

基于混合域注意力 ResNeSt 的结肠息肉分割模型

周孟然,刘思怡,卞凯,王宁,高立鹏

安徽理工大学 电气与信息工程学院,安徽 淮南 232001

摘要:目的 针对因息肉大小不一,边界不清,光线影响,在图片中所占比例较小导致的分割精度不高的问题,提出了一种改进的 U 型结构网络 BMR-Net。方法 该模型的框架为编码器-解码器形式,在编码器部分采用 ResNeSt 提取特征,在计算成本增加很少的情况下改善了特征提取效果;在编码器和解码器之间设计边界预测生成模块(BPGM)来聚合高层特征并加入改良空间金字塔池化模块,在其中引入注意力机制,提升多尺度信息融合效果,获得更精确的全局特征图表示;针对不清晰的边缘部分采用反向注意力模块,删除已预测区域,校正边界信息。结果 在 CVC-ClinicDB、Kvasir-SEG、CVC-ColonDB、ETIS-Larib、EndoScene 数据集上进行测试,mDice 值分别达到了 0.930、0.903、0.743、0.712、0.874。结论 该方法分割性能和泛化性能均优于其他的先进方法,并且可以更加精确和完整地分割出小尺寸息肉,可以为结肠息肉患者提供早期预后信息。

关键词:图像分割;结肠息肉;ResNeSt;编解码网络;注意力机制

中图分类号:TP391.4 文献标识码:A doi:10.16055/j.issn.1672-058X.2025.0001.012

Colon Polyp Segmentation Model Based on Mixed-domain Attention ResNeSt

ZHOU Mengran, LIU Siyi, BIAN Kai, WANG Ning, GAO Lipeng

School of Electrical and Information Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective To address the challenges posed by polyps of varied sizes, unclear boundaries, lighting effects, and their relatively small proportions in images that result in lower segmentation accuracy, an improved U-shaped network structure, BMR-Net, was proposed. **Methods** The model adopted an encoder-decoder architecture. The encoder partially utilized ResNeSt for feature extraction, enhancing the feature extraction performance with only a slight increase in computational cost. Between the encoder and the decoder, a boundary prediction generation module (BPGM) was designed to aggregate high-level features and incorporate a modified spatial pyramid pooling module, in which an attention mechanism was introduced. This promoted multi-scale information fusion, obtaining a more accurate global feature map representation. For unclear edge areas, a reverse attention module was applied to remove previously predicted areas and correct the boundary information. **Results** Tests were conducted on the CVC-ClinicDB, Kvasir-SEG, CVC-ColonDB, ETIS-Larib, and EndoScene datasets, with mDice values reaching 0.930, 0.903, 0.743, 0.712, and 0.874, respectively. **Conclusion** This method outperforms other advanced methods in terms of segmentation performance and

收稿日期:2023-09-16 修回日期:2023-11-18 文章编号:1672-058X(2025)01-0085-09

基金项目:安徽省科技重大专项项目(201903A07020013);教育部产学研创新基金(2019ITA01010);安徽理工大学引进人才科研启动基金项目(2022YJRC43)。

作者简介:周孟然(1965—),男,安徽淮南人,教授,博士,从事矿山机电系统监测、光电信息处理、煤矿安全监测研究。

通讯作者:刘思怡(1999—),女,安徽淮南人,硕士研究生,从事图像处理研究。Email:daisy641ccc@163.com。

引用格式:周孟然,刘思怡,卞凯,等.基于混合域注意力 ResNeSt 的结肠息肉分割模型[J].重庆工商大学学报(自然科学版),2025,42(1):85-93.

ZHOU Mengran, LIU Siyi, BIAN Kai, et al. Colon polyp segmentation model based on mixed-domain attention ResNeSt [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(1): 85-93.

generalization ability. Furthermore, it can segment small-sized polyps more precisely and completely, providing early prognosis information for patients with colon polyps.

Keywords: image segmentation; colonic polyps; ResNeSt; encoder-decoder network; attention mechanism

1 引言

肿瘤(腺瘤性)结直肠息肉是起源于结肠上皮细胞的良性肿瘤,十分常见。它具有恶性潜能,随着时间的推移,可能会向周围的器官扩散,发展成结直肠癌^[1]。结直肠癌占有所有癌症发病率的 9% 以上,是世界上第三大最常见的癌症和第四大最常见的死亡原因^[2]。我国的结直肠癌患者在近些年持续增多,癌症统计报告显示,2020 年中国结直肠癌新发病例数已有 55.5 万,死亡病例数已有 28.6 万^[3]。息肉的准确识别对大肠癌的早期诊断和治疗至关重要,目前的结肠镜检查中,人工复查费力费时,需要胃肠病学家的全神贯注,且容易出现漏检。因此,研究可以帮助自动识别息肉的计算机辅助算法将极大地提高结肠镜检查的诊断准确性和效率^[4]。

早期的息肉分割方法要先通过手工提取,将息肉的颜色、形状、或者是一些其他的特征因素送入分类器训练,再通过分类器区分息肉及其周边环境。但是这种方法导致了较高的漏检率,究其原因手工提取的特征表示能力十分有限。随着近年来深度学习的发展,这一策略被成功应用在了息肉分割中,涌现出了一大批基于深度学习的息肉检测网络^[5-6]。Brandao 等^[7]扩展传统 CNN,首先将 FCN 架构应用到分割息肉上,第一次完整地分割出了息肉,实现了高分割精度。Akbari 等^[8]在 FCN 的基础上加以改进,在图像增强阶段中使用 patch 选择方法,并采用后处理方法降低了假阳性率,以此得到了更好的分割结果。U-Net^[9]作为最有代表性的分割网络,采用了编解码结构,在它们中间引入了跳跃连接,使得高级语义信息可以在编码器和解码器之间传递,有助于像素级的精确分割。Zhou 等^[10]提出了 U-Net 的变体 UNet++,将单一的连接形式改为了密集跳跃连接,有效地捕获不同层次的语义信息,提高了模型的分割性能。ResUNet^[11]将残差连接形式引入到 UNet 的子模块中,使网络可以学习更复杂的特征表示。但因为引入残差连接,网络更加复杂,需要更多计算资源。Deng 等^[12]在网络的解码器中应用反向注意力,设计出了 PraNet,通过渐进整合的策略,将多个尺度的特征聚合起来,对于不同尺寸的目标达到了更好的分割效果。Feng 等^[13]使用阶梯融合模块构建了一种阶梯型网络,使用跳过连接、双注意力模块和最终的多

尺度融合模块融合不同尺度的特征。Wang 等^[14]运用了 DoubleU-Net 对息肉图像进行分割。类似于两个平行的 U 型网络,将它们的底部进行连接,同时进行初始分割和文本指导的增强分割,进行更全面的帮助。以上方法对于息肉的分割结果都起到了一定改善,但由于息肉大小不统一,许多模型只关注了大体区域部分,忽略了边缘信息,导致分割出来的结果不完整或分割过度,难以检测出小目标对象。

本文提出了一种结合边界预测生成模块(Boundary Prediction Generate Module, BPGM)和反向注意力模块(Reserve Attention, RA)^[15]的新的 U 型网络用于改进息肉分割效果和提高小目标分割性能,取名为 BMR-Net。在编码器末端设计加入 BPGM 模块,聚合了高级特征,去除了效率较低的浅层特征,防止计算资源冗余,充分结合了多尺度信息,改善对小息肉的分割;加入 SE 通道注意力机制,让模块在提取特征时分重点性地关注不同特征,改善 ASPP 因空洞率变大导致的空洞卷积效果变差的问题;引进 RA 模块,通过从主干网络的侧输出特征中删除目前已预测到的明显区域,补充边界信息,迭代和细化预测结果,从而提高分割精度。BMR-Net 改进了特征提取网络,融合了多尺度特征信息,即使是形状较小的息肉也可以准确分割,能够对医生的诊断提供很大帮助。

2 BMR-Net 网络分割算法

BMR-Net 网络的整体框架如图 1 所示,该框架分为编码器、解码器和连接在编解码器之间的 BPGM 模块。在编码器部分采用了 ResNeSt,输入图像($H \times W \times C$)通过 ResNeSt 进行 5 个阶段($f_i, i = 1, 2, \dots, 5$)的特征提取,得到分辨率为 $\left(\frac{H}{2^{i-1}}, \frac{W}{2^{i-1}}\right)$ 的不同级别特征;接下来的 BPGM 模块选择了三组高级特征($f_i, i = 3, 4, 5$)进行多尺度信息融合,改善特征提取效果,生成边界预测图 BPM,作为后续步骤的初始指导;在解码器部分采用一组反向注意力模块 RA,每个 RA 接受了 ResNeSt 的侧输出和上一层 RA 的输出,计算后得到反向注意力特征图,用来补充和界定边缘细节,得到的新的特征图用于对下一层的 RA 进行指导。经过 3 次 RA 操作后,末端的输出经过 Sigmoid 激活函数,得到最终的分割图。每个组成部分将在下面进行详细阐述。

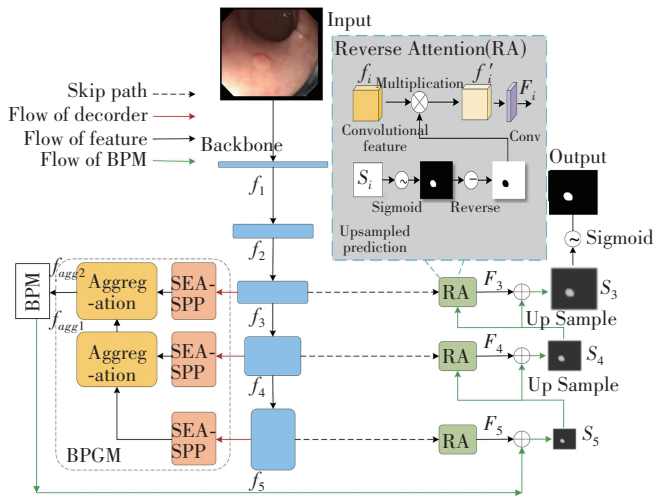


图 1 BMR-Net 网络整体框架

Fig. 1 Overview of BMR-Net network

2.1 ResNeSt

ResNeSt 是基于 ResNet 改进的网络模型^[16]。它提出了一种新的分散注意力模块,按照 ResNet 网络的方式堆叠这些分散注意力块,使注意力能够跨越特征图组。ResNeSt 优于其他具有相似复杂度的网络,普遍地提高了图像处理各领域如分类、检测和分割的性能。

将输入特征图分为 K 个支路 (Cardinal 1, Cardinal 2, ..., Cardinal k) 在每个支路里再次进行划分,得到 R 个子分支 (Split 1, Split 2, ..., Split r)。因此,特征组总数为 $G=KR$ 。在每个支路里,对每个子分支进行 $1 \times 1 + 3 \times 3$ 卷积操作后送入分散注意力模块,由分散注意力模块为子分支分配好权重后,输出支路特征,将 K 个支路的输出进行拼接。拼接后的特征图经过 1×1 卷积调整通道数,与原图进行残差连接。ResNeSt 模块如图 2 所示。

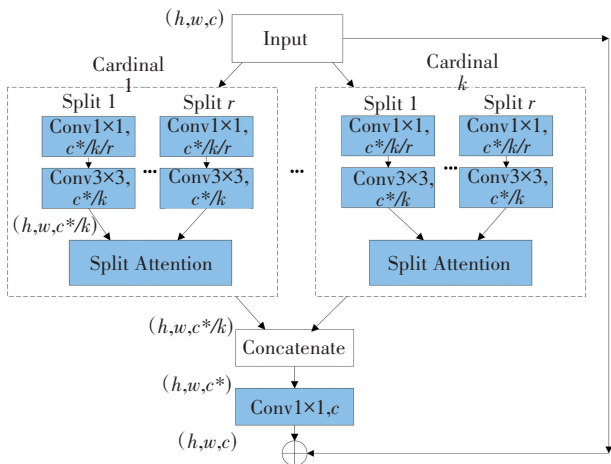


图 2 ResNeSt block 结构

Fig. 2 Structure of ResNeSt block

分散注意力模块起到为子分支特征图分配注意力权重的作用,其结构如图 3 所示。将输入的每组特征先 Sum 操作后进行全局平均池化,将通道权重参数分

别与 R 个子分支进行点乘操作后再相加,最后得到某一个支路的输出特征。其中的通道权重参数是通过全局平均池化获得的,后续还有空间金字塔池化介绍。

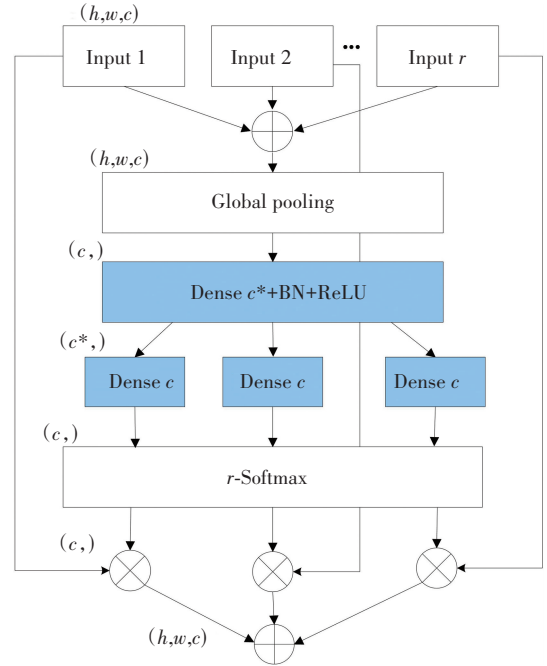


图 3 分散注意力模块

Fig. 3 Split attention module

2.2 边界预测生成模块 BPGM

研究发现,浅层特征分辨率高,占用计算资源,深层信息也能较好地表示浅层信息的空间细节^[18]。因此,决定抛弃浅层特征,对更深层的特征进行更多地计算。BPGM 融合了编码器后 3 个阶段的特征,逐步提高了特征分辨率。先将捕获的 3 组特征送入改良 ASPP 模块进行降维,降维后将 3 个级别的特征按照顺序进行聚合,使用一种轻量且高效的聚合方式^[19],先聚合第四层和第五层编码器特征 f_4 和 f_5 ,输出得到 f_{agg1} ;然后将 f_{agg1} 与第 3 层编码器特征 f_3 相融合,得到 f_{agg2} ,将其双线性插值上采样 8 倍后生成边界预测图 BPM。

聚合方式如图 4 所示。高分辨率特征通过平均池化下采样得到 f_{hl} ,低分辨率特征通过双线性插值上采样得到 f_{lh} ,将它们都进行 3×3 卷积、批归一化 (Batch Normalization, BN) 操作和线性整流函数 (Rectified Linear Unit, ReLU) 激活,得到的结果分别与卷积后的 f_l 和 f_h 相加,得到了特征 f'_l 和特征 f'_h 。将 f'_l 进行上采样并做 3×3 卷积,与 3×3 卷积后的 f'_h 相加,得到最终输出特征。

计算公式如下:

$$f'_h = \text{Conv}(f_h) + \text{Conv}(f_{lh}) \quad (1)$$

$$f'_l = \text{Conv}(f_l) + \text{Conv}(f_{hl}) \quad (2)$$

$$f_{\text{out}} = \text{Conv}(f'_h) + \text{Conv}(\text{Up}(f'_l)) \quad (3)$$

其中,Conv()表示卷积核为 3×3 的卷积操作,后面附加 BN 和 ReLU 操作。Up()表示双线性插值上采样。

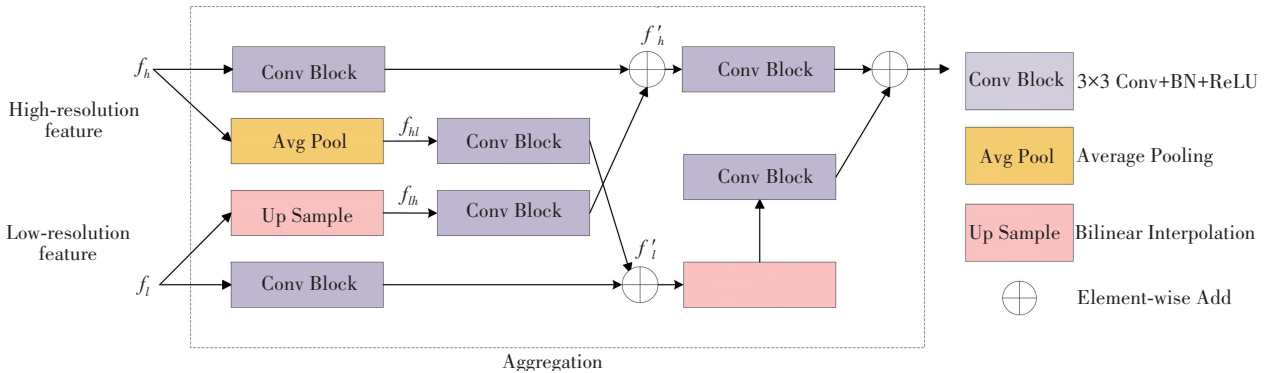


图 4 特征聚合模块

Fig. 4 Feature aggregation module

2.3 改进 ASPP 模块

ASPP(Atrous Spatial Pyramid Pooling),称为空间金字塔池化。它是一种改良版池化层,对输入的特征使用几个扩张率不同的卷积核进行采样,得到多尺度的特征信息,再将这些信息进行拼接,可以更全面地提取特征。

通道注意力机制(SE-NET)通过全局平均池化,将 $h \times w \times c$ 尺度的特征图压缩,只保留通道尺度上的大小 c ,即转换为 $1 \times 1 \times c$,再经过压缩、ReLU、恢复操作,最后使用 Sigmoid 函数进行激活,将各通道的值控制在 $0 \sim 1$ 的范围内,相当于将每个通道的特征转换为权重值。

将 SE 模块融入 ASPP 模块中去,得到改进后的 SE-ASPP 模块。将输入特征图用扩张率分别为 1, 6, 12, 18 的 3×3 卷积核进行空洞卷积,提取出 h' 和 w' 相同的 4 组特征,将它们在通道维度上进行拼接。拼接后得到的特征图与 SE 模块产生的权重值相乘,对特征逐通道加权后,完成对特征图的重标定。改进后的 SE-ASPP 模块如图 5 所示。

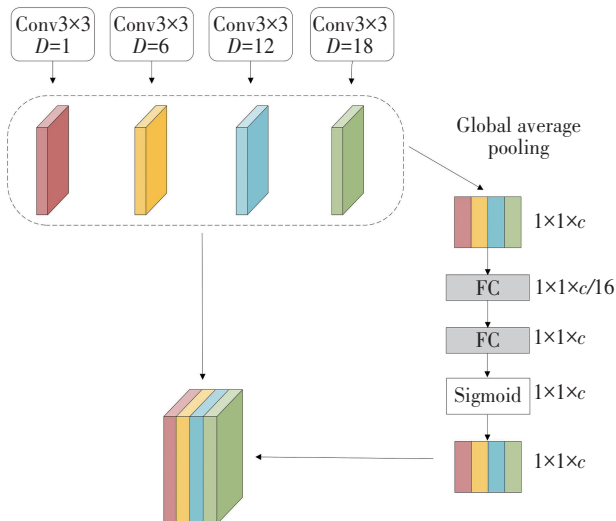


图 5 SE-ASPP 模块

Fig. 5 SE-ASPP module

2.4 反向注意力模块 RA

获得的 BPM 引导图来自最深的 CNN 层,只能捕捉到息肉所在的粗略位置,对于需要进行像素级预测的分割任务来说,缺失了结构、边缘细节。因此,在模型中引入反向注意力模块 RA,将其应用在解码器部分。RA 在高层侧输出特征中擦去已预测到的区域部分,引导网络将重点放在更难辨认的目标边缘和细节上,从粗到精逐步完善,将粗略的显著图一步步细化为精细的高分辨率显著图。RA 的结构如图 6 所示。

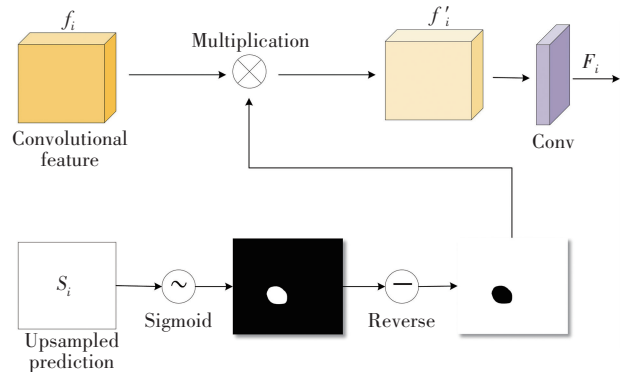


图 6 反向注意力模块

Fig. 6 Reverse attention module

RA 模块的输出特征 F_i 是高级侧输出特征 f_i 和反向注意力权重 A_i 相乘得到的,计算公式如式(4):

$$F_i = f_i \odot A_i \tag{4}$$

反向注意力权重的计算公式如式(5):

$$A_i = 1 - \text{Sigmoid}(\text{Up}(S_{i+1})) \tag{5}$$

前一个 RA 模块输出的特征 S_{i+1} 上采样后得到 S_i ,通过 Sigmoid 计算和取反计算,得到反向注意力权重 A_i 。

3 实验结果与分析

3.1 实验设置

本文实验基于 E5-2680V3 CPU, NVIDIA GeForce

RTX 3090 独立显卡,显存 24 G 的计算设备,配置环境版本为 Python3.8、Pytorch1.12.1、CUDA 版本为 11.6。输入图像的尺寸全部设置为 352×352 ,训练过程中使用的损失函数是 IoU 损失与二元交叉熵损失加权求和,对解码器的三阶段输出(即 S_3, S_4 和 S_5)和 BPM 采用深度监督,每个输出映射向上采样到与标签图 G 相同的大小。所以,总损失函数可以表示为 $L_{\text{total}} = L(G, S_{\text{BPM}}) + \sum_{i=3}^5 L(G, S_i^{\text{up}})$ 。使用 Adam 优化器,设置初始学习率为 $1e^{-4}$,批处理大小为 8,训练轮次为 100 轮。当训练损失在 50 轮内保持不变时,采取学习率衰减策略,衰减指数设置为 0.1。

3.2 数据集

实验使用了 5 个息肉分割数据集,分别为 ETIS-Larib, CVC - ClinicDB, CVC - ColonDB, EndoScene, 和 Kvasir-SEG,均为公开数据集。

ETIS-Larib 数据集包含 196 张结肠镜图像,其中分布了 36 种不同类型息肉,分辨率为 $1\,225 \times 966$,息肉大多很小,很难找到,因此分割难度较大。

CVC-ClinicDB 也称为 CVC-612,包含从 25 个结肠镜视频中提取的 612 张图像,分辨率为 384×288 。

CVC-ColonDB 是一个小型数据库,有 380 幅图像,来源于 15 个短结肠镜视频。

Endoscene 是 CVC-612 和 CVC300 的组合,为了避免与 CVC-ClinicDB 的数据重合,这里只使用了 CVC300 部分的 60 张图像。

Kvasir 是近年发布出的一个大数据集,范围很广,Kvasir-SEG 是在 Kvasir 的息肉类里筛选出的 1 000 幅

图像,尺寸跨度从 332×487 到 $1\,920 \times 1\,072$ 。

其中训练集使用来自 CVC-ClinicDB 的 612 张图像和 Kvasir-SEG 的 1000 张图像,共 1 612 张图片,将它们按照 8 : 1 : 1 随机划分,得到 1 450 张图像用于训练和验证。测试集包含了五个数据集,包括 CVC-ClinicDB 剩下的 62 张图像和 Kvasir-SEG 剩下的 100 张图像,以及 CVC-ColonDB、ETIS-Larib、EndoScene 的全部图像。后 3 个数据集被用来测试模型的泛化能力。

3.3 评价指标

实验中使用了 6 个评价指标,除了经常用在图像分割中作为评价标准的 mDice 系数 (m_{Dice}) 和平均交并比 mIoU (m_{IoU}),还加入了 4 个在目标检测领域中经常使用的度量指标^[20-22]:加权 Dice 度量 F_{β}^{ω} ^[23],它给位置不同的误差值赋予不同的权重,以此修正 Dice 中的“同等重要性缺陷”;MAE 度量 (Y_{MAE})^[24]将预测值和标签值之间的绝对误差取平均,来评估像素级精度; E_{ϕ}^{max} 是增强对准度量^[25],将局部像素值和图像级均值结合在一起,评测像素级和全局级的相似性;为了不忽略结构上的相似度,还加入结构相似性度量 S_{α} 作为评价指标^[26]。

3.4 实验结果对比与分析

3.4.1 分割效果

为了验证 BMR-Net 的分割效果和性能,将 BMR-Net 分别与 UNet、UNet++、ResUNet-mod^[27]、ResUNet++、SFA^[28]、PraNet 几个具有代表性的模型在相同的数据集上进行实验。此处使用了两个数据集:CVC - ClinicDB 和 Kvasir-SEG。本文模型与其他先进方法实验后得到的分割结果评价如表 1 所示。

表 1 在 Kvasir-SEG 和 CVC-ClinicDB 数据集上的量化结果

Table 1 Quantification of different methods on Kvasir-SEG and CVC-ClinicDB datasets

Methods	m_{Dice}	m_{IoU}	F_{β}^{ω}	S_{α}	E_{ϕ}^{max}	Y_{MAE}
UNet	0.818	0.746	0.794	0.858	0.893	0.055
UNet++	0.821	0.743	0.808	0.862	0.910	0.048
ResUNet-mod	0.791	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
Kvasir-SEG	ResUNet++	0.813	0.793	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
	SFA	0.723	0.611	0.670	0.782	0.849
	PraNet	0.898	0.840	0.885	0.915	0.948
BMR-Net	0.903	0.847	0.896	0.919	0.956	0.029

续表(表1)

	Methods	m_{Dice}	m_{IoU}	F_{β}^{ω}	S_{α}	E_{ϕ}^{\max}	Y_{MAE}
	UNet	0.823	0.755	0.811	0.889	0.954	0.019
	UNet++	0.794	0.729	0.785	0.873	0.931	0.022
	ResUNet-mod	0.779	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
CVC-ClinicDB	ResUNet++	0.796	0.796	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>	<i>n/a</i>
	SFA	0.700	0.607	0.647	0.793	0.885	0.042
	PraNet	0.899	0.849	0.896	0.936	0.979	0.009
	BMR-Net	0.930	0.885	0.929	0.954	0.980	0.007

从表 1 中可以看出,本文提出的模型 BMR-Net 在两个数据集上都实现了良好的分割效果,并且优于其他的先进模型。在 Kvasir-SEG 数据集上的 mDice 值 (m_{Dice}) 达到 0.903, mIoU 值 (m_{IoU}) 达到 0.847, F_{β}^{ω} 、 S_{α} 、 E_{ϕ}^{\max} 、MAE 值 (Y_{MAE}) 比其余模型的精度都要高,总体达到了最优的评价。在 CVC-ClinicDB 数据集上的分割精度也达到了各模型中最高, mDice 值 (m_{Dice}) 达到 0.930, mIoU 值 (m_{IoU}) 达到 0.885, F_{β}^{ω} 达到 0.929。比起精度第二高的 PraNet, mDice 值 (m_{Dice}) 提升了 3.1 个百

分点, mIoU 值 (m_{IoU}) 提升了 3.6 个百分点, F_{β}^{ω} 提升了 3.3 个百分点, 其余的评价指标也得到了显著提升。这表明 BMR-Net 在提升息肉图像的分割精度上是有效的。

图 7 展示了一组可视化分割结果, 是 BMR-Net 与其他模型在五个数据集上得到的效果图, 可以看出, BMR-Net 分割出的结果是最接近真实标签图的, 尤其在针对小目标和边界不清的情况时, BMR-Net 体现出了明显的优越性。

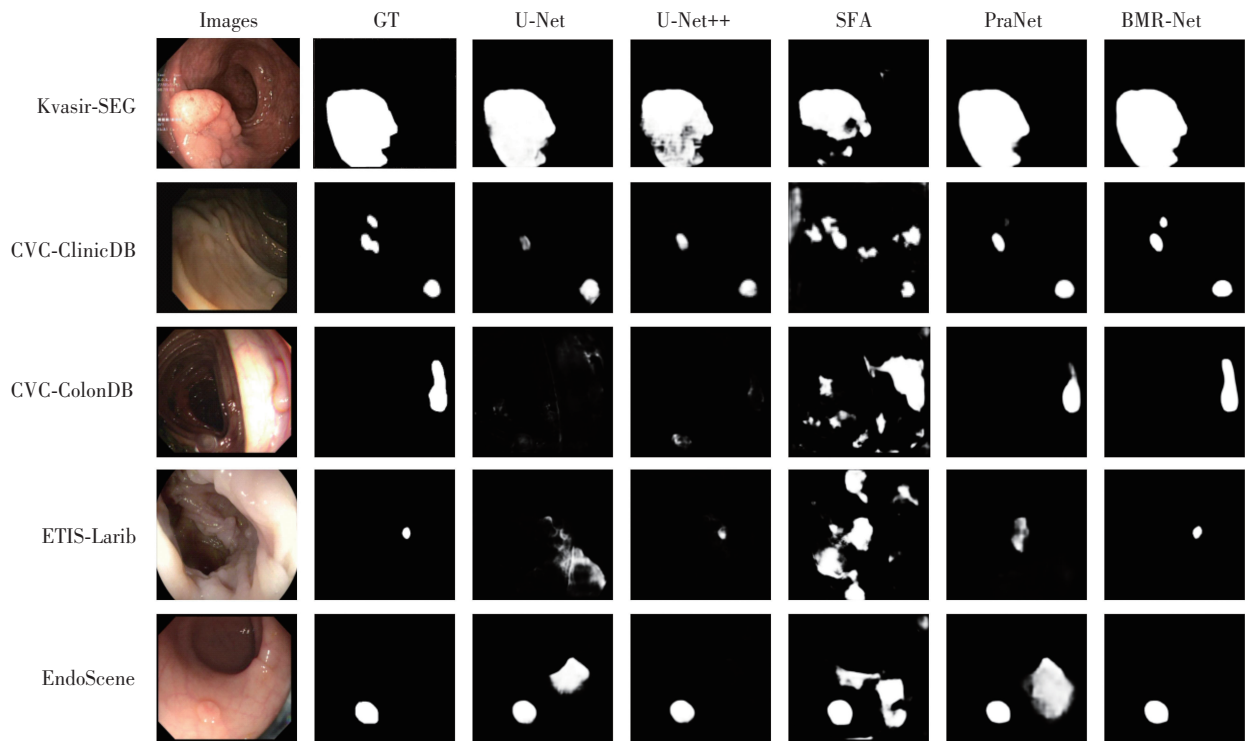


图 7 在五个数据集上分割结果对比图

Fig. 7 Comparison of segmentation results on five datasets

3.4.2 泛化性能

训练集中未被使用的 3 个数据集: CVC-ColonDB、ETIS-Larib、EndoScene 在最后测试时使用, 用来检验 BMR

-Net 在不同数据集上的通用性。实验结果如表 2 所示。

从表 2 中可以看到, BMR-Net 在 3 个数据集上都取得了显著的进步, 除了个别情况下 E_{ϕ}^{\max} 不是最优结果

之外,其余指标都优于其他模型。可以发现,UNet、UNet++、SFA 在更换了数据集后,各自的指标参数都有

所下降,其中 SFA 的数值动荡程度更为剧烈,这说明了它们的泛化性能较弱。

表 2 泛化性能测试结果

Table 2 Test results of generalization capability

	Methods	m_{Dice}	m_{IoU}	F_{β}^{ω}	S_{α}	E_{ϕ}^{\max}	Y_{MAE}
CVC-ColonDB	UNet	0.512	0.444	0.498	0.712	0.776	0.061
	UNet++	0.483	0.410	0.467	0.691	0.760	0.064
	SFA	0.469	0.347	0.379	0.634	0.765	0.094
	PraNet	0.709	0.640	0.696	0.819	0.869	0.045
	BMR-Net	0.756	0.682	0.743	0.843	0.865	0.035
ETIS-Larib	UNet	0.398	0.335	0.366	0.684	0.740	0.036
	UNet++	0.401	0.344	0.390	0.683	0.776	0.035
	SFA	0.297	0.217	0.231	0.557	0.633	0.109
	PraNet	0.628	0.567	0.600	0.794	0.841	0.031
	BMR-Net	0.712	0.631	0.673	0.835	0.853	0.021
EndoScene	UNet	0.710	0.627	0.684	0.843	0.876	0.022
	UNet++	0.707	0.624	0.687	0.839	0.898	0.018
	SFA	0.467	0.329	0.341	0.640	0.817	0.065
	PraNet	0.871	0.797	0.843	0.925	0.972	0.010
	BMR-Net	0.874	0.802	0.850	0.926	0.950	0.009

BMR-Net 除了在 3 个数据集上都展示出了优越的性能之外,值得注意的是,它在最具挑战性的小息肉数据集 ETIS-Larib 上,评价数值提升很大, mDice 值 (m_{Dice}) 达到了 0.712, mIoU 值 (m_{IoU}) 达到了 0.631, 达到了 0.673, 相对于精度第二的 PraNet 分别提升了 8.4 个百分点、6.4 个百分点和 7.3 个百分点。其余的 F_{β}^{ω} 、 S_{α} 、 E_{ϕ}^{\max} 、MAE 值 (Y_{MAE}) 各项指标也与其他模型拉开了巨大的差距。这说明 BMR-Net 不仅泛化能力较强,在分割小目标息肉时比起其他模型也更具有优势。

3.4.3 消融实验结果分析

为了验证 ResNeSt、BPGM 模块和 SE-ASPP 模块对实验结果的影响,设定一个初始基线模型:使用 ResNet 作为编码器,RA 模块作为解码器,在编码器和解码器之间只加入 ASPP 模块用以提取特征。逐步在基线模型上加入部分模块,构建出 4 个不同的改进模型,用以测试单独模块对模型的影响效果。如表 3 所示。

基于表 3 的模型设置,采用相同的参数对这四个模型进行训练,得到一系列消融实验在 CVC-ColonDB 和 ETIS-Larib 数据集上的结果。如表 4 所示。

表 3 消融实验列表

Table 3 List of ablation experiments

序号	实验描述
1	ResNet+ASPP
2	ResNeSt+ASPP
3	ResNeSt+BPGM+ASPP
4	ResNeSt+BPGM+SE-ASPP

表 4 各模型在两个数据集上的实验结果对比

Table 4 Comparison of experimental results of each model on two datasets

序号	CVC-ColonDB		ETIS-Larib	
	m_{Dice}	m_{IoU}	m_{Dice}	m_{IoU}
1	0.667	0.603	0.562	0.506
2	0.742	0.666	0.673	0.601
3	0.744	0.673	0.674	0.609
4	0.756	0.682	0.712	0.631

对比消融实验结果可知,每个模块均对网络的性能起到了改善,实验 2 将主干网络换成 ResNeSt 后,

Dice 指标在 CVC-ColonDB 和 ETIS-Larib 数据集上分别提升了 7.5%, 11.1%, 实验 3 在编码器和解码器之间加入 BPGM 模块, Dice 指标分别上升了 0.2%, 0.1%, 实验 4 将 ASPP 模块换成 SE-ASPP 模块, Dice 指标分别上升了 1.2%, 3.8%。经过分析各模块的优势如下:

(1) ResNeSt 网络的分散注意力机制将特征分成不同的组, 在每组中捕捉到不同尺度的信息, 将这些特征整合后, 进一步地学习到图像不同大小的结构和模式, 提高了模型学习复杂任务的能力。

(2) BPGM 模块将高层次特征聚合起来, 涵盖了丰富的语义信息, 在保持模型性能的同时, 减少了浅层特征的干扰并排除了冗余的计算量, 增强了对不同质量图像的鲁棒性。

(3) ASPP 通过多尺度卷积核得到丰富的上下文信息, SE-ASPP 模块通过对不同信息的鉴别和重标定, 进一步增强了这种上下文捕获能力, 使模型在特征融合时, 不仅考虑到了多尺度信息, 还考虑到了特征的重要性, 从而得到更加鉴别性的融合特征。

4 结 论

本文提出了一种新型的结直肠息肉分割网络 BMR-Net, 用来改善息肉图像中颜色、形状、纹理差异较大时, 难以区分息肉与背景, 导致分割结果不清晰的问题。该网络中, ResNeSt 作为主干特征提取网络, 其独有的分散注意力模块构造改善了学习到的特征表示; 边界预测生成模块 BPGM 通过轻量且高效的聚合方式融合高层特征, 得到的边界预测图 BPM 充分结合了多尺度信息, 改善了不同大小息肉的分割, 尤其提高了小息肉的分割准确性; 在聚合高层特征时, 加入改良得到的 SE-ASPP 模块, 在提高感受野的同时引导网络将注意力放在更重要的通道上, 提高了特征捕捉能力, 改善了模型分割性能; 解码器部分使用的 RA 模块可以对边缘信息进行更深入细致的挖掘, 进一步提升分割精度。不同模块在整体网络上的迭代配合, 共同辅助网络达成了更加优良的分割效果。

BMR-Net 在 5 个具有挑战性的数据集上, 在 6 个评价指标上, 总体表现都优于对比的方法, 其中在 Kvasir-SEG 的 mDice 值、mIoU 值达到了 0.903、0.847, 在 CVC-ClinicDB 的 mDice 值、mIoU 值达到了 0.930、0.885。而且 BMR-Net 不需要经过任何预处理, 计算复杂度很低, 还具有很强的泛化性能, 综合来看, BMR-

Net 具有突出的优势。在今后的研究工作中, 可以尝试引入更多有效的模块, 进一步提升分割精度; 也可以针对模型轻量化进行改进; 除了息肉分割, 还可以探索 BMR-Net 在其他医学图像分割中的效果, 如肺部感染分割、脑肿瘤分割等。

参考文献(References):

- [1] SHUSSMAN N, WEXNER S D. Colorectal polyps and polyposis syndromes[J]. *Gastroenterology Report*, 2014, 2(1): 1-15.
- [2] MARMOT M, ATINMO T, BYERS T, et al. World cancer research fund, American institute for cancer research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective[M]. Washington DC, US: American Institute for Cancer Research, 2007: 1-10.
- [3] 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*, 2021, 71(3): 209-249.
- [4] JIA X, XING X, YUAN Y, et al. Wireless capsule endoscopy: a new tool for cancer screening in the colon with deep-learning-based polyp recognition[J]. *Proceedings of the IEEE*, 2020, 108(1): 178-197.
- [5] PATEL K, BUR A M, WANG G. Enhanced U-net: a feature enhancement network for polyp segmentation[J]. *Proceedings of the International Robots & Vision Conference International Robots & Vision Conference*, 2021, 2021: 181-188.
- [6] 李胜, 曹婧, 叶淑芳, 等. 边缘先验信息下的多类型肠道息肉图像分类网络[J]. *中国生物医学工程学报*, 2022, 41(4): 431-442.
- [7] LI Sheng, CAO Jing, YE Shu-fang, et al. Multi-category intestinal polyp image classification network based on edge prior information[J]. *Chinese Journal of Biomedical Engineering*, 2022, 41(4): 431-442.
- [7] BRANDAO P, MAZOMENOS E, CIUTI G, et al. Fully convolutional neural networks for polyp segmentation in colonoscopy[C]//*Proceedings of the Medical Imaging 2017: Computer-Aided Diagnosis*. SPIE, 2017: 101-107.
- [8] AKBARI M, MOHREKESH M, NASR-ESFAHANI E, et al. Polyp segmentation in colonoscopy images using fully convolutional network[J]. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual*

- International Conference, 2018, 2018: 69–72.
- [9] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234–241.
- [10] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHS N, et al. UNet ++: A nested U-net architecture for medical image segmentation[C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3–11.
- [11] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Res UNet: An advanced architecture for medical image segmentation[C]//Proceedings of the IEEE International Symposium on Multimedia. Piscataway: IEEE Press, 2019: 225–2255.
- [12] FAN D P, JI G P, ZHOU T, et al. PraNet: Parallel reverse attention network for polyp segmentation[C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 263–273.
- [13] FENG R, LEI B, WANG W, et al. SSN: a stair-shape network for real-time polyp segmentation in colonoscopy images[C]//Proceedings of the IEEE 17th International Symposium on Biomedical Imaging. Piscataway: IEEE Press, 2020: 225–229.
- [14] WANG P, CHUNG A C S. DoubleU-net: Colorectal cancer diagnosis and gland instance segmentation with text-guided feature control[C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 338–354.
- [15] CHEN S, TAN X, WANG B, et al. Reverse attention for salient object detection[C]//FERRARI V, HEBERT M, SMINCHISESCU C, et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 236–252.
- [16] ZHANG H, WU C, ZHANG Z, et al. ResNeSt: split-attention networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2022: 2735–2745.
- [17] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848.
- [18] WU Z, SU L, HUANG Q. Cascaded partial decoder for fast and accurate salient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3902–3911.
- [19] QIU Z, WANG Z, ZHANG M, et al. BDG-Net: Boundary distribution guided network for accurate polyp segmentation[C]//Proceedings of the Medical Imaging 2022: Image Processing. SPIE, 2022: 792–799.
- [20] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented polyp dataset[C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 451–462.
- [21] 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: ACM, 2021: 2167–2175.
- [22] FAN D P, JI G P, SUN G, et al. Camouflaged object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 2774–2784.
- [23] MARGOLIN R, ZELNIK-MANOR L, TAL A. How to evaluate foreground maps[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 248–255.
- [24] PERAZZI F, KRÄHENBÜHL P, PRITCH Y, et al. Saliency filters: Contrast based filtering for salient region detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 733–740.
- [25] FAN D P, GONG C, CAO Y, et al. Enhanced-alignment measure for binary foreground map evaluation[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2018: 698–704.
- [26] CHENG M M, FAN D P. Structure-measure: A new way to evaluate foreground maps[J]. International Journal of Computer Vision, 2021, 129(9): 2622–2638.
- [27] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-net[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 749–753.
- [28] FANG Y, CHEN C, YUAN Y, et al. Selective feature aggregation network with area-boundary constraints for polyp segmentation[C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 302–310.