

基于 MSBRAU-Net++ 的多尺度息肉分割方法^{*}

付金丽 褚丽莉 李 波

(辽宁工业大学电子与信息工程学院 锦州 121001)

摘 要: 针对息肉分割中普遍存在的共现现象、边界模糊和分割不足等挑战,提出了一种新型多尺度注意力融合网络 MSBRAU-Net++。采用多尺度门控注意力融合,将多尺度与注意力形成交互结构处理上下文信息,增强前景特征响应,抑制背景干扰,显著改善息肉与相似组织的区分能力;利用混合空间通道模块,通过深层特征校准与局部细节恢复,解决边界模糊问题,提升边缘分割精度;设计全新的多路径特征聚合块,融合低阶细节与高阶语义特征,避免信息损失,确保分割结果的完整性。在 Kvasir-SEG 和 CVC-ClinicDB 数据集上对 MSBRAU-Net++ 进行评估,IoU 分别为 84.65% 和 88.87%,相似性系数 DSC 分别为 90.63% 和 91.99%。实验结果表明,MSBRAU-Net++ 优于其他对比模型,能准确分割图像,尤其对复杂边界和小息肉区域的分割效果显著。

关键词: 医学图像分割;息肉分割;边界模糊;多尺度融合;特征提取

中图分类号: TP391;TN911 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Multi-scale polyp segmentation method based on MSBRAU-Net++

Fu Jinli Chu Lili Li Bo

(School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China)

Abstract: To address the challenges prevalent in polyp segmentation, such as co-occurrence phenomena, boundary ambiguity, and under-segmentation, a novel multi-scale attention fusion network, MSBRAU-Net++, is proposed. It adopts multi-scale gated attention fusion to create an interactive structure between multi-scale features and attention, processing contextual information to enhance foreground feature responses and suppress background interference, thereby significantly improving the ability to distinguish polyps from similar tissues. By utilizing a hybrid spatial channel module, it addresses the issue of boundary ambiguity through deep feature calibration and local detail recovery, thereby enhancing the precision of edge segmentation. A novel multi-path feature aggregation block is designed to fuse low-level details with high-level semantic features, preventing information loss and ensuring the integrity of the segmentation results. MSBRAU-Net++ was evaluated on the Kvasir-SEG and CVC-ClinicDB datasets, achieving IoU scores of 84.65% and 88.87%, and DSC scores of 90.63% and 91.99%, respectively. The experimental results demonstrate that MSBRAU-Net++ outperforms other comparative models and can accurately segment images, showing particularly significant results in segmenting regions with complex boundaries and small polyps.

Keywords: medical image segmentation; polyp segmentation; boundary ambiguity; multi-scale fusion; feature extraction

0 引 言

结直肠癌(colorectal cancer, CRC)是全球第三大癌症,也是男性和女性癌症相关死亡的第二大原因^[1]。CRC 通常始于结肠直肠粘膜中的良性息肉,腺瘤性息肉占有结肠直肠息肉的约 70%,并被认为是 CRC 的前体。并非所有的良性息肉都会进展为 CRC,小于 10 mm 的息肉往往被医生忽略。虽然结肠镜检查是有效的诊断手段,但由于肠

道皱褶和视野受限导致的可见性差,加上操作者的依赖性,小息肉可能会被遗漏。计算机辅助分割方法能够减少漏诊风险,特别是对于小的或被肠道皱褶包裹的息肉,其精准的分割能力对提高诊断准确性至关重要^[2]。

相较于传统的局部阈值、区域生长、边缘检测和基于聚类等方法^[3],深度学习技术的发展极大地加快了息肉分割的研究。U-Net^[4]在医学图像分割领域成功确立了编码器-解码器架构的主导地位,从而开发出多种增强方法,包括

A-Dense UNet^[5]、N-Unet^[6]、RA-Unet^[7]和TransResUNet^[8]。注意力机制的集成^[9-12]进一步提高了语义分割中的边界精度。其他方法^[13-15]也证明了通过利用多尺度特征聚合的显著改进。最近,基于变换器的主干,例如Swin transformer^[16]和Pvt v2^[17],在增强息肉分割模型的泛化能力方面取得了显著的成功。

然而,现有息肉分割架构仍存在3个关键局限性:1)传统的跳跃连接机制未能充分补偿编码器与解码器之间的语义差异,难以有效传递图像特征,限制了模型在共现现象下的分割精度提升。Lei等^[18]提出Condseg框架,通过一致性强化、语义信息解耦、对比驱动特征聚合和尺寸感知解码器,有效解决了医学图像分割中的共现问题。2)在处理息肉的形态不规则、边界模糊等问题时,难以有效融合局部细节与全局语义信息,导致分割边界不准确;Yue等^[19]提出BRNet,通过边界生成模块融合细节与概念生成边界图,并利用边界引导模块结合边界线索与高层预测逐层细化特征,最终实现清晰边界的息肉区域精确定位。由于缺少全局语义的分析,Lan等^[20]提出BRAU-Net++,以双层路由注意力为核心构建模块,并设计了U形编码器-解码器结构。该结构的编码器和解码器均采用层次化设计,旨在降低计算复杂度的同时学习全局语义信息。然而,该方法在分割的准确性与速度方面仍有不足。Zhang等^[21]在BRAU-Net++基础上提出MCBTNet在跳跃连接上引入了新颖的“频率-通道-空间”三维度注意力机制,通过最大化多尺度特征的利用来提升分割的准确性与速度。3)在特征提取方面,多数模型依赖固定的感受野缺乏对不同息肉的自适应感知能力,导致对小尺度息肉的分割敏感性不足。Lan等^[22]提出MSLAU-Net,一种混合CNN-Transformer架构,通过多尺度线性注意力和轻量级特征聚合机制,显著提升了小尺度息肉分割的精度与效率。张攀峰等^[23]提出一种融合局部和全局特征的息肉分割模型。该模型创新性地采用卷积神经网络和Transformer作为并行编码器,以兼顾图像的多尺度局部细节与全局语义特征,实现了对息肉区域的高精度分割。

针对模型面临的共现现象、模糊边界及特征提取不足三大挑战,本文以改进的BRAU-Net++作为主干,提出了MSBRAU-Net++模型。本文的主要贡献包括:

1)在跳跃连接处引入多尺度门控注意力融合(multi-scale gated attention fusion,MSGAF)通过并行设置不同尺度的卷积和注意力分支,分别处理来自编码器的多级特征映射。各分支输出通过可学习的门控权重进行动态融合,既补偿了采样导致的空间信息损失,又增强了跨尺度特征的全局交互能力有效解决息肉分割中的共现现象问题。

2)在Bottleneck处引入混合空间通道模块(hybrid spatial channel module, HSCM)通过空间双路径可分离(spatial dual-path separable, SDPS)和通道感知上下文(channel aware context, CAC)的串联作用,对编码器的预

测图像进行边界定位,然后通过比对预测图像与输入图像的边界一致性进行边界校验,最终将优化后的边界特征送入解码器进行后续处理。

3)针对原特征提取模块Biformer在全局特征建模方面的局限性——其特有的注意力机制难以全面捕捉长程依赖关系,提出多路径特征聚合块(multi-path feature aggregation block, MPFAB)。采用多路径并行架构协同融合全局上下文与局部细节特征,并且引入动态参数控制机制自适应调整各路径特征权重,增强模型对复杂场景的适应能力。通过分层特征处理策略,在扩大感受野的同时保持较高的计算效率,实现全局与局部特征的优势互补。

1 息肉图像基本分割框架

1.1 MSBRAU-Net++网络结构

BRAU-Net++是一种用于医学图像分割的U型混合CNN-Transformer网络结构,其整体架构如图1(a)所示。该模型结合了CNN与Transformer的优势,并引入动态稀疏注意力机制与通道-空间注意力机制,在维持高分割精度的同时有效降低了计算复杂度。编码器采用分层结构,由多个BiFormer块组成。其中,BiFormer块以双层路由注意力机制(bi-level routing attention, BRA)为核心,能够高效建模长距离依赖并减少计算量;Bottleneck部分由多个BiFormer块组成,保持特征图的分辨率与维度不变,进一步提取深层特征。解码器与编码器结构对称,也包含多个BiFormer块。此外,解码器通过引入“跳跃连接通道-空间注意力(skip connection channel-spatial attention, SCCSA)”模块,与编码器进行特征融合,从而有效补偿因下采样导致的空间信息损失,并增强跨维度交互能力。

MSBRAU-Net++网络模型结构如图1(b)所示。对于编码器,给定一幅医学图像,将输入图像划分为 4×4 的补丁块,每个块的大小是原始图像的 $1/4$,通过小块嵌入将每个小块的特征维度映射到任意指定维度(记为C)。变换后的补丁块通过多个MPFAB和Patch Merging来生成成分层特征表示。具体而言,Patch Merging通过合并相邻小块来缩减特征图的空间分辨率和扩展其通道维度,而MPFAB用于学习特征表示。然后经过Bottleneck层进一步处理,分辨率和维数保持不变,通过MPFAB和HSCM来进一步增强多尺度特征提取,并为分割结果输出做好准备。

为确保特征有效传递且避免信息丢失,在跳跃连接处引入多尺度门控注意力融合模块(MSGAF),依据其层级传递特征大小分别为 $1/4$ 、 $1/8$ 和 $1/16$ 。通过MSGAF的多尺度特征融合机制,能够有效补偿下采样过程导致的空间信息损失,同时增强全局维度间的交互作用。

与编码器类似,解码器也基于MPFAB块构建,用Patch Expanding来对解码器中提取的深度特征进行上采样,主要用于将特征图整形为更高分辨率的特征图,即分辨

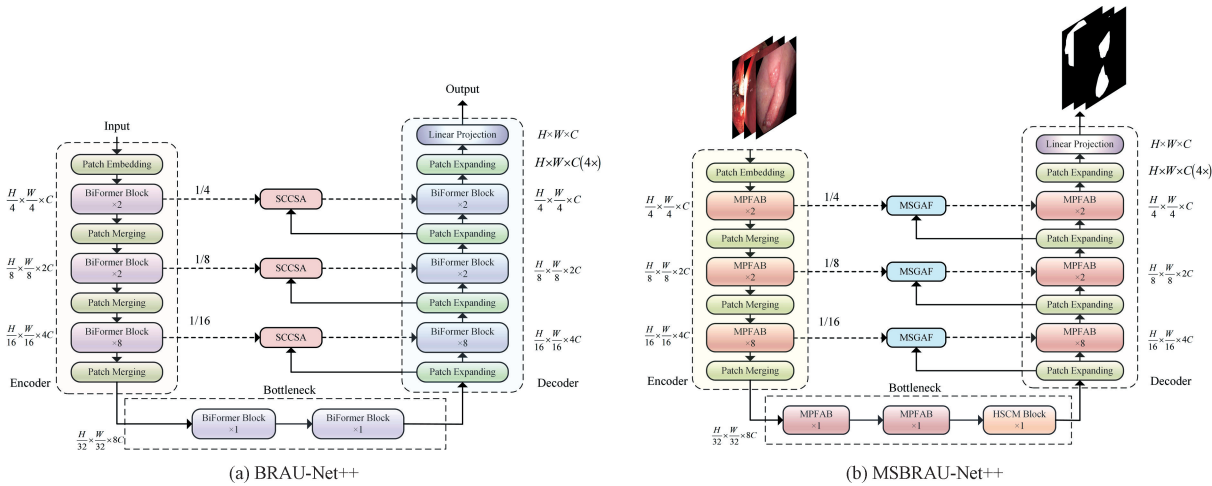


图 1 BRAU-Net++ 和 MSBRAU-Net++ 网络模型结构
Fig. 1 BRAU-Net++ and MSBRAU-Net++ network model structure

率提高 2 倍,特征尺寸减半。最后一个 Patch Expanding 进行 4 倍上采样后经过 Linear Projection,输出分辨率为 $R^{H \times W \times C}$ 的特征图,用于预测像素级分割。

1.2 MSGAF 网络结构

在结肠镜检查图像中,较小的息肉经常同时出现多个,较大的息肉在其附近经常伴有较小的息肉。受文

献[22, 24-25]的启发,将多尺度深度可分离卷积和注意力机制相融合可以增强模型捕捉到更丰富的上下文特征,从而提出了 MSGAF,如图 2 所示。MSGAF 模块旨在有效补偿编码器在下采样过程中丢失的空间信息,通过增强解码器各层多尺度特征的全局维度交互,在生成输出掩模的同时,显著恢复图像的细粒度细节。

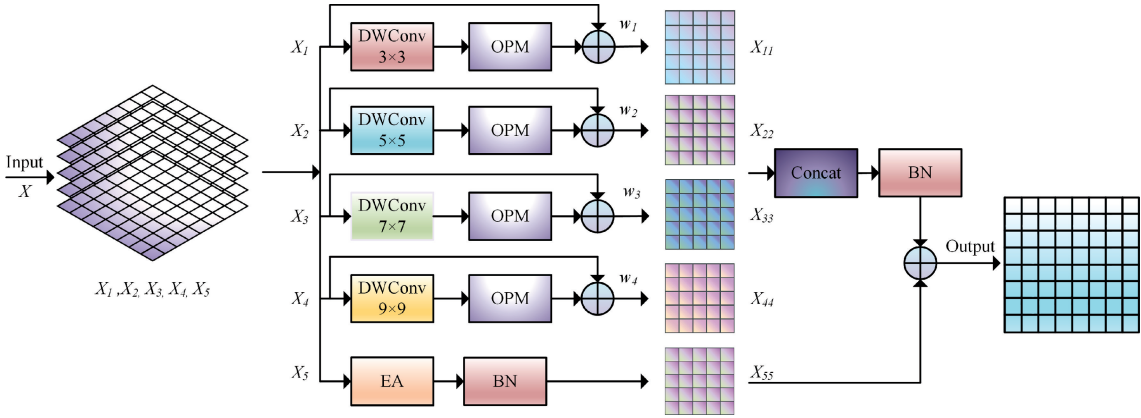


图 2 MSGAF 网络模型结构
Fig. 2 MSGAF network model structure

1) 多尺度特征提取

首先将输入图片沿着通道维度分成 5 个部分。在重塑和分裂之后,这些特征然后被馈送到 4 个平行的深度卷积分支及 EA 中,这些分支设计分别使用 $3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9$ 的深度可分离卷积和 EA 注意力探索多尺度表示。较小的卷积核擅长检测医学图像中的细粒度细节,而较大的卷积核则擅长更有效地捕获更广泛的结构,包括息肉的整体轮廓。随后,通过残差连接将来自不同尺度的卷积特征与原始输入特征进行整合,这个过程有利于更丰富的多尺度特征表示,可以表述如式(1)。然后将 X_{11}, X_{22}, X_{33} 和 X_{44} 进行 Concatenate 拼接,经过归一化之后与 X_{55} 进行相

加最后输出图片。

$$X_{ii} = w_i(OPM(\varphi_{k_i \times k_i}^{dwc}(X_i)) + X_i) \quad (1)$$

式中: i 为 $\{1, 2, 3, 4\}$, $k_i = 2i + 1$, $\varphi_{k_i \times k_i}^{dwc}$ 为 k_i 的核大小进行深度卷积。

2) OPM 正交化模块

正交池化模块(orthogonal pooling module, OPM)结构如图 3 所示。首先采用水平池化 x_h 和垂直池化 x_v 两种方法在两个方向上捕获轴向全局上下文。即通过对水平池化和垂直池化生成两个不同的轴矢量进行广播加法得到 F , OPM 可以有效地对感兴趣的矩形区域进行建模,然后,设计形状自校准函数 F_{ex} 来校准感兴趣区域,这样可以

使感兴趣区域更接近前景物体,形状自校准函数可以表述如式(2)。

$$F_{ex} = \sigma(\varphi_{k \times 1}(\phi(\varphi_{1 \times k}(F(x_v + x_h)))))) \quad (2)$$

式中: x_v 为垂直池化, x_h 为水平池化, F 为水平池化跟垂直池化相加之后的输出, φ 为大条核卷积, k 为条卷积的核大小, ϕ 为 BN 后接 Relu 函数, σ 为 Sigmoid 函数, F_{ex} 为自校准函数。

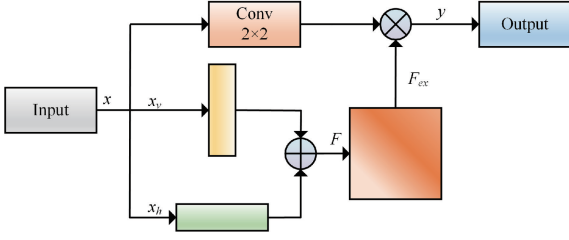


图 3 OPM 网络模型结构

Fig. 3 OPM network model structure

此外,将校准特征与输入特征进行融合。具体来说,使用 2×2 卷积来进一步提取输入特征的局部细节。通过相乘将校准后的特征加权到输入特征,最后得到输出。整个过程可以表示为式(3)。

$$y = F_{ex} \times \varphi_{2 \times 2}(x) \quad (3)$$

式中: x 为输入, $\varphi_{2 \times 2}$ 为 2×2 的卷积, y 为输出。

3)EA 高效注意力

高效注意力(efficient attention,EA)结构如图 4 所示。单个特征向量 $\mathbf{X} \in \mathbf{R}^{n \times d}$ 经过 3 个线性层,形成查询特征 $\mathbf{Q} \in \mathbf{R}^{n \times d_k}$, 键特征 $\mathbf{K} \in \mathbf{R}^{n \times d_k}$, 值特征 $\mathbf{V} \in \mathbf{R}^{n \times d_v}$, 使用这

些特征映射中的每一个作为对所有位置的加权,并通过加权求乘将所有位置的值特征聚合起来,形成全局上下文向量,高效注意机制的特征如式(4)。

$$E(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \rho_q(\mathbf{Q})(\rho_k(\mathbf{K})^T \mathbf{V}) \quad (4)$$

式中: ρ_q 和 ρ_k 分别为查询特征和键特征的归一化函数。

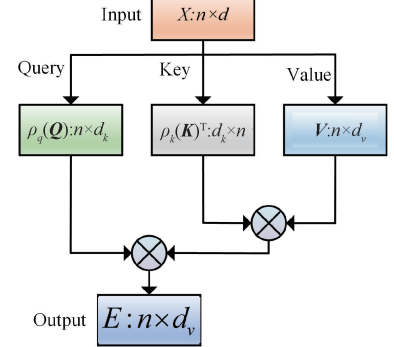


图 4 EA 网络模型结构

Fig. 4 EA network model structure

1.3 HSCM 网络模型结构

医学领域中,结肠息肉的边界模糊对诊断有着重要依据。受文献[26]的启发,本文利用通道-空间注意对 Bottleneck 和解码器的连接处进行了重新设计,提出 HSCM,其整体结构如图 5 所示。该模块主要分为 SDPS 和 CAC 两部分。通过 SDPS 和 CAC 的串联作用,HSCM 模块能够自适应地建模长距离空间依赖关系并动态校准通道间的相互依赖性,从而显著增强模型的特征表示能力,促进编码器与解码器之间更有效的信息交互。

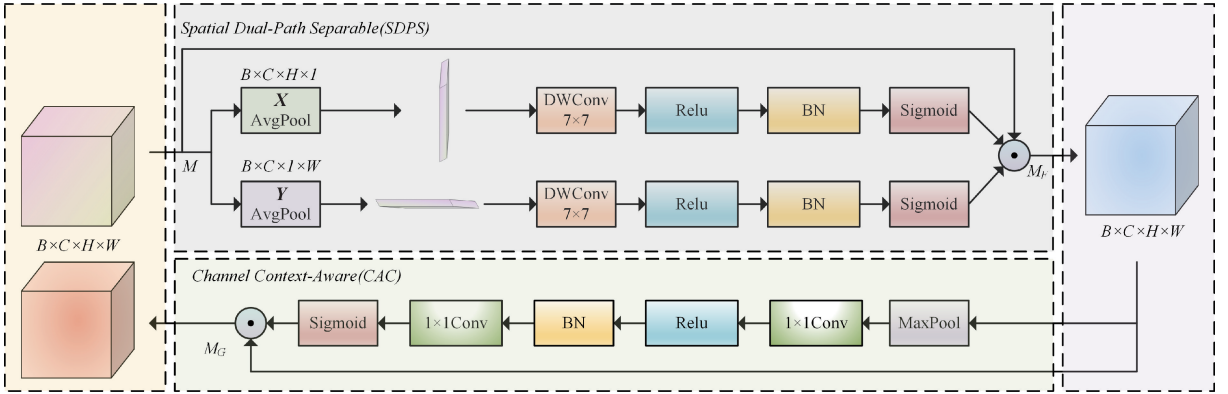


图 5 HSCM 网络模型结构

Fig. 5 HSCM network model structure

在 SDPS 部分,首先,将输入特征图分别沿着高度(H)和宽度(W)维度分解。具体而言,沿高度方向进行全局平均池化,得到一个表征垂直方向全局信息的张量 $\mathbf{X}$,沿宽度方向进行全局平均池化,得到一个表征水平方向全局信息的张量 $\mathbf{Y}$ 。接着,对这两个方向特征张量分别应用一个 7×7 的深度可分离卷积进行特征提取,以捕获各自方向上的长距离依赖关系。将卷积后的特征首先通过 Relu 激活

函数进行非线性化,再通过 Sigmoid 函数进行归一化,分别生成空间高度注意力权重图和空间宽度注意力权重图。最后,将这两个注意力权重图分别与原始输入特征图 $\mathbf{X}$ 进行逐元素相乘,得到经过空间方向注意力加权后的特征图,作为 SDPS 的输出。整个过程如式(5)。

$$M_F = \sigma(\phi(\int_{7 \times 7}^{duc} (\mathbf{X}^{AvgPool} \odot \mathbf{Y}^{AvgPool}))) \odot M \quad (5)$$

式中: M 为输入, M_F 为 SDPS 的输出, $\odot$ 为逐元素相乘。

在 CAC 部分,对 SDPS 的输出特征图进行最大平均池,将其压缩为一个表征全局通道信息的向量。首先,通过一个 1×1 卷积层进行降维。接着,应用 ReLU 激活函数,再通过一个 1×1 卷积恢复到原始通道数。最后,通过 Sigmoid 函数生成通道注意力权重向量。将生成的通道注意力权重向量与 SDPS 的输出特征图进行逐元素相乘,得到最终的 HSCM 模块输出。整个过程如式(6)。

$$M_G = \sigma(\varphi_{1 \times 1}(\phi(\varphi_{1 \times 1}(M_F^{MaxPool})))) \odot M_F \quad (6)$$

式中: M_G 为 HSCM 的输出。

1.4 MPFAB 网络模型结构

在 BRAU-Net++ 中采用 BiFormer 块提取特征,虽然通过 BRA 提升了计算效率,但在高分辨率或复杂场景下对图像的特征提取能力不足,且性能高度依赖于区域划分策略的合理性。在此基础上,本文引入 LSK 模块^[27]代替传统的 BRA,提出 MPFAB 整体结构如图 6 所示。旨在通过增强特征提取的局部感知能力与自适应调节机制,进一步降低计算复杂度,提升模型对复杂场景的特征提取能力。MPFAB 由一个 3×3 的卷积、2 个 LN 层、一个 LSK 块、动态控制参数(α 、 β 和 δ)、3 个残差连接和 MLP 组成。整个过程可表示为式(7)~(9)。

$$Z_{l-2} = \text{Conv}_{3 \times 3}(Z_{l-3}) + \alpha Z_{l-3} \quad (7)$$

$$Z_{l-1} = \text{LSK}(\text{LN}(Z_{l-2})) + \beta Z_{l-2} \quad (8)$$

$$Z_l = \text{MLP}(\text{LN}(Z_{l-1})) + \delta Z_{l-1} \quad (9)$$

式中: Z_{l-3} 为输入, Z_{l-2} 、 Z_{l-1} 和 Z_l 分别为卷积、LSK 和 MLP 的输出, α 、 β 和 δ 为动态参数。

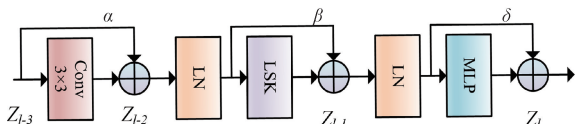


图 6 MPFAB 网络模型结构

Fig. 6 MPFAB network model structure

MPFAB 具备以下设计优势:1)借助 LSK 模块替代传统的 BRA 注意力机制,可以在获得更大感受野和更稳定训练的同时,显著提升计算效率;2)运用多路径,将全局信息与局部细节加以融合,进而实现达成长距离依赖关系的建模;3)通过动态参数控制,使模型更加灵活、高效和强大,特别适合处理多样化输入和复杂场景的任务。

LSK 通过一系列卷积核和空间选择机制,能够根据目标所处的复杂背景动态调整感受野大小。这种设计使模型在检测不同尺度、方向或遮挡程度的目标时,能更灵活地聚焦关键特征,减少背景干扰,LSK 网络结构如图 7 所示。具有多个分支和跳跃连接的复杂网络结构,由包含不同大小的卷积核(F_x 和 F_y)、池化层以及激活函数组成,旨在从输入数据中提取多尺度特征并进行有效处理,最终输出结果。整个过程如式(10)~(13)。

$$F = \sigma(\eta(F_x \oplus F_y)) \quad (10)$$

$$y_1 = F \times F_x \quad (11)$$

$$y_2 = F \times F_y \quad (12)$$

$$y = y_1 + y_2 \quad (13)$$

式中: F_x 为 1×1 和 7×7 的卷积, F_y 为 1×1 的卷积, $\oplus$ 为 Concatenate 拼接, η 为平均池化和最大池化, F 和 y 为输出。

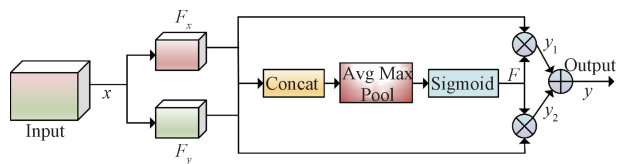


图 7 LSK 网络模型结构

Fig. 7 LSK network model structure

2 实验结果分析与讨论

2.1 数据集及预处理

为验证本文所提算法的有效性采用 Kvasir-SEG 数据集和 CVC-ClinicDB 数据集。Kvasir-SEG 数据集是胃息肉肉分割领域的权威数据集,囊括了不同大小、形态的息肉类型。该数据集包含 1 000 张高质量内窥镜图像及其专家标注的分割掩码,严格按照 8:2 比例划分为 800 张训练图像和 200 张测试图像。CVC-ClinicDB 数据集是 MICCAI 2015 自动息肉检测挑战赛的训练数据集,覆盖了多种息肉类型和不同黏膜背景条件。该数据集共包含 612 张图像,按照 8:2 的比例随机划分为训练集与测试集,其中训练集包含 490 张图像,测试集包含 122 张图像。

预处理包括将所有图像和标签统一缩放到 256×256 尺寸,然后通过随机水平翻转、垂直翻转和亮度对比度调整等多数据增强手段来扩充数据集,以提升模型的泛化能力并防止过拟合。

2.2 实验设置

本文实验环境基于 Python 3.10 和 Pytorch 框架实现,训练环境包括 CUDA 11.8.0 和 NVIDIA RTX 4060 显卡(内存 16 GB)。采用 Adam 优化器进行训练,批量大小设置为 4,初始学习率设定为 0.000 5,总训练周期为 120 个 epoch。

2.3 评估指标

为了评价所提出的 MSBRAU-Net++ 的分割性能,采用医学影像分析中公认的 4 个核心指标,对模型的分割精度进行量化评估。

1) 交并比(intersection over union, IoU),用于衡量两个集合相似度的指标,计算的是“预测结果”和“真实标签”这两个区域之间的重叠程度。计算公式如式(14)。

$$IoU = \frac{TP}{TP + FP + FN} \quad (14)$$

2) 相似性系数(dice similarity coefficient, DSC),是一种集合相似度量函数,通常用于计算两个样本的相似度,取值范围在 $[0, 1]$ 。计算公式如式(15)。

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{15}$$

3) 精确度(Precision),指预测正确的数据与预测为正例的数据的数量比值,能够反映出模型预测结果的正确率。计算公式如式(16)。

$$Prcision = \frac{TP}{TP + FP} \tag{16}$$

4) 召回率(Recall),指预测正确的数据与真实为正例的数据的数量比值,能够反映出模型对所有目标的检出率。计算公式如式(17)。

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

式中: TP 为实际为正例,且被分类器正确判断为正例的样本数, FN 为实际为负例,但被分类器错误判断为正例的样本数, FP 为实际为正例,但被分类器错误判断为负例的样本数。

2.4 损失函数

在训练过程中,为有效应对样本类别不平衡问题并提升模型对目标边界的定位精度,本文构建了一种混合损失函数。该函数将二元交叉熵损失 BCE 与 Dice 系数损失进行加权融合,旨在同时利用 BCE 在像素级别上的精确分类能力和 Dice 系数对区域重叠度的优化优势。二元交叉熵损失式(18)、Dice 损失式(19)和混合损失式(20)。

$$\xi_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{18}$$

式中: N 为样本数量, y_i 为第 i 个样本的真实标签, p_i 为模型预测的第 i 个样本为正类的概率。

$$\xi_{Dice} = 1 - \frac{2 \times \sum_i P_i G_i}{\sum_i (P_i^2 + G_i)} \tag{19}$$

式中: P_i 和 G_i 分别为预测结果和真实标签在第 i 个像素点上的值。

$$\xi_{loss} = \alpha \times \xi_{Dice} + (1 - \alpha) \times \xi_{BCE} \tag{20}$$

式中: α 为权重系数,用于平衡二元交叉熵损失和 Dice 损失的影响。

2.5 实验结果分析

为验证本文提出的 MSBRAU-Net++ 模型的性能优势,选取了 7 种具有代表性的模型进行对比,包括经典基准模型、结构相似模型以及最新研究进展。基准模型: U-Net(2015)^[4] 是息肉分割领域的经典基准; BRAU-Net++(2024)^[20] 是本文所参照的基准模型。结构相似模型: CASF-Net(2023)^[28] 与本文模型同样采用多分支融合多尺度信息,但其多分支交叉融合模块计算复杂度高,实时性较差。DCSAU-Net(2023)^[29] 的模型结构与本文相似,它通过初级特征保留模块和紧凑分割注意力模块提取有效特征,但在目标被遮挡或背景复杂时,分割精度会下降。最新研究进展: DTAN(2024)^[30]、MSLAU-Net(2025)^[22] 和 Condseg(2025)^[18] 为息肉分割领域的最新成果。DTAN 利用文本注意力机制引导分割,但难以捕捉目标细节; MSLAU-Net 结合了 CNN 与 Transformer 的优势,但模型复杂度较高; Condseg 通过分解特征图和两阶段训练提升精度,但训练时间较长。通过与上述密切相关的模型对比,可以全面体现本文算法的先进性。所有实验均在相同条件下进行,结果如表 1 所示(加粗数值为最优)。

表 1 不同算法在 Kvasir-SEG 与 CVC-ClinicDB 数据集上的性能对比

Table 1 Experimental comparison of datasets Kvasir-SEG and CVC-ClinicDB under different algorithms									%
算法	Kvasir-SEG 数据集				CVC-ClinicDB 数据集				
	IoU	DSC	Recall	Precision	IoU	DSC	Recall	Precision	
U-Net	65.75	75.80	83.61	77.68	80.89	86.82	88.51	87.98	
BRAU-Net++	77.45	86.64	86.56	88.15	84.45	87.74	88.56	88.65	
CASF-Net	81.77	88.76	89.25	88.21	85.11	89.96	90.05	90.11	
DCSAU-Net	82.01	88.98	89.04	88.76	86.80	89.99	90.02	89.76	
DTAN	84.18	90.48	91.25	90.21	86.77	90.56	91.33	90.89	
MSLAU-Net	83.98	90.55	90.51	92.01	88.68	91.62	91.22	92.34	
Condseg	84.62	90.51	92.32	91.71	88.69	91.89	92.50	92.61	
MSBRAU-Net++	84.65	90.63	92.51	92.33	88.87	91.99	92.64	92.77	

在 Kvasir-SEG 数据集上, IoU、DSC、Precision 和 Recall 分别达到了 84.65%、90.63%、92.51% 和 92.33%。与基线模型 BRAU-Net++ 相比,各项指标分别显著提升了 7.2%、3.99%、5.95% 和 4.18%,充分验证了 MSBRAU-Net++ 架构的优越性。在 CVC-ClinicDB 数据

集上,这 4 项指标分别达到了 88.87%、91.99%、92.64% 和 92.77%,比 BRAU-Net++ 分别提高了 4.42%、4.25%、4.08% 和 4.12%。实验结果充分表明,本文所提出的模型相对于其他模型(如 U-Net、DTAN 和 DCSAU-Net 等),分割性能均表现最佳,显著优于对比模型。本文

算法在息肉分割任务中具有更高的临床可用性,有效提升了息肉分割的精度,并展现出更高的结构完整性和轮廓保留能力,表明其泛化能力较强。

Kvasir-SEG 数据集和 CVC-ClinicDB 数据集的不同模型分割可视化结果如图 8、9 所示。由图 8 和 9 可知,各模型在不同息肉分割场景下的表现存在显著差异。在复杂背景下的多息肉分割任务中,多数模型难以精确分割息肉边界,仅 MSBRAU-Net++(图 8)以及 MSLAU-Net 与 MSBRAU-Net++(图 9)的分割结果接近真实标签。对于简单背景下的单息肉分割,大部分模型表现良好,但图 8 中的 BRAU-Net++ 出现了明显偏差,错误地将部分背景

区域分割为息肉。挑战性最大的复杂背景单息肉分割任务,因背景与息肉颜色高度相似,导致模型易将背景误判为息肉。从图 8 和 9 可以看出,U-Net、BRAU-Net++、CASF-Net 及 DCSAU-Net 均出现严重偏差;MSLAU-Net 和 Condseg 在边界处理上存在微小误差,结果较为接近真实标签;而 MSBRAU-Net++ 的表现则与真实标签高度一致。最后,在简单背景下的多息肉分割任务中,仅有 Condseg 和 MSHSCM-Net++ 达到了理想效果,其他模型均存在不同程度的偏差。结果表明,本文提出的 MSHSCM-Net++ 模型在各类息肉分割场景中均展现出优异的泛化能力与分割精度。

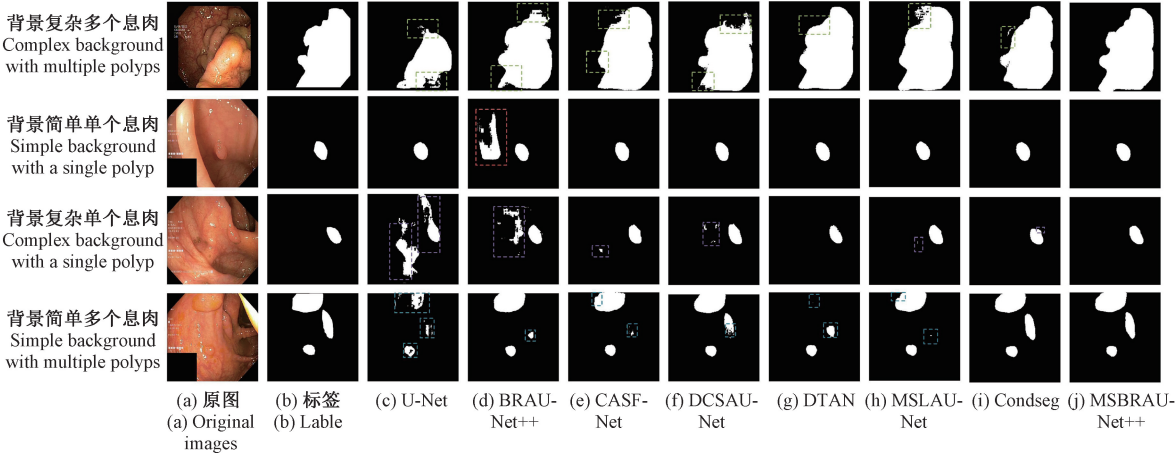


图 8 不同算法在 Kvasir-SEG 数据集分割结果
Fig. 8 Segmentation results of Kvasir-SEG dataset

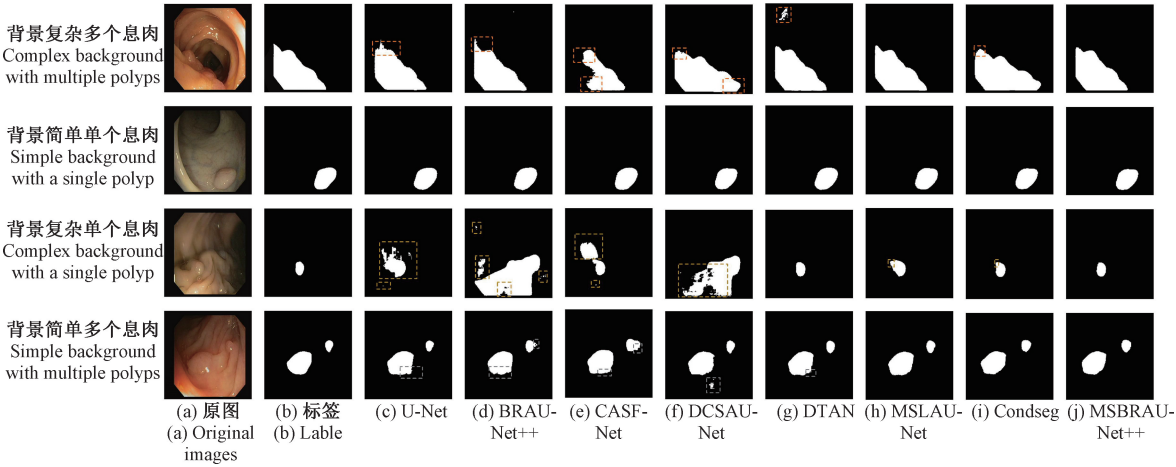


图 9 不同算法在 CVC-ClinicDB 数据集分割结果
Fig. 9 Segmentation results of CVC-ClinicDB dataset

2.6 消融实验

为验证 MSHSCM-Net++ 中各关键模块的贡献程度,在 Kvasir-SEG 数据集和 CVC-ClinicDB 数据集上设计了一系列消融实验。通过逐步引入各个改进模块并分析其整体性能的影响,结果如表 2 所示(表 2 中加粗数值为最优值)。以 BRAU-Net++ 为基准,单独引入各模块后,

性能均有提升:引入 MPFAB 模块后(IoU 上升至 80.39% 和 86.37%、DSC 上升至 88.12% 和 89.15%);引入 HSCM 模块后(IoU 上升至 81.78% 和 87.26%、DSC 上升至 88.56% 和 89.79%)引入 MSGAF 模块后(IoU 上升至 82.46% 和 87.91%、DSC 上升至 88.99% 和 90.32%)。这表明各模块均能显著增强信息提取能力,但在局部特征的

表 2 Kvasir-SEG 数据集和 CVC-ClinicDB 数据集消融指标对比

Table 2 Comparison of ablation metrics in Kvasir-SEG dataset and CVC-ClinicDB dataset

%

模块			Kvasir-SEG 数据集				CVC-ClinicDB 数据集			
MSGAF	HSCM	MPFAB	IoU	DCS	Recall	Precision	IoU	DCS	Recall	Precision
×	×	×	77.45	86.64	86.56	88.15	84.45	87.74	88.56	88.65
×	×	✓	80.39	88.12	88.98	88.33	86.37	89.15	90.32	90.74
×	✓	×	81.78	88.56	90.11	90.34	87.26	89.79	90.11	90.45
✓	×	×	82.46	88.99	90.58	90.15	87.91	90.32	91.72	90.88
×	✓	✓	83.15	89.36	90.78	91.45	88.01	90.23	92.11	91.49
✓	×	✓	83.85	89.78	91.60	90.67	88.25	90.81	91.37	92.20
✓	✓	×	84.39	90.48	92.21	91.33	88.65	91.37	92.18	92.51
✓	✓	✓	84.65	90.63	92.51	92.33	88.87	91.99	92.64	92.77

细粒度建模方面仍存在局限。因此,将各模块两两结合以验证其协同效应。组合 HSCM 与 MPFAB(IoU 上升至 83.15% 和 88.01%、DSC 上升至 89.36% 和 90.23%)、MSGAF 与 MPFAB(IoU 上升至 83.85% 和 88.25%、DSC 上升至 89.78% 和 90.81%)、MSGAF 与 HSCM(IoU 上升至 84.39% 和 88.65%、DSC 上升至 90.48% 和 91.37%)、各项指标分较于单模块均有提升,表明其可以有效缓解特征融合过程中的通道不匹配问题,从而提升全局信息表达能力,但在边界处理方面仍需进一步优化。为全面增强模型的特征提取能力,进一步将 3 个模块融合,这不仅弥补了各模块的不足,还提升了整体分割精度(IoU 上升至 84.65% 和 88.87%、DSC 上升至 90.63% 和 91.99%),实现了准确性与边界平滑度的双重优化。由此可知,本文所提出的方法,性能均达到最优,在分割精度、边界准确性和特征提取能力上均优于原始的 BRAU-Net++。

Kvasir-SEG 数据集和 CVC-ClinicDB 数据集上的消融结果如图 10、11 所示。从图 10 和 11 中可以看出,在面对背景复杂、存在多个息肉的情况下,引入 MPFAB 后,相较于原始的 BRAU-Net++,模型对息肉的分割性能显著提

升,图像特征提取能力也更加突出。引入 HSCM 后,模型对息肉边界的处理效果较好,但对内部某些息肉组织的分割仍不够明确。而引入 MSGAF 后,模型能有效区分前景(息肉)与背景,但对边界的处理仍存在不足。为进一步验证各模块的有效性,进行两两组合实验。HSCM + MPFAB:能够大致划分出息肉,但对边界和共现现象的处理效果欠佳;MSGAF + MPFAB:显著提升了前景与背景的区分能力,但边界分割问题仍未完全解决;MSGAF + HSCM:共现现象得到较好解决,但边界处理仍存在轻微不足。在简单背景下的单个息肉分割任务中,各模块均表现良好,分割结果与真实标签高度一致。然而,在背景复杂、单个息肉较小且边界模糊的情况下,分割难度较大,各模块均难以精确分割息肉;仅 MSGAF + HSCM 组合及 MSBRAU-Net++ 能较好解决共现现象,但对小息肉边界的处理仍有不足。对于背景简单的多个息肉分割,单一模块的分割效果存在差异,但模块组合后均能实现相对准确的分割。实验结果证明,本文提出的算法在共现现象处理、边界分割精度以及图像特征提取方面均表现出优异的性能。

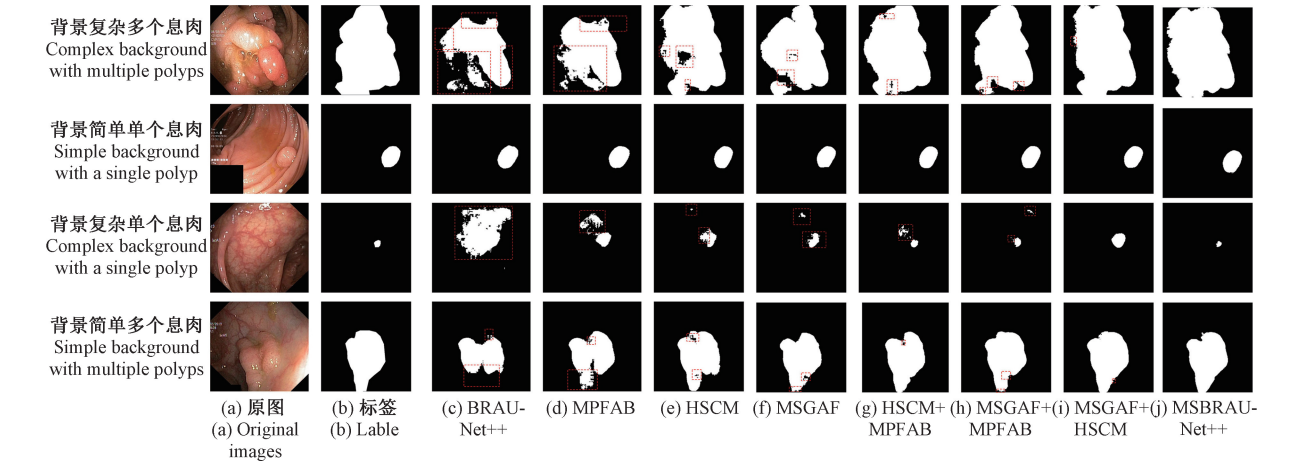


图 10 Kvasir-SEG 数据集消融实验分割结果

Fig. 10 Ablation experiment segmentation results on the Kvasir-SEG dataset

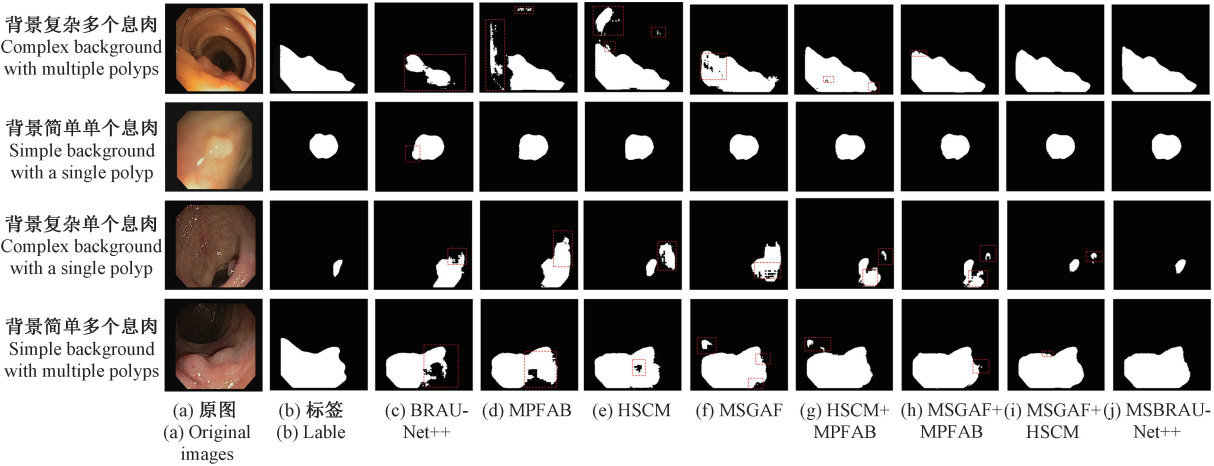


图 11 CVC-ClinicDB 数据集消融实验分割结果

Fig. 11 Ablation experiment segmentation results on the CVC-ClinicDB dataset

3 结 论

本研究提出了一种通用的医学图像分割框架 MSBRAU-Net++。首先,针对息肉分割任务中尤为突出的共现现象,引入 MSGAF。该模块通过在多个尺度上聚合特征,并利用注意力机制动态抑制无关背景信息的干扰,同时显著增强对前景目标(息肉)的特征响应。通过在编码器与解码器之间构建高效的信息交互通路,MSGAF 能够有效打破背景与前景的特征混淆。其次,针对边界模糊问题,设计 HSCM。该模块从深层特征中提取和恢复更精细的局部细节与边缘信息,通过在空间维度和通道维度上分别进行特征校准与增强,能够更精确地捕捉息肉的细微轮廓和不规则边缘。最后,为全面提升模型的特征提取能力并解决分割不足问题,构建全新的特征聚合块 MPFAB。作为网络中的基础特征提取单元,MPFAB 通过并行的多路径卷积结构同时捕获图像中的低阶细节信息和高阶语义信息。这种多粒度的特征提取方式极大地丰富了特征表示的多样性,使模型能够更全面地理解图像内容,从而有效避免因特征提取不充分导致的分割断裂或遗漏,确保分割结果的完整性和连续性。在两个公开的挑战性息肉分割数据集(Kvasir-SEG 和 CVC-ClinicDB)上进行了充分实验。结果表明,与其他模型相比,MSBRAU-Net++ 在所有评价指标上均取得了最优性能。在未来的工作中,将专注于为多模态医学图像分割任务设计更复杂且通用的架构,以进一步提升其在处理复杂病理结构时的计算效率和鲁棒性。

参考文献

[1] SIEGEL R L, WAGLE N S, CERCEK A, et al. Colorectal cancer statistics, 2023[J]. CA: A Cancer Journal for Clinicians, 2023, 73(3): 233-254.

[2] LALINIA M, SAHAFI A. Colorectal polyp detection

in colonoscopy images using YOLO-V8 network[J]. Signal, Image and Video Processing, 2024, 18(3): 2047-2058.

[3] 陈业航, 李智, 冯宝, 等. 基于改进的活动轮廓模型的胸膜接触型肺结节分割[J]. 仪器仪表学报, 2019, 40(11): 107-116.

CHEN Y H, LI ZH, FENG B, et al. Pleural contact pulmonary nodule segmentation based on improved active contour model[J]. Chinese Journal of Scientific Instrument, 2019, 40(11): 107-116.

[4] 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 International Publishing, 2015: 234-241.

[5] SAFAROV S, WHANGBO T K. A-Dense UNet: Adaptive densely connected UNet for polyp segmentation in colonoscopy images with atrous convolution[J]. Sensors, 2021, 21(4): 1441.

[6] 夏平, 何志豪, 雷帮军, 等. 密集级联卷积与自注意力特征聚合的视网膜血管分割算法[J]. 电子测量与仪器学报, 2024, 38(9): 36-44.

XIA P, HE ZH H, LEI B J, et al. Retinal vascular segmentation algorithm based on full scale dense convolutional u-shaped networks [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 36-44.

[7] 邱拴虎, 杨文瀚, 廖苗, 等. 基于 RA-Unet 的 CT 图像肝脏肿瘤分割[J]. 仪器仪表学报, 2022, 43(8): 65-72.

- DI SH H, YANG W H, LIAO M, et al. Liver tumor segmentation from CT images based on RA-Unet[J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 65-72.
- [8] TOMAR N K, SHERGILL A, RIEDERS B, et al. TransResU-Net: Transformer based ResU-Net for real-time colonoscopy polyp segmentation[J]. ArXiv preprint arXiv:2206.08985, 2022.
- [9] 刘肇隆, 范馨月. 基于全尺度跳跃连接的 TransUNet 医学图像分割网络[J]. 国外电子测量技术, 2023, 42(11): 42-48.
- LIU ZH L, FAN X Y. TransUNet medical image segmentation network based on full-scale jump connectivity [J]. Foreign Electronic Measurement Technology, 2023, 42(11): 42-48.
- [10] 陈广秋, 任天蓉, 段锦, 等. 基于双分支多尺度特征融合的跨模态语义分割算法[J]. 电子测量与仪器学报, 2025, 39(5): 144-154.
- CHEN G Q, REN T R, DUAN J, et al. Cross-modal semantic segmentation algorithm based on dual-branch multi-scale feature fusion[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 144-154.
- [11] 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.
- [12] 薛钟毫, 姜金刚, 孙健鹏, 等. 基于多尺度卷积和选择性核双注意力机制的半监督全景 X 射线图像龋齿分割[J]. 仪器仪表学报, 2025, 46(6): 241-250.
- XUE ZH H, JIANG J G, SUN J P, et al. Semi-supervised caries segmentation of panoramic X-ray images based on multi-scale convolution and selective kernel dual-attention[J]. Chinese Journal of Scientific Instrument, 2025, 46(6): 241-250.
- [13] WANG W Y, ZHOU X CH, YANG L J. Multimodal medical image fusion based on parameter adaptive PCNN and latent low-rank representation [J]. Instrumentation, 2023, 10(1): 45-58.
- [14] SIDDIQUI N A, QADRI M T, AKHTER M O, et al. High-precision brain tumor segmentation using a progressive layered U-Net(PLU-Net) with multi-scale data augmentation and attention mechanisms on multimodal magnetic resonance imaging [J]. Instrumentation, 2025, 12(1): 77-92.
- [15] LIU J ZH, CHEN Q S, ZHANG Y, et al. Multi-level feature fusion network combining attention mechanisms for polyp segmentation[J]. Information Fusion, 2024, 104: 102195.
- [16] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [17] WANG W H, XIE EN Z, LI X, et al. Pvt v2: Improved baselines with pyramid vision transformer [J]. Computational Visual Media, 2022, 8(3): 415-424.
- [18] LEI M Q, WU H CH, LYU X H, et al. Condseg: A general medical image segmentation framework via contrast-driven feature enhancement [C]. AAAI Conference on Artificial Intelligence, 2025, 39(5): 4571-4579.
- [19] YUE G H, LI Y Y, JIANG W CH, et al. Boundary refinement network for colorectal polyp segmentation in colonoscopy images [J]. IEEE Signal Processing Letters, 2024, 31: 954-958.
- [20] LAN L B, CAI P ZH, JIANG L, et al. BRAU-Net++: U-shaped hybrid CNN-transformer network for medical image segmentation[J]. Arxiv preprint arXiv: 2401.00722, 2024.
- [21] ZHANG B H, ZHENG Z L, ZHAO Y Q, et al. MCBTNet: Multi-feature fusion CNN and Bi-Level routing attention transformer-based medical image segmentation network[J]. IEEE Journal of Biomedical and Health Informatics, 2025, 29(7): 5069-5028.
- [22] LAN L B, LI Y X, LIU X J, et al. MSLAU-Net: A hybrid CNN-Transformer network for medical image segmentation[J]. Arxiv preprint arXiv:2505.18823, 2025.
- [23] 张攀峰, 杨贺, 神显豪, 等. 融合局部和全局特征的息肉分割模型[J]. 电子测量技术, 2024, 47(16): 100-109.
- ZHANG P F, YANG H, SHEN X H, et al. Polyp segmentation model based on fusion of local and global features [J]. Electronic Measurement Technology, 2024, 47(16): 100-109.
- [24] SHEN ZH R, ZHANG M Y, ZHAO H Y, et al. Efficient attention: Attention with linear complexities[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3531-3539.
- [25] NI ZH L, CHEN X H, ZHAI Y J, et al. Context-guided spatial feature reconstruction for efficient semantic segmentation[C]. European Conference on Computer Vision. Cham: Springer Nature

Switzerland, 2024: 239-255.

[26] SI Y ZH, XU H Y, ZHU X ZH, et al. SCSA: Exploring the synergistic effects between spatial and channel attention [J]. Neurocomputing, 2025, 634: 129866.

[27] LI Y X, LI X, DAI Y M, et al. LSKNet: A foundation lightweight backbone for remote sensing[J]. International Journal of Computer Vision, 2025, 133 (3): 1410-1431.

[28] ZHENG J W, LIU H, FENG Y CH, et al. CASF-Net: Cross-attention and cross-scale fusion network for medical image segmentation [J]. Computer Methods and Programs in Biomedicine, 2023, 229: 107307.

[29] XU Q, MA ZH CH, DUAN W T. DCSAU-Net: A deeper and more compact split-attention U-Net for

medical image segmentation[J]. Computers in Biology and Medicine, 2023, 154: 106626.

[30] ZHAO Y Y, LI J J, REN L, et al. DTAN: Diffusion-based text attention network for medical image segmentation[J]. Computers in Biology and Medicine, 2024, 168: 107728.

作者简介

付金丽,硕士研究生,主要研究方向为信号与信息处理技术。

E-mail:3488920098@qq.com

褚丽莉(通信作者),博士,教授,研究生导师,主要研究方向为信号与信息处理技术。

E-mail:chulili902@126.com

李波,博士,教授,研究生导师,主要研究方向为智能信息处理。

E-mail:leebo@yeah.net