

基于边界感知与特征融合的病理图像分割网络

陈海鹏¹, 孔 鸣¹, 张洪语¹, 孙宝胜²

(1. 吉林大学 计算机科学与技术学院, 长春 130012; 2. 吉林省肿瘤医院 放疗科, 长春 130012)

摘要: 针对病灶形态多样性引发的病理图像识别精度不足及特征融合过程中的语义鸿沟问题, 提出一种融合 Transformer 与注意力机制的改进型 U-Net 架构. 首先, 设计边界感知模块强化病理图像的病灶边缘特征表达, 提升模型对复杂结构的感知能力; 其次, 在瓶颈层引入正则化大核注意力模块以建模长程依赖, 并通过逐层正则化策略缓解过拟合风险; 最后, 进一步引入可学习的视觉中心模块, 增强全局与局部特征之间的互补. 在数据集 MoNuSeg 和 GlaS 上的实验结果表明, 该方法在分割精度、边界清晰度方面均优于当前主流模型.

关键词: 边界感知; 特征融合; 病理图像; 图像分割; 卷积神经网络; Transformer 架构

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1671-5489(2026)03-0581-10

Pathological Image Segmentation Network Based on Boundary-Aware and Feature Fusion

CHEN Haipeng¹, KONG Ming¹, ZHANG Hongyu¹, SUN Baosheng²

(1. College of Computer Science and Technology, Jilin University, Changchun 130012, China;

2. Department of Radiotherapy, Jilin Province Cancer Hospital, Changchun 130012, China)

Abstract: Aiming at the problems of insufficient accuracy of pathological image recognition and semantic gaps in feature fusion process caused by the diversity of lesion morphology, we proposed an improved U-Net architecture that integrated Transformer and attention mechanisms. Firstly, we designed a boundary-aware module to enhance the expression of lesion edge features in pathological images, thereby improving the model's ability to perceive complex structures. Secondly, we introduced a regularized large-kernel attention module at the bottleneck layer to model long-range dependencies, and mitigated overfitting risk through a layer-wise regularization strategy. Finally, we further introduced a learnable visual center module to strengthen the complementarity between global and local features. Experimental results on the MoNuSeg and GlaS datasets show that the proposed method outperforms current mainstream models in terms of segmentation accuracy and boundary clarity.

Keywords: boundary-aware; feature fusion; pathological image; image segmentation; convolutional neural network; Transformer architecture

医学成像技术是辅助临床评估疾病、优化预防与控制策略的重要工具. 其中, 目标区域的精确分

收稿日期: 2025-02-24.

第一作者简介: 陈海鹏(1978—), 男, 汉族, 博士, 教授, 从事图像处理与模式识别的研究, E-mail: chenhp@jlu.edu.cn. 通信

作者简介: 孙宝胜(1972—), 男, 汉族, 博士, 教授, 从事肿瘤放射医学的研究, E-mail: 1575164354@qq.com.

基金项目: 吉林省科技发展计划重点研发项目(批准号: YDZJ202502CXJD068)和国家自然科学基金面上项目(批准号:

62276112).

割与定量分析为病理研究提供关键信息,尤其在治疗决策、疾病监测和预后预测中具有重要意义.当前主流的语义分割方法^[1-3]多采用基于卷积的编码器-解码器架构:编码器提取低分辨率特征,解码器将其上采样生成像素级分割图. U-Net^[4]是其中应用最广泛的结构之一,通过跳跃连接融合多层次语义与空间特征,有效提升分割精度和空间还原能力.

尽管 U-Net 在医学图像分割中被广泛应用,但其仍存在若干关键局限^[5]. 首先,卷积操作受限于感受野大小,难以有效建模全局上下文;其次,语义特征通常分布广泛,且背景与目标差异不显著,仅依赖局部卷积难以从复杂背景中准确提取分割区域;再次, U-Net 的编码器与解码器之间存在语义鸿沟,简单拼接或加和难以实现不同尺度特征的有效融合,易导致细粒度信息丢失^[6]. 因此,当前 U-Net 扩展面临两个挑战: 1) 如何有效连接编码器与解码器特征层,缓解语义差异并增强全局上下文建模; 2) 如何在保留其局部特征提取能力的同时,引入对全局信息的建模,提升整体分割性能.

为缓解编码器与解码器间特征不兼容的问题,近年来已提出了多种改进方法. TA-MOSC^[7]通过注意力机制,关注全局特征信息提高分割性能; UNet++^[8]通过嵌套密集跳跃路径替代传统跳跃连接,有效缩小语义差距,显著提升了分割性能;非线性变换也被用于增强跳跃连接,平衡语义不一致^[9]. 此外, Transformer^[10]凭借其强大的全局建模能力,成为近年来的研究热点^[11],但其高计算成本和对大规模预训练数据的依赖限制了其在医学图像中的应用. 为提升效率, Swin Transformer^[12]引入分层与移位窗口机制,增强多尺度建模能力. 特征金字塔网络(FPN)^[13]通过多尺度特征融合,缓解语义丢失. Uctransnet^[5]结合 Transformer 长程依赖建模与跳跃连接优化,提升小目标识别性能,并行融合卷积神经网络(CNN)与 Transformer 的编码结构^[14]提供了融合卷积与 Transformer 的新思路,增强了特征表达的鲁棒性. H-DenseUNet^[15]将分层聚合体积上下文信息,在保持计算效率的同时提升分割性能. TransFuse^[16]结合 U-Net 与 Transformer,兼顾局部与全局建模,在保证效率的同时取得了优异分割效果.

然而,现有方法在处理病理图像中复杂边界与背景区分方面仍面临挑战. 因此,本文提出一种基于边界感知与特征融合的病理图像分割网络(pathology image segmentation network based on boundary sensing and feature fusion, BFFNet),旨在优化 U-Net 架构中的特征连接与融合策略. BFFNet 引入边界感知编码器,通过可变形卷积^[17]提供多样化感受野以适应非刚性结构,并结合空间组增强模块(SGE)^[18]有效抑制噪声和不相关模式的干扰. 在跳跃连接路径中,设计可学习的显示视觉中心(explicit visual center, EVC)模块^[19],增强全局与局部特征交互. 解码阶段引入注意力融合模块,以缓解语义鸿沟,提升特征整合效果. 同时,在瓶颈层构建正则化大核注意力模块,结合大核深度卷积与空洞卷积^[20]扩展感受野,有效提取关键的大尺度特征. 整体架构实现了对复杂变形结构的更强感知能力,显著提升分割性能. 实验结果表明, BFFNet 在数据集 MoNuSeg 上的 Dice 系数和交并比(IoU)分别达 82.51%和 72.02%,在数据集 GlaS 上的 Dice 系数和 IoU 分别达 92.20%和 85.53%,显示出优异的分割效果和良好的泛化能力.

1 本文方法

图 1 为 BFFNet 的整体架构. 这是一种为提升病理图像分割性能而设计的改进版 U-Net 框架,针对复杂背景和非固定结构特征的医学图像分割挑战提供了解决方案. 该架构基于 U-Net 的编码器-解码器结构,在特征提取和特征融合过程中表现出色.

在编码器部分,每层依次采用边界感知编码器模块,对医学图像的多尺度特征进行提取和增强. 普通卷积负责提取局部边缘和纹理等基本特征,而可变形卷积则通过引入动态偏移量自适应调整卷积核的位置,使其更灵活地捕捉复杂的非固定几何形状(如非固定形状细胞和器官等). 这一特性增强了模型在复杂医学场景下的特征捕获能力. 此外,空间组增强模块根据每个空间位置动态分配注意力权重,强化感兴趣区域的表达,抑制背景噪声和不相关信息.

在瓶颈层设计中,本文采用大核注意力模块,使用卷积层模拟全局建模能力. 该模块结合了 7×7 深度卷积和 5×5 空洞卷积,通过扩展感受野捕捉全局信息,并加强对全局上下文的理解. 此外,引入

L_2 正则化以防止过拟合, 特别是在医学图像分割任务中, 数据集通常较小且样本不平衡, 正则化技术有效地提高了模型的泛化能力, 减少了对小数据集的过拟合影响。

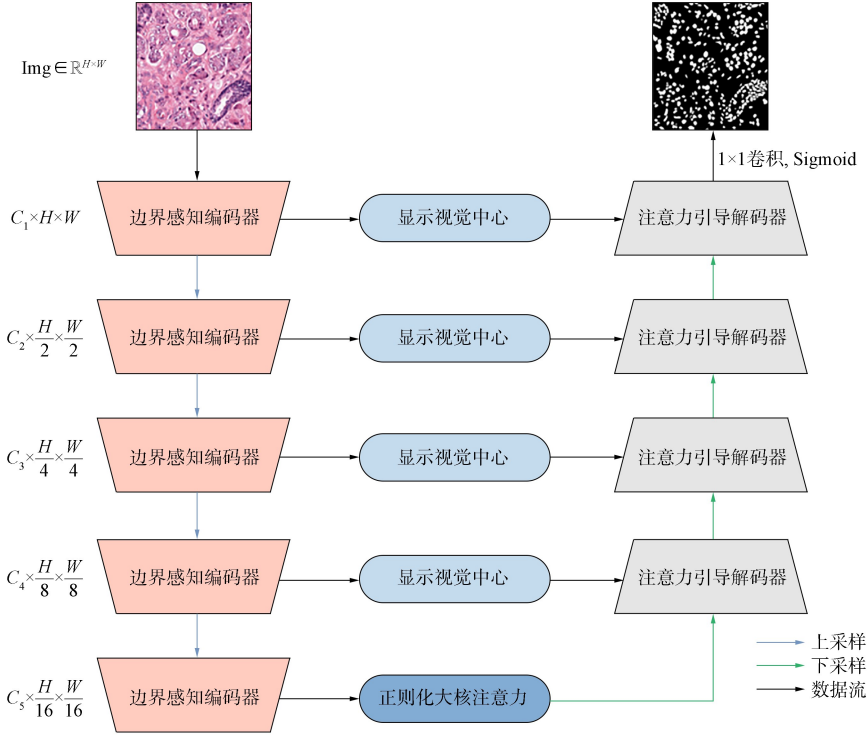


图 1 基于边界感知与特征融合的病理解图像分割网络架构

Fig. 1 Architecture of boundary-aware and feature fusion-based pathological image segmentation network

在跳跃连接路径中, 设计了显式视觉中心模块. 显示视觉中心模块利用轻量级的视觉 Transformer 结构和可学习视觉中心双流结合网络, 通过对全局特征和通道重要性的学习, 强化了全局信息的建模能力. 相比于传统 U-Net 中的跳跃连接, 显示视觉中心模块有效缓解了编码器和解码器之间的语义鸿沟, 确保特征在不同阶段的融合更一致, 减少了细粒度特征的丢失。

BFFNet 架构中各模块注重医学图像中具有复杂背景的目标特征提取, 以及局部特征和全局特征的提取与融合, 使其在医学图像分割任务中具有卓越的性能, 尤其在复杂、非固定形状目标和小数据集条件下展现了强大的适应性和稳健性。

1.1 边界感知模块

1.1.1 边界感知编码器

本文提出的边界感知编码器模块旨在解决传统卷积神经网络在处理背景特征不明显与复杂几何结构图像时的局限性. 该编码器整合了普通卷积、可变形卷积和空间群智能增强模块, 旨在提升网络识别复杂背景和处理形变特征的感知能力, 增强分割精度和鲁棒性. 图 2 为边界感知编码器的架构。

初始阶段, 采用 3×3 的普通卷积核对输入图像进行初级特征提取. 该过程的重点是捕获局部纹理及边缘特征, 为后续的复杂特征处理奠定基础. 普通卷积层提供了可靠的局部特征表达, 但在面对具有复杂形变或非刚性结构的目标时性能有限, 易忽略局部几何变形特征. 针对上述问题, 本文通过引用可变形卷积, 利用可学习偏移量使卷积核的采样位置可自适应地调整, 从而更精确地捕捉具有非固定形状的目标特征. 其核心机制通过预测动态偏移值实现, 公式化描述如下:

$$\text{offset} = \text{Conv}(\mathbf{F}), \quad (1)$$

$$\mathbf{F}_{\text{output}} = \sum_{k=1}^n \mathbf{W}_k \cdot \mathbf{F}(p_0 + p_k + \text{offset}(p_k)), \quad (2)$$

其中 $\mathbf{F}$ 为输入特征图, offset 为从偏移预测层学习到的偏移量, Conv 为卷积操作, $\mathbf{W}_k$ 为卷积核权重,

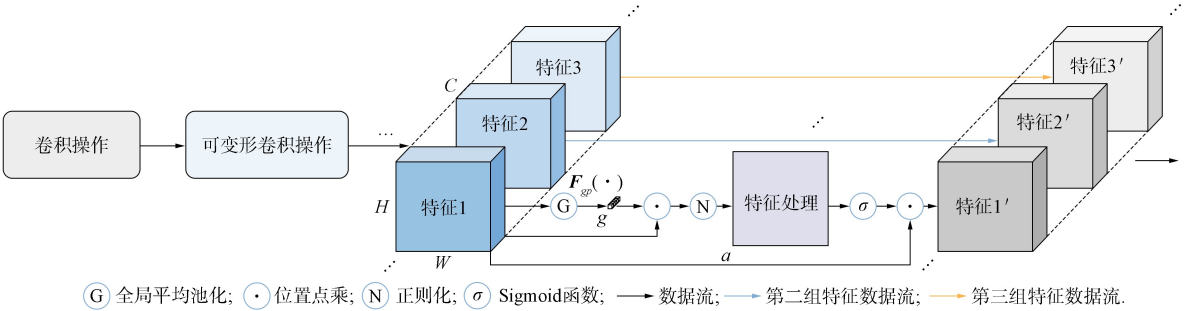


图 2 边界感知编码器结构

Fig. 2 Structure of boundary-aware encoder

p_k 为卷积核中的位置偏移, $\mathbf{F}_{\text{output}}$ 为矩阵的输出结果.

为进一步提升特征的语义表达能力, 空间群智能增强模块将特征图按通道分组, 并在每组内部通过全局信息整合生成缩放因子, 从而动态调整每组特征的重要性. 该模块不仅增强了关键区域的特征表达, 还有效抑制了噪声和不相关模式的干扰. 其过程可公式化表示为

$$G_i = \text{Norm}\left(\sum_{i=1}^n F_i\right), \tag{3}$$

$$S_i = \sigma(G_i \cdot F_i), \tag{4}$$

其中: F_i 表示第 i 组的特征; G_i 是全局池化后的组特征, 用于计算缩放因子; σ 表示 Sigmoid 激活函数; S_i 表示最终缩放后的特征图. 通过集成普通卷积、可变形卷积和空间群智能增强模块, 该方法在理论上提出了创新的多视野特征提取机制, 并在实际应用中表现出卓越的性能. 在医学图像分割和自然场景解析等高级任务中, 该方法展现了良好的适应性和实用性, 尤其在处理具有复杂几何和多尺度变化的目标时, 能显著提升分割精度和模型的鲁棒性.

1.1.2 正则化大核注意力模块

正则化大核注意力模块的设计目的是引入正则化大核注意力(regularized large kernel attention, RLKA), 并在全网络中应用 L_2 正则化. 该架构旨在解决传统 U-Net 在处理复杂场景和目标多样性时瓶颈层处理细粒度特征的局限性, 通过增大感受野捕获大尺度特征, 该模块尤其适用于特征形状不固定及具有复杂背景的医学图像分割任务. 正则化大核注意力模块结构如图 3 所示. RLKA 模块通过创新的卷积核分解方法, 增强了对大尺度特征和远程依赖关系的处理能力. 具体实现包括:

$$\mathbf{X}_{\text{local}} = \text{DC}(\mathbf{X}, \text{kernelsize} = (2d - 1), (2d - 1)), \tag{5}$$

$$\mathbf{X}_{\text{dilated}} = \text{DDConv}(\mathbf{X}_{\text{local}}, \text{dilation} = d), \tag{6}$$

$$\mathbf{X}_{\text{channel}} = \text{Conv}(\mathbf{X}_{\text{dilated}}, \text{kernelsize} = (1, 1)), \tag{7}$$

其中: DC 为深度卷积, 使用深度卷积提取捕获局部特征; d 为窗口大小; DDConv 为深度空洞卷积; dilation 为空洞系数. 输入特征经过深度卷积得到局部特征 $\mathbf{X}_{\text{local}}$, 之后经过深度空洞卷积, 扩张感受野, 得到扩张感受野特征 $\mathbf{X}_{\text{dilated}}$, 最后经过一个 1×1 卷积层进行通道整合, 实现远程特征扩展, 进一步扩大感受野, 提升语义表达能力. 通过 L_2 正则化, 抑制大权重值以防止过拟合:

$$L_{\text{total}} = L_{\text{data}} + \lambda \sum_i w_i^2, \tag{8}$$

其中 L_{data} 表示基于训练数据的损失, λ 为正则化系数, 求和项表示所有模型权重的平方和, L_{total} 表示

得到的总损失函数. 本文通过逐层正则化策略, 在降低高分辨率计算负担的同时, 显著抑制了模型在

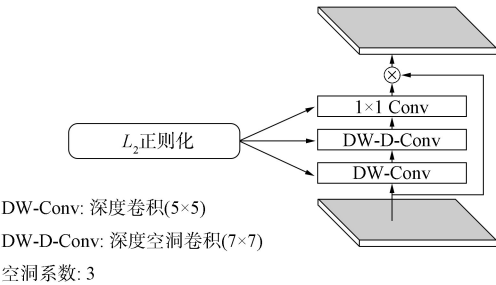


图 3 正则化大核注意力模块结构

Fig. 3 Structure of regularized large kernel attention module

小数据集上的过拟合风险. 该策略通过分层约束(如权重衰减或稀疏化)参数空间, 迫使模型学习更泛化的特征表示, 而非过度依赖训练数据噪声.

1.2 全局-局部特征平滑融合模块

1.2.1 显式视觉中心

为解决传统跳跃连接在特征融合过程中存在的局部特征限制和语义鸿沟问题, 本文提出一种创新的显式视觉中心(explicit visual center, EVC)模块, 旨在从全局和局部两个视角增强特征表达, 特别适用于复杂结构和语义信息密集的医学图像分割任务. 总体思想是利用轻量级 Transformer^[21]的全局特征提取能力与可学习视觉中心的局部特征提取能力相融合, 改善跳跃连接中产生的语义鸿沟问题. 图 4 为显式视觉中心模块结构及特征传递路径. 首先对特征进行增强, 然后对跳跃连接的输入特征采用卷积操作提取初步特征, 再通过深度卷积(DConv)和组归一化(GN)强化特征表达, 公式为

$$\mathbf{X}_{out} = \text{DConv}(\text{GN}(\mathbf{X})) + \mathbf{X}, \tag{9}$$

其中 $\mathbf{X}$ 为输入特征, $\mathbf{X}_{out}$ 为输出特征. 通过深度卷积过程增强了特征的深度表达能力, 并保留连接中的语义信息. 对于增强后的特征, 利用轻量级的 Vision Transformer^[21] 和可学习视觉中心机制分别提取全局和局部视觉信息. EVC 中的可学习视觉中心(learnable visual center, LVC)模块先对输入特征进行编码, 生成一组固有的码本和对应的可学习比例因子; 然后利用这些比例因子对输入特征进行加权融合, 增强特征表达能力, 并以较低的计算开销完成全局信息的补充. 公式化为

$$\text{LVC}(\mathbf{X}_{out}) = \text{FFN}(\text{Norm}(\text{MHA}(\text{SR}(\text{Norm}(\mathbf{X}_{out}))) + \mathbf{X}_{out})) + \mathbf{X}_{out}, \tag{10}$$

$$\mathbf{X} = \text{Cat}(\text{LVT}(\mathbf{X}_{out}), \text{LVC}(\mathbf{X}_{out})), \tag{11}$$

其中: $\mathbf{X}$ 为经过线性层初步提取特征, 并进行归一化调整数值分布; SR 为空间归约, 降低空间维度提取关键的空间信息并减少复杂度; MHA 为多头注意力, 同时关注输入数据的不同部分, 捕捉输入序列中不同位置的依赖关系; FFN 为前馈网络; Cat 为拼接操作, 将双路特征进行融合, 优化了特征传递过程的效果.

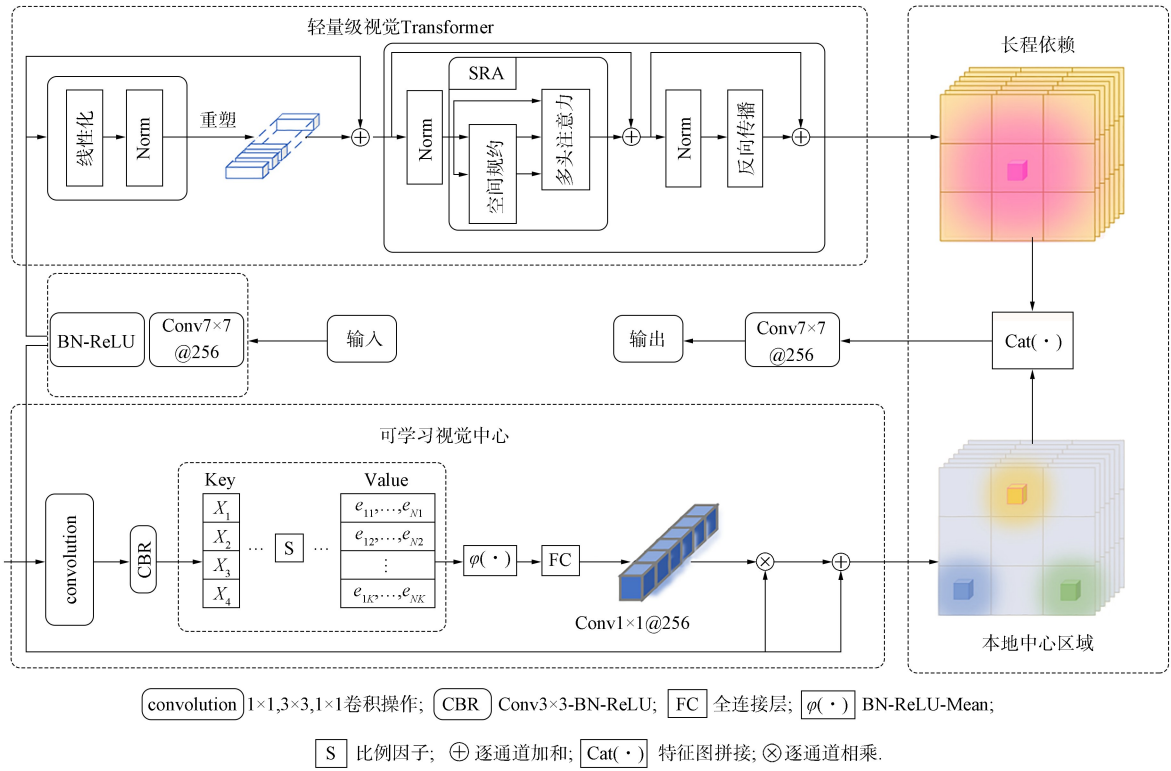


图 4 显式视觉中心模块结构及特征传递路径

Fig. 4 Explicit visual center module structure and feature transmission path

EVC 模块有效提升了模型对图像复杂区域的特征捕获能力,同时保持较低的计算开销.这种改进提高了分割任务的精度和模型整体性能,适用于需要高精度和强鲁棒性的医学图像分割任务.

1.2.2 注意力引导的特征上采样方法

在解码器部分,本文引入了 Stoken Attention 模块^[22],注意力引导的特征上采样方法受启发于注意力特征融合^[23]技术,旨在缓解跳跃连接和上采样导致的特征语义鸿沟.该模块通过动态采样和注意力机制提升全局与局部特征的融合质量.其公式为

$$\mathbf{S} = \text{SP}(\mathbf{X}, \mathbf{Y}; \theta), \quad (12)$$

$$\mathbf{S}' = \text{SA}(\mathbf{S}), \quad (13)$$

$$\mathbf{Z} = \text{MP}(\mathbf{S}', \mathbf{X}, \mathbf{Y}; \varphi), \quad (14)$$

$$\mathbf{Z} = \mathbf{X} + \mathbf{Y} + \text{M}(\mathbf{X} + \mathbf{Y}), \quad (15)$$

其中: SP 为 Stoken Attention 的快速采样算法,使用动态采样算法学习超标记空间($\mathbf{S}$),自适应提取局部特征并压缩冗余信息; SA 表示自注意力,通过自注意力对 $\mathbf{S}$ 进行处理,捕获特征间的长程依赖关系,增强特征表达一致性; MP 为映射回原始空间,基于学习参数的映射函数将增强特征重新分布到原始标记空间,实现细节补充; $\text{M}(\mathbf{X} + \mathbf{Y})$ 表示融合特征,实现特征权重的动态分配,优化全局与局部特征的融合质量.通过设计精细的特征融合框架,本文实现了上下文感知与线性特征调整的统一处理.该方法全面提升了图像的融合质量,并确保特征映射的准确性.

2 实验结果与分析

2.1 数据集和实验环境

本文采用公开数据集 Gland Segmentation(GlaS)^[24] 和 MoNuSeg^[25] 对提出的方法进行全评估.数据集 Gland Segmentation 包含 85 张训练图像和 80 张测试图像,主要用于腺体分割任务,是评估模型在组织结构分割中的重要基准.数据集 MoNuSeg 包含 30 张训练图像和 14 张测试图像,专注于细胞核分割任务,为复杂形态的生物医学图像分割提供了典型测试场景.通过实验可系统验证本文方法在多种医学图像分割任务中的适用性和鲁棒性.实验仿真基于百度云平台(AI Studio)进行,实验环境配置为 PaddlePaddle 2.2.2,计算平台为 NVIDIA 3090 GPU,配备 32 GiB 内存和 16 GiB 显存.模型训练采用交叉熵损失函数,并使用 AdamW 优化器,实验迭代次数设为 13 000,训练批量大小为 16,初始学习率为 1×10^{-4} .一阶和二阶矩估计的指数衰减率分别设为 0.9 和 0.99,权重衰减系数为 1×10^{-6} .为应对过拟合问题,采用水平翻转、垂直翻转和随机旋转在线数据增强策略,并且未使用任何预训练权重训练本文模型.在实验设置中,对于数据集 GlaS 和 MoNuSeg,批处理大小设为 4.两个数据集的输入图像都被调整到 224×224 像素,且 Token 尺寸设为 16.

2.2 定性比较

图 5 为在数据集 MoNuSeg 和 GlaS 上不同网络分割结果的定性比较.由图 5 可见,本文方法在医学图像分割任务中优势明显.与 U-Net 相比,尽管 U-Net 在整体分割结构上有一定精度,但其在细节捕捉和边界处理方面表现不足,易出现细节丢失和边界模糊的情况,而本文方法能更准确地还原细节特征,边界处理更清晰.相较于 MultiResUNet,虽然 MultiResUNet 通过多分辨率特征融合提升了分割性能,但其在复杂纹理区域仍存在信息丢失现象,而本文方法则能更好地保留目标区域的细节和纹理特征,分割结果更贴近真实标签.与 TransUNet 相比,TransUNet 在全局特征建模方面有一定优势,但在小目标分割和细节处理上存在不连续问题,而本文方法在全局一致性和小目标区域分割能力上均表现更优.相比 MedT^[26],尽管 MedT 在大目标区域分割上较稳定,但其边界处理存在模糊现象,而本文方法在边界的完整性和精确性方面显著优于 MedT,能更贴近真实标签.SwinUNet^[27] 的提取结果在复杂背景下对细粒度特征效果不理想,在与 UCTransNet 的对比中,虽然 UCTransNet 的分割性能已非常接近真实标签,但在复杂边界区域仍存在轻微不连续现象,而本文方法则表现出更强的适应能力和细节还原能力.总体可见,本文方法通过高效的特征提取和融合策略,显著提升了医学图像分割的精度,特别是在细节处理、小目标分割和边界完整性方面,展示了其卓越的性能和实用价值.

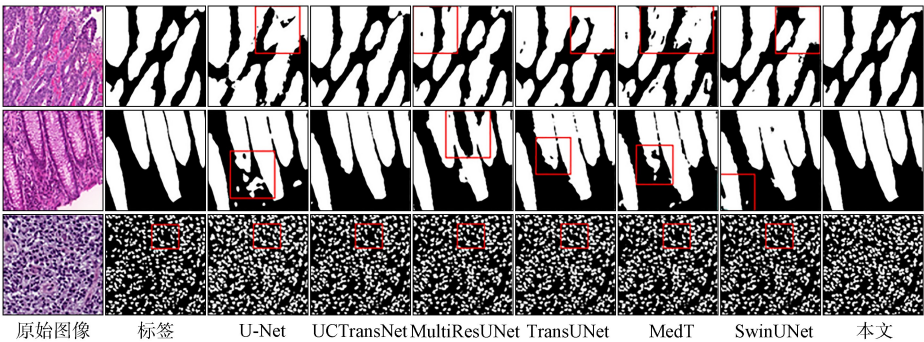


图 5 在数据集 MoNuSeg 和 GlaS 上不同网络分割结果定性比较

Fig. 5 Qualitative comparison of segmentation results of different networks on MoNuSeg and GlaS datasets

2.3 定量比较

表 1 列出了不同网络在数据集 MoNuSeg 和 GlaS 上分割结果的定量比较. 由表 1 可见, 本文方法在数据集 GlaS 和 MoNuSeg 上的分割性能显著优于现有主流模型, 充分展示了其在医学图像分割任务中的卓越性能和强大的鲁棒性. 在数据集 GlaS 上, 本文方法的 Dice 系数达 92.2%, IoU 为 85.53%, 分别比目前性能最好的 UCTransNet 和 USCT-UNet 高 2.02 百分点和 1.04 百分点. 该结果表明, 本文方法在病理图像分割任务中不仅能更精确地捕捉目标区域的细节特征, 还能更好地处理复杂边界, 确保分割结果与真实标签高度一致. 其他方法(如 UNet++, MRUNet, TransUNet 等)尽管在一定程度上提升了图像的分割性能, 但仍存在细节丢失和边界模糊的现象, 而本文方法在这些方面的显著改进突显了其对病理图像分割任务的适配性和优越性. 在数据集 MoNuSeg 上, 本文方法的 Dice 系数为 82.51%, IoU 为 70.20%, 同样优于其他对比方法. 与 MBUTransNet^[28]相比, 本文方法的 Dice 系数提升 0.96 百分点, IoU 提升 1.35 百分点. 与性能较好的 USCT-UNet^[29]相比, 本文方法的 Dice 系数提升 2.44 百分点, IoU 提升 3.44 百分点. 数据集 MoNuSeg 的分割任务聚焦于复杂形态的细胞核分割, 而本文方法能有效处理细胞核在形态、大小及分布上的复杂性, 表明其在多尺度特征提取和语义信息融合方面的优势. 尤其是在预测结果的细节还原能力和边界清晰度方面, 本文方法验证了其在生物医学图像分割任务中的强大适用性.

表 1 不同网络在数据集 MoNuSeg 和 GlaS 上分割结果定量比较

方法	GlaS		MoNuSeg	
	Dice	IoU	Dice	IoU
U-Net	85.45	74.60	76.45	61.88
U-Net++	87.56	77.87	77.01	62.61
AttUNet	88.80	79.86	76.47	61.90
MRUNet	88.73	79.74	78.22	64.23
TransUNet	88.40	79.21	78.53	64.65
SwinUNet	89.58	82.12	77.69	63.52
UCTransNet	90.18	82.96	79.08	65.40
MBUTransNet	90.20	82.15	81.55	68.85
USCT-UNet	91.16	83.76	80.07	66.76
本文	92.20	85.53	82.51	70.20

通过对比经典模型 U-Net 和 UNet++ 以及目前性能较好的先进模型(如 AttUNet, SwinUNet, TransUNet, MedT 等)可见, 本文方法在 Dice 和 IoU 这两个关键指标上均领先, 不仅反映了本文方法在全局语义建模和局部细节保留上的创新性设计, 同时也展现了其对复杂场景的医学图像分割任务的高适配性.

尽管部分方法在性能上接近于本文方法, 但在极端条件下或复杂目标区域的分割任务中, 仍无法达到本文方法的细节还原效果和边界精确度. 本文方法在数据集 GlaS 和 MoNuSeg 上的实验结果表

明, 其能在不同的医学图像分割任务中超越现有主流方法. 无论是在 Dice 系数还是 IoU 指标上, 本文方法均实现了高精度和强鲁棒性的分割性能, 从而进一步验证了其在医学图像分割领域的广泛适用性和实用价值.

使用数据集 Synape 对本文方法进行泛化性分析实验, 与 U-Net, MedT, TransUNet, SwinUNet 和 UCTransnet 等 5 种方法进行比较. 采用平均 Dice 系数(mDice)和平均 Hamming 距离(mHD)作为评价指标. 表 2 列出了不同方法在数据集 Synape 上的分割结果. 由表 2 可见, 本文方法没有在多器官分割上达到最佳性能, 说明本文模型是一种专为病理图像设计的模型, 泛化能力不足. 首要原因是当前模型针对对比度较高的复杂边界区域进行分割, 对低对比度的病灶, 器官的分割泛化性不足; 次要原因是不同医疗中心的设备参数、成像协议差异导致数据分布不一致, 模型未进行针对性优化.

表 2 不同方法在数据集 Synape 上的分割结果

Table 2 Segmentation results of different methods on Synapse dataset

%

方法	mDice	mHD
U-Net	68.81	—
MedT	76.87	39.70
UCTransnet	77.48	31.69
TransUNet	77.49	31.68
SwinUNet	79.12	21.55
本文	78.98	25.89

2.4 消融实验

表 3 列出了本文方法在数据集 MoNuSeg 上的消融实验结果. 由表 3 可见, 通过在 U-Net 架构中逐步引入正则化大核注意力模块、边界感知编码器模块(boundary-aware encoder, BAE)、注意力融合模块(attention fusion, AF)和显式视觉中心模块, 模型的分割性能得到显著提升. 在基础 U-Net 上仅添加 RLKA 模块后, Dice 系数由 76.45%提升至 78.51%, 说明 RLKA 模块在增强特征捕捉能力和全局语义信息建模方面发挥了重要作用. 进一步加入 BAE 模块后, Dice 系数提升至 79.12%, 表明深度卷积模块在优化局部特征表达和细节保留方面有明显贡献. 当结合 RLKA, BAE 和 AF 模块时, Dice 系数达到 81.07%, 展现了注意力融合模块在多尺度特征整合和语义信息提取中的关键作用. 在完整引入 EVC 模块后, 模型的 Dice 系数进一步提升至 82.51%, 验证了 EVC 模块在改进全局与局部特征交互方面的有效性.

表 3 本文方法在数据集 MoNuSeg 上的消融实验结果

Table 3 Ablation experiment results of proposed method on MoNuSeg dataset

RLKA	BAE	AF	EVC	Dice/%
				76.45
✓				78.51
✓	✓			79.12
✓	✓	✓		81.07
✓	✓	✓	✓	82.51

本文通过消融实验考察了 L_2 正则化系数对模型性能的影响. 实验结果表明, 正则化系数在控制模型复杂度和防止过拟合中起关键作用. 表 4 列出了 L_2 正则化系数消融实验结果. 由表 4 可见, 当 L_2 系数为 0.5 时, 模型性能最佳, Dice 系数达 82.51%, 表明中等强度的正则化能在抑制过拟合与保持模型学习能力之间取得平衡. 较小或较大的正则化系数(如 0.1, 0.9)均导致模型性能下降, 可能是由于过拟合抑制不足或模型复杂度被过度限制所致. 该研究为图像分割任务中优化正则化参数提供了重要依据.

综上所述, 针对医学图像分割中存在的特征表征局限性与多尺度信息融合难题, 本文提出了一种 BFFNet 模型, 系统性地构建了多维度优化框架. BFFNet 模型引入了边界感知编码器, 动态适配多样

化感受野,有效提升了对非固定几何结构和感兴趣区域边界特征的捕捉能力.正则化大核注意力模块通过大核深度卷积和深度空洞卷积扩展感受野,全面建模全局和局部特征,并通过正则化策略有效抑制小样本数据集上的过拟合现象.跳跃连接部分采用了显式视觉中心模块,通过轻量级视觉 Transformer 和可学习视觉中心建模特征依赖关系,显著提升了全局与局部信息融合的效果.解码阶段,注意力引导的特征上采样方法优化了局部与全局特征的交互过程,实现了边界细节和小目标区域的精准分割,显著增强了对复杂背景下细粒度语义信息的捕捉能力.实验结果表明,BFFNet 模型在多个公开医学图像数据集上显著优于现有先进方法,展现出卓越的分割性能、鲁棒性和泛化能力.该结果进一步证明了多尺度动态建模、正则化大核注意力和视觉中心模块等机制的有效性,并显示了全局语义信息与局部细节特征融合的重要性.

表 4 L_2 正则化系数消融实验结果

Table 4 Ablation experiment results of L_2 regularization coefficient			%
正则化系数值	Dice	IoU	
0.1	81.92	69.43	
0.3	82.11	69.64	
0.5	82.51	70.22	
0.7	82.22	69.85	
0.9	81.89	69.32	

参 考 文 献

[1] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3431-3440.

[2] NI Z L, BIAN G B, WANG G A, et al. Pyramid Attention Aggregation Network for Semantic Segmentation of Surgical Instruments [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 11782-11790.

[3] WEN Y, XIE K, HE L H. Segmenting Medical MRI via Recurrent Decoding Cell [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 12452-12459.

[4] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]//18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer International Publishing, 2015: 234-241.

[5] WANG H N, CAO P, WANG J Q, et al. Uctransnet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022: 2441-2449.

[6] IBTEHAZ N, RAHMAN M S. MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation [J]. Neural Networks, 2020, 121: 74-87.

[7] LUO Z C, ZHU X S, ZHANG L, et al. Rethinking U-Net: Task-Adaptive Mixture of Skip Connections for Enhanced Medical Image Segmentation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2025: 5874-5882.

[8] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Berlin: Springer International Publishing, 2018: 3-11.

[9] 汪慎文,周瑶. 基于改进 U-Net 的肝脏 MRI 分割方法 [J]. 吉林大学学报(理学版), 2022, 60(5): 1153-1160. (WANG S W, ZHOU Y. Liver MRI Segmentation Method Based on Improved U-NET [J]. Journal of Jilin University (Science Edition), 2022, 60(5): 1153-1160.)

[10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [EB/OL]. (2017-06-12) [2025-01-01]. <https://arxiv.org/abs/1706.03762>.

[11] CHEN J N, LV Y Y, YV Q H, et al. Transunet: Transformers Make Strong Encoders for Medical Image

- Segmentation [EB/OL]. (2021-02-08)[2025-01-15]. <https://arxiv.org/abs/2102.04306>.
- [12] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 10012-10022.
- [13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2117-2125.
- [14] 赵欣, 李森, 李智生. 基于 CNN 和 Transformer 并行编码的腹部多器官图像分割 [J]. 吉林大学学报(理学版), 2024, 62(5): 1145-1154. (ZHAO X, LI S, LI Z S. Abdominal Multi-organ Image Segmentation Based on Parallel Coding of CNN and Transformer [J]. Journal of Jilin University (Science Edition), 2024, 62(5): 1145-1154.)
- [15] LI X M, CHEN H, QI X J, et al. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes [J]. IEEE Transactions on Medical Imaging, 2018, 37(12): 2663-2674.
- [16] ZHANG Y D, LIU H Y, HU Q. Transfuse: Fusing Transformers and CNNs for Medical Image Segmentation [C]//24th International Conference on Medical Image Computing and Computer Assisted Intervention. Berlin: Springer International Publishing, 2021: 14-24.
- [17] DAI J P, QI H Z, XIONG Y W, et al. Deformable Convolutional Networks [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 764-773.
- [18] LI X, HU X L, YANG J. Spatial Group-Wise Enhance: Improving Semantic Feature Learning in Convolutional Networks [EB/OL]. (2019-05-23)[2025-02-10]. <https://arxiv.org/abs/1905.09646>.
- [19] QUAN Y, ZHANG D, ZHANG L Y, et al. Centralized Feature Pyramid for Object Detection [J]. IEEE Transactions on Image Processing, 2023, 32: 4341-4354.
- [20] GUO M H, LU C Z, LIU Z N, et al. Visual Attention Network [J]. Computational Visual Media, 2023, 9(4): 733-752.
- [21] WANG W H, XIE E Z, LI X, et al. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2021: 568-578.
- [22] HUANG H B, ZHOU X Q, CAO J, et al. Vision Transformer with Super Token Sampling [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 22690-22699.
- [23] DAI Y M, GIESEKE F, OEHMCKE S, et al. Attentional Feature Fusion [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway, NJ: IEEE, 2021: 3560-3569.
- [24] SIRINUKUNWATTANA K, PLUIM J P W, CHEN H, et al. Gland Segmentation in Colon Histology Images: The Glas Challenge Contest [J]. Medical Image Analysis, 2017, 35: 489-502.
- [25] KUMAR N, VERMA R, SHARMA S, et al. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology [J]. IEEE Transactions on Medical Imaging, 2017, 36(7): 1550-1560.
- [26] QI Q F, LIN L Y, ZHANG R, et al. MEDT: Using Multimodal Encoding-Decoding Network as in Transformer for Multimodal Sentiment Analysis [J]. IEEE Access, 2022, 10: 28750-28759.
- [27] CAO H, WANG Y Y, CHEN J, et al. Swin-UNet: UNet-Like Pure Transformer for Medical Image Segmentation [C]//European Conference on Computer Vision. Berlin: Springer, 2022: 205-218.
- [28] QIAO J, WANG X, CHEN J, et al. MBUTransNet: Multi-branch U-Shaped Network Fusion Transformer Architecture for Medical Image Segmentation [J]. International Journal of Computer Assisted Radiology and Surgery, 2023, 18: 1895-1902.
- [29] XIE X S, YANG M. USCT-UNet: Rethinking the Semantic Gap in U-Net Network from U-Shaped Skip Connections with Multichannel Fusion Transformer [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024, 32: 3782-3793.