

基于局部上下文融合的息肉语义分割模型

蔡体健, 蒋嘉豪, 刘遵雄, 赵师明, 易晟权

华东交通大学信息与软件工程学院, 江西 南昌 330013

【摘要】为了准确分割息肉,提出了局部上下文融合的分割模型,该模型引入局部上下文注意力机制来过滤掉无关的特征信息,并增强对重要区域的关注。通过多核扩展卷积捕获不同尺度的特征,以提高息肉边界分割的精度。引入金字塔上下文选择模块,利用较浅层编码器特征补偿深层编码器丢失的低级信息,使模型能够适应各种大小的息肉。该模型在Kvasir-SEG、EndoScene和CVC-ClinicDB数据集上分别达到了97.67%、97.19%和99.23%的准确率,平均交并比分别为91.2%、88.31%和94.75%,比现有经典方法有更好的准确性和泛化性,验证了该模型在息肉分割任务中的优越性能。本文模型在息肉分割准确性方面得到提升,为息肉分割提供更为精准的辅助。

【关键词】结直肠癌;息肉分割;深度学习;扩展卷积;上下文信息

【中图分类号】R318

【文献标志码】A

【文章编号】1005-202X(2025)01-0128-07

Polyp semantic segmentation model based on local context fusion

CAI Tijian, JIANG Jiahao, LIU Zunxiong, ZHAO Shiming, YI Shengquan

School of Information and Software Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract: A local context fusion based segmentation model which uses a local context attention mechanism to filter out irrelevant feature information and enhance the attention to important regions is presented for accurate polyp segmentation. The features at different scales are captured by multi-kernel dilated convolution for improving the accuracy of polyp boundary segmentation. Pyramid context selection module utilizes shallow encoder features to compensate for the low-level information lost by the deeper encoder, enabling the model to adapt to polyps of various sizes. The proposed model achieves accuracies of 97.67%, 97.19% and 99.23% on Kvasir-SEG, EndoScene and CVC-ClinicDB datasets, respectively, with mIoU of 91.20%, 88.31% and 94.75%, respectively, exhibiting higher accuracy and generalizability than the existing classical methods and validating its superior performance in polyp segmentation. The proposed model can improve polyp segmentation accuracy and provide a more accurate aid for polyp segmentation.

Keywords: colorectal cancer; polyp segmentation; deep learning; dilated convolution; context information

前言

结直肠癌是一种常见的癌症,其死亡率在所有癌症中位居第二位^[1]。研究表明,结直肠息肉的存在会增加结直肠癌的风险,而结肠镜检查是检测和切除结直肠息肉的最佳方法。然而对于较小的息肉,结肠镜检查的漏诊率显著增加^[2]。在结肠镜检查中,所有息肉的漏诊率为25%,一些结直肠癌患者可能是由于早期检查漏诊而进展为癌症^[3]。研究还表明,息肉

的大小以及位置是影响漏诊率的关键因素,息肉越小漏诊率越高^[4]。

由于手工检测息肉存在误诊率高和人工成本高的问题,基于深度学习的结直肠息肉分割作为辅助检测方法具有重要的现实意义^[5]。U-Net在医学图像分割领域表现出卓越性能,其扩展路径输出的特征包含了位置信息,而收缩路径进一步对位置进行了详细刻画^[6]。此外,U-Net网络通过金字塔池化等技术缓解了物体尺寸变化的问题。因此,基于U-Net的骨架模型变体已经成为医学图像分割模型的主流^[7-9]。SegNet通过在编码阶段运用最大池化层下采样,在解码阶段根据位置信息进行上采样,从而解决了物体边界不清晰的问题^[10]。近年来,有许多专门用于息肉分割的网络模型被提出,如ACSNet利用局部上下文模块和全局上下文模块分别关注难以分割

【收稿日期】2024-09-26

【基金项目】国家自然科学基金(62166018);江西省研究生创新专项资金项目(YC2023-S531);江西省自然科学基金(20232BAB202055)

【作者简介】蔡体健,博士,副教授,硕士生导师,研究方向:计算机视觉、深度学习、稀疏表示,E-mail: cai2017@ecjtu.edu.cn

的区域和全局上下文特征,以提高性能的鲁棒性^[11]。然而,ACSNet仅依赖一个全局上下文特征为解码器提供补偿,可能忽视了较低级的细节信息。传统的卷积神经网络侧重于提取局部特征,容易忽略全局特征,而通过注意力机制在卷积网络中引入全局特征提取能够弥补这一不足^[12-14]。PraNet主要通过反向注意力模块关注前景与息肉边界的差异,从而实现更精确的分割^[15]。CCBANet提出了级联上下文和注意力平衡模块,将其集成到编码器和解码器中实现背景、息肉和边界3个区域的注意机制,以构建高效的语义分割模型^[16]。LDNet合理运用了自注意力机制来捕获长范围的上下文关系,以增强息肉和背景区域之间的特征对比度,进一步提高了分割精度,然而却忽视了低层特征给息肉轮廓和边缘定位带来的作用^[17]。因此,本文提出了一种新模型,在该模型的编码器和解码器的跳连接处,引入了局部上下文融合(Local Context Fusion, LCF)模块。笔者认为,通过获得局部上下文权重值,可以在不同感受野上捕获信息,这有助于解码器对息肉区域的预测^[18-19]。因此,笔者将局部上下文权重值与扩展卷积相结合,形成LCF模块。该模块能够在低层编码器块中提取更多关键特征信息,并将这些信息传递到解码器块。此外,为了应对编码器进行更深层卷积操作可能带来的空间信息丢失问题^[20-21],在模型中引入了金字塔上下文选择(Pyramidal Context Selection, PCS)模块,该模块利用了3层不同尺度的编码器块融合后的特征图来进行上下文选择,以弥补深层结构中细节信息的缺失^[22-23]。通过对不同尺度特征进行融合,能够更好地保留空间信息,从而提高模型的整体性能。本文主要贡献总结如下:(1)提出了LCF模块。该模块中的局部上下文注意(LCA)模块用于精确定位和识别息肉边界,多核扩展卷积(MKDC)模块则能够捕获更加健壮的特征信息,从而提高息肉分割的准确率。(2)提出了PCS模块,PCS模块通过池化融合来自不同尺度的特征金字塔中的上下文信息,通过选择和融合来提取不同层级的特征,用低层编码器的细节信息去弥补多次卷积导致丢失的信息,这样能将更多的息肉特征传递给解码器提供更准确的预测。(3)本文提出的LCFNet在Kavsir-SEG、EndoScene和CVC-ClinDB公开数据集上进行了广泛的实验评估,取得了更高的评价指标。实验结果表明,LCFNet具有更强的泛化能力,在不同数据集上都能取得优异的性能。

1 文中所采用的方法

本文提出的LCFNet网络架构如图1所示,骨干

框架是编码器-解码器结构架构。编码器部分采用了ResNet34结构^[24],其中,每个编码器块连接到LCA模块,LCA用于增强当前层的局部上下文,并将特征图分别传给当前层和上一层的MKDC模块。LCA和MKDC组成了LCF模块,通过融合上下层的局部上下文信息,实现高效的特征细节捕获,从而准确地确定息肉边界。为了减少区域之间的上下文信息丢失,受金字塔池化模块(Pyramid Pooling Module)的启发^[7],提出了金字塔上下文选择(Pyramidal Context Selection, PCS)模块,该模块利用池化操作对各编码器块不同尺度的特征图进行上采样,得到分辨率一致的特征图,然后对处理过后的特征图进行特征融合,以实现特征信息的互补。这样做可以有效地利用不同层次的特征图,提高网络对上下文信息的理解和表达能力^[25]。在整个网络结构中,笔者还引入了全局上下文模块(Global Context Module, GCM)和自适应选择模块(Adaptive Selection Module, ASM)。GCM帮助网络感知全局特征和上下文关系,而ASM则有助于动态选择和调整不同模块的重要性。本文网络模型采用了深度监督的优化策略,通过在网络内部的多个层级上引入监督信号,引导网络的学习过程,从而实现高效的模型优化。

1.1 局部上下文融合模块

息肉分割任务中提取有效的上下文信息能精准地定位息肉位置并识别息肉边界^[26-27],为了充分捕捉息肉特征,提出了LCF模块,该模块包括LCA模块和MKDC模块。LCF模块允许网络在关注局部结构和上下文关系的同时,融合不同感受野捕捉到的息肉特征信息^[28-29],使模型能够从不同角度和尺度理解特征信息,从而提供更全面和丰富的表示能力。LCF模块的结构如图2所示。

该模块首先对当前层编码器输出的特征图传入到LCA模块,与第 $i+1$ 个解码器块生成的预测图 $D_{i+1} \in R^{1 \times H \times W}$ 进行像素乘,帮助网络捕捉息肉边界周围的微小特征,这些特征对于准确定位边界非常关键。LCA模块的计算公式如下:

$$LCA(E_i, D_{i+1}) = E_i \oplus \left(E_i \otimes \left(1 - \frac{|D_{i+1} - T|}{\max(T, 1 - T)} \right) \right) \quad (1)$$

其中, T 为超参数,用于控制像素区域是前景还是背景,为了实验方便,笔者将 T 的值其设置为固定的0.5。

然后,将LCA模块的输出特征 L_i 与下层LCA模块的输出特征 L_{i+1} 相融合,作为MKDC的输入。MKDC模块对特征图进行了4个卷积核大小分别为 1×1 、 3×3 、 7×7 、 11×11 的普通卷积操作,并将特征联合

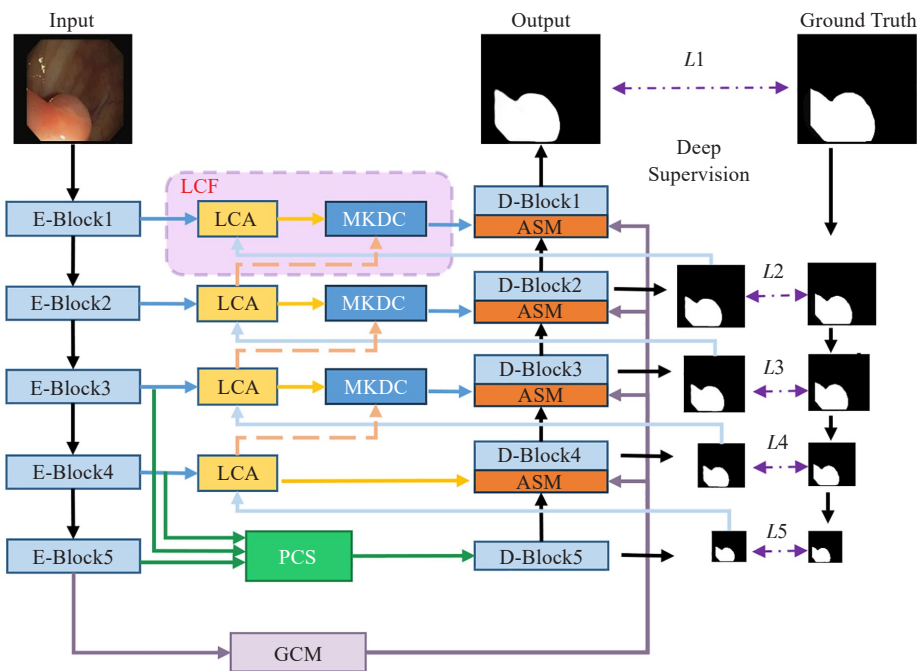


图1 LCFNet整体结构
Figure 1 Overall structure of LCFNet

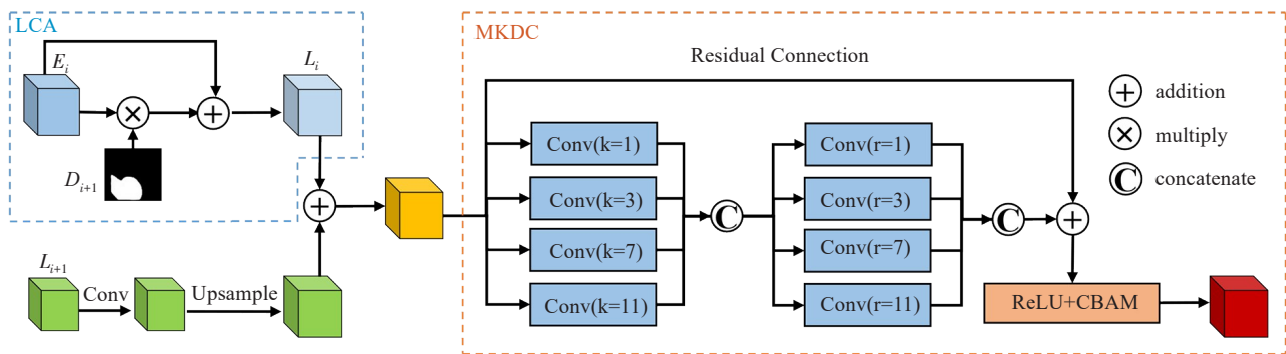


图2 局部上下文融合模块
Figure 2 Local context fusion module

起来构建更复杂的特征表示^[30],产生更接近真实的息肉与背景分割特征,这样的设计允许模型在不同的感受野上捕捉细粒度的局部特征和粗粒度的全局特征,这对于检测不同大小的息肉非常重要^[31]。同时,通过使用扩张率 r 为1、3、7、11的扩展卷积,进一步增强了模型不同感受野下对息肉区域的特征表示能力。较小扩张率的卷积层可以更好地捕捉细节和局部特征,而较大扩张率的卷积层则有助于捕捉更远的上下文信息。最后,引入了卷积块注意力模块(CBAM)^[32],以增强对重要特征的关注并抑制无关信息。CBAM结合了通道注意力和空间注意力机制,通过自适应地调整通道权重和空间权重,帮助网络更好地感知和利用特征信息。综上所述,LCF模块结合了LCA模块和MKDC的优势,并引入了

CBAM,能提高模型的感知能力和泛化能力。

LCF模块的工作流程如式(2)所示:

$$LCF = CBAM \left(\begin{matrix} \text{ReLU} (r_{1,3,7,11} (K_{1,3,7,11} ((LCA(E_i, D_{i+1}) + \\ \text{Up}(\text{Conv}(L_{i+1})))))) \end{matrix} \right) \quad (2)$$

1.2 金字塔上下文选择模块

随着神经网络的加深,经过更多的卷积等一系列操作之后,高层特征更多地关注语义和抽象表征,低层特征会被逐渐抽象和合并,导致丢失一部分原始输入的细节信息。为此,笔者采用金字塔池化思想设计了PCS模块。该模块位于编码器顶层,并利用平均池化和全局池化操作提取3、4、5层编码器块的特征图,以整合不同通道的特征信息并调整图片

尺寸。为了弥补上下文信息的损失,引入了全局注意力机制(GAM)^[33],将其插入特征融合后的阶段。GAM通过增强全局交互表征的权重,进一步提升网络模型的性能。这种设计使得模型能够更准确地地理

解输入数据,充分利用不同层次的特征和上下文信息,以提高网络在各种任务中的表现。PCS模块结构如图3所示。

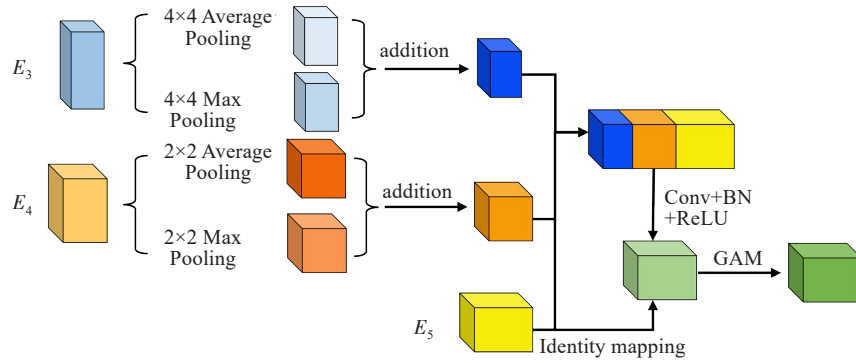


图3 金字塔上下文选择模块

Figure 3 Pyramid context selection module

PCS模块详细流程如下:

$$PCS = GAM \left(\text{Conv} \left(\text{Concat} \left(PL(E_3), PL(E_4), E_5 \right) \oplus E_5 \right) \right) \quad (3)$$

其中,Conv表示卷积、批归一化处理和ReLU的组合。PL(E_i)如下表示:

$$PL(E_i) = \text{AvgPool}(E_i) \oplus \text{MaxPool}(E_i) \quad (4)$$

GAM中的通道注意力模块利用了双层的多层感知器将跨维度通道空间依赖性放大,防止丢失大量的跨维度信息。而空间注意力子模块能够集中空间信息,通过对不同位置的注意力加权使模块聚焦于更重要和相关性更高的区域,从而能够更好地利用上下文信息,增强模型的鲁棒性和泛化能力。综上所述,PCS模块应用了金字塔池化的思想,结合多层次特征融合和上下文信息的充分利用,有助于在分割边界处进行细化,以增强模型的鲁棒性和泛化能力。通过合理设计特征处理和注意力机制,笔者能够提取到关键信息,从而提升整体模型的性能和效果^[25]。

2 实验分析

2.1 数据集

为了评估LCFNet的性能,笔者使用公开数据集Kvasir-SEG^[34]、EndoScene^[35]和CVC-ClinicDB进行实验验证。这些数据集都由经验丰富的内窥镜医生完成图像的注释和验证。Kvasir-SEG数据集包含了1 000张带有Ground Truth的息肉图像,图像的分辨

率大小不同。EndoScene数据集由CVC-ClinicDB的612张分辨率为332×487的图像和CVC-endoscenestill的300张分辨率为574×500的图像组成。在评估阶段,笔者将自己设计的LCFNet与经典模型和最先进的近期模型进行了比较。通过在各个数据集上进行实验,评估了模型在息肉分割任务上的性能表现,并与其他模型进行性能对比。通过以上实验设置和评估方法,能够全面地了解LCFNet在不同数据集上的表现,并与其他模型进行客观比较,以验证LCFNet的优越性和有效性。

2.2 实验设置

本文实验硬件配置为Intel i7-8700K CPU, NVIDIA GeForce RTX 3080 GPU。软件环境为Win10操作系统,Python3.8,PyTorch1.8.1。为了公平比较,对所有的模型使用了相同的超参数集。在进行训练之前,对Kvasir-SEG数据集中的图像进行了预处理,统一将分辨率调整为320×320。对于EndoScene和CVC-ClinicDB数据集,将图像的分辨率统一调整为288×384。整个数据集被划分为训练集、验证集和测试集。为了增加数据的多样性和泛化能力,对图像进行了数据增强操作,包括随机旋转、缩放、水平翻转和垂直翻转等。在网络的超参数设置方面,初始learning_rate=0.001,weight_decay=0.000 01,momentum=0.9,batch_size=6,笔者使用SGD优化器对模型进行训练。学习率会随着训练轮数增加而调整,调整公式为: $lr = 1 - \left(\frac{\text{epoch}}{\text{nEpoch}} \right)^{\text{power}}$, epoch表示当前训练轮数,nEpoch表示总训练轮数,

将nEpoch设定为150, power设为0.9进行训练。模型引入深度监督优化策略,使用二元交叉熵损失和骰子损失结合损失函数。

2.3 评价指标

选择运用最为广泛的6个指标来评估各种神经网络的性能和准确性:召回率(Recall)、特异度(Specificity)、精确率(Precision)、平均交并比(mean Intersection over Union, mIoU)、准确率(Accuracy)和Dice相似系数(Dice Similarity Coefficient, DSC)。部分评估度量指标公式如下:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{mIoU} = \frac{1}{k + 1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \tag{8}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$\text{DSC} = \frac{2TP}{2TP + FP + FN} \tag{10}$$

2.4 实验分析

将本文模型(LCFNet)在Kvasir-SEG、EndoScene和CVC-ClinicDB数据集与其他传统方法以及先进方法进行了比较,包括U-Net^[6]、PraNet^[15]、ACSNet^[11]、CCBANet^[16]、LDNet^[17]。在实验中,使用了他们的默认设置以保证比较的公平性。实验结果如表1所示,LCFNet在Kvasir-SEG、EndoScene和CVC-ClinicDB数据集取得了较好的结果,在Kvasir-SEG数据集上的准确率能达到97.67%,与性能较高的LDNet模型相比提升了0.51%;在EndoScene数据集上的准确率能达到97.19%,超过其他模型。在Kvasir-SEG和EndoScene数据集上的mIoU指标分别高出LDNet模型0.59%和0.93%,优于其他模型。在CVC-ClinicDB数据集上也是略优于其他5个模型。3个数据集的实验结果表明,本文方法相较于其他方法有着更好的泛化性能。

图4展示了各个模型在150轮迭代训练过程中的准确率曲线,从图中可以看出不同模型的表现都随着训练次数的增加逐渐趋于稳定。从图4可以看出,LDNet在前期准确率快速提升,这得益于其使用了经过预训练的ResNet模型。本文模型LCFNet的准确率曲线在训练的中后期继续稳步提升,最终超过LDNet并且保持在一个较高且稳定的水平,相比于其他模型显示出更好的收敛性。

图5展示了各种当前息肉分割方法在同一数据集

表1 不同算法在Kvasir-SEG、EndoScene和CVC-ClinicDB数据集上的结果(%)

Table 1 Results of different algorithms on Kvasir-SEG, EndoScene and CVC-ClinicDB datasets (%)

模型	召回率	特异度	精确率	mIoU	准确率	DSC
Kvasir-SEG数据集						
U-Net	87.04	97.25	84.28	83.64	95.05	82.60
PraNet	88.51	97.78	90.80	89.44	96.67	90.47
ACSNet	92.18	98.06	92.76	90.72	96.86	90.83
CCBANet	91.82	98.95	90.44	89.99	96.88	90.06
LDNet	91.39	98.64	92.87	90.61	97.16	89.50
LCFNet	91.65	99.20	94.41	91.20	97.67	90.98
EndoScene数据集						
U-Net	85.54	98.75	83.56	83.29	96.25	80.54
PraNet	81.99	98.62	90.94	86.56	96.53	84.40
ACSNet	85.75	99.28	91.68	87.48	96.93	84.95
CCBANet	86.48	99.04	91.89	86.37	97.13	86.59
LDNet	86.40	99.40	91.42	87.38	96.98	83.26
LCFNet	86.80	99.44	92.16	88.31	97.19	86.09
CVC-ClinicDB数据集						
U-Net	88.61	98.70	85.10	87.74	97.95	85.12
PraNet	95.44	99.22	89.23	91.91	98.75	93.22
ACSNet	94.04	99.41	94.37	93.80	98.95	93.22
CCBANet	94.79	99.23	91.39	92.87	99.12	93.83
LDNet	94.76	99.59	95.00	94.58	99.15	93.84
LCFNet	95.06	99.60	95.07	94.75	99.23	94.80

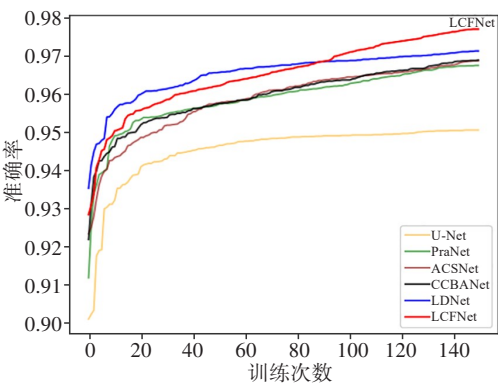


图4 准确率曲线图
Figure 4 Accuracy curves

上进行息肉分割测试的实验结果。从图中可以看出,LCFNet能够更准确地识别出息肉的轮廓形状,其划分的边界更为清晰,且结果也更接近真实值,表现优于其他方法。图中不仅展示出本文方法对大块息肉的准确分割,也对形状难以识别的小块息肉能够进行准确的

识别和边界定位。这说明LCFNet在处理息肉分割任务时具有更强的能力,这种优越性可以归因于本文提出的LCF模块和PCS模块,它们共同增强了模型对复杂边界和细节区域的识别能力,从而提升了整体分割

效果。综合以上分析,图中的数据清晰地证明了LCFNet相较于其他方法的显著优势,无论在准确率还是边界区分或其他评价指标的表现上,都显示出其在息肉分割任务中的卓越性能。

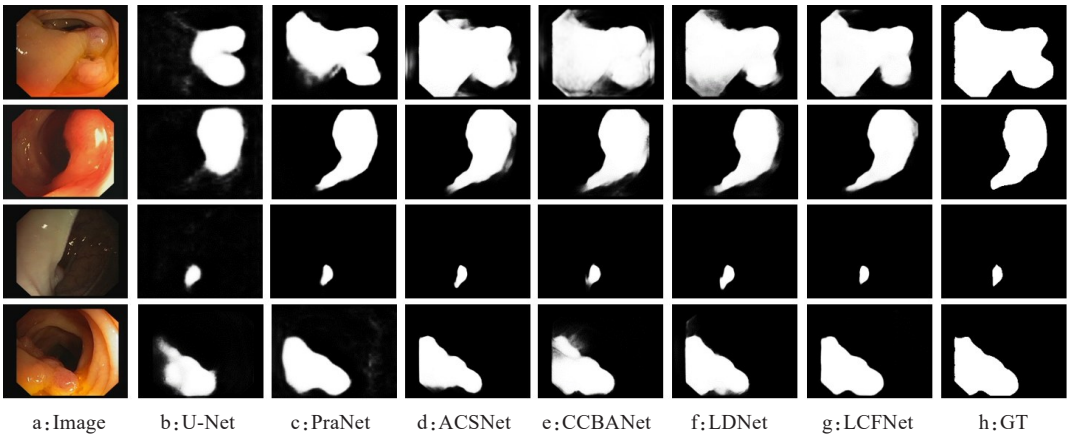


图5 不同方法的实验结果

Figure 5 Experimental results of different methods

白色部分为息肉分割区域,黑色部分为背景区域,GT为真实值

2.5 消融实验

为了评估本文方法的有效性,笔者在Kvasir-SEG、EndoScene、CVC-ClinicDB数据集上对本文方法中两个关键模块进行了消融实验。将没有加入LCF、PCS模块的架构当做基线,分别进行了加入LCF模块和PCS模块的实验。实验结果如表2所示,笔者在基线上加入PCS模块时,模型在3个数据集上表现出来的性能都较优于基线性能,这表明加入PCS模块后模型算法对息肉的分割能够更接近真实的分割结果;当笔者在基线模型上加入LCF模块后,图像分割模型的主要评价指标准确率、mIoU和DSC指标都有所提升,加入LCF模块能够增强息肉细节特征的捕获,提升算法对息肉定位的能力。笔者以ResUNet结合ASM模块和GCM模块作为基线模型(Baseline),当笔者在基线中加入LCF和PCS时,即本文模型LCFNet,与基线模型相比较,本文模型LCFNet在Kvasir-SEG、EndoScene和CVC-ClinicDB上mIoU性能分别提升了0.58%、0.72%、1.71%,DSC性能分别提升了0.44%、0.98%、1.66%,模型效果在CVC-ClinicDB数据集上提升最为显著。消融实验结果表明,所提出的两个模块对网络性能都有显著提升,以此证明本文提出的模块有效性。

3 结论

在本文中,笔者提出了一种基于上下文信息融

表2 消融实验结果(%)

Table 2 Ablation study results (%)

模型	召回率	特异度	精确率	mIoU	准确率	DSC
Kvasir-SEG数据集						
Baseline	92.35	98.13	92.67	90.62	97.13	90.54
Baseline+PCS	91.38	98.69	93.81	90.95	97.15	90.66
Baseline+LCF	91.55	98.41	93.42	91.03	97.26	90.79
LCFNet	91.65	99.20	94.41	91.20	97.67	90.98
EndoScene数据集						
Baseline	85.44	99.36	91.95	87.59	96.88	85.11
Baseline+PCS	85.65	99.39	92.03	87.78	96.95	85.64
Baseline+LCF	86.13	99.3	92.08	88.09	97.01	85.47
LCFNet	86.80	99.44	92.16	88.31	97.19	86.09
CVC-ClinicDB数据集						
Baseline	94.40	99.04	92.58	93.04	98.64	93.14
Baseline+PCS	94.98	99.02	94.34	94.61	99.01	94.48
Baseline+LCF	94.65	99.58	94.97	94.52	99.07	94.64
LCFNet	95.06	99.60	95.07	94.75	99.23	94.80

合的分割模块。本文方法利用跨层级编码器输出特征图进行融合,然后输入到多个并行的不同膨胀率的扩展卷积中,以捕获不同大小感受野的上下文信息,并将其拼接融合,以获得强特征信息传输到解码器。这样的设计使得网络架构在息肉的识别和边界

判定方面获得更充分的特征信息,从而提高网络的学习能力。笔者还引入了PCS模块,利用深层的3层编码器块在不同大小分辨率情况下利用池化操作获取图像的高级特征信息。这样做可以实现更为精确的像素级目标识别和定位。在3个公开数据集上进行了大量实验,结果表明LCFNet模型在对息肉的分割上具有更好的准确性。然而,笔者也意识到目前方法在参数计算量以及网络架构上仍有改进空间。未来,笔者计划在尽可能保持网络性能的前提下对网络架构进行轻量化优化。这将使得本文方法更加高效且适用于更广泛的应用场景。

【参考文献】

- [1] Siegel RL, Miller KD, Wagle NS, et al. Cancer statistics, 2023[J]. CA Cancer J Clin, 2023, 73(1): 17-48.
- [2] van Rijn JC, Reitsma JB, Stoker J, et al. Polyp miss rate determined by tandem colonoscopy: a systematic review[J]. Am J Gastroenterol, 2006, 101(2): 343-350.
- [3] Leufkens AM, van Oijen MG, Vleggaar FP, et al. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study[J]. Endoscopy, 2012, 44(5): 470-475.
- [4] Bensen S, Mott LA, Dain B, et al. The colonoscopic miss rate and true one-year recurrence of colorectal neoplastic polyps. Polyp Prevention Study Group[J]. Am J Gastroenterol, 1999, 94(1): 194-199.
- [5] 康帅, 奚峥皓, 黄陈, 等. 基于DPU-Net的直肠肿瘤分割与T分期研究[J]. 中国医学物理学杂志, 2023, 40(10): 1189-1197.
Kang S, Xi ZH, Huang C, et al. Rectal tumor segmentation and T staging based on DPU-Net[J]. Chinese Journal of Medical Physics, 2023, 40(10): 1189-1197.
- [6] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241.
- [7] Zhao HS, Shi JP, Qi XJ, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6230-6239.
- [8] Fu J, Liu J, Wang YH, et al. Adaptive context network for scene parsing [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 6747-6756.
- [9] Fu J, Liu J, Tian HJ, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 3141-3149.
- [10] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(12): 2481-2495.
- [11] Zhang RF, Li GB, Li Z, et al. Adaptive context selection for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer International Publishing, 2020: 253-262.
- [12] Zhao QJ, Sheng T, Wang YT, et al. M2Det: a single-shot object detector based on multi-level feature pyramid network[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2019: 9259-9266.
- [13] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.
- [14] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2016-05-19). <https://arxiv.org/abs/1409.0473>.
- [15] Fan DP, Ji GP, Zhou T, et al. PraNet: parallel reverse attention network for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2020. Cham: Springer International Publishing, 2020: 263-273.
- [16] Nguyen TC, Nguyen TP, Diep GH, et al. CCBANet: cascading context and balancing attention for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer International Publishing, 2021: 633-643.
- [17] Zhang RF, Lai PW, Wan X, et al. Lesion-aware dynamic kernel for polyp segmentation[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2022. Cham: Springer Nature Switzerland, 2022: 99-109.
- [18] Chen LC, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Trans Pattern Anal Mach Intell, 2018, 40(4): 834-848.
- [19] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2016-04-30). <https://arxiv.org/abs/1511.07122>.
- [20] Tan MX, Pang RM, Le QV. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 10778-10787.
- [21] Huang G, Chen DL, Li TH, et al. Multi-scale dense networks for resource efficient image classification[EB/OL]. (2018-06-07). <https://arxiv.org/abs/1703.09844>.
- [22] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 936-944.
- [23] Yuan YH, Huang L, Guo JY, et al. OCNet: object context network for scene parsing [EB/OL]. (2021-03-15). <https://arxiv.org/abs/1809.00916>.
- [24] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.
- [25] Xie Q, Lai YK, Wu J, et al. MLCVNet: multi-level context VoteNet for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 10444-10453.
- [26] Mottaghi R, Chen XJ, Liu XB, et al. The role of context for object detection and semantic segmentation in the wild [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 891-898.
- [27] Choi MJ, Lim JJ, Torralba A, et al. Exploiting hierarchical context on a large database of object categories[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2010: 129-136.
- [28] Wu ZF, Shen CH, van den Hengel A. High-performance semantic segmentation using very deep fully convolutional networks[EB/OL]. (2016-04-15). <https://arxiv.org/abs/1604.04339>.
- [29] Zhou BL, Zhao H, Puig X, et al. Semantic understanding of scenes through the ADE20K dataset[J]. Int J Comput Vis, 2019, 127(3): 302-321.
- [30] Wang PQ, Chen PF, Yuan Y, et al. Understanding convolution for semantic segmentation [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2018: 1451-1460.
- [31] Chen LC, Yang Y, Wang J, et al. Attention to scale: scale-aware semantic image segmentation [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 3640-3649.
- [32] Woo S, Park J, Lee JY, et al. CBAM: convolutional block attention module [C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [33] Liu YC, Shao ZR, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10). <https://arxiv.org/abs/2112.05561>.
- [34] Jha D, Smedsrud PH, Riegler MA, et al. Kvasir-SEG: a segmented polyp dataset [C]//MultiMedia Modeling. Cham: Springer International Publishing, 2020: 451-462.
- [35] Vázquez D, Bernal J, Sánchez FJ, et al. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. J Healthc Eng, 2017, 2017: 4037190.

(编辑:薛泽玲)