

结合多尺度信息和特征增强的息肉分割方法

汪琴韵, 于 璩*

(安徽理工大学 计算机科学与工程系, 安徽 淮南 232001)

摘要: 计算机辅助诊断技术在结肠息肉图像分割中具有十分重要的作用。本研究针对结肠息肉分割存在的边缘检测分割比较复杂、息肉与正常组织对比度较低等问题, 使用DeepLabv3+模型进行息肉分割并对其进行改进。该模型将残差网络作为主干网络进行特征信息提取, 引入特征强化模块对低层次特征进行处理, 并加入混合注意力机制对高层次特征进行关键信息的捕捉。得到改进模型在Kvasir-SEG与CVC-ClinicDB两个数据集上的mIoU为82.19%和90.40%。实验结果表明, 改进的模型息肉分割效果优于初始DeepLabv3+模型, 在息肉分割的精度以及边缘分割上有一定提升。

关键词: DeepLabv3+; 息肉分割; 混合注意力; 特征增强; 残差网络

中图分类号: TP391

文献标识码: A

文章编号: 1008-9659(2025)01-0001-10

随着经济的快速发展及生活水平的不断提高, 人们比以往更加关注身体健康状况。国际癌症组织提供的全球癌症负担情况指出, 在五大恶性肿瘤中, 结肠直肠癌的死亡率为9.4%, 位于第二位, 并且结肠癌的发病率在人群中不断上升^[1]。目前对抗结肠癌最有效的方法是对结肠息肉进行早期检测。通过早期检测, 判断疾病的发展状况, 进而采取进一步的治疗措施。临床通过结肠镜检查结肠部分来检测结肠息肉。结肠镜的前端有一个电子摄像探头, 它能够把拍摄到的结肠黏膜图像记录下来并且保存在计算机处理中心, 通过计算机的监视器屏幕可以观察直肠和回肠黏膜的息肉或病变区域^[2]。结肠镜检查技术作为当前临床诊断结肠癌常规有效的方法, 在一定程度上能够控制结肠癌的发展。对于大小不一或结构复杂的息肉, 利用肉眼进行观察不仅耗时耗力, 还会导致漏检。因此, 研究自动化结肠息肉分割技术以提升计算机辅助诊断能力, 对提高结肠息肉的早期诊断效率和准确性具有重要意义。

在结肠息肉分割领域, 传统的息肉分割方法包括利用像素灰度值与阈值的关系进行分类的阈值分割, 利用高斯滤波器进行处理的Canny边缘检测方法或利用将息肉的纹理和形状特征结合在一起的上下文形状方法能够可靠地定位具有曲线边界的息肉^[1-2]。近年来, 随着卷积神经网络(CNN)的发展, 在结肠息肉分割领域基于深度学习的方法比传统方法更具优势。其中, Olaf等人^[3]提出在医学图像处理方面有较好效果的U-Net模型, 其独特的编码器-解码器结构以及跳跃连接设计实现了对图像分割任务的高效处理, 但其用于息肉分割时需要大量的训练数据才能获得较好的性能。Chen等人^[4]提出的DeepLabv1模型结合DCNNs和DenseCRFs模型并且利用空洞卷积扩大感受野, 降低了分辨率的损失, 而后续的DeepLab系列模型又在DeepLabv1上进行改进^[5], 在多个尺度上捕获信息, 从而提高分割的准确性, 但由于采用信息融合的方式^[6], 模型在整合多尺度信息时可能会存在局部信息丢失或全局信息未充分利用等问题, 特别是在处理复杂结构的息肉时可能存在一定的限制。Dong等人^[7]提出的Polyp-PVT分割网络采用金字塔视觉Transformer作为编码器进行息肉分割, 但当训练数据不够多或者数据质量不佳时, 该模型可能无法很好地泛化到其他类似数

[收稿日期] 2024-03-12

[修回日期] 2024-03-29

[基金项目] 2021年安徽省重点研究与开发计划项目(202104d07020010)。

[作者简介] 汪琴韵(2000-), 女, 硕士研究生, 主要从事图像处理方面研究, E-mail: 512154983@qq.com.

* [通讯作者] 于 璩(1973-), 女, 教授, 主要从事图像处理方面研究, E-mail: 3550707289@qq.com.

据上,致使其性能下降。Fan 等人^[8]提出的 PraNet 模型通过并行部分解码器聚合高层特征并组合特征来获取上下文信息,但该网络在小息肉分割上仍有提升空间。因此,Ange 等人^[9]针对小医学对象分割问题在 PraNet 模型的基础上进行改进,提出 CaraNet 模型,该模型将轴向反向注意力和通道特征金字塔进行结合,在分割性能上有一定的提升,但其在向上采样过程中会丢失一些有用的信息,从而导致边界分割粗糙。Chen 等人^[10]提出的 TransResU-Net 模型将 Transformer 与残差网络进行结合,在保证避免梯度消失的同时能够更好地理解解图像中全局上下文的信息,但其存储开销过大,易造成存储空间浪费。

针对上述问题,分析各个分割模型的优缺点,文章最终选用 DeepLabv3+ 网络作为初始网络进行分割,并基于 DeepLabv3+ 网络进行改进,使该网络更适用于结肠息肉分割。该网络模型创新之处与意义在于:(1)利用残差网络替换 Xception 模型应用于分割任务,在有效避免梯度消失、加速模型收敛的同时可以利用残差连接保留原始特征,能够重用和共享特征信息,从而改善模型性能;(2)使用膨胀卷积对低层次的特征进行增强,能够在不丢失信息的情况下组合不同感受野的语义信息,即通过不同尺度的卷积核对特征进行采样,能够准确、高效地对任意尺度的区域进行分类,从而提高分割精度;(3)引入结合通道和空间的注意力机制,即轻量级卷积注意力(CBAM),该注意力机制不仅兼顾通道维度上的信息,还可以对不同空间位置的重要性进行关注,有效提升了卷积神经网络的表达能力。

1 相关工作

1.1 DeepLabv3+网络

2018年,DeepLabv3+模型被提出,该模型在 DeepLab 系列的基础上发展而来。该系列从2015年开始,到2018年,已经发布了四个版本,包括 DeepLabv1, DeepLabv2, DeepLabv3 和 DeepLabv3+。其中,DeepLabv3+模型在 DeepLabv3 的基础上继续使用编码器-解码器结构,编码器部分的作用是逐步减少特征图并提取更高的语义信息,而解码器部分的作用则是使分割结果恢复至原来的大小,这样的结构进一步提升了语义分割的性能。其中编码器的主体是深度卷积神经网络 Xception。Xception 的核心思想是深度可分离卷积操作^[11],即将普通卷积操作分解为深度卷积和逐点卷积两个步骤,这样虽然提高了计算效率,但是无法保证提高分割精度。与主干网络相连的是 ASPP,即带有空洞卷积的金字塔池化模块,它增强了网络对息肉分割的鲁棒性,在多个尺度上捕获目标与上下文信息,同时利用不同的感受野提取输入特征^[12]。解码器部分首先接受特征提取网络的低层次信息并通过 1×1 的卷积降低通道数,将编码器部分所获结果进行上采样并经过卷积之后的高层次特征信息进行拼接,经过 3×3 的卷积后,上采样将特征图恢复至原来的大小。DeepLabv3+模型结构如图1所示。

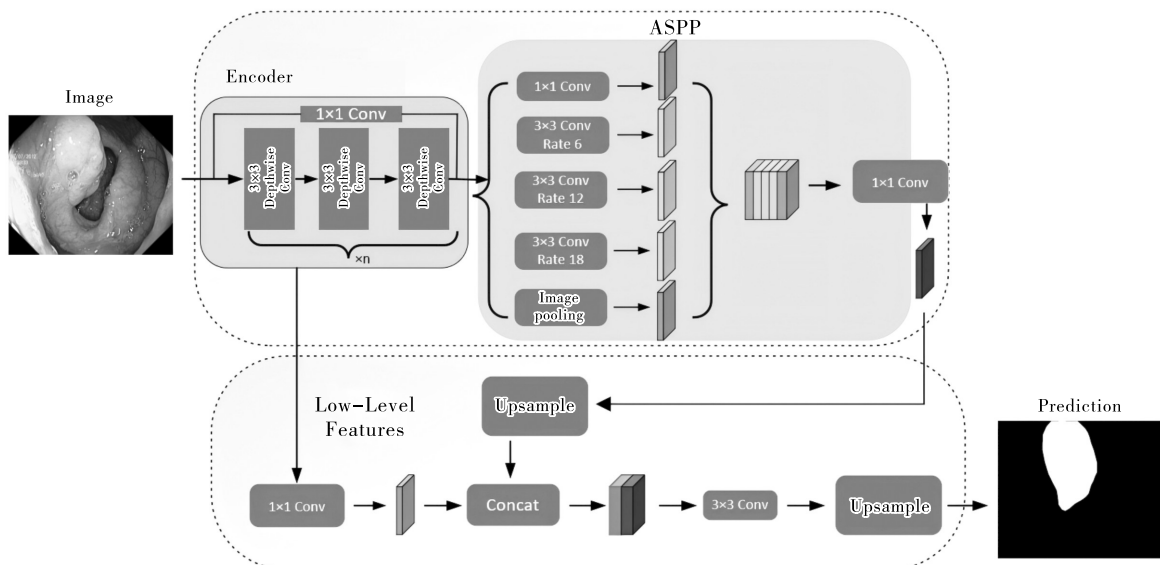


图1 DeepLabv3+模型

然而,结肠息肉不仅有不规则的形状,而且息肉与周围正常组织的颜色相近,对比度较低。除此之外,息肉大小不一也给模型分割增加了难度。这些问题都使分割息肉变得困难。如果仅仅使用初始的DeepLabv3+网络进行息肉分割,模型就无法准确地检测息肉边缘并界定息肉的形状,这会使息肉分割的准确率降低。因此,文章在初始DeepLabv3+模型的基础上进行改进,从而提高息肉分割的精度。

1.2 ResNet

2015年,残差网络(ResNet)被提出,使用残差网络解决图像分割任务有出色的效果。它有更深的网络结构,网络的深度可以增加模型的鲁棒性,其通过增加网络层数来进行更加复杂的特征提取。但是,仅仅简单地增加网络深度会导致梯度爆炸。为了避免这个问题,ResNet利用残差学习的思想,使用短跳跃连接来帮助训练深层网络,这样不仅能够解决深度神经网络中的退化问题,还可以实现更高的准确度^[14]。ResNet的网络结构主要由残差模块组成。残差模块分为常规残差模块和瓶颈残差模块。文章所使用的残差模块为瓶颈残差模块,包括三个卷积层,常用于构建ResNet50、ResNet101等。瓶颈残差模块结构如图2所示。

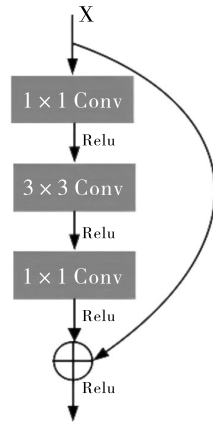


图2 瓶颈残差模块结构

2 改进方法

2.1 改进模型总体结构

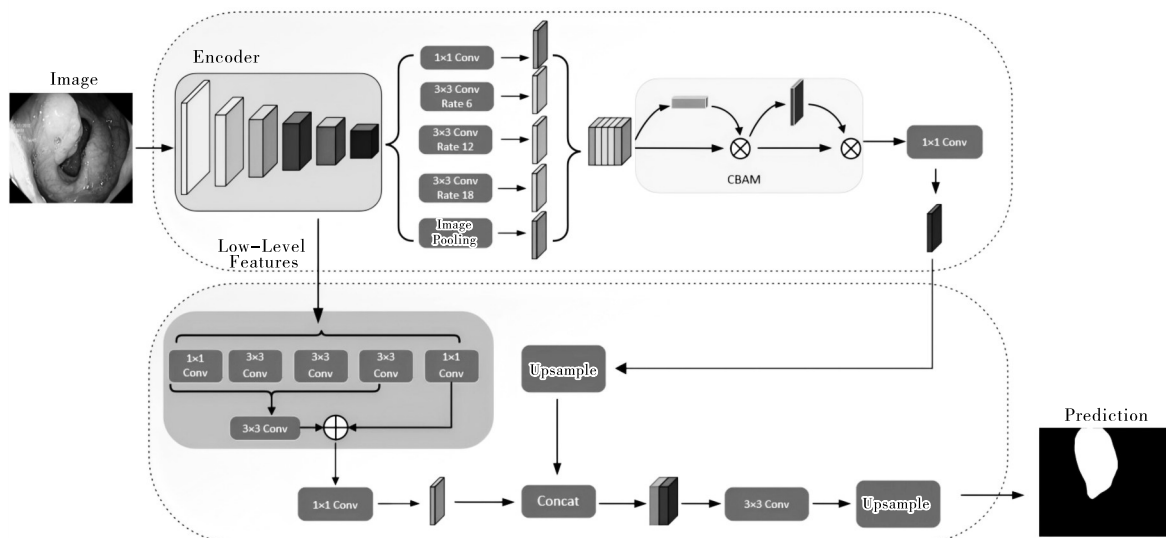


图3 改进DeepLabv3+模型

为解决息肉的复杂边缘检测以及周围正常组织背景分割等问题,文章在初始DeepLabv3+模型的基础上,使用ResNet101、CBAM注意力机制以及特征增强进行模型改进。首先利用ResNet101代替Xception模型

作为主干网络进行息肉的特征提取,得到不同层次的特征。利用残差连接不仅可以更好地提取图像中息肉的细节信息,还可以加速模型收敛。输入图片,经过主干网络后可以得到一个高层次特征和低层次特征。ASPP模块接受高层次特征并对其进行处理,由于不同的采样比例可以获得不同的感受野,因此可以在多个尺度上捕获目标特征以及更全面的上下文信息。然后将获得的特征通过CBAM注意力机制进行通道和空间上的处理,可以获得图像的重要特征并抑制不重要的背景信息,以区分息肉和周围正常组织,最后通过一个卷积层即获得编码器部分的输出。解码器部分接受两个输入,一个为经过ResNet101所得到的低层次特征,通过特征强化部分来提高图像的对比度和息肉的边缘信息,使图像中息肉边缘更加明显,然后输入该特征与另一个,即经过上采样的编码器部分的特征,进行融合,随后通过 3×3 的卷积并进行上采样恢复至原来的宽和高,其改进模型如图3所示。

2.2 主干网络构建

由于结肠息肉有大有小,小而圆滑的息肉、大而不规则的息肉都可能存在,尺寸上的多样性使模型在准确分割上存在一定难度。而使用ResNet作为主干网络进行特征提取对于分割大小不一的结肠息肉具有显著优势。第一,其残差连接机制允许信息直接传递,这有助于更深层网络对低层次特征进行学习。第二,ResNet能够更好地提取层级特征,这种多层特征提取的能力对于捕获不同大小息肉的特征至关重要。第三,利用残差网络在解决梯度爆炸问题上的优势还可以维持网络性能,保证其不退化,且更易进行网络优化,加快模型收敛,从而提高息肉分割的精度^[15]。

因此,本模型将初始模型中的Xception网络替换为ResNet101。输入的息肉图片首先通过 7×7 的卷积进行初步的特征提取和降维,然后通过多个堆叠的残差模块进行多层次特征学习。不同深度的残差模块可以学习不同级别的特征,浅层特征用于捕获小尺寸息肉的细节,深层特征则更适合提取大尺寸息肉的整体结构特征,最后通过全局平均池化层和全连接层将特征向量映射至最终输出,其网络结构如图4所示。

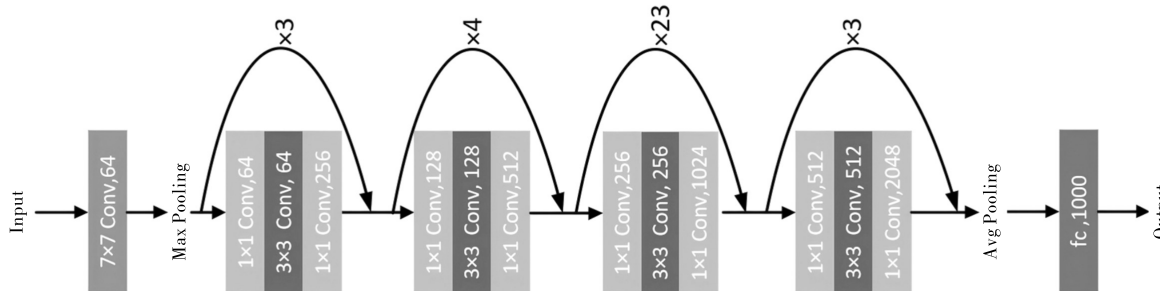


图4 ResNet101结构图

因此,使用ResNet替换Xception可以更有效地从浅层到深层逐步提取不同尺寸和层次的特征,ResNet能更好地适应不同尺寸和形状的息肉。相比之下,Xception需要使用复杂的正则化技术和学习率调度策略来控制模型的训练过程,其性能很大程度上取决于调整的超参数选择。对于不同任务,需要进行详细调优,这增加了模型使用的复杂度和难度。综上所述,该项改进有助于捕获结直肠息肉的各种特征,包括形状和尺寸等。值得注意的是,虽然ResNet网络使模型的深度在提取不同层次和尺寸的息肉特征上发挥了优势,但是在网络模型的宽度上,包括一些细微特征的捕捉,Xception仍旧比ResNet具有优势。

2.3 特征强化

观察息肉图片可知,结肠息肉与肠道周围的正常组织,如内壁等颜色相近,两者对比度较低,这将很难清晰完整地分割出息肉组织。因此,增强息肉的特征能够凸显息肉与周围组织的界限,有助于分割模型更精准地检测息肉边缘,进而实现更精确的分割结果。

膨胀卷积是一种特殊的卷积操作,通过引入膨胀率来扩大卷积核的感受野,有助于更好地处理低层次特征^[16]。低层次特征中通常包含图像的局部信息,其与全局或更广泛的上下文信息同等重要。除此之外,膨胀卷积可以使卷积核有更大的感受野,因此能够涵盖更大范围的输入图像区域,有助于捕获更广泛的上下文信息,使低层次特征更具语义。这样处理能够突出息肉的形状和纹理,更清晰完整地分割息肉与周围

组织的边界。因此,对低层次特征进行强化在息肉分割中有一定的作用。

受Tomar等人^[16]和HDCFE-Net分割模型^[17]的启发,本模型对通过主干网络ResNet101后所获得的低层次特征进行强化,输入特征分别通过 1×1 卷积层和3个 3×3 膨胀卷积层,将所得结果进行拼接并通过通道注意力机制对不同通道的特征响应动态增强或减弱,区分息肉与周围组织的不同特征。将该特征再次进行特征提取并相加,随后利用空间注意力作用,即根据不同区域的关注度,加强感兴趣区域的特征,提高息肉区域与周围组织的对比度,使特征更易被分辨。特征强化结构如图5所示。

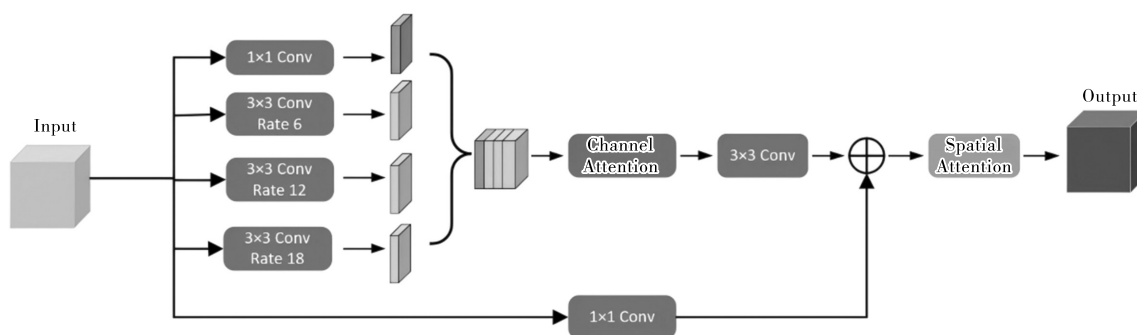


图5 特征强化模块

2.4 混合注意力机制

2018年,混合注意力机制CBAM^[18]被提出并被广泛应用于计算机视觉领域。该机制接收特征图后,沿着通道和空间两个不同维度和顺序进行推断,帮助模型关注图像中的重要区域或特征。值得注意的是,CBAM的内部没有大量的卷积结构、池化层以及特征融合等操作,这样就避免了大量计算以及资源的占用,使该注意力机制本身的轻量化程度较高。

图片经过主干网络所得的高层次特征在通过金字塔池化之后利用CBAM进行信息处理。轻量级卷积注意力分为两部分,即通道注意力机制(CAM)和空间注意力机制(SAM)。其中,通道注意力机制可以将输入特征图进行最大池化和平均池化,得到两个不同维度的特征,然后通过卷积进行降维升维,将特征图叠加和归一化后,使权重乘以输入特征图,可以得到各个通道的重要性^[18]。其计算公式如下

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

然后将所得结果送入空间注意力部分进行处理,形成区域的重要性,从而适应不同的数据分布。空间注意力部分计算公式如下

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \quad (2)$$

因此,CBAM使网络更加专注于边缘特征的学习。这有助于提高模型对息肉边缘的捕获和表征能力,尤其在边缘不清的情况下。CBAM结构如图6所示。

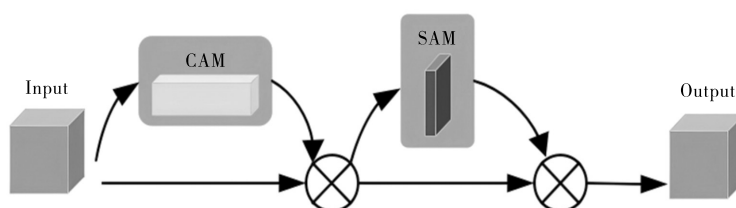


图6 混合注意力机制结构图

由于主干网络ResNet101的参数量大,在考虑控制参数量的同时还可以提高分割精度,轻量级的CBAM注意力机制是较为合适的选择,不仅可以控制使用注意力机制所带来的参数量,也提高了模型的性能,还不会给模型增加太多负担。这样不仅能够关注边缘特征的有效通道,提高对息肉边缘特征的代表和识别的准确性,而且能够增强局部区域的特征表示,特别是对于息肉边缘的细微特征,网络能够更准确地定位以及分割边缘。这有助于从全局角度捕获息肉边缘特征,提高对整体图像中边缘的感知能力,使得网络更全面地

理解图像的结构和边缘信息。

3 实验结果与分析讨论

3.1 实验环境与数据集

实验环境基于深度学习 PyTorch 框架,操作系统为 64 位 Win11,处理器为 AMD EPYC 7543,显卡为 24GB 显存的 NVIDIA RTX 3090,软件环境为 Pycharm、Pytorch1.10、Python3.8、CUDA11.3。

实验采用的数据集为 Kvasir-SEG 数据集和 CVC-ClinicDB 数据集。其中,Kvasir-SEG 数据集由挪威奥斯陆大学医学中心提供^[19],是 Kvasir 数据集的一个分割版本,其包含 1000 张来自内窥镜检查的胃肠道图像,主要用于胃肠道疾病的图像分割任务,每个图像都有对应的分割掩码,由一位或多位医学专家仔细注释,其中包含的图像分辨率从 332×487 像素到 1920×1072 像素不等。CVC-ClinicDB 数据集是 MICCAI 2015 结肠镜检查视频自动息肉检测挑战赛训练阶段使用的官方数据库,包含 612 张原始图像与对应的分割掩码,其中,原始图像均是从结肠镜检查视频中提取的帧^[20]。

由于两个数据集并没有划分训练集与验证集,所以本实验按照 9:1 的比例对此进行划分,同时在两个数据集上测试模型的性能。

3.2 评价标准与损失函数

为了验证针对 DeepLabv3+ 的改进是否有效,文章采用平均交并比(mIoU)、类别平均像素准确率(mPA)、准确率(Accuracy)这三个在医学图像分割中常用的性能指标对实验结果进行评估,其计算公式如下

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (3)$$

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

其中, $k+1$ 表示类别数, p_{ii} 表示正确分类的像素数量, p_{ij} 表示 i 类被预测为 j 类的像素数, p_{ji} 表示 j 类被预测为 i 类的像素数。TP 表示模型正确预测为正类的样本数,FP 为模型错误预测为正类的样本数,FN 为模型错误预测为负类的样本数,TN 表示模型正确预测为负类的样本数^[23]。

mIoU 表示模型对每一类预测结果与真实值的交集与并集的比值,求和再平均的结果,即对每个类别的 IoU 进行平均后得到的值。mIoU 的取值范围在 0~1 之间,值越高表示模型对每个类别的分割性能越好。它是广泛应用于图像分割任务中的重要评价指标之一,能够综合考虑模型在各个类别上的性能,并给出整体评价。mPA 是累加每个类被正确分类像素数的比例后求平均所得的结果,是像素级别的准确度的平均值,表示模型正确分类的像素数占总像素数的比例。Accuracy 表示模型正确分类的样本数占总样本数的比例,准确度的取值范围在 0~1 之间,数值越接近 1 表示模型性能越好,通常用于二分类或多分类任务。

损失函数是评价模型的预测值和真实值不一样的程度。文章使用的损失函数是 Dice Loss 与 Focal Loss 的组合^[21]。Dice Loss 由 Dice 系数而得名,Dice 系数是一种用于评估两个样本相似性的度量函数。Dice Loss 常用于语义分割问题中,其计算公式如下

$$Dice\ Loss = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \quad (6)$$

其中, X 表示真实分割图像的像素标签, Y 表示模型预测分割图像的像素类别, $|X|$ 表示 X 中元素的个数, $|Y|$ 表示 Y 中元素的个数, $|X \cap Y|$ 表示 X 与 Y 之间交集元素的个数。Dice Loss 在处理图像分割任务时的优点之一是对不平衡类别的鲁棒性较强,在训练模型时使用 Dice Loss 可以优化模型的预测结果,使其更加精准地匹配实际分割标签。

Focal Loss 是一种用于解决类别不平衡问题的损失函数,它在平衡交叉熵损失函数的基础上,增加一个调节因子降低易分类样本权重,聚焦于困难样本的训练。其定义如下

$$Focal\ Loss(p_i) = -\alpha_i (1 - p_i)^\gamma \log(p_i) \quad (7)$$

其中, p_i 表示模型对样本属于正类的预测概率, α 是平衡因子, 用于调整正负样本之间的权重, γ 是可调整指数参数。

将 Dice Loss 和 Focal Loss 组合使用可以在图像分割任务中取得一些优势。例如, 在应对类别不平衡的情况时, Focal Loss 专注于减轻容易分类的样本对损失的贡献, 对于类别不平衡的问题有良好的效果, 结合 Dice Loss 可以进一步强调对难以分类的样本的学习, 提高对正类别上的关注度。而在针对边缘细节保留的问题上, Dice Loss 在边缘分割方面表现出色, 有助于保留目标的边缘细节, 当与 Focal Loss 结合时, 可以同时兼顾模型对边缘的敏感性和对类别不平衡的鲁棒性, 提高模型的全局性能。

3.3 实验结果与分析

为了验证改进的 DeepLabv3+模型, 将其主干网络改为残差网络是有效的, 需要通过实验对改进模型的分割效果进行展示。模型在两个数据集上的结果如表1所示。

表1 主干网络对比

数据集	模块	mIoU	mPA	Accuracy
Kvasir-SEG	Baseline+ResNet101	0.7645	0.8614	0.9305
	Baseline+Xception	0.7289	0.8527	0.9099
CVC-ClinicDB	Baseline+ResNet101	0.8661	0.9323	0.9723
	Baseline+Xception	0.8648	0.9303	0.9721

由表1可知, 将 DeepLabv3+ 的主干网络从 Xception 替换为 ResNet101 进行特征提取是有效的, 将主干网络进行替换后在 Kvasir-SEG 数据集上指标提高较多, mIoU、mPA、Accuracy 分别提高了 3.56%、0.8%、2.06%, 而在 CVC-ClinicDB 数据集上虽然未有大幅提升, 但指标也有少量的提高。

为了验证特征强化模块与混合注意力机制对息肉分割是否有效, 通过消融实验进行对比验证, 其效果如表2所示。

表2 消融实验对比

数据集	模块	mIoU	mPA	Accuracy
Kvasir-SEG	Baseline(ResNet101)	0.7645	0.8614	0.9305
	Baseline+CBAM	0.7858	0.8623	0.9389
	Baseline+FEM(特征强化)	0.8033	0.8707	0.9451
	Baseline+CBAM+FEM	0.8219	0.8830	0.9510
CVC-ClinicDB	Baseline(ResNet101)	0.8661	0.9323	0.9723
	Baseline+CBAM	0.8963	0.9530	0.9790
	Baseline+FEM(特征强化)	0.8983	0.9510	0.9796
	Baseline+CBAM+FEM	0.9040	0.9551	0.9808

分析消融实验结果可知, 在主干网络是 ResNet101 的基础上, 加入特征强化模块与混合注意力机制对于息肉的分割是有用的。在 Kvasir-SEG 数据集上, 单独使用特征强化模块或混合注意力机制三个评价指标都有一定的提升, 组合使用则大幅提升了评价指标, mIoU、mPA、Accuracy 分别提高了 5.74%、2.16%、2.05%。在 CVC-ClinicDB 数据集上, 单独使用特征强化模块或混合注意力机制三个评价指标都有比较明显的提升, 而组合使用则在单独使用特征强化模块或混合注意力机制的基础上有一些小幅提升, mIoU、mPA、Accuracy 对比基础模型分别提高了 3.79%、2.28%、0.85%。

针对两个不同的数据集, 模型的实验数据差异较大, 原因分析如下: (1) CVC-ClinicDB 中的图片分辨率比 Kvasir-SEG 的分辨率高, 较高的分辨率可以提供更多的图像细节。(2) 仔细对比两个数据集中的图片, 可以发现 Kvasir-SEG 中的图片背景比 CVC-ClinicDB 中的背景简单。但是两个数据集均具有权威性。将本模型与其他模型进行实验对比, 结果如表3所示。由表3的实验数据可以看出本模型改进后的结果优于其他模型。

表3 模型对比

数据集	方法	mIoU	mPA	Accuracy
Kvasir-SEG	DeepLabv3+(改进)	0.8219	0.8830	0.9510
	U-Net	0.7562	0.8359	0.9422
CVC-ClinicDB	DeepLabv3+(改进)	0.9040	0.9551	0.9808
	U-Net	0.7730	0.9344	0.9690

此外,对于文章中改进模型的复杂度是否在可接受范围内,文章使用参数量(parameters)、浮点运算次数(FLOPs)以及每秒传帧数(FPS)这三个评价指标进行评估^[22]。参数量是指神经网络中可学习的权重和偏置的数量。参数量的多少直接决定了模型的复杂度和容量,其描述的是模型的大小。浮点运算次数是衡量计算速度的一个常用指标,表示计算机系统每秒执行的浮点运算次数。每秒传帧数是图像领域中的定义,是指画面每秒传输帧数。本模型的算法复杂度结果如表4所示。

表4 算法复杂度结果对比

模块	方法	参数量(M)	浮点运算次数(GMac)	每秒传帧数
DeepLabv3+	Xception	32.9	43.41	
DeepLabv3+	ResNet101	32.1	42.86	

分析表4的算法复杂度结果可知,改进的DeepLabv3+模型在参数量、浮点运算次数以及每秒传帧数上对比初始DeepLabv3+均有降低。与其他分割网络进行比较后可知,DeepLabv3+模型修改后的复杂程度在可控范围内,其针对初始DeepLabv3+模型的改进是有效的。

将文章提出的改进后的DeepLabv3+模型分别在Kvasir-SEG数据集以及CVC-ClinicDB数据集上进行训练,并将所得分割结果进行比较,如图7所示。其中,图7(a)表示分割原图,图7(b)表示标签图,图7(c)表示初始模型进行分割所得结果,图7(d)表示改进模型分割结果。根据分割结果可以看出,改进模型比原模型在边缘分割能力上有一定的提升,包括周围组织的辨别上也有一定进步。除此之外,模型对于大小不一的息肉进行分割也有一定的适应性。

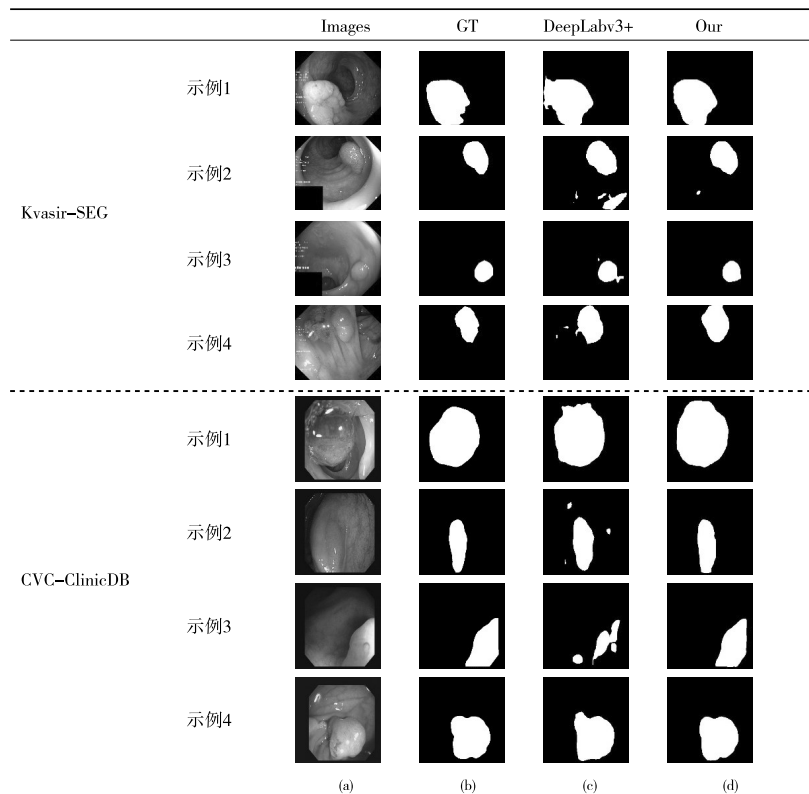


图7 结肠息肉分割对比图

4 结论

文章在DeepLabv3+分割模型的基础上进行改进,将ResNet101作为特征提取网络,更有效地提取不同尺寸和层次的特征,并加入混合注意力机制和特征强化以帮助息肉分割。混合注意力机制不仅能够关注边缘特征的有效通道,提高对息肉边缘特征的特征和识别准确性,而且还能够增强局部区域的特征表示。对特征进行强化,增大感受野,使卷积核涵盖更大范围的输入图像区域,有助于捕获更广泛的上下文信息,能够对息肉的形状和纹理加以关注。同时将Dice损失函数和Focal损失函数进行结合,更好地训练模型。将改进模型在两个数据集上进行测试,实验结果表明改进DeepLabv3+模型在mIoU等指标上有明显的提升,即改进DeepLabv3+模型在息肉分割的效果上有一定的提高。模型复杂程度还有待改进,例如,在考虑减轻模型复杂度的基础上仍保证息肉分割能力。为了实现这一目标,下一步可以考虑采用一些方法来降低模型的复杂度,如使用轻量级卷积神经网络、剪枝或量化等技术。这些方法能够在不影响模型性能的前提下减少计算量和存储空间需求。

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]. *A Cancer Journal for Clinicians*, 2021, 71(03):209–249.
- [2] LUISA F S, LUIS B C, ARTZAI P, et al. Deep Learning to Find Colorectal Polyps in Colonoscopy: A Systematic Literature Review[J]. *Artificial Intelligence in Medicine*, 2020, 108:101923.
- [3] OLAF R, PHILIPP F, THOMAS B. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//IEEE Conference on Medical Image Computer and Computer-assisted Intervention, Boston, Massachusetts, USA, 2015, 18(03):234–241.
- [4] CHEN L C, GEORGE P, IASONAS K, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[C]//International Conference on Learning Representations, 2015, arXiv:1412.7062.
- [5] CHEN L C, GEORGE P, IASONAS K, 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*, 2017, 40(04):834–848.
- [6] CHEN L C, GEORGE P, FLORIAN S, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J]. *Computer Vision and Pattern Recognition*, 2017, arXiv:1706.05587.
- [7] DONG B, WANG W H, FAN D P, et al. Polyp-PVT: Polyp Segmentation with Pyramid Vision Transformers[J]. *Image and Video Processing*, 2023, arXiv:2108.06932.
- [8] FAN D P, JI G P, ZHOU T, et al. PraNet: Parallel Reverse Attention Network for Polyp Segmentation[J]. *Medical Image Computing and Computer Assisted Intervention*, 2020, 12266:263–273.
- [9] ANGE L, GUAN S Y, HANSEOK K, et al. CaraNet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects[J]. *Medical Imaging: Image Processing*, 2022, 12032:81–89.
- [10] CHEN J N, LU Y Y, YU Q H, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation[J]. *Computer Vision and Pattern Recognition*, 2021, arXiv:2102.04306.
- [11] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions[C]. *IEEE Conference on Computer on Computer Vision and Pattern Recognition*, 2017, 1251–1258.
- [12] CHEN L C, ZHU Y K, GEORGE P, et al. Encoder-decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//European Conference on Computer Vision(ECCV), 2018, 801–818.
- [13] LIU T T, YE X J, HU K, et al. Polyp Segmentation with Distraction Separation[J]. *Expert Systems with Applications*, 2023, 228:120434.
- [14] HE K M, ZHANG X Y. Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016:770–778.
- [15] SRENG S, MANEERAT N, HAMAMOTO K, et al. Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images[J]. *Computerized Medical Imaging and Graphics*, 2022, 98:102072.
- [16] TOMAR N K, JHA D, BAGCI U, et al. TGANet: Text-guided Attention for Improved Polyp Segmentation[J]. *Medical Image Computing and Computer Assisted Intervention(MICCAI)*, 2023:151–160.

- [17] DEBESH J, PIA H S, MICHAEL A R, et al. Hybrid Dilated Convolution Guided Feature Filtering and Enhancement Strategy for Hyperspectral Image Classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19:5508105.
- [18] WOO S, PARK L, LEE J Y, et al. CBAM: Convolutional Block Attention Module[C]//European Conference on Computer Vision (ECCV), 2018, 3-19.
- [19] DEBESH J, PIA H S, MICHAEL A R, et al. Kvasir-SEG: A Segmented Polyp Dataset[J]. *MultiMedia Modeling*, 2020:451-462.
- [20] 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.
- [21] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. *IEEE International Conference on Computer Vision (ICCV)*, 2017:2980-2988.
- [22] FENG R W, WANG B W, CHEN T T, et al. SSN: A Stair-shape Network for Real-time Polyp Segmentation in Colonoscopy Images[J]. *IEEE 17th International Symposium on Biomedical Imaging*, Iowa City, IA, USA, 2020.
- [23] MICHAEL Y, EVIS S, LEONARDO R, et al. Focus U-Net: A Novel Dual Attention-gated CNN for Polyp Segmentation During Colonoscopy[J]. *Computers in Biology and Medicine*, 2021, 137:104845.
- [24] MARIN B, IRENA G, MARIJA H, et al. Training on Polar Image Transformations Improves Biomedical Image Segmentation[J]. *IEEE Access*, 2021, 9:133365-133375.
- [25] WANG M, AN X W, PEI Z C, et al. An Efficient Multi-task Synergetic Network for Polyp Segmentation and Classification[J]. *IEEE Journal of Biomedical and Health Informatics*, 2023:1-12.
- [26] SAMIR J, ROHAN A, ANUBHAV G, et al. ColnNet: A Convolution-involution Network with a Novel Statistical Attention for Automatic Polyp Segmentation[J]. *IEEE Transactions on Medical Imaging*, 2023:1.
- [27] LIU Y, YANG Y, JIANG Y Q, et al. Multi-view Orientational Attention Network Combining Point-based Affinity for Polyp Segmentation[J]. *Expert Systems with Applications*, 2024, 249:123663.
- [28] YIN X H, ZENG J, HOU T X, et al. RSAFormer: A Method of Polyp Segmentation with Region Self-attention Transformer[J]. *Computers in Biology and Medicine*, 2024, 172:108268.

Polyp Segmentation Method Combining Multiscale Information and Feature Enhancement

WANG Qin-yun, YU Li*

(School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui, 232001, China)

Abstract: Computer-aided diagnosis technology plays a very important role in colon polyp image segmentation. In view of the problems of colon polyp segmentation such as complicated edge detection and segmentation and low contrast between polyps and normal tissue, an improved segmentation model based on DeepLabv3+ is proposed. This model uses the residual network as the backbone network to extract feature information, introduces a feature enhancement module to process low-level features, and adds a hybrid attention mechanism to capture key information for high-level features. The average intersection and union ratios of this model reached 82.19% and 90.40% respectively in experiments on the Kvasir-SEG data set and the CVC-ClinicDB data set. Experimental results show that the polyp segmentation effect of the improved model is better than that of the initial DeepLabv3+ model, and there has been a certain improvement in the accuracy of polyp segmentation and edge segmentation.

Keywords: DeepLabv3+; Polyp segmentation; Convolutional block attention module; Feature enhancement; ResNet