

基于 U-Net 改进的内窥镜息肉图像分割算法

邓晓青¹, 李 征^{1,2}, 王雁林¹

(1. 四川大学计算机学院, 成都 610065; 2. 四川大学天府工程数值模拟与软件创新中心, 成都 610207)

摘要: 息肉图像的分割在临床医疗和计算机辅助诊断技术等领域具有广泛的研究和应用价值,但是就目前的研究和应用需求来看,准确的息肉分割仍然是一项挑战. 针对内窥镜息肉图像中出现的息肉与黏膜边界不清晰、息肉的大小和形状差异较大等影响分割质量的问题,该文提出了一种基于 U-Net 改进的息肉图像分割算法(SBF-Net). 首先,在 U-Net 架构上引入了边界特征加强模块(BFEM),考虑到息肉边界和内部区域的关键线索,该模块利用编码器高层特征生成额外的边界补充信息,在解码器阶段进行融合,提升模型处理边界特征的能力. 其次,该模型的解码器(GFBD)采用了从上至下逐步融合特征的方式,将编码器阶段的输出特征经过局部加强(LE)模块之后再逐步融合边界特征,这种多尺度特征融合方式有效缓解了编码器和解码器之间的语义差距问题. 最后,在后处理阶段采用测试时数据增强(TTA)来进一步对分割结果进行细化. 该模型在 CVC-300、CVC-ClinicDB、Kvasir-SEG、CVC-ColonDB 和 ETIS-LaribPolypDB 等 5 个公开数据集上进行了对比实验和消融实验,实验结果证明了该文所改进方法的有效性,并在内窥镜息肉图像上表现出更好的分割性能和更强的稳定性,为息肉图像的处理和分析提供了新的参考.

关键词: 内窥镜息肉图像; 息肉分割; U-Net; 边界加强

中图分类号: TP391.41

文献标志码: A

DOI: 10.19907/j.0490-6756.2024.013004

An improved endoscopic polyp image segmentation algorithm based on U-Net

DENG Xiao-Qing¹, LI Zheng^{1,2}, WANG Yan-Lin¹

(1. College of Computer Science, Sichuan University, Chengdu 610065, China;

2. Tianfu Engineering-Oriented Numerical Simulation & Software Innovation Center, Sichuan University, Chengdu 610207, China)

Abstract: The segmentation of polyp images has extensive research and application value in the fields of clinical treatment and computer-aided diagnostic technology, but accurate polyp segmentation is still a challenge in terms of current research and application needs. In order to solve the problems that affect the segmentation quality of endoscopic polyp images, such as the unclear boundary between polyps and mucous membranes, and the large difference in the size and shape of polyps, this paper proposed an improved U-Net polyp segmentation algorithm. Firstly, the boundary feature enhancement module was introduced on the U-Net architecture. Considering the key clues of polyp boundary and internal area, this module used the high-level features of the encoder to generate additional boundary supplementary information, which is fused at the decoder stage to improve the ability of the model to process boundary features. Secondly, the decoder of the model adopts the method of gradually fusing features from the top to

收稿日期: 2023-01-05

基金项目: 国家重点研发计划项目(2020YFA0714003); 国家重大项目(GJXM92579); 四川省科技厅重点研发项目(2021YFQ0059)

作者简介: 邓晓青(1999-), 女, 四川资阳人, 硕士研究生, 研究方向为计算机视觉. E-mail: dengxq0123@163.com

通讯作者: 李征. E-mail: lizheng@scu.edu.cn

the bottom. After the output features of the encoder stage are passed through local emphasis module, the boundary features are gradually fused. This multi-scale feature fusion method effectively reduces the semantic gap between the encoder and the decoder. Finally, test-time augmentation was used in the post-processing stage to further refine the segmentation results. The model has been compared and ablated on five public datasets: CVC-300, CVC-ClinicDB, Kvasir-SEG, CVC-ColonDB and ETIS-LibPolyp-DB. The experimental results prove the effectiveness of the modified method, and it shows better segmentation performance and stronger stability in the endoscopic polyp image, which provides a new reference for the processing and analysis of the polyp image.

Keywords: Endoscopic polyp image; Polyp segmentation; U-Net; Boundary strengthening

1 引言

结直肠癌是世界上第三大最常见的癌症类型^[1],大多数结直肠癌通常是由于结直肠息肉未及时发现和治疗而转变成为恶性肿瘤所引起的^[2].研究表明,在息肉转为恶性肿瘤之前,可以通过结肠镜检查和治疗使结直肠癌的发病率和死亡率得到有效降低.结直肠息肉检查中,内窥镜检查是最重要也是最常用的一种方式,内窥镜检查是利用结肠镜深入到患者的结直肠内部,通过内窥镜拍摄大量的图像或者实时视频的方式获取病灶的信息,这种方式不仅可以确定结直肠息肉的位置,还能明确息肉的大小、体积和质地等信息,为诊断治疗提供参考,使医生能够在其发展成结直肠癌之前将其切除^[3],迄今为止,结肠镜检查仍然是最有效的结直肠癌预防技术.但是,此过程高度依赖于临床医生的技能和经验,检查效果非常不稳定,这也是息肉检测错误率和漏检率高的原因之一,漏检率甚至高达 25%^[4].用于息肉分割的计算机辅助诊断系统可以通过提高准确性和减少人工干预来减少医生的工作量并改善临床工作流程^[5],因此,能够在早期检测到所有可能的息肉并进行自动和准确的分割,这对于预防结直肠癌具有重要意义^[6].

早期的息肉图像分割方法大多依赖于阈值、形状和边缘检测等传统方法,虽然实现简单,但是精度不高^[7].图像分割是医学图像处理的重要研究方向之一,随着深度学习技术的迅速发展,各种深度神经网络模型被广泛应用于医学图像分割领域^[8],并取得了显著的成果. Ronneberger 等人^[9]提出的 U-Net 是一个用于医学图像分割的全卷积网络,因其能够采用多级特征来重建图像的预测结果而引起了医学图像分割领域的广泛关注,该网络使用跳跃连接将编码器的每一层特征都拼接至对应的解码器端,以实现浅层特征中的空间信息的

最大利用. Zhou 等人^[10]提出的 UNet++ 是 U-Net 的改进版本,与 U-Net 不同的是,UNet++ 在编码器和解码器之间使用了密集的跳跃连接,通过这种连接间接地融合了多个层次的特征,这对于医学图像中各种形状和大小不同的病灶分割效果好. Jha 等人^[11]提出的 ResUnet++ 是在基于 U-Net 结构的基础上增加了 SE(Squeeze and Excitation)模块、ASPP(Atrous Spatial Pyramidal Pooling)模块、残差模块和注意力块来增强模型的特征提取能力,实现更快更准确的息肉分割. Wei 等人^[12]提出了浅层注意力网络 SANet,使用颜色交换的方式来消除颜色对分割结果的影响,以及一种概率校正策略来减少负样本的干扰并提高正样本的预测能力. 这些方法专注于分割息肉的整体区域,忽略了边界和内部区域的关系,而这对于准确的分割非常重要. 因此, Fan 等人^[13]提出了 PraNet,该网络采用了两阶段分割方法,使用反向注意模块来恢复息肉的边界和内部结构从而实现更加准确的息肉分割. 受此启发, Kim 等人^[14]提出的 UACANet 整合了前景、背景以及不确定区域的特征来细化边界的线索. 然而这些方法提出的息肉边界和内部结构信息都未在特定的监督下生成,因而隐藏层的特征透明度和辨别力不高. BDG-Net^[15]模型添加了边界分布图来指导解码器分割,但是未考虑来自编码器的浅层特征与解码器的深层特征存在的信息差距问题,并且未充分利用编码器的浅层信息.

基于强化息肉边界的关键线索以及减少特征层之间信息差距的思想,本文提出了一种基于 U-Net 改进的内窥镜息肉图像分割网络,称为逐步融合边界特征网络(Stepwise Boundary Fusion Net, SBF-Net). 其整体上基于 U-Net 框架,在编码器阶段利用高层特征生成息肉边界补充信息,提升模型对边界处特征的学习能力,辅助解码器恢复出更准

确的分割结果. 在解码器阶段, 将编码器的输出特征经过局部加强(Local Emphasis, LE)^[16] 模块后与当前阶段的上一层输出特征和息肉边界特征进行逐步融合, 最后将 TTA^[17] (Test-Time Augmentation) 应用于模型的后处理阶段以实现分割结果的进一步细化.

2 SBF-Net 模型结构

2.1 SBF-Net 模型

SBF-Net 的整体架构如图 1 所示, 该网络主体上是 U 型网络结构, 中间部分是编码器, 右侧是逐步融合边界特征的解码器 (Gradually Fused Boundary Feature Decoder, GFBD), 跳跃连接部分加入了局部加强 (LE) 模块, 编码器左侧部分是

边界特征加强模块 (Boundary Feature Enhancement Module, BFEM). 由于内窥镜息肉图像中息肉的颜色与周围黏膜非常相似, 因此边界的区分非常不明显^[18]. 所以为了让模型更好地学习边界特征, 本文参考了 BDG-Net^[15] 中以高斯分布近似真实的息肉边界分布. 但是为了充分利用浅层特征的有效信息, 本文将编码器最后四个阶段的输出用于生成额外的边界补充信息. 同时, 考虑到 U 型网络中可能产生的语义差距问题, 本文的解码器使用局部加强的跳跃连接来融合编码器特征、息肉边界特征和当前阶段的上一层输出特征, 使高维特征和低维特征之间融合的信息差距更小, 同时还能更好地减少了息肉边界处的信息损失, 恢复出高分辨率的预测结果.

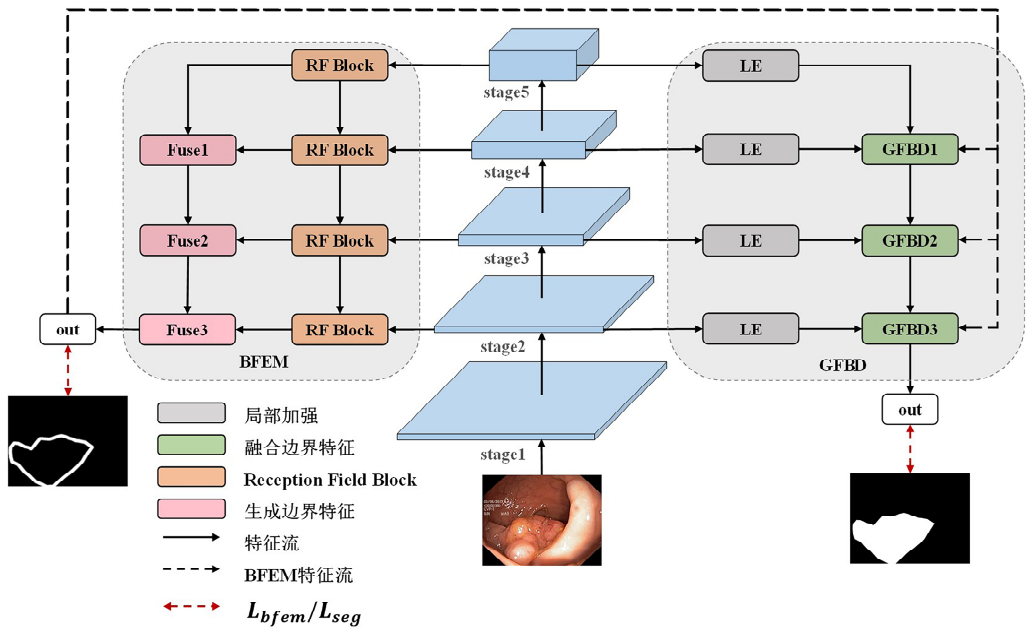


图 1 SBF-Net 结构

Fig. 1 The structure of SBF-Net

2.2 边界特征加强模块

医学图像与普通图像的差异较大, 例如在内窥镜息肉图像中, 部分病灶边缘模糊不清, 边缘与周围黏膜组织的过渡区域附近存在细小的纹理及噪声, 在息肉病灶局部区域, 息肉与周围邻近像素点间存在较高的颜色相似性和亮度, 但整体区域的一致性却各有不同. 目前许多优秀的息肉分割算法在息肉内部区域的分割效果良好, 但是息肉边界上往往无法得到很好的处理^[13]. 为了更好地让模型学习边界处的特征, 本文提出了一种基于残差结构和通道注意力机制的边界特征加强模块, 即图 1 中 BFEM 部分, 为充分利用编码器浅层的细节信息, 该模块由编码器的最后 4 个阶段生成.

BDG-Net 中添加了边界分布生成模块来生成息肉边界分布图指导解码器实现更准确的分割, 但是该模块在保持模型轻量化的情况下只融合了编码器的高三层特征, 缺少局部细节信息, 恢复出的息肉边界精度有限. 本文将编码器四个阶段的输出进行融合输出息肉边界, 在提升精度的同时分割的息肉结果拥有更丰富的边缘细节. 考虑到浅层特征可能包含更多噪声, BDG-Net 中简单的卷积融合并不能很好地抑制息肉黏膜周围的干扰信息, 因此本文引入了通道注意力机制来融合多尺度特征, 以尽可能抑制边界处的无用噪声, 增强有用特征. 本文重新设计了边界生成的方式, BFEM 的设计旨在

加强该模块的学习能力和特征复用能力以生成包含更多细节信息的息肉边缘,总体上是基于 Res-Net^[19]中提出的残差机制以及通道注意力机制^[20]的思想. 首先,使用 RF Block^[21]对来自编码器的特征进行处理,RF Block 是在 Inception^[22]的基础上引入了不同膨胀系数的卷积层,可以在减少特征图的通道数的同时,增大模型的感受野,辅助 BFEM 提取出包含更多全局信息的特征. 其次,为了使得深层网络中包含更多的特征信息,并缓解其可能存在的退化问题,我们添加了如图 2 所示的残差连接,使模型在原有特征的基础上学习出新的特征,从而提升特征复用能力. 最后,由于内窥镜息肉图像中在息肉病灶以外的区域也可能表现出诸如毛细血管、褶皱等复杂的结构特征,这些信息会对息肉边界的分割造成干扰. 因此我们在 BFEM 融合过程中引入了通道注意力机制的挤压激励(Squeeze and Excite, SE)模块,以达到强化有效特征,抑制边界处无效噪声的目的. BFEM 以此机制作为一个学习单元,可以使模型恢复出更加准确的息肉边缘,BFEM 的具体融合过程如图 2 所示.

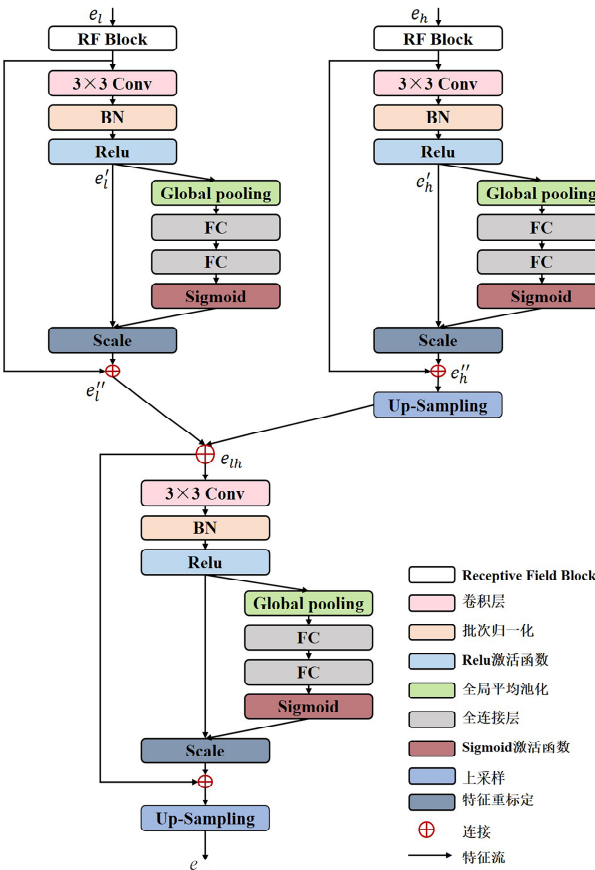


图 2 生成边界特征的融合过程

Fig. 2 The fusion process for generating boundary features

首先,我们将编码器最后两个阶段的输出特征进行融合,低层特征表示为 e_l ,高层特征表示为 e_h , e_l 经过 RF Block 后进行 C_3 BR 模块处理得到 e_l' , C_3 BR 包括 3×3 卷积层(Convolution)、批次归一化^[23](Batch Normalization)和 Relu^[24] 激活函数层, e_l' 再传入 SE 模块对不同通道的特征进行加权,在恒等映射和残差相加之前对残差分支进行了“特征重标定”. 然后与经过 RF Block 后的 e_l 进行残差连接得到 e_l'' . 类似地, e_h 进行同样的处理得到 e_h'' ,接着将 e_h'' 进行双线性插值上采样后与 e_l'' 相加得到融合了高层和低层特征的 e_{th} ,接着 e_{th} 经过 C_3 BR 处理,再传入 SE 模块后与输入特征 e_{th} 进行残差连接,最后再作上采样处理得到最终的输出 e_{out} . 输出特征 e_{out} 与上一阶段编码器特征继续同样的融合过程,直到融合至编码器的第二个阶段为止. 最终的输出 e_{out} 可以描述为式(1)所示, e_{th} 表达式如式(2)所示, e_h'' 和 e_l'' 表达式分别如式(3)和式(4)所示, e_h' 和 e_l' 的生成表达式分别如式(5)和式(6)所示. 其中, $RFB(\cdot)$ 表示 RF Block, $SE(\cdot)$ 表示挤压激励操作, $Up(\cdot)$ 表示双线性插值上采样.

$$e_{out} = Up(SE(C_3BR(e_{th}))) + e_{th} \quad (1)$$

$$e_{th} = Up(e_h'') + e_l'' \quad (2)$$

$$e_h'' = SE(e_h') + RFB(e_h) \quad (3)$$

$$e_l'' = SE(e_l') + RFB(e_l) \quad (4)$$

$$e_h' = C_3BR(RFB(e_h)) \quad (5)$$

$$e_l' = C_3BR(RFB(e_l)) \quad (6)$$

由于内窥镜息肉图像存在的反光、粗边缘等原因,息肉与黏膜界限很难确定. 为了更加精确地定位息肉边界附近的像素是否属于息肉,使模型训练出更加真实的边界分布,本文使用高斯分布来近似息肉边界像素的分布,以此作为息肉与黏膜的理想界限. 高斯分布是一种随机分布,用于描述测量值在真值附近的分布,距离真值越远,观测实例就越稀少. 所以本文参考了 BDG-Net 的真实边界分布图的生成方式,假设息肉的边界是遵循均值为 0, 标准差为 σ 的高斯分布^[15],息肉边界分布的真实值可以近似为一元高斯分布,表达式如下式所示.

$$Gs(b_{ij}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\epsilon(b_{ij})^2}{2\sigma^2}} \quad (7)$$

式中, $\epsilon(b_{ij})$ 表示 (i, j) 像素点到边界的最短欧氏距离. 首先将息肉分割标签 $mask$ 先进行预处理转换为前景为 1,背景为 0 的二值化图像,然后分别计算出前景点和背景点到息肉边界像素点的距离,得到距离转换图,并整合前景和背景到边界的距离,

我们将式(7)作用于距离转换图像即可估计出边界附近的像素点属于边界的概率映射, 最终的息肉边界分布可以表示为如下式所示.

$$BFEM(mask, \sigma) = \lambda \cdot Gs(trans(mask) + trans(1 - mask), \sigma) \quad (8)$$

式中, $mask$ 表示息肉分割的标签图像; λ 表示增强因子, 用于增强特征响应, 本文实验中将 λ 设为 σ^2 ; $trans$ 表示距离转换操作.

2.3 逐步融合边界特征解码器

由于 U-Net 在医学图像分割领域的优秀性能, 现已广泛用于各种病灶分割. 虽然 U-Net 因其能够融合多尺度特征, 充分利用图像的空间语义信息, 在小型医学数据集上也能表现良好. 但是类似于 U-Net 这样的编码器-解码器结构的一个主要问题就是来自编码器的浅层特征和来自解码器的深层特征存在较大的语义差距. 这种来自编码器和解码器之间的语义鸿沟, 限制了分割性能, 甚至有一些跳跃连接会对分割产生负面效应.

基于减少特征层语义差距的思想, 本文提出的解码器采用了 U 型网络中从上至下的方式逐步恢复高分辨率的息肉分割结果, 在逐步融合边界特征的过程中尽可能使特征融合平滑进行, 如图 1 中 GFBD 部分. 首先, 我们通过在编码器输出的特征上引入了局部加强的非线性变换来加强跳跃连接, 以缩小编码器和解码器之间的语义差距, 利用卷积的固定感受野来混合每个像素点相邻的图像块, 从而增加相邻图像块到图像中心的相关权重, 以强调每个图像块关键的局部信息. 其次, 整个解码器端采用 U-Net 中从上到下的方式逐步融合特征金字塔中不同层次的特征, 避免直接聚合具有显著深度差异的特征影响分割性能. 最后, 将 BFEM 生成的息肉边界特征在解码器端辅助 GFBD 生成更加准确的分割结果. 这种特征融合方式旨在保持轻量 and 高效, 并且可以使融合的高维和低维特征之间的信息差距更小.

以图 1 中 GFBD2 为例, 解码器融合息肉边界特征过程如图 3 所示. GFBD2 融合过程的输入包括上一阶段解码器的输出 d_{i-1} 、BFEM 输出的息肉边界 b 和经过 LE 的编码器输出特征 e'_i , 输出是 d_i . GFBD1 和 GFBD3 的结构与此类似, 由于 GFBD1 为解码器端首个融合单元, 该部分输入包括 e'_{i-1} , e'_i 和 b , 在 GFBD3 中, 在输出 d_i 后还经过了一个由卷积层和两倍双线性插值上采样组成的分

割头来预测最终的分割结果.

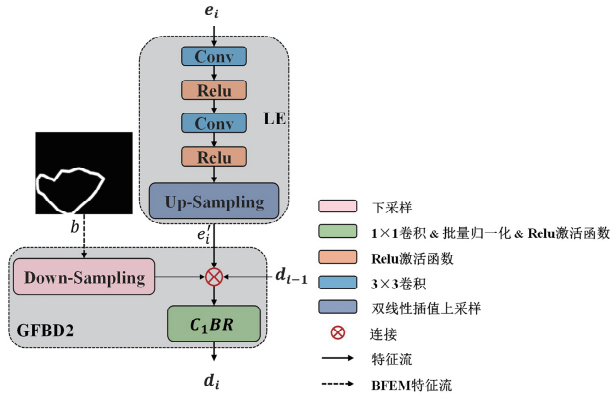


图 3 GFBD2 融合边界特征过程

Fig. 3 The process of GFBD2 fusion of boundary features

LE 模块是两个 3×3 卷积和 ReLU 激活函数, 再加上一个双线性插值上采样组成, 编码器的输出 e_i 经过 LE 后得到 e'_i . 接着 e'_i 再与 d_{i-1} 和 b 进行连接, 最后再经过 C_1BR 块得到输出特征 d_i , C_1BR 包括 1×1 卷积、批次归一化和 ReLU 激活函数. 解码器融合各个特征的操作输出 d_i . d_i 作为 GFBD3 的输入继续下一阶段融合. d_i 表达式可以描述为式(9)所示, $LE(\cdot)$ 表示局部加强, 表达式如式(10)所示, $Down(\cdot)$ 表示下采样, 本文中采用的是双线性插值法, $Conv(\cdot)$ 表示 3×3 卷积, $Concat(\cdot)$ 表示连接操作.

$$d_i = C_1BR(Concat(LE(e_i), d_{i-1}, Down(b))) \quad (9)$$

$$LE(e_i) = Up(ReLu(Conv(ReLu(Conv(e_i)))))) \quad (10)$$

2.4 测试时数据增强

测试时数据增强(TTA)广泛应用于各类视觉任务上提升精度, 但同时会带来测试时间的增加, 在非实时的视觉任务中应用更多. TTA 是一种通过对测试数据集进行裁剪、翻转、旋转等增强手段^[17], 对每个测试图像进行增强, 并对增强后的图像进行预测, 将预测结果的均值作为最终的输出预测. 在本文中, 我们在测试时应用到了水平翻转和旋转的增强方法.

2.5 损失函数

本文的损失函数由边界特征加强模块的损失 L_{bfem} 与分割损失 L_{seg} 的和来作为最终的损失 L . L_{bfem} 采用 BDG-Net 的辅助损失^[15] 用于计算边界特征加强模块的中间结果与真实值之间的损失, 如式(11)所示, $f(b_{ij})$ 用于忽略损失较低的部分, 表

达式如式(12)所示.

$$L_{\text{bfem}} = \sum_{(i,j)} (b_{ij} - \hat{b}_{ij})^2 \cdot f(b_{ij}) \quad (11)$$

$$f(b_{ij}) = \delta[(b_{ij} - \hat{b}_{ij})^2 < \kappa] \quad (12)$$

其中, b_{ij} 表示息肉边界的预测值; \hat{b}_{ij} 表示理想的真实值. k 表示忽略损失区间的阈值, k 的初始值设置为 0.002, 随着训练次数的增加, 损失越来越低, k 的值也会逐渐降低. 当预测点与真实值之间的欧式距离的平方小于 k 时, $\delta[\cdot]$ 的值为 0, 即忽略掉这部分损失, 当大于 k 时, $\delta[\cdot]$ 的值为 1. 最终的损失函数 L 如式(13)所示. L_{seg} 用于计算最终分割结果和真实值之间的损失, 由加权二元交叉熵(Binary Cross Entropy, BCE)和加权交并比(Intersection over Union, IoU)损失组成, 如式(14)所示.

$$L = L_{\text{bfem}} + L_{\text{seg}} \quad (13)$$

$$L_{\text{seg}} = L_{\text{wbce}} + L_{\text{wiou}} \quad (14)$$

3 实验结果和分析

3.1 实验环境

本文提出的模型基于 PyTorch 1.12.1 实现. 在实验中, 使用了 Intel(R) Core(TM) i7-4790K CPU @ 4.00 GHz 的处理器, 并在两张 NVIDIA GeForce GTX 1080 Ti(11 GB) 设备上训练, 使用 Adam(Adaptive Moment Estimation) 优化器^[25]来调整模型参数, 初始学习率为 10^{-4} , 随着训练次数的增加逐渐降低学习率.

3.2 数据集

为了更加全面地评估模型的性能, 本文在 CVC-300^[26]、CVC-ClinicDB^[27]、Kvasir-SEG^[28]、CVC-ColonDB^[29] 和 ETIS-LaribPolypDB^[30] 等 5 个公开数据集进行了对比实验. CVC-300 数据集是 EndoScene^[26] 数据集的一部分, 我们选取了 60 张图像作为测试集, EndoScene 包含了来自 36 例患者的 44 个结肠镜检查序列的 912 张图像, 分辨率为 574×500 . CVC-ClinicDB 包含了来自 31 个结肠镜检查序列的 612 张图像, 分辨率为 384×288 . Kvasir-SEG 是胃肠息肉图像及其相应标签的数据集, 由医生手动标注, 然后由经验丰富的胃肠病学家验证, 图像的分辨率各不相同. CVC-ColonDB 数据集中包含了来自 15 个结肠镜检查序列的 380 张图像, 图像大小为 574×500 . ETIS-LaribPolypDB 数据集包含了从 34 个结肠镜视频中采集的一共 196 张图像, 图像大小为 1225×966 ,

是几个数据集中分辨率最高的, 并且由于 ETIS-LaribPolypDB 是早期的息肉图像, 所以该数据集中息肉与黏膜极其相似, 并且形状和大小多样化, 这使得该数据集更具有挑战性.

我们参考了 PraNet^[13] 的数据集分割方案, 为了使实验结果更加客观, 这与现在许多息肉分割模型所采用的数据集保持一致, 在 CVC-ClinicDB 选取了 550 张图像, 在 Kvasir-SEG 数据集中选取了 900 张图像, 一共 1450 张图像作为训练集, CVC-ClinicDB 剩余的 62 张图像和 Kvasir-SEG 剩余的 100 张图像作为测试集. 同时, 为了验证模型的泛化能力, 我们还在 CVC-300、CVC-ColonDB 和 ETIS-LaribPolypDB 这三个模型未见过的数据集上也进行了测试.

3.3 评估指标

实验中使用了医学图像分割领域中使用频率最高的度量指标 Dice 系数(Dice Similarity Coefficient, DSC)以及语义分割中的系统性能评价指标平均交并比 MIoU (Mean Intersection over Union). Dice 系数常用于医学图像分割领域中计算两个样本的相似度, 阈值为 $[0, 1]$, 分割的结果最好时为 1, 最差时为 0, 计算公式如式(15)所示. MIoU 是用于计算真实值和预测值两个集合的交集与并集之比, 如式(16)所示. 其中, A 表示模型的预测结果集; B 表示原始息肉图像的真实标签结果集; TP 表示被模型预测为正类的正样本数; FN 表示被模型预测为负类的正样本数; FP 表示被模型预测为正类的负样本数.

$$Dice(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|} = \frac{2 \times TP}{FP + FN + 2 \times TP} \quad (15)$$

$$MIoU(A, B) = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN} \quad (16)$$

3.4 与现有方法的对比实验

为了验证本文提出的 SBF-Net 的性能, 本文将 SBF-Net 在息肉图像上的分割结果与 UNet^[9], UNet++^[10], PraNet^[13], SANet^[12], UACANet-S^[14] 和 BDG-Net^[15] 模型分别在五个公开数据集上进行对比. 算法对比结果如表 1 所示.

我们提出的经过 SBF-Net 模型经过 TTA 后在 CVC-ClinicDB 和 Kvasir-SEG 数据集上均取得了最优的 Dice 和 MIoU. 在 CVC-ClinicDB 数据集中, SBF-Net 在 Dice 和 MIoU 上分别达到了 92.32% 和 88.53%, 经过 TTA 处理后(SBF-Net+

TTA)分别达到了 92.60%和 88.81%,均优于对比方法. 在 Kvasir-SEG 数据集上, SBF-Net 在 MIoU 上达到了 86.79%, 优于所有对比方法, Dice 系数达到了 91.04%, BDG-Net 的 Dice 仅高于 SBF-Net 约 0.004 5, 达到几乎可以忽略的优势, 并且 SBF-Net 经过 TTA 处理后 Dice 和 MIoU 分别达到了 91.60%和 87.92%, 高于 BDG-Net 和其他对比方法.

表 1 不同方法在 CVC-ClinicDB 和 Kvasir-SEG 数据集上的分割结果对比

Tab. 1 Comparison of different methods on CVC-ClinicDB and Kvasir-SEG datasets

Methods	CVC-ClinicDB		Kvasir-SEG	
	Dice/%	MIoU/%	Dice/%	MIoU
U-Net ^[9]	82.30	75.50	81.80	74.60
UNet++ ^[10]	79.40	72.90	82.10	74.30
PraNet ^[13]	89.90	84.90	89.80	84.00
SANet ^[12]	91.60	85.90	90.40	84.70
UACANet-S ^[14]	91.60	—	90.50	85.20
BDG-Net ^[15]	91.60	86.40	91.50	86.50
SBF-Net	92.32	88.53	91.04	86.79
SBF-Net+TTA	92.60	88.81	91.60	87.92

为了评估 SBF-Net 的稳定性和泛化能力, 我们将 SBF-Net 和 SBF-Net + TTA 在 CVC-300、CVC-ClonDB 和 ETIS-LaribPolypDB 三个模型未见过的数据集上进行了测试, 算法对比结果如表 2 所示.

表 2 不同方法在 CVC-300、CVC-ColonDB 和 ETIS-LaribPolypDB 数据集上的分割效果对比

Tab. 2 Comparison of different methods on CVC-300, CVC-ColonDB and ETIS LibPolypDB datasets

Methods	CVC-300		CVC-ColonDB		ETIS-LaribPolypDB	
	Dice /%	MIoU /%	Dice /%	MIoU /%	Dice /%	MIoU /%
U-Net ^[9]	71.00	62.70	51.20	44.40	39.80	33.50
UNet++ ^[10]	70.70	62.40	48.30	41.00	40.10	34.40
PraNet ^[13]	87.10	79.70	70.90	64.00	62.80	56.70
SANet ^[12]	—	—	75.30	67.00	75.00	65.40
UACANet-S ^[14]	—	—	78.30	70.40	69.40	61.50
BDG-Net ^[15]	89.90	83.10	80.40	72.50	75.60	67.90
SBF-Net	90.18	84.55	82.40	75.42	76.24	69.86
SBF-Net + TTA	90.22	84.69	82.34	75.62	77.49	71.29

由表 2 可知, 我们提出的模型在 CVC-300、CVC-ColonDB 和 ETIS-LaribPolypDB 上也表现出优秀的性能, Dice 和 MIoU 均高于其他方法. 其中, ETIS-LaribPolypDB 数据集为早期息肉图像, 大部分息肉还未出现明显的凸起, 因此表现为小型的扁平状, 我们的模型分割结果仍然取得了显著提高. 该测试结果可以反映出我们提出的方法具有更加优秀的泛化能力和准确的预测能力.

图 4 展示了 SBF-Net 与其他分割模型在部分息肉图像上的分割结果对比. 图 4 每一列分别是原息肉图像、真实标签图、SBF-Net、U-Net、UNet++、PraNet、SANet、UACANet-S 和 BDG-Net 网络的分割结果图.

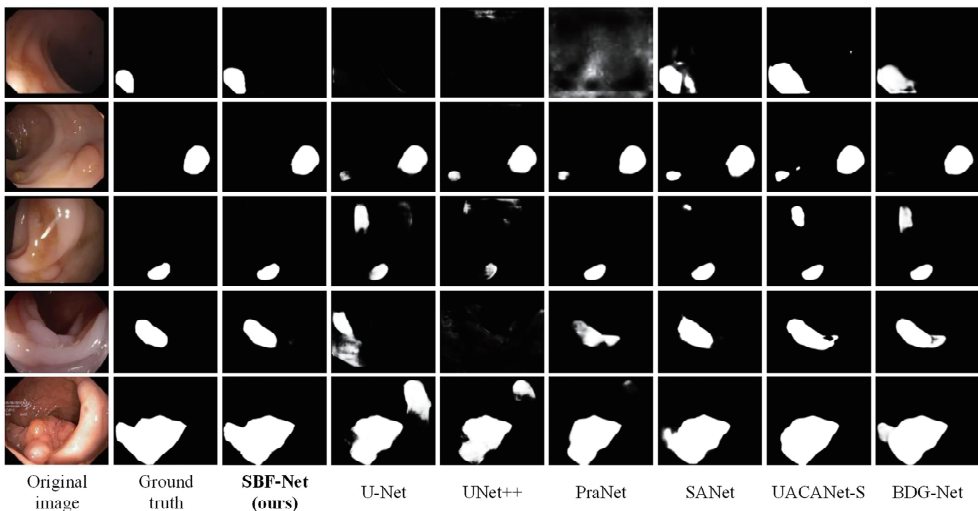


图 4 SBF-Net 与其他模型的息肉分割结果对比

Fig. 4 The results of polyp segmentation were compared between SBF-Net and other model

从图 4 可以看出, SBF-Net 的分割结果相比于其他方法更接近于真实标签。如图 4 中第 1 行所示, 息肉由于过度曝光导致 U-Net、UNet++ 和 PraNet 模型根本无法识别, 剩余的几个方法的分割结果也与真实标签出现较大差异, SBF-Net 通过加强局部特征并融合多尺度特征的解码器能够有效捕获更多深层和浅层信息, 对过度曝光区域的处理更好。如图 4 第 2 行和第 3 行所示, 由于周围黏膜与息肉颜色极其相似, 许多对比方法都将黏膜误识别为息肉, 我们的模型可以在关注关键局部信息的同时, 还抑制了其他无用信息, 并且在息肉外部无伪影出现。如图 4 第 4 和第 5 行所示, SBF-Net 都比其他算法表现出更好的边界处理效果, 在其他对比方法中均出现了过度分割或分割不足的情况, 我们的模型通过边界加强模块让模型对边界处特征有更好的学习能力, 有效地对息肉边缘进行了更准确的分割。

3.5 消融实验

为了验证边界特征加强模块、逐步融合边界特征解码器和 TTA 的有效性, 我们设计了从 Base 模型逐渐过渡到 SBF-Net 模型的消融实验来评价

各个模块对实验结果的影响, 并分别在五个数据集上进行了对比, 消融实验结果如表 3 所示。本文主要做了 5 组实验, 第 1 组是使用 EfficientNet-B5 编码器^[31] 加上一个语义分割头 Segmentation-Head 作为 Base 模型。第 2 组是在 Base 的基础上加上 U-Net 的解码器, 并加入边界特征加强模块 (BFEM)。第 3 组是在 Base 的基础上加入了逐步融合边界特征解码器 (GFBD)。第 4 组则是为了研究边界特征加强模块与逐步融合边界特征解码器的组合是否有效, 即本文提出的模型 SBF-Net。第 5 组为加上 TTA 作为后处理的 SBF-Net。通过对比可以发现, 本文提出的边界特征加强模块和逐步融合边界特征解码器对模型有着积极的影响, 从第一组到第 4 组实验相对于前一组实验均有不同程度的提升。第 4 组 (SBF-Net) 相比于第 1 组 (Base) 在 CVC-ColonDB 和 ETIS-LaribPolypDB 数据集上的 Dice 和 MIoU 有 4.19%~6.37% 的明显上升, 在其他数据集上也有显著上升。而经过 TTA 处理后的 SBF-Net 模型也在大部分数据集上都有精度上的提升。实验结果表明, 本文方法生成的息肉分割图像在分割质量和泛化能力上都具有优秀的表现。

表 3 SBF-Net 在不同数据集上的消融研究

Tab. 3 Ablation of SBF-Net on different datasets

Methods	CVC-ClinicDB		Kvasir-SEG		CVC-300		CVC-ColonDB		ETIS-LaribPolypDB	
	Dice/%	MIoU/%	Dice/%	MIoU/%	Dice/%	MIoU/%	Dice/%	MIoU/%	Dice/%	MIoU/%
Base	90.62	85.46	90.13	85.05	88.62	80.58	76.08	69.05	72.05	65.21
Base + BFEM	92.23	88.24	90.85	86.11	89.46	83.42	78.64	71.23	73.81	66.89
Base + GFBD	92.30	88.48	90.75	86.43	90.16	84.50	78.86	72.20	75.84	69.85
Base + GFBD + BFEM	92.32	88.53	91.04	86.79	90.18	84.55	82.40	75.42	76.24	69.86
Base+GFBD+BFEM+TTA	92.60	88.81	91.60	87.92	90.22	84.69	82.34	75.62	77.49	71.29

4 结 论

针对内窥镜息肉图像分割中存在的边界分割过度或不足, 以及在不同图像上息肉形状差异较大导致分割效果差等问题, 本文提出了一种基于 U-Net 改进的 SBF-Net 模型。该模型通过引入边界特征加强模块来提升模型对边界特征的学习能力, 在编码器的跳跃连接中间加入一个特征加强模块用来加强关键的局部特征和浅层特征, 在解码器阶段, 通过逐步融合编码器特征、上一阶段的输出特征和息肉边界特征, 指导生成更加准确的分割, 并减少高维和低维特征之间的信息差距, 改善了 U-

Net 中直接并行聚合深度差异较大的不同阶段特征而产生的信息差距问题, 最后再使用 TTA 来进一步提高模型在某些数据集上的整体预测性能。在 5 个公开数据集上的实验结果表明, SBF-Net 的 Dice 和 MIoU 在 5 个数据集上都在不同程度上优于其他方法, 较为客观地说明本文方法在分割性能和泛化能力上更有优势。但是我们在测试时发现经过 TTA 处理的 SBF-Net 在推理效率上有显著降低, 因此我们未来将进一步对本文方法进行优化, 以及考虑使用数据预处理等手段来改善因内窥镜图像高光对分割造成的不良影响等。

参考文献:

- [1] Bray F, Ferlay J, Soerjomataram I, *et al.* Global cancer statistics 2018; GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA-Cancer J Clin*, 2018, 68: 394.
- [2] Gschwantler M, Kriwanek S, Langner E, *et al.* High-grade dysplasia and invasive carcinoma in colorectal adenomas; a multivariate analysis of the impact of adenoma and patient characteristics [J]. *EUr J Gastroen Hepat*, 2002, 14: 183.
- [3] Hagggar F A, Boushey R P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors [J]. *Clin Colon Rect Surg*, 2009, 22: 191.
- [4] 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: 470.
- [5] Rahim T, Usman M A, Shin S Y. A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging [J]. *Comput Med Imag Grap*, 2020, 85: 101767.
- [6] 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]. *P IEEE*, 2019, 108: 178.
- [7] Wang Y G, Xi Y Y, Pan X Y. Method for intestinal polyp segmentation by improving DeepLabv3+ network [J]. *FCST*, 2020, 14: 1243. [王亚刚, 郝怡媛, 潘晓英. 改进 DeepLabv3+ 网络的肠道息肉分割方法 [J]. *计算机科学与探索*, 2020, 14: 1243.]
- [8] Li D, Wang Y, Ma Z Q, *et al.* Ultrasound image segmentation based on Dense ASPP model [J]. *J Sichuan Univ(Nat Sci Ed)*, 2020, 57: 741. [李嶝, 王艳, 马宗庆, 等. 基于 DenseASPP 模型的超声图像分割 [J]. *四川大学学报(自然科学版)*, 2020, 57: 741.]
- [9] 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.
- [10] Zhou Z, Rahman S M M, Tajbakhsh N, *et al.* Unet ++: A nested u-net architecture for medical image segmentation [M]//Deep learning in medical image analysis and multimodal learning for clinical decision support. Berlin: Springer, 2018: 3.
- [11] 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). [S. l.]: IEEE, 2019.
- [12] Wei J, Hu Y, Zhang R, *et al.* Shallow attention network for polyp segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021.
- [13] 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.
- [14] Kim T, Lee H, Kim D. Uacnet; uncertainty augmented context attention for polyp segmentation [C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM Press, 2021.
- [15] Qiu Z, Wang Z, Zhang M, *et al.* BDG-Net; boundary distribution guided network for accurate polyp segmentation [C]//Medical Imaging 2022; Image Processing. [S. l.]: SPIE, 2022.
- [16] Wang J, Huang Q, Tang F, *et al.* Stepwise feature fusion; local guides global [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2022.
- [17] Jha D, Smedsrud P H, Johansen D, *et al.* A comprehensive study on colorectal polyp segmentation with ResUNet ++, conditional random field and test-time augmentation [J]. *IEEE J Biomed Health*, 2021, 25: 2029.
- [18] Dong B, Wang W, Fan D P, *et al.* Polyp-pvt: polyp segmentation with pyramid vision transformers [EB/OL]. [2022-11-20]. <https://arxiv.org/abs/2108.06932>.
- [19] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2016.
- [20] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018.
- [21] Liu S, Huang D. Receptive field block net for accurate and fast object detection [C]//Proceedings of

- the European Conference on Computer Vision (ECCV). Cham: Springer, 2018.
- [22] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2015.
- [23] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//International Conference on Machine Learning. New York: PMLR, 2015.
- [24] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks [C]//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA: JMLR, 2011.
- [25] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. [2022-11-24]. <https://arxiv.org/abs/1412.6980v6>.
- [26] Vázquez D, Bernal J, Sánchez F J, *et al.* A benchmark for endoluminal scene segmentation of colonoscopy images [J]. *J Healthc Eng*, 2017, 2017: 4037190.
- [27] 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]. *Comput Med Imag Grap*, 2015, 43: 99.
- [28] 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.
- [29] Tajbakhsh N, Gurudu S R, Liang J. Automated polyp detection in colonoscopy videos using shape and context information [J]. *IEEE T Med Imaging*, 2015, 35: 630.
- [30] 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 Ass Rad*, 2014, 9: 283.
- [31] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks [C]//International Conference on Machine Learning. New York: PMLR, 2019.