\mathbb{C}\left(f_a^{j'}\right)\right)\right) \tag{10}$$

式中:Conv3 和 Conv1 分别表示 3×3 卷积和 1×1 卷积,$\mathbb{C}$ 表示拼接操作.最终,为了得到各级预测图 $P_i(i \in \{1,2,3\})$,对各级 MCAM 的输出 f_c^i 使用 1×1 卷积将通道数缩减为 1.

2 数据来源与实验细节

2.1 数据集设置

数据集:本算法在 3 个息肉分割数据集上进行了评估:ETIS[28]、CVC-Clinic[29]和 Kvasir-SEG[30].其中,ETIS 包含 196 张息肉图像,分辨率为 1 255×966,常用于结直肠癌的早期诊断.CVC-Clinic 又称 CVC-612,包括来自 31 个结肠镜检查片段的 612 张开放性图像,分辨率均为 384×288.Kvasir-SEG 是一个具有挑战性的数据集,包含 1 000 张图像,分辨率不等.

实验中对 Kvasir-SEG 和 CVC-Clinic 两个数据集进行随机划分,80% 用于训练,10% 用于验证,10% 用于测试模型的学习能力.另外 ETIS 同样被用作测试集,用于测试模型的泛化能力.

2.2 损失函数设计

本算法使用的损失函数为 $L = L_{\mathrm{IoU}}^{\mathrm{w}} + L_{\mathrm{BCE}}^{\mathrm{w}}$,其中 $L_{\mathrm{IoU}}^{\mathrm{w}}$ 代表加权的 IoU 损失,$L_{\mathrm{BCE}}^{\mathrm{w}}$ 代表二进制交叉熵损失,前者提供全局性的监督而后者重点关注像素级的损失.两种损失都更加关注难分类像素,对难分类像素分配较大权重.初始分割图 S_g 和三个侧输出 $P_i(i \in \{1,2,3\})$ 均上采样恢复到原分辨率,并和 Ground-Truth 采用上述损失函数进行深度监督.因此,损失函数表述为:

$$L_{\mathrm{total}} = \lambda L\left(S_g^{\mathrm{up}}, G\right) + \sum_{i=1}^{3} L\left(P_i^{\mathrm{up}}, G\right) \tag{11}$$

式中:λ 是一个平衡参数,本算法设定为 3.

2.3 实施细节和评估指标

本文提出的 PRNet 基于 PyTorch 框架实现,并使用 NVIDIA Geforce RTX 3080 GPU 进行训练.输入模型的图像统一调整为 224×224 分辨率,并通过随机水平翻转和垂直翻转进行增强.使用在 ImageNet 上预训练的 Res2Net50 模型初始化编码器网络参数,其他层随机初始化.使用多尺度训练策略{1,1.5}进行模型训练[31],Adam 优化器[32]进行损失优化,批次规模设置为 8,迭代次数设置为 60.学习率初始化为 1e-4,采用 poly 策略进行学习率的调整,power 为 0.9.

采用 8 个广泛使用的评估指标评估模型性能,包括“准确率”(Acc)、“Dice 系数”(Dice)、“召回率”(Rec)、“查准率”(Prec)、“息肉交并比”(IoUp)、“背景交并比”(IoUb)、“平均交并比”(Mean Intersection over Union,mIoU)、“平均绝对误差”(Mean Absolute Error, MAE).评估指标定义如式(12)所示.

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$\mathrm{Dice} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

$$\mathrm{IoUp}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

$$\mathrm{IoUb}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

$$\mathrm{mIoU}=\left(\mathrm{IoUp}+\mathrm{IoUb}\right)/2$$

$$\mathrm{Prec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

$$\mathrm{Rec}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\mathrm{MAE}=\frac{1}{w\times h}\sum_{x}^{w}\sum_{y}^{h}\left|P\left(x,y\right)-G\left(x,y\right)\right| \tag{12}$$

式中:TP、TN、FP和FN均在像素级别上定义,分别代表真阳性、真阴性、假阳性和假阴性的数量;w、h分别为图像的宽和高;$P\left(x,y\right)$和$G\left(x,y\right)$分别表示预测结果和真实标签中(x,y)处的值.Acc指被正确分类的像素的百分比,该指标易受类别不均衡的影响.Prec表示预测结果为阳性中真阳性所占的比例;Rec则关注真实标签被正确预测的概率.指标IoU和Dice衡量预测结果与真实标签之间的相似度.其中,IoUp和IoUb分别代表网络模型对于前景像素和背景像素的分类准确度.当二者差异较大时(通常发生在息肉与周围黏膜相似度较高的场景),单一指标难以评估网络模型的分割效果.相比之下,mIoU能够更加客观地展示网络性能.Dice系数在评估小目标的分割性能方面更具表现力.MAE通过计算预测值与真实标签之间距离的平均值来评估像素级精度.前7个指标数值越高越好,最后一个指标MAE,其值越小,表明预测值与真实值越吻合.

3　实验结果与分析

为了证明PRNet对于结肠镜图像中息肉分割的有效性,将其与7种目前最具代表性的医学图像分割方法进行学习能力和泛化能力的比较,包括U-Net、U-Net++[33]、PraNet、ACSNet、HarD-MSEG[34]、UACANet、LDNet[35].为了公平比较,所有实验均在相同条件下进行,所有预测结果都由开放源代码重新训练产生.

3.1　学习能力分析

如表1和表2所示,PRNet在Kvasir-SEG和CVC-Clinic数据集上都取得最好结果.在Kvasir-SEG数据集上,PRNet达到92.09%的相似性系数和92.20%的平均交并比,同时在CVC-Clinic数据集上Dice系数和mIoU分别为93.05%和93.61%.这表明PRNet具有更强大的学习能力,可以在复杂情况下精确地分割息肉.

表1　不同模型在Kvasir-SEG数据集上的分割结果

Tab.1　The segmentation results of different models on the Kvasir-SEG dataset

模型	Dice	mIoU	IoUp	IoUb	Acc	Prec	Rec	MAE
U-Net[7]	0.830 6	0.845 7	0.749 7	0.941 8	0.953 0	0.824 8	0.901 2	0.049 0
U-Net++[33]	0.853 6	0.856 9	0.760 8	0.952 9	0.963 6	0.874 5	0.890 5	0.051 5
PraNet[11]	0.904 0	0.905 6	0.848 3	0.962 8	0.970 5	0.912 0	0.927 3	0.033 1
ACSNet[13]	0.900 3	0.899 6	0.840 8	0.958 5	0.968 2	0.921 9	0.903 3	0.035 0
HarD-MSEG[34]	0.904 4	0.906 2	0.848 6	0.963 8	0.970 4	0.942 3	0.897 0	0.031 7
LDNet[35]	0.899 7	0.902 2	0.843 1	0.961 4	0.969 9	0.916 9	0.911 7	0.033 4
UACANet[15]	0.908 3	0.906 7	0.854 6	0.958 8	0.968 9	0.916 9	**0.921 5**	0.027 8
本文	**0.920 9**	**0.922 0**	**0.872 4**	**0.971 6**	**0.978 3**	**0.950 4**	0.914 0	**0.023 9**

表2　不同模型在CVC-Clinic数据集上的分割结果

Tab.2　The segmentation results of different models on the CVC-Clinic dataset

模型	Dice	mIoU	IoUp	IoUb	Acc	Prec	Rec	MAE
U-Net[7]	0.870 5	0.892 9	0.805 2	0.980 6	0.982 8	0.877 4	0.899 4	0.017 4
U-Net++[33]	0.881 4	0.894 2	0.814 8	0.973 6	0.983 0	0.880 1	0.899 0	0.026 2
PraNet[11]	0.905 3	0.917 2	0.848 5	0.986 0	0.987 8	0.876 1	**0.965 8**	0.013 4
ACSNet[13]	0.903 6	0.917 7	0.849 0	0.986 5	0.988 6	0.904 5	0.927 5	0.012 5
HarD-MSEG[34]	0.917 9	0.923 2	0.858 6	0.987 8	0.989 8	0.911 9	0.937 5	0.010 4
LDNet[35]	0.893 4	0.911 2	0.835 3	0.987 2	0.988 7	0.873 7	0.952 8	0.012 9
UACANet[15]	0.903 8	0.918 8	0.851 1	0.986 5	0.988 0	0.888 0	0.941 4	0.008 5
本文	**0.930 5**	**0.936 1**	**0.880 6**	**0.991 6**	**0.992 7**	**0.928 3**	0.938 6	**0.007 5**

图 8 展示了 PRNet 与其他 7 种模型在 Kvasir-SEG 和 CVC-Clinic 数据集上一些复杂情况下的视觉分割结果.可以观察到:1)当息肉与周围黏膜组织易于区分时(第一行、第三行),U-Net、U-Net++、PraNet 和 HarD-MSEG 不能精准地定位息肉,出现分割结果内部不连续、错分割现象;而 ACSNet 和 LDNet 能有效地分割出病变区域,但病灶内有明显的伪影;2)当息肉与周围黏膜组织极为相似时(第二行、第六行),前 6 种模型均未能正确识别出息肉,出现了分割过度和分割缺失的情况;而 PRNet 在前景和背景具有极大相似性的情况下,仍然能够正确定位图像中的息肉区域,说明 PRNet 具有更强的息肉特征建模能力;3)总体上,PRNet 能够应对结肠镜中息肉尺寸差异大的情况,这主要得益于模型中使用的多尺度特征融合策略,使得模型具有多个感受野,能够捕捉到大小不等的息肉.

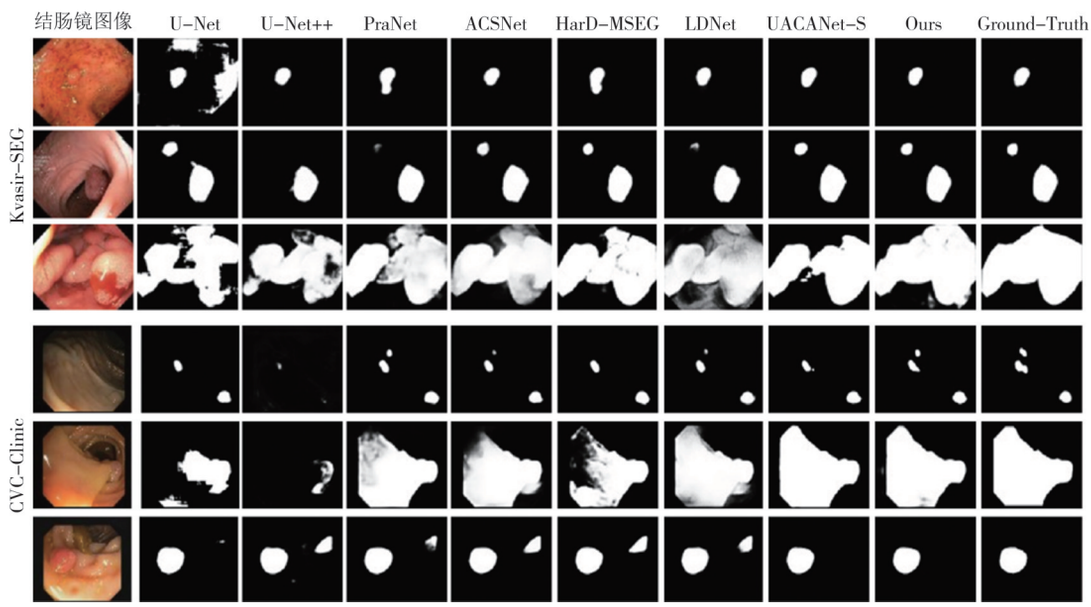


图 8 不同模型在 Kvasir-SEG 和 CVC-Clinic 数据集上分割预测对比图像

Fig.8 Comparison of segmentation predictions by different models on Kvasir-SEG and CVC-Clinic datasets

3.2 泛化能力分析

为了适应临床场景中不同类型的息肉,需要检测模型对未知数据的预测能力.表 3 列出了在 ETIS 数据集上各模型的评估结果.由表 3 数据可以看出,PRNet 在 ETIS 数据集上取得了最优结果,Dice 系数和平均交并比为 74.19% 和 82.66%,表明 PRNet 在未知数据上的泛化能力较强,相比其他模型更能够提取出病灶区域的一般化特征.

表 3 不同模型在 ETIS 数据集上的分割结果比较

Tab.3 Comparison of segmentation results of different models on the ETIS dataset

模型	Dice	mIoU	IoUp	IoUb	Acc	Prec	Rec	MAE
U-Net[7]	0.415 8	0.650 5	0.959 1	0.343 5	0.957 5	0.440 0	0.508 6	0.045 3
U-Net++[33]	0.495 1	0.694 9	0.972 6	0.418 3	0.971 4	0.590 0	0.521 6	0.033 4
PraNet[11]	0.665 0	0.762 2	0.940 9	0.585 4	0.939 0	0.614 9	**0.856 6**	0.062 7
ACSNet[13]	0.620 9	0.742 9	0.941 8	0.546 5	0.939 3	0.590 4	0.799 5	0.062 3
HarD-MSEG[34]	0.727 8	0.814 4	0.978 5	0.651 4	0.977 3	0.695 4	0.815 6	0.022 1
LDNet[35]	0.666 5	0.766 3	0.953 4	0.588 1	0.951 7	0.630 7	0.826 2	0.051 6
UACANet[15]	0.605 8	0.687 7	0.865 4	0.512 8	0.862 5	0.556 0	0.823 7	0.068 5
本文	**0.741 9**	**0.826 6**	**0.981 3**	**0.673 0**	**0.980 1**	**0.736 3**	0.790 4	**0.019 5**

3.3 消融实验分析

为了验证各模块的有效性,在数据集 Kvasir-SEG 和 ETIS 上进行消融实验,结果如表 4 所示.基线模型(B),删除了所有附加模块(即 MCFM、FFM、UPM 和 MCAM),只保留了 FFM 中 Reduce Channel 的 1×1 卷积以减少通道数,以及 MCAM 中基础的拼接

操作.

首先,与基线模型相比,使用FFM在Kvasir-SEG和ETIS数据集上Dice系数分别提升了2.20%和8.32%.这表明对编码特征进行重标定是有效的,减少了冗余和干扰信息对特征建模的影响.其次,对比第一行、第三行的实验结果,使用MCFM在两个数据集上Dice系数有明显的提高,分别提高了1.45%和8.61%,验证了MCFM能有效地聚焦图像中的病灶区域,从不同尺度捕获息肉特征,进而提升分割精度.再次,通过将UPM引入B+MCFM模型来验证UPM的有效性.对比第三行和第四行结果可以发现,B+MCFM+UPM模型取得了更好的整体性能,在测试的2个数据集上所有指标均有一定程度的提升,特别是Dice系数在ETIS数据集上提升了1.98%.这表明使用UPM对不确定区域进行提取并建立病灶区域的长期依赖关系对提升分割病灶区域的准确度和稳定性是有帮助的.最后,为了验证MCAM的有效性,将基线模型中使用的简单通道拼接换为MCAM中的具体结构.通过对比第一行和第五行的结果,Dice系数在Kvasir-SEG数据集提升了1.73%.这表明MCAM能减少解码过程中层级特征之间的语义差异,同时能够挖掘出病灶区域周围潜在的上下文语义信息,从而更好地实现模型对病灶区域的建模.

另外,对于UPM中的阈值设置也进行了对比实验,结果如表5所示.由表中结果可以看到,第一组{0.8,0.75,0.7,0.65}虽然在Kvasir-SEG数据集上取得了和本模型选取的{0.75,0.75,0.7,0.7}最为接近的结果,但其在ETIS数据集上相较于{0.75,0.75,0.7,0.7}结果相差较大.这说明第一组的阈值设置只学习到了训练集中的一些数据特点,但对未知数据的概括能力较差.所以,在阈值选择上,本算法选择{0.75,0.75,0.7,0.7}.

表4　在Kvasir-SEG和ETIS数据集上的消融实验

Tab.4　Ablation experiments on the Kvasir-SEG and ETIS datasets

模型	Kvasir-SEG				ETIS			
	Acc	Dice	mIoU	MAE	Acc	Dice	mIoU	MAE
B	0.973 1	0.893 7	0.906 4	0.029 4	0.983 6	0.634 2	0.781 4	0.021 4
B+FFM	0.975 7	0.915 7	0.915 8	0.026 8	0.984 1	0.717 4	0.817 9	0.017 2
B+MCFM	0.975 8	0.908 2	0.912 8	0.026 5	**0.984 9**	0.720 3	0.814 9	**0.016 3**
B+MCFM+UPM	0.976 3	0.910 3	0.913 6	0.025 7	0.979 7	0.740 1	0.824 4	0.020 6
B+MCAM	0.971 7	0.911 0	0.911 4	0.030 5	0.965 4	0.633 6	0.767 6	0.037 1
PRNet	**0.978 3**	**0.920 9**	**0.922 0**	**0.023 9**	0.980 1	**0.741 9**	**0.826 6**	0.019 5

表5　在Kvasir-SEG和ETIS数据集上对阈值设置的消融研究

Tab.5　Ablation studies of threshold settings on Kvasir-SEG and ETIS datasets

阈值设置	Kvasir-SEG				ETIS			
	Acc	Dice	mIoU	MAE	Acc	Dice	mIoU	MAE
{0.8,0.75,0.7,0.65}	0.974 1	0.911 8	0.915 3	0.028 1	0.964 9	0.720 6	0.807 9	0.035 4
{0.75,0.7,0.65,0.6}	0.972 4	0.906 9	0.909 1	0.030 0	0.944 2	0.674 9	0.776 7	0.057 5
{0.7,0.65,0.5,0.5}	0.972 0	0.904 9	0.910 7	0.030 0	0.972 2	0.697 5	0.799 5	0.029 0
{0.75,0.75,0.7,0.7}	**0.978 3**	**0.920 9**	**0.922 0**	**0.023 9**	**0.981 3**	**0.741 9**	**0.826 6**	**0.019 5**

4　结 论

本文提出一个通过渐进式消减不确定区域提升息肉分割精度的模型PRNet.PRNet包含多尺度跨级融合模块、特征过滤模块、不确定区域处理模块和多尺度上下文感知模块四部分.首先,借助多尺度跨级融合模块整合注意融合机制和多尺度特征,解决单一感受野难以同时捕捉不同大小息肉的难题.其次,通过特征过滤模块提取有价值的信息并与多尺度跨级融合模块的输出一起作为不确定区域处理模块的输入,并自上而下设置递减的阈值,逐步增强对病灶区域边缘的特征提取和特征以建模能力,解决息肉边界分割准确度低的难题.最后,多尺度上下文感知模块融合相邻层级特征缩小语义鸿沟,并结合多尺度感受野共同探索目标区域潜在的上下文语义信息.在Kvasir-SEG和CVC-Clinic数据集上的实验结果表明,PRNet相比其他模型具有更强的学习能力,

能够在病灶区域边缘模糊的情况下实现精准分割;在ETIS数据集上,PRNet相比其他7种分割算法在Dice系数上有所提升,7种算法系数不同,表明所提算法具有更好的泛化能力.

由于肺部感染分割任务与结直肠息肉分割面临的挑战类似,所以未来会进一步将PRNet迁移到肺部感染分割任务上,观察PRNet在其他数据上的特征学习和泛化能力,在此基础上优化边界提取,进一步提升在病灶区域边缘处的分割精度.

参考文献

[1] 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.

[2] LEUFKENS A M, VAN OIJEN M G H, VLEGGAAR F P, et al. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study[J]. Endoscopy, 2012, 44(5): 470-475.

[3] LI Q L, YANG G Y, CHEN Z W, et al. Colorectal polyp segmentation using a fully convolutional neural network[C]//2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Shanghai, China: IEEE, 2017: 1-5.

[4] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.

[5] BRANDAO P, MAZOMENOS E, CIUTI G, et al. Fully convolutional neural networks for polyp segmentation in colonoscopy[C]//SPIE Proceedings, Medical Imaging 2017: Computer-Aided Diagnosis. Orlando, Florida, USA. SPIE, 2017.

[6] 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 (ISM). San Diego, CA, USA: IEEE, 2019: 225-232.

[7] 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. Cham: Springer, 2015: 234-241.

[8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.

[9] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[10] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision-ECCV 2018: 15th European Conference, Munich, Germany, September 8 - 14, 2018, Proceedings, Part VII. ECOCV, ACM, 2018: 833-851.

[11] FAN D P, JI G P, ZHOU T, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2020: 263-273.

[12] CHEN S H, TAN X L, WANG B, et al. Reverse attention for salient object detection[C]//European Conference on Computer Vision. Cham: Springer, 2018: 236-252.

[13] ZHANG R F, LI G B, LI Z, et al. Adaptive context selection for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2020: 253-262.

[14] NGUYEN T C, NGUYEN T P, DIEP G H, et al. CCBANet: cascading context and balancing attention for polyp segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 633-643.

[15] KIM T, LEE H, KIM D. UACANet: uncertainty augmented context attention for polyp segmentation[C]//Proceedings of the 29th ACM International Conference on Multimedia. October 20-24, 2021. Virtual Event, China: ACM, 2021: 2167-2175.

[16] GUO Q Q, FANG X Y, WANG L B, et al. Polyp segmentation of colonoscopy images by exploring the uncertain areas[J]. IEEE Access, 2022, 10: 52971-52981.

[17] DAI Y M, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2021: 3559-3568.

[18] HALL J R, CUTHILL I C, BADDELEY R, et al. Camouflage, detection and identification of moving targets[J]. Proceedings Biological Sciences, 2013, 280(1758): 20130064.

[19] GAO S 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.

[20] WU Z, SU L, HUANG Q M. Cascaded partial decoder for fast and accurate salient object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 3902-3911.

[21] CHEN Z Y, XU Q Q, CONG R M, et al. Global context-aware progressive aggregation network for salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10599-10606.

[22] WANG Q L, WU B G, ZHU P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 11531-

11539.

[23] MURUGESAN B, SARVESWARAN K, SHANKARANARAYANA S M, et al. Psi-Net: shape and boundary aware joint multi-task deep network for medical image segmentation [C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Berlin, Germany: IEEE, 2019: 7223-7226.

[24] FANG Y Q, ZHU D L, YAO J H, et al. ABC-net: area-boundary constraint network with dynamical feature selection for colorectal polyp segmentation [J]. IEEE Sensors Journal, 2021, 21(10): 11799-11809.

[25] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 603-612.

[26] 朱新山,卢俊彦,甘永东,等. 融合多尺度特征与多分支预测的多操作检测网络[J]. 湖南大学学报(自然科学版),2023,50(8):94-105.
ZHU X S, LU J Y, GAN Y D, et al. Multi-manipulation detection network combining multi-scale feature and multi-branch prediction [J]. Journal of Hunan University (Natural Sciences), 2023, 50(8): 94-105. (in Chinese)

[27] MEI H Y, LIU Y Y, WEI Z Q, et al. Exploring dense context for salient object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1378-1389.

[28] 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.

[29] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPARRACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: validation *vs*. saliency maps from physicians [J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.

[30] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: a segmented polyp dataset [C]//International Conference on Multimedia Modeling. Cham: Springer, 2020: 451-462.

[31] ZHENG D H, ZHENG X C, YANG L T, et al. MFFN: multi-view feature fusion network for camouflaged object detection [C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2023: 6221-6231.

[32] KINGMA D P, BA J. Adam: a method for stochastic optimization [J]. ArXiv e-Prints, 2014: arXiv: 1412.6980.

[33] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. UNet++ : a nested U-net architecture for medical image segmentation [J]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain: 2018, 11045: 3-11.

[34] HUANG C H , WU H Y , LINY L .HarDNet-MSEG: a simple encoder-decoder polyp segmentation neural network that achieves over 0.9 Mean Dice and 86 FPS[J], 2021.

[35] ZHANG R F, LAI P W, WAN X, et al. Lesion-aware dynamic kernel for Polyp segmentation [M]//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022: 99-109.