

聚焦边界和多尺度特征融合的脑卒中病灶分割

刘晨红, 李凤莲, 阳佳, 王夙喆, 陈桂军

引用本文

刘晨红, 李凤莲, 阳佳, 王夙喆, 陈桂军. 聚焦边界和多尺度特征融合的脑卒中病灶分割[J]. 计算机科学, 2026, 53(2): 264-272.

LIU Chenhong, LI Fenglian, YANG Jia, WANG Suzhe, CHEN Guijun. [Boundary-focused Multi-scale Feature Fusion Network for Stroke Lesion Segmentation](#) [J]. Computer Science, 2026, 53(2): 264-272.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[多模态水声图像目标视觉检测](#)

Multimodal Visual Detection for Underwater Sonar Target Images

计算机科学, 2026, 53(2): 227-235. <https://doi.org/10.11896/jsjcx.241200082>

[融合多尺度特征和注意力机制的时间序列预测模型](#)

Time Series Forecasting Model Integrating Multi-scale Features and Attention Mechanism

计算机科学, 2026, 53(2): 180-186. <https://doi.org/10.11896/jsjcx.250100113>

[GTKT:融合联通主义学习和多层时序图Transformer的知识追踪模型](#)

GTKT:Knowledge Tracing Model Integrating Connectivism Learning and Multi-layer Temporal Graph Transformer

计算机科学, 2026, 53(2): 78-88. <https://doi.org/10.11896/jsjcx.250700188>

[个性化教育资源推荐综述](#)

Review of Personalized Educational Resource Recommendations

计算机科学, 2026, 53(2): 1-15. <https://doi.org/10.11896/jsjcx.250700184>

[EvR-DETR:融合事件与RGB图像的轻量级端到端目标检测](#)

EvR-DETR:Event-RGB Fusion for Lightweight End-to-End Object Detection

计算机科学, 2026, 53(1): 153-162. <https://doi.org/10.11896/jsjcx.250300021>

聚焦边界和多尺度特征融合的脑卒中病灶分割

刘晨红¹ 李凤莲¹ 阳佳² 王凤喆¹ 陈桂军¹

1 太原理工大学电子信息工程学院 太原 030024

2 中国运载火箭技术研究院研究发展中心 北京 100076

(liuchenhong0402@link.tyut.edu.cn)

摘要 借助计算机辅助诊断技术定位脑卒中发病区域,有助于提升临床医生的诊断与治疗效率。目前,医学图像里脑卒中病变与健康组织边界常不清晰,而现有多数基于深度学习的分割方法在识别小尺寸病变及处理模糊边界方面存在不足。为此,提出了一种创新的边界感知多尺度特征集成网络(Boundary-Aware Multi-Scale Feature Integration Network, BAMFNet),用于更准确地进行脑卒中病灶分割。BAMFNet中设计了多尺度特征提取模块,该模块利用卷积神经网络和Transformer的混合架构来捕获局部和全局多尺度特征,通过内卷积机制有效减少冗余信息。此外,提出了边界增强和融合模块,其在特征融合过程中有效增强了边界区域的特征。融合模块集成了多层次的信息交互机制,增强了边界特征表示,实现了深层特征和浅层特征的有效结合。在ATLAS v1.2, ATLAS v2.0和ISLES 2022卒中数据集上的实验证明, BAMFNet的骰子相似系数分别达到了62.93%, 61.79%和86.66%, 优于对比方法。

关键词: 深度学习; 病灶分割; 多尺度特征融合; 边界增强; Transformer

中图分类号 TP391

Boundary-focused Multi-scale Feature Fusion Network for Stroke Lesion Segmentation

LIU Chenhong¹, LI Fenglian¹, YANG Jia², WANG Suzhe¹ and CHEN Guijun¹

1 College of Electronic Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China

2 R&D Department, China Academy of Launch Vehicle Technology, Beijing 100076, China

Abstract Computer-aided diagnosis helps clinicians locate stroke-affected brain regions, improving diagnostic and therapeutic efficiency. Currently, the boundaries between stroke lesions and healthy tissues in medical images are often unclear, and most existing deep learning-based segmentation methods lack effectiveness in identifying small-sized lesions and handling blurred boundaries. To address this, the boundary-aware multi-scale feature integration network (BAMFNet) is proposed for more accurate stroke lesion segmentation. In BAMFNet, the multi-scale feature extraction module combines convolutional neural networks and Transformers to capture local and global features at multiple scales and uses involution to reduce information redundancy. The boundary enhancement and fusion module strengthens boundary-region features during fusion and integrates a multi-level information interaction mechanism. This enhances the boundary feature representation and combines deep and shallow features effectively. Experiments on the ATLAS v1.2, ATLAS v2.0 and ISLES 2022 stroke datasets show BAMFNet achieves Dice similarity coefficients of 62.93%, 61.79%, and 86.66% respectively, outperforming other methods.

Keywords Deep learning, Lesion segmentation, Multi-scale feature fusion, Boundary enhancement, Transformer

1 引言

脑卒中是一种因脑部血液供应骤然中断或破裂引发脑部供血不足的病症。根据世界卫生组织的数据,全球每年约出现1700万脑卒中病例,约600万人因此死亡,该病症已成为全球第二大死因^[1]。随着人口老龄化加剧及不健康生活方式

的普遍化,全球脑卒中发病率呈上升态势^[2]。医学图像分割在计算机辅助诊断中起着关键作用,可以提供量化、准确、快速的脑部病变信息,帮助医生更好地检测、诊断和治疗患者,并辅助制定个性化的治疗方案和预后方案。相较于计算机断层扫描,磁共振成像(Magnetic Resonance Imaging, MRI)可以提供更为详细的脑部解剖结构。因为卒中往往涉及到脑部的

到稿日期:2025-03-25 返修日期:2025-05-22

基金项目:国家自然科学基金(62171307);山西省自然科学基金(202103021224113)

This work was supported by the National Natural Science Foundation of China(62171307) and Natural Science Foundation of Shanxi Province, China(202103021224113).

通信作者:李凤莲(lifenglian@tyut.edu.cn)

软组织和血管系统, MRI对脑梗死和脑出血等病变的诊断准确性更高。然而,在传统卒中诊疗中,医生依赖肉眼判读影像进行病灶定位费时费力,且低对比度病灶和微小出血灶等非常容易被漏诊,这对临床精准决策构成挑战。

随着人工智能技术的突飞猛进,基于深度学习的方法在医学图像分割领域取得了显著进步。在过去十年中,卷积神经网络(Convolutional Neural Network, CNN)在医学图像处理中得到了广泛应用。U-Net^[3]利用其对称的U型结构,在编码器-解码器中使用跳跃连接捕获图像中的细节信息,使得神经网络可以高效利用有限的医学图像数据进行训练。后续许多方法都是基于U-Net进行改进的。UNet++^[4]集成不同深度的U型结构来探索适合的网络深度,并设计了剪枝方案来加快推理速度。Attention U-Net^[5]将注意力门和U-Net结合,增强模型对前景像素的敏感度。类似的操作还有使用可变形卷积^[6]代替常规卷积,将2D卷积和3D卷积进行结合^[7]等。

尽管U-Net系列模型在医学分割任务中表现优异,但其仍面临挑战——传统CNN的局部卷积特性导致其全局上下文建模能力不足^[8],这一缺陷在卒中病灶分割中尤为突出。由于卒中病灶具有形态不规则、空间分布弥散及解剖位置随机等特点,模型须同时具备局部细节捕捉与全局语义理解能力。为突破CNN的固有局限,研究者进行了多种创新。文献^[9]使用CBAM注意力模块,将其输出作为特征细化模块的输入,通过缩小语义距离来改进U-Net分割模型。文献^[10]将深度拉普拉斯金字塔配准与注意力增强U-Net相结合,提高了卒中病灶的分割精度。MRAU-Net^[11]在解码端创新性地设计了一个多尺度残差注意块,利用不同尺度的感受野捕捉丰富的上下文信息,更准确地定位目标区域。当前研究证实,CNN的局部归纳偏置特性虽有利于医学图像的纹理特征提取,但其有限的感受野仍制约着对复杂解剖结构的整体理解。

基于自注意力机制的Transformer^[12]模型凭借其全局感受野特性,在长距离依赖建模方面展现出显著优势。混合CNN-Transformer的结构成为目前医学图像分割领域的首选模型。TransUNet^[13]的混合CNN-Transformer设计,在多器官分割和心脏分割中超越了各种竞争方法。nnFormer^[14]交错使用卷积层和Transformer层,以学习3D医学图像的表达。但是,这些混合架构仅通过解码阶段的浅层特征拼接实现模态融合,致使CNN捕获的局部细节与Transformer建模的全局上下文难以协同。平衡局部细节保留与全局语义推理,仍是亟待解决的核心问题。

针对医学图像中病灶形态的复杂性与边界模糊性,基于深度学习的边界感知分割方法取得了显著进展。BA-Net^[15]构建金字塔式边缘提取模块,通过交互式注意力桥将多尺度边缘特征与分割主干网络深度融合,能有效利用边界信息。MDCNet^[16]设计了双阶段分割框架,首阶段采用参数化损失抑制不确定边界,次阶段引入高分辨率特征补偿机制,从而产生平滑的轮廓边界。MCI-Net^[17]设计了一个边界校正模块,

可消除多个分支捕获的多尺度特征中的冗余。CTO^[18]首次实现了CNN-Transformer-边界检测器的三重架构耦合,其自生成的边界掩码监督机制指导解码学习过程,从而提高了分割的准确性和效率。

为应对卒中病灶分割的挑战,本文提出一种创新的边界感知多尺度特征融合网络(BAMFNet)。该网络基于经典编码器-解码器框架构建,融合CNN和Transformer架构,实现局部-全局特征互补。进一步地,提出多尺度特征提取模块(Multi-Scale Feature Extraction Module, MSVIT)和边界增强与融合模块(Boundary-Focused Fusion Module, OC-Fusion)。MSVIT通过不同尺寸卷积核和Transformer模块提取多尺度特征,引入内卷积操作代替部分传统的卷积操作,自适应地生成空间位置相关的卷积核,动态调整感受野范围,增强模型对不同尺度特征的捕捉能力。OC-Fusion负责融合辅助编码路径与多尺度特征提取路径的特征,通过频域分解策略将特征图分解为高频边界分量与低频语义分量,采用通道加权融合机制强化病灶边缘响应。此外,在融合过程中融入跨层级多尺度信息交互机制,强化模型的特征表达,实现深浅特征的有效融合。为验证BAMFNet的有效性,在ATLAS v1.2, ATLAS v2.0和ISLES 2022卒中数据集上开展了广泛实验。结果一致表明,BAMFNet相较于对比方法,显著提升了分割性能。

本文主要贡献有3点:

1)针对脑卒中病灶形态复杂、尺度多变的特点,提出了边界感知多尺度特征融合网络(BAMFNet)。该架构通过深度整合局部细节与全局语义信息,有效解决了传统方法在病灶边缘模糊区域的分割难题。

2)设计了多尺度特征提取模块(MSVIT),创新性地将在内卷积机制与改进型Transformer结合。通过并行多分支卷积捕获不同感受野的特征,并利用Transformer的长程依赖建模能力增强上下文信息融合,实现局部细节保留与全局语义推理的平衡。

3)设计了边界增强与融合模块(OC-Fusion),不仅实现了多尺度的信息融合,还通过频域分解策略来强化病灶边缘特征的表达,提升了模型对复杂边界的区分能力。

2 BAMFNet

BAMFNet的结构如图1所示,网络采用编码器-解码器结构。其中,图1(a)为主体分割结构BAMFNet;图1(b)为多尺度特征提取模块MSVIT;图1(c)为特征融合模块OC-Fusion。编码端由辅助编码路径、多尺度特征提取路径和融合路径组成。在辅助编码路径中,通过卷积和下采样提取丰富的局部特征。多尺度特征提取路径中,MSVIT模块负责提取多尺度特征,通过内卷积减少卷积核冗余,利用改进的Transformer结构提取全局信息。融合路径里,OC-Fusion负责实现两条编码路径的多级融合,增强边界特征的表达。解码器包含卷积和上采样过程,如图2(b)所示,最终通过 1×1 卷积实现分割结果的输出。

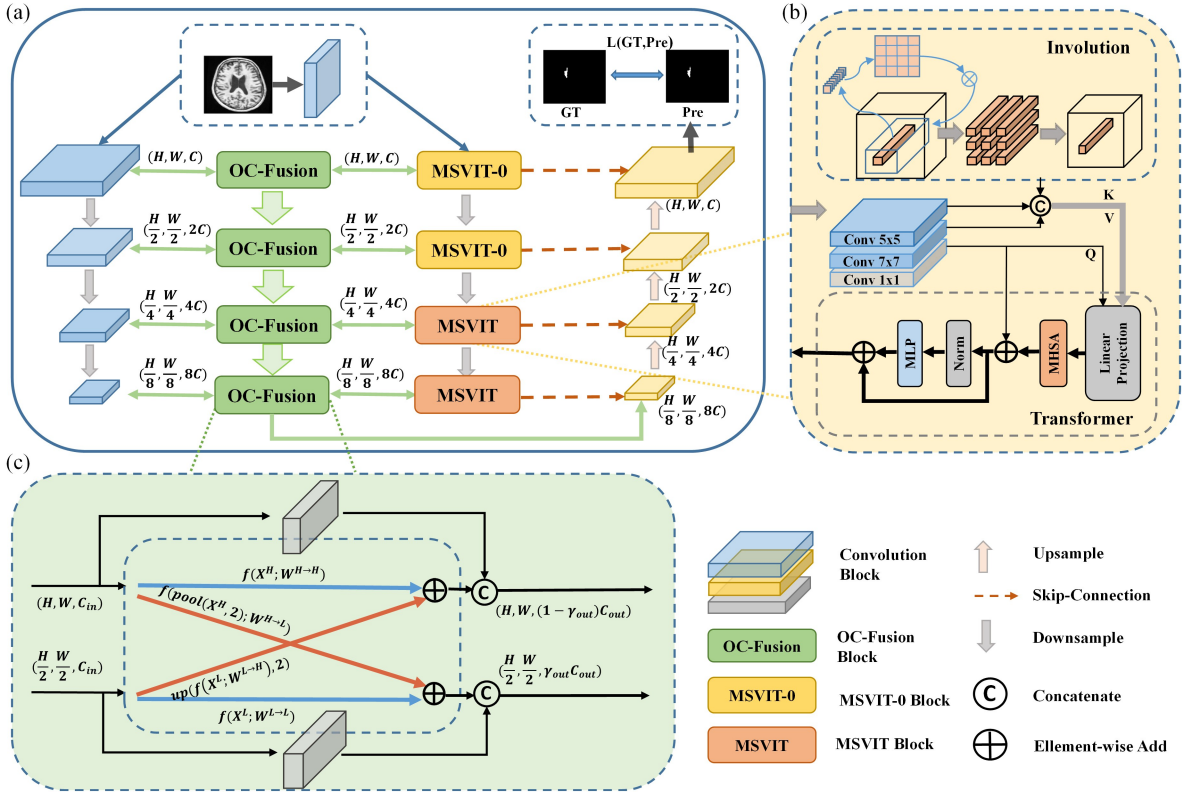


图1 BAMFNet 的整体框图

Fig. 1 Overall block diagram of BAMFNet

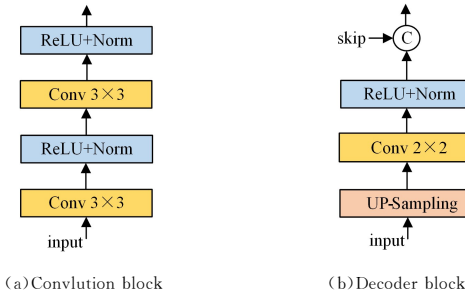


图2 编码模块和解码模块

Fig. 2 Encoding module and decoding module

2.1 辅助编码器

U-Net 凭借其卷积基础架构,在医学图像分割领域取得了显著成就。本文选用 4 层卷积模块构成辅助编码器,以提取图像的局部特征。每个卷积模块包含两组 3×3 核大小的卷积层,搭配 ReLU 层和 Norm 层,且每个模块后紧跟一个 2×2 的池化层用于调整分辨率。卷积模块的结构如图 2(a) 所示,其定义如下:

$$\text{Conv_block}(\mathbf{X}) = \sigma(\text{BN}(\text{Conv}(\mathbf{X}))) \quad (1)$$

其中, $\text{Conv}(\cdot)$ 表示核大小为 3×3 且步长为 1 的二维卷积, $\sigma(\cdot)$ 表示 ReLU 线性函数, $\text{BN}(\cdot)$ 表示批量归一化。输入 $\mathbf{X} \in R^{H \times W \times C}$, 其中 $H \times W$ 表示空间分辨率, C 表示通道数。值得注意的是,输入图像经过第一个卷积模块后没有改变分辨率,只是增加了通道数,丰富了初始输入图像的特征表示。

每一层级辅助编码的输出都作为特征融合模块的输入,为整个编码端提供局部特征,并和多尺度特征提取路径进行信息融合。

2.2 多尺度特征提取模块

卒中病灶大小不一,且呈弥散分布,使用多尺度卷积可以使模型更好地适应不同大小和形状的病灶或器官,减少细节信息的丢失。特别是对于弥散型病灶,通过集成多尺度特征,复杂结构的病灶能得到更全面的特征表示。本文提出的多尺度特征提取模块将输入的特征图进行核大小为 5×5 和 7×7 的卷积来提取多样化的特征。多尺度卷积的计算如下:

$$\mathbf{F}_{\text{cat}} = \text{concat}(\text{Conv}^m(\mathbf{F}_{\text{in}}), \text{Conv}^n(\mathbf{F}_{\text{in}})) \quad (2)$$

其中, $\mathbf{F}_{\text{in}} \in R^{H \times W \times C}$ 表示输入特征, $\text{Conv}^m(\cdot)$ 和 $\text{Conv}^n(\cdot)$ 表示二维卷积运算, m 和 n 表示不同的核大小, $\text{concat}(\cdot)$ 表示通道维度的拼接操作。

使用多尺度卷积会提高模型对输入的理解能力,但也会造成特征冗余。这是由于常规卷积本身具有局限性。针对这一问题,在多尺度卷积的基础上增加内卷积,其核心思想是通过动态生成空间特异性的核参数,实现输入特征的自适应聚合。内卷积中,基于输入特征图动态生成核。对于每个输入特征图,在每个位置生成对应的核。通过两个线性变换来压缩和拓展通道数。核生成函数 $\phi(\cdot)$ 将输入特征映射为动态的卷积核:

$$\mathbf{H}_{i,j} = \phi(\mathbf{X}_{i,j}) = \mathbf{W}_1 \sigma(\mathbf{W}_0 \mathbf{X}_{i,j}) \quad (3)$$

其中, $\mathbf{X}_{i,j}$ 表示位置 (i, j) 处的特征向量; $\mathbf{W}_0(\cdot)$ 和 $\mathbf{W}_1(\cdot)$ 都是线性变换, $\mathbf{X}_0(\cdot)$ 用于压缩通道数, $\mathbf{X}_1(\cdot)$ 用于扩展通道数; $\sigma(\cdot)$ 表示 $\text{ReLU}(\text{BN}(\cdot))$, 即批归一化和非线性激活函数。对于每个位置 (i, j) ,使用核 $\mathbf{H}_{i,j}$ 对以当前点为中心的 $K \times K$ 区域进行逐元素相乘操作,然后将区域内的信息进行求和,得到输出特征 $\mathbf{Y}_{i,j}$:

$$Y_{i,j} = \sum_{u,v} H_{i,j}[u,v] X_{i+u,j+v} \quad (4)$$

其中, (u, v) 是相对于中心点 (i, j) 的偏移量。

内卷积的核在空间上是特异的, 即在不同的位置生成不同的核, 能够自适应地提取更多的视觉信息。与此同时, 在输出的通道维度共享核, 可以减少核的冗余。与传统的卷积核不同, 内卷积的核是根据输入特征动态生成的, 是一个数据依赖的操作, 能够更好地适应不同的输入特征。增加内卷积机制后, 多尺度特征提取路径的输出特征如下:

$$F'_{\text{cat}} = \text{concat}(\text{Conv}^m(F_{\text{in}}), \text{Conv}^n(F_{\text{in}}), H(F_{\text{in}})) \quad (5)$$

其中, $H(\cdot)$ 表示内卷积操作。

本文设计的多尺度特征提取模块 MSVIT-0 的结构如图 3(a) 所示。使用内卷积和多尺度卷积, 不仅实现了多尺度特征的获取, 还能够在更广阔的空间聚合上下文, 在一定程度上克服远距离建模的困难。进一步地, 为了充分利用输入的信息, 也为了缓解网络退化问题, 引入了残差结构。输入特征 F_{in} 通过一个核大小为 1×1 的卷积, 与拼接后的特征 F'_{cat} 进行相加, 得到 MSVIT-0 最终的输出特征为:

$$F_{\text{out}} = \sigma(\text{Conv}(F'_{\text{cat}} + \text{Conv}_{1 \times 1}(F_{\text{in}}))) \quad (6)$$

其中, $\sigma(\cdot)$ 表示 $\text{ReLU}(\text{BN}(\cdot))$, $\text{Conv}(\cdot)$ 表示卷积操作。

为了进一步增强特征提取能力, 同时整合局部和全局信息, 在 MSVIT-0 结构的基础上引入了改进的 Transformer, 设计了 MSVIT 结构, 如图 3(b) 所示。MSVIT 的详细内部结构如图 1(b) 所示。

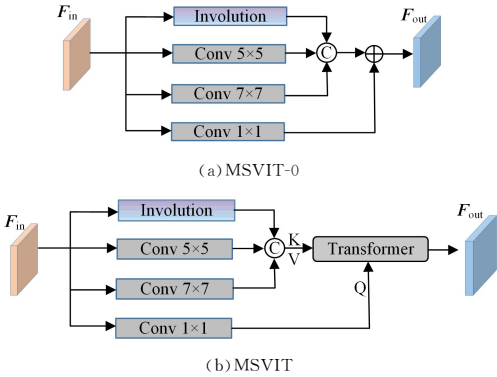


图 3 多尺度特征提取模块

Fig. 3 Multi-scale feature extraction module

Transformer 结构的查询、键和值通常是固定的输入矩阵经过线性投影得到, 所得特征维度较高。为了更积极地关注全局特征, 提高 Transformer 结构对输入中不同层次特征的理解能力, MSVIT 模块将分别从不同层次的特征中提取查询和键, 以综合利用全局和局部信息。

改进后的 Transformer 的输入由拼接特征 $F_{\text{cat}} \in \mathbf{R}^{H \times W \times C}$ 和初始输入特征 $F_{\text{in}} \in \mathbf{R}^{H \times W \times C}$ 组成, 二者经过一个线性变换和位置编码得到嵌入矩阵 $F_{\text{embedat}} \in \mathbf{R}^{N \times D}$ 和 $F_{\text{embedin}} \in \mathbf{R}^{N \times D}$, 其中 $N = H/P \times W/P$, $D = P^2 C$, H 和 W 表示特征图的高和宽, C 表示通道数, P 表示补丁的大小。 F_{embedat} 作为多头自注意力机制 (Multi-Head Self-Attention, MHSA) 的键 \mathbf{K} 和值 \mathbf{V} , F_{embedin} 作为查询 \mathbf{Q} 。MHSA 每个头执行如下操作:

$$\text{MHSA}_i(F_{\text{embedat}}, F_{\text{embedin}}) = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{D_h}}\right) \mathbf{V}_i \quad (7)$$

其中, $\mathbf{Q}_i = F_{\text{embedin}} \mathbf{W}_i^Q$, $\mathbf{K}_i = F_{\text{embedat}} \mathbf{W}_i^K$ 和 $\mathbf{V}_i = F_{\text{embedat}} \mathbf{W}_i^V$ 是输入的线性变换; \mathbf{W}_i^Q , \mathbf{W}_i^K 和 \mathbf{W}_i^V 是权重矩阵 (每个头 i 的权重不同), 且 $D_h = D/h$ 。Transformer 的输出可以表示为:

$$(\mathbf{Z}_i)_l = \text{MHSA}(\text{LN}(F_{\text{embedin}}, F_{\text{embedat}})) + F_{\text{embedin}} \quad (8)$$

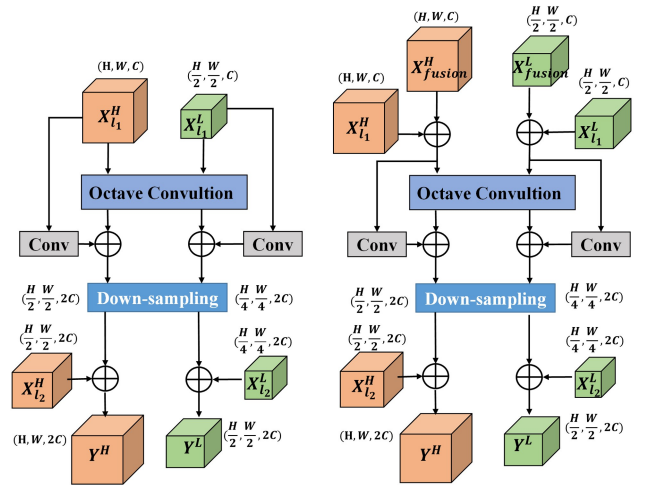
$$(\mathbf{Z}_i)_{l+1} = \text{MLP}(\text{LN}((\mathbf{Z}_i)_l)) + (\mathbf{Z}_i)_l \quad (9)$$

其中, $(\mathbf{Z}_i)_l$ 和 $(\mathbf{Z}_i)_{l+1}$ 表示第 l 层 MHSA 和 MLP 的输出, $\text{LN}(\cdot)$ 表示层归一化操作, $\text{MLP}(\cdot)$ 表示多层感知机。

采用以上结构, 利用多尺度特征提取模块实现特征提取, 并使用内卷积机制减少核冗余。利用改进的 Transformer 结构不仅强化了模型对全局特征的关注, 还提高了对局部细节的捕捉能力, 有助于增强特征提取的准确性和鲁棒性。多尺度特征提取模块的输出特征被送入融合模块 OC-Fusion, 结合辅助编码路径的特征, 提高模型对输入的建模能力。

2.3 特征融合模块

高频特征图蕴含了轮廓和边缘等关键细节, 这对于病灶检测至关重要; 而低频特征图包含相对较少的信息。为了增强边缘特征的表达, 集来自两个编码路径的信息, 设计了 OC-Fusion 模块, 其结构如图 1(c) 所示。OC-Fusion 模块主要利用八度卷积^[19], 通过集成多个层次的信息, 显著增强了特征融合过程中的表达能力。根据不同阶段的网络需求, OC-Fusion 模块接收数量和分辨率差异多样的输入特征。如图 4(a) 所示, 在模型的浅层阶段 (BAMFNet 编码路径的第一层), OC-Fusion 模块接收来自两个不同层次 l_1 和 l_2 的特征输入, 这两个层级的分辨率和通道数均不相同; 如图 4(b) 所示, 在模型的深层阶段 (BAMFNet 编码路径中第一层以下的层级), OC-Fusion 模块接收来自 3 个不同层次的特征输入, 即 l_1 , l_2 以及上一层级融合模块的输出, 这 3 个层级的分辨率和通道数均不相同。这种设计允许 OC-Fusion 模块在不同的网络深度灵活地融合不同分辨率和通道数的特征, 从而能够更好地处理多尺度信息, 增强网络对特征的捕捉能力。



(a) OC-Fusion module at the start of the model (b) OC-Fusion module in the middle of the model

图 4 不同阶段的 OC-Fusion

Fig. 4 OC-Fusion at different stages

OC-Fusion 模块将 CNN 辅助编码路径输出的特征作为低频输入, 将多尺度特征提取路径的特征作为高频输入。这

样划分,是由于多尺度特征提取路径捕获的远距离信息对于小病灶和弥散型病灶的检测更为重要。低频和高频的特征经过八度卷积进行信息交流。低频特征图 $\mathbf{X}^L \in \mathbf{R}^{H/2 \times W/2 \times C_m}$ 的分辨率是高频特征图 $\mathbf{X}^H \in \mathbf{R}^{H \times W \times C_m}$ 的一半,所以不同频率在进行信息交流之前需要进行分辨率调整。高频输入 \mathbf{X}^H 经过卷积核 $\mathbf{W}^{H \rightarrow H}$ 后,得到输出 $\mathbf{Y}^{H \rightarrow H}$;同时, \mathbf{X}^H 经过下采样后,再经过卷积核 $\mathbf{W}^{H \rightarrow L}$,得到输出 $\mathbf{Y}^{H \rightarrow L}$ 。计算过程如下:

$$\mathbf{Y}_{p,q}^{H \rightarrow H} = \sum_{i,j \in N_k} (\mathbf{W}_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{H \rightarrow H})^T \mathbf{X}_{p+i, q+j}^H \quad (10)$$

$$\mathbf{Y}_{p,q}^{H \rightarrow L} = \sum_{i,j \in N_k} (\mathbf{W}_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{H \rightarrow L})^T \text{pool}(\mathbf{X}_{p+i, q+j}^H) \quad (11)$$

其中, (p, q) 表示特征图上的位置, $N_k = \{(i, j); i, j = \{-(k-1)/2, \dots, (k-1)/2\}\}$ 是一个集合, k 为奇数, $\text{pool}(\cdot)$ 表示池化操作。

同理,低频输入 \mathbf{X}^L 经过两条卷积路径得到 $\mathbf{Y}^{L \rightarrow L}$ 和 $\mathbf{Y}^{L \rightarrow H}$ 。计算过程如下:

$$\mathbf{Y}_{p,q}^{L \rightarrow L} = \sum_{i,j \in N_k} (\mathbf{W}_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{L \rightarrow L})^T \mathbf{X}_{p+i, q+j}^L \quad (12)$$

$$\mathbf{Y}_{p,q}^{L \rightarrow H} = \sum_{i,j \in N_k} (\mathbf{W}_{i+\frac{k-1}{2}, j+\frac{k-1}{2}}^{L \rightarrow H})^T \text{upsample}(\mathbf{X}_{p+i, q+j}^L) \quad (13)$$

其中, $\text{upsample}(\cdot)$ 表示上采样操作。经八度卷积后,输出为:

$$\mathbf{Y}^H = \mathbf{Y}^{H \rightarrow H} + \mathbf{Y}^{L \rightarrow H} \quad (14)$$

$$\mathbf{Y}^L = \mathbf{Y}^{L \rightarrow L} + \mathbf{Y}^{H \rightarrow L} \quad (15)$$

每个特征融合模块的输出为:

$$(\mathbf{Y}^H)' = (\mathbf{Y}^{H \rightarrow H} + \mathbf{Y}^{L \rightarrow H}) \oplus (\text{Conv}_{1 \times 1}(\mathbf{X}^H)) \quad (16)$$

$$(\mathbf{Y}^L)' = (\mathbf{Y}^{L \rightarrow L} + \mathbf{Y}^{H \rightarrow L}) \oplus (\text{Conv}_{1 \times 1}(\mathbf{X}^L)) \quad (17)$$

其中,符号 \oplus 表示元素点加, $\text{Conv}_{1 \times 1}(\cdot)$ 表示核大小为 1×1 的卷积操作。

OC-Fusion 模块不仅实现了多尺度的信息融合,强化了高频特征所包含的轮廓和边缘特征的表达,还加强了网络中多层次之间的信息交流。在深部结构中使用 Transformer 会造成注意力崩溃,而 OC-Fusion 模块通过促进多层之间的信息交换,降低了这种风险^[20]。

3 实验及分析

3.1 实验数据及预处理

在本文实验中,使用了 ATLAS v1.2, ATLAS v2.0 和 ISLES 2022 数据集评估本文方法的有效性。

ATLAS 系列数据集是一个缺血性卒中病变的公开数据集,其中 ATLAS v1.2 包含 240 张 T1 加权磁共振成像, ATLAS v2.0 包含 955 张 T1 加权磁共振成像^[21] (655 个具备真值标签, 316 个隐藏了标签信息)。ISLES 2022 数据集包含了 400 个急性和亚急性缺血性卒中的 MRI 数据 (250 个具备真值标签, 150 个没有标签信息)。ISLES 2022 数据集中每个病例都包含 DWI, ADC 和 FLAIR 这 3 种成像序列^[22]。这 3 个数据集包含了广泛的缺血性卒中疾病谱,包括来自多个中心的数据、更复杂的梗塞模式和不同解剖位置的病变,可以更加有效地验证所提方法的泛化性能。

对于所涉及的各个数据集,均采用随机划分的方式,按照 7:2:1 的比例将其划分为训练集、验证集以及测试集。本文中,将所有三维 MRI 脑卒中数据均从 Z 轴方向进行切片,所有样本在输入网络前均经过 Z-score 标准化处理和裁剪

操作。对于 ATLAS v1.2, 每个病例的大小均为 233×197 , 将每张切片调整为 192×192 的分辨率; 对于 ATLAS v2.0, 原始数据的大小均不固定, 将每张切片调整为 224×224 的分辨率; 对于 ISLES 2022, 将其大小调整至 96×96 。此外, 在训练阶段, 采用随机水平/垂直平移 (幅度为图像宽高的 $\pm 10\%$)、旋转变换 (角度范围为 $[-15^\circ, +15^\circ]$) 及缩放操作 (缩放因子为 $0.8 \sim 1.2$) 进行数据增强。

3.2 实验设置

所有实验都基于深度学习框架 TensorFlow-GPU 2.5 和 Python 3.7, 并在 Nvidia RTX3080 (10 GB) 上进行。选择 Adam 优化器进行参数优化。实验中, 初始学习率、权重衰减系数和训练周期分别设置为 1×10^{-3} , 1×10^{-5} 和 100。使用早停策略防止过拟合, 使用 MultiStepLR 作为学习率衰减策略。此外, 在训练时设置随机种子确保结果可复现。

为了综合考虑区域级别和像素级别的差异, 采用骰子损失和二元交叉熵损失相结合的方法来训练网络。其计算式如下:

$$L_{\text{seg}} = \sum \left(\alpha \left(1 - \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|} \right) + \beta (-y \log(\hat{y}) - (1 - y) \log(1 - \hat{y})) \right) \quad (18)$$

其中, y 代表地面真值, \hat{y} 表示预测结果, α 和 β 为权重系数。

3.3 评价指标

本文采用了 Dice、精确度和召回率作为评价指标, 来评估拟议方法的性能。这些指标的计算式如下:

$$\text{Dice} = \frac{2TP}{FP + 2TP + FN} \times 100\% \quad (19)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (20)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (21)$$

其中, TP 表示真阳性, FP 表示假阳性, TN 表示真阴性, FN 表示假阴性。Dice 计算两个区域之间的相似度, 是上述指标中最重要的。本研究给出的评估结果基于所有切片的平均值。

3.4 实验结果及分析

3.4.1 参数实验 1: MSVIT 的核尺寸选择

卷积核尺寸的选取直接影响模型对局部细节与全局语义特征的捕获能力。较小尺寸的卷积核更擅长捕捉细粒度纹理特征, 而较大尺寸的卷积核则能有效建模病灶区域的空间上下文关系。基于 ATLAS v1.2 数据集, 系统对比了不同卷积核配置下模型的分割性能, 并引入内卷积以验证其与常规卷积的互补性。实验结果如图 5 所示。

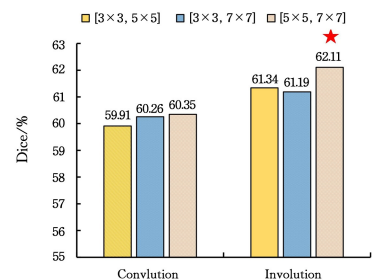


图 5 MSVIT 中不同卷积核尺寸的实验结果

Fig. 5 Experimental results of different convolution kernel sizes in MSVIT

在未引入内卷积的情况下,使用 $[5 \times 5, 7 \times 7]$ 卷积核配置时,Dice系数为60.35%,显示出较好的细粒度特征提取能力; $[3 \times 3, 5 \times 5]$ 和 $[3 \times 3, 7 \times 7]$ 配置下的Dice系数分别为59.91%和60.26%,分割性能有所下降。引入内卷积后,分割性能普遍提升,例如在 $[3 \times 3, 5 \times 5]$ 配置下,Dice系数从59.91%提升至61.34%。这表明内卷积能够有效增强模型对局部细节和全局语义特征的捕获能力。本文选择 $[5 \times 5, 7 \times 7]$ 卷积核尺寸,并结合内卷积进行特征提取。

3.4.2 参数实验2:MSVIT的放置位置选择

为了探究MSVIT模块在不同层级对分割性能的影响,将其放置在骨干网络的不同位置(其余编码模块采用MS-VIT-0)。实验结果如表1所列。可以看出,当MSVIT模块仅放置在单一层级时,分割性能较为平稳,其中第4层的Dice系数达到61.54%,Precision和Recall分别为62.06%和61.96%。然而,当MSVIT模块同时放置在编码端的第3和第4层时,分割性能得到了显著提升,Dice系数提升至62.59%,Precision达到63.88%,Recall为65.53%。这表明,在第3和第4层同时放置MSVIT模块,能够更有效地利用多尺度特征,从而显著提升图像分割的性能。

表1 MSVIT模块放置在不同位置上的性能对比

Table1 Comparison of performance when placing MSVIT modules in different locations

放置层级	Dice/%	Precision/%	Recall/%
1	61.15	58.89	55.68
2	61.58	57.56	58.77
3	61.46	62.11	58.73
4	61.54	62.06	61.96
3+4	62.59	63.88	65.53

3.4.3 参数实验3:OC-Fusion的参数选择

在融合模块OC-Fusion中,通过调整参数 γ 控制低频特征的使用比例。图6展示了不同 γ 值对分割性能的影响。随着 γ 值从0逐渐增加到0.75,分割性能逐渐提升,Dice系数从59.75%上升到62.59%,Precision从61.30%上升到63.88%,Recall从60.29%上升到65.53%。这一趋势表明,在初始阶段,随着低频特征的逐渐引入,模型能够更好地利用辅助编码路径提供的细节信息,从而更准确地还原分割结果。然而,当 γ 进一步增加到1,即仅使用低频特征时,各项指标均有所下降。这表明,在病灶分割任务中,完全依赖低频特征反而会降低分割性能。综上,在融合模块OC-Fusion中,适度引入低频特征($\gamma=0.75$)可以提高病灶分割的准确性。

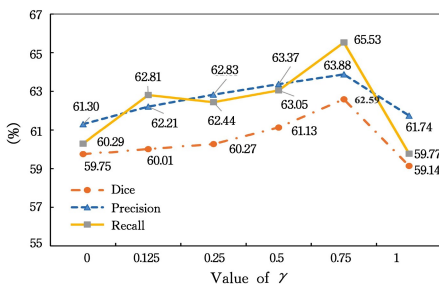


图6 融合模块中低频控制参数 γ 的性能比较

Fig. 6 Performance comparison of low-frequency control parameter γ in fusion modules

3.4.4 对比实验

为了验证所提出的卒中病灶分割方法的优越性,将其与已经在ATLAS和ISLES数据集上测试过的最先进的分割方法进行比较。由于部分对比方法未提供源码,为确保实验的公平性与可重复性,本文仅复现了已公开代码的方法,并确保这些方法与本文方法在相同的实验条件下运行。对于未复现的方法,本文直接引用了原始文献中报告的评估结果。不同方法在ATLAS v1.2,ATLAS v2.0和ISLES 2022数据集上的评估结果如表2—表4所列。

表2 不同分割方法在ATLAS v1.2上的评价结果

Table 2 Evaluation results of different segmentation methods on ATLAS v1.2 dataset

方法	Dice	Precision	Recall
U-Net ^[3]	48.34	54.45	53.68
Attention U-Net ^[5]	49.60	49.25	62.53
X-Net ^[23]	48.67	60.00	47.52
D-UNet ^[7]	53.49	63.31	52.43
CLCI-Net ^[24]	58.10	64.90	58.10
TransUNet ^[13]	56.23	57.15	65.95
TransFuse ^[25]	58.13	57.64	60.06
TransRender ^[26]	59.79	63.91	62.08
BAMFNet	62.93	65.95	67.33

表3 不同分割方法在ATLAS v2.0上的评价结果

Table 3 Evaluation results of different segmentation methods on ATLAS v2.0 dataset

方法	Dice	Precision	Recall
U-Net ^[3]	41.73	58.01	36.04
Residual U-Net ^[27]	45.64	59.25	42.05
Attention U-Net ^[5]	48.72	63.68	43.93
TransAttn U-Net ^[28]	57.24	66.02	56.58
U-Net Transformer ^[29]	58.35	65.91	59.16
BAMFNet	61.79	64.66	60.33

表4 不同分割方法在ISLES 2022上的评价结果

Table 4 Evaluation results of different segmentation methods on ISLES 2022 dataset

方法	Dice	Precision	Recall
U-Net ^[3]	82.04	85.31	81.44
Attention U-Net ^[5]	81.45	84.70	80.98
TransUNet ^[13]	84.23	86.88	84.19
TransFuse ^[25]	84.39	87.36	84.15
TransRender ^[26]	85.37	86.48	83.24
BAMFNet	86.66	87.39	85.71

实验结果显示,在ATLAS v1.2,ATLAS v2.0以及ISLES 2022数据集上,BAMFNet方法的综合表现均优于其他对比方法。特别是在ISLES 2022数据集上,其Dice系数高达86.66%,显著优于其他方法。在ATLAS v1.2数据集上,BAMFNet的Dice系数为62.93%,Recall指标达到67.33%,显示出该方法强大的病灶召回能力。尽管BAMFNet在ATLAS v2.0数据集上的Precision为64.66%,略低于部分对比方法(这可能与该数据集来自多个站点、不同扫描设备导致的图像质量差异有关),但其在3个数据集上的Dice

系数和 Recall 均为最优,显示出了多尺度特征提取的优势和边界特征增强的有效性。与同样采用 CNN-Transformer 混合架构进行语义分割的 TransFuse 相比,本文方法在各个评估指标上均取得了更好的效果。

3.4.5 对比实验可视化分析

图 7 和图 8 展示了不同卒中分割方法在 ATLAS v1.2 和 ISLES 2022 卒中数据集上获得的视觉分割结果。其中,红色

是真实标签,绿色是模型的分割结果。通过参考地面真值,可以看出本文方法有很大的视觉优势。本节实验在两个数据集各可视化了 4 个卒中患者切片。这些病灶的位置、形状均不相同,根据每张切片包含病灶像素的个数进行大小分类,根据是否包含多处病灶进行弥散型判定。在两张可视化结果中,(a)行均为小病灶,(b)行均为大病灶,(c)行和(d)行均为弥散型病灶。

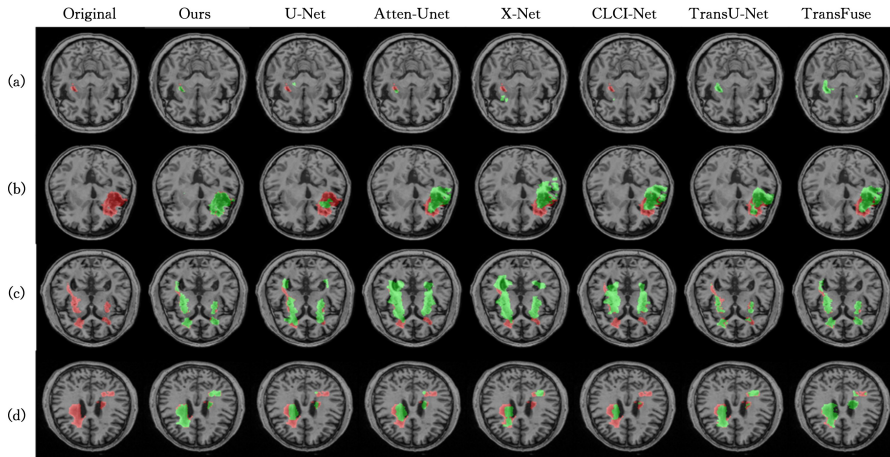


图 7 不同方法在 ATLAS v1.2 数据集上的可视化分割结果(电子版为彩图)

Fig. 7 Visualization segmentation results of different methods on ATLAS v1.2 dataset

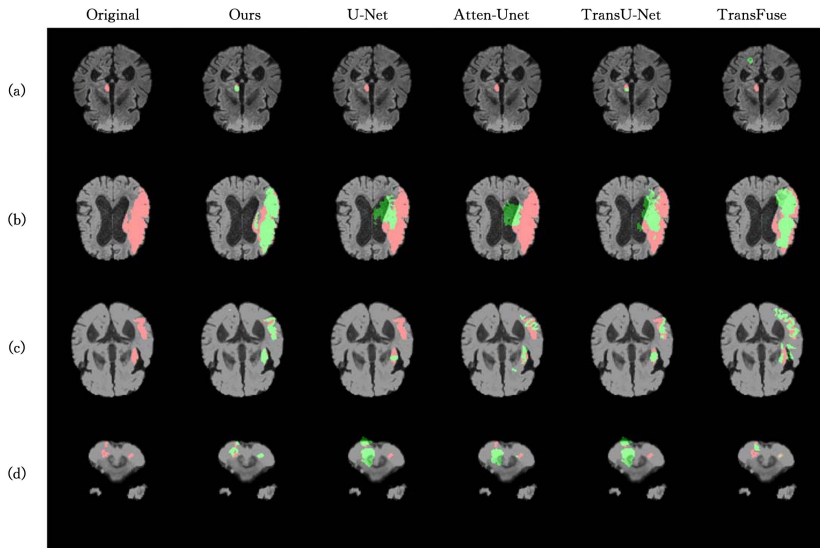


图 8 不同方法在 ISLES 2022 数据集上的可视化分割结果(电子版为彩图)

Fig. 8 Visualization segmentation results of different methods on ISLES 2022 dataset

以图 7 为例,本文方法对于小病灶(a)均实现了正确检测,且检测范围相对准确;基于 CNN 的方法(U-Net,X-Net 和 CLCI-Net)获得的分割结果和地面真值相距甚远;基于 CNN-Transformer 的方法(TransUNet 和 TransFuse)虽然覆盖了地面真值,但存在过分割的问题。对于大病灶(b),所列方法均能和地面真值部分重叠,但本文方法除了实现病灶区域的覆盖,对边界区域的检测也较为精确。对于弥散型病灶(c)和(d),其边界区域极不规则,且病灶位置的关联性差,使得分割具有挑战性。基于 CNN-Transformer 的方法的分割表现优于基于 CNN 的方法,但仍然存在边界处分割效果差的问题,且对于部分病灶存在漏检。本文方法可以检测到分割范围内

的全部病灶,虽然在边界处仍然存在一定程度的欠分割,但边界距离依旧是最小的。

总体而言,本文方法在不同大小的病灶分割中均表现出显著优势。对于具有挑战性的弥散型病灶,本文方法也能保证病灶的检测。综合来看,本文方法具有分割不同类型病灶的能力。

3.4.6 消融实验

为了探究 MSVIT 模块和 OC-Fusion 模块的有效性,本节进行了消融实验研究。实验中采用 U-Net 作为基线模型,逐个添加不同的模块来验证每个模块的有效性。主要比较了以下几种组合结构的性能。

U-Net:使用 U-Net 进行卒中病灶分割。将其设置为基线模型。

U-Net+MSVIT:用 MSVIT 代替 U-Net 的编码模块,其余部分不变。

U-Net+MSVIT+Fusion:在上述模型的基础上,添加一条 CNN 辅助编码路径,并对两条编码路径在通道维度进行简单的拼接。

U-Net+MSVIT+OC-Fusion:在上述模型的基础上,引入 OC-Fusion 模块,用于两条编码路径的特征融合。

实验结果如表 5 所列。

表 5 消融实验结果

Table 5 Results of ablation experiment

					(%)		
基线	MSVIT	Fusion	OC-Fusion	Dice	Precision	Recall	
				48.34	54.45	53.68	
U-Net	✓			56.45	55.73	54.72	
	✓	✓		57.38	56.26	57.93	
	✓		✓	60.38	61.28	63.34	

通过在基线模型上添加本文提出的模块,验证每个模块对基线模型性能的提升作用。结果表明,引入 MSVIT 模块后,U-Net 的分割精度有一定提升,Dice 系数、准确率和召回率均有所增长,说明 MSVIT 能增强模型对病变特征的提取能力。在加入 MSVIT 的基础上,添加 CNN 辅助编码路径并进行简单拼接融合,模型性能进一步提高,表明多路径特征提取和简单融合有助于提升分割精度。而采用 OC-Fusion 模块后,模型性能更优,Dice 系数提升至 60.38%,准确率和召回率分别为 61.28%和 63.34%,这说明 OC-Fusion 模块在特征融合方面更为有效,能够更好地综合多路径特征,提高分割的准确性。

结束语 本文提出了一种创新的边界感知多尺度特征融合网络 BAMFNet,显著提升了脑卒中病灶分割的性能。BAMFNet 结合了 CNN 和 Transformer 技术,能够有效地捕捉图像中的局部和全局特征,突破了传统深度学习方法在识别小型且弥散病变方面的局限性。设计的多尺度特征提取模块 MSVIT 采用混合架构,实现了多尺度下的局部和全局特征的有效提取,并通过内卷积机制优化了信息处理流程,减少了核冗余。为了进一步加强特征融合,引入了边界增强融合模块 OC-Fusion,它不仅促进了不同尺度信息的整合,还特别增强了边界区域特征的表达,确保了深层和浅层特征的有效融合。在多个数据集上的实验验证了 BAMFNet 的优越性,可视化结果进一步证实了 BAMFNet 在识别小病灶和处理模糊边界方面的卓越性能。展望未来,计划探索更轻量级的网络结构以降低计算成本;同时,将继续优化特征提取过程,以进一步提升分割性能;此外,还将扩大 BAMFNet 的应用范围,使其适应更多样化的语义分割任务。期望 BAMFNet 能够成为临床医生更精准的辅助诊断工具,从而显著提高脑卒中病灶的检测和治疗效率。

参考文献

[1] LIU M B, HE X Y, YANG X H, et al. Summary of "China Cardiovascular Health and Disease Report 2023": Prevalence of

Cardiovascular Diseases and Status of Interventional Diagnosis and Treatment [J]. Chinese Journal of Interventional Cardiology, 2024, 32(10): 541-550.

- [2] FEIGIN V, BRAININ M, NOPRVING B, et al. World Stroke Organization(WSO): Global Stroke Fact Sheet 2022 [J]. International Journal of Stroke: Official Journal of the International Stroke Society, 2022, 17(1): 18-29.
- [3] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C] // Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [4] ZHOU Z, SIDDIQUEE M, TAJBAKHAS N, et al. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation [J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.
- [5] OKTAY O, SCHLEMPE J, FOLGOC L, et al. Attention U-Net: Learning Where to Look for the Pancreas [C] // Medical Imaging with Deep Learning, 2022.
- [6] LEI T, WANG R, ZHANG Y, et al. DefED-Net: Deformable Encoder-Decoder Network for Liver and Liver Tumor Segmentation [J]. IEEE Transactions on Radiation and Plasma Medical Sciences, 2022, 6(1): 68-78.
- [7] ZHOU Y, HUANG W, DONG P, et al. D-UNet: A Dimension-Fusion U Shape Network for Chronic Stroke Lesion Segmentation [J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021, 18(3): 940-950.
- [8] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(4): 640-651.
- [9] WANG L B, WANG S M. Fundus vascular image segmentation algorithm based on attention mechanism [J]. Computer Science, 2019, 51(S2): 359-364.
- [10] LI X W, XU W X, CHEN Y, et al. Joint image registration for automatic segmentation of stroke lesions [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2024, 36(4): 729-737.
- [11] SHU X, LI X, ZHANG X, et al. MRU-net: Multi-scale residual attention U-shaped network for medical image segmentation [J]. Computers and Electrical Engineering, 2024, 118: 109479.
- [12] VASWANI A, SHAZEER N, PARMARR N, et al. Attention is All you Need [C] // Advances in Neural Information Processing Systems, 2017.
- [13] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation [J]. arXiv: 2102.04306, 2021.
- [14] ZHOU H Y, GUO J, ZHANG Y, et al. nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer [J]. IEEE Transactions on Image Processing, 2023, 32: 4036-4045.
- [15] WANG R, CHEN S, JI C, et al. Boundary-aware context neural network for medical image segmentation [J]. Medical Image Analysis, 2022, 78: 102395.
- [16] LI C, MAO Y, GUO Y, et al. Multi-Dimensional Cascaded Net with Uncertain Probability Reduction for Abdominal Multi-Organ Segmentation in CT Sequences [J]. Computer Methods and

Programs in Biomedicine, 2022, 221:106887.

- [17] XIE X, PAN X, SHAO F, et al. MCI-Net: Multi-scale context integrated network for liver CT image segmentation[J]. Computers and Electrical Engineering, 2022, 101:108085.
- [18] LIN Y, ZHANG D, FANG X, et al. Rethinking Boundary Detection in Deep Learning Models for Medical Image Segmentation [C]// Information Processing in Medical Imaging. Cham: Springer, 2023:730-742.
- [19] CHEN Y, FAN H, XU B, et al. Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks With Octave Convolution[C]// IEEE/CVF International Conference on Computer Vision. 2019:3434-3443.
- [20] ZHOU D, KANG B, JIN X, et al. DeepViT: Towards Deeper Vision Transformer[J]. arXiv:2103.11886, 2021.
- [21] LIEW S, LO B, DONNELLY M, et al. A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms[J]. Scientific Data, 2022, 9(1):320.
- [22] HERNANDEZ M, DE L, HANNING U, et al. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset[J]. Scientific Data, 2022, 9(1):762.
- [23] QI K, YANG H, LI C, et al. X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies[C]// Medical Image Computing and Computer Assisted Intervention. 2019:247-255.
- [24] YANG H, HUANG W, QI K, et al. CLCI-Net: Cross-Level Fusion and Context Inference Networks for Lesion Segmentation of Chronic Stroke[C]// Medical Image Computing and Computer Assisted Intervention. 2019:266-274.
- [25] ZHANG Y, LIU H, HU Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation[C]// Medical Image

Computing and Computer Assisted Intervention. 2021:14-24.

- [26] WU Z, ZHANG X, LI F, et al. TransRender: a transformer-based boundary rendering segmentation network for stroke lesions[J]. Frontiers in Neuroscience, 2023, 17:1259677.
- [27] NI Z, BIAN G, ZHOU X, et al. RAUNet: Residual Attention U-Net for Semantic Segmentation of Cataract Surgical Instruments [C]// Neural Information Processing. 2019:139-149.
- [28] CHEN B, LIU Y, ZHANG Z, et al. TransAttUnet: Multi-Level Attention-Guided U-Net With Transformer for Medical Image Segmentation[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(1):55-68.
- [29] PETIT O, THOME N, RAMBOUR C, et al. U-Net Transformer: Self and Cross Attention for Medical Image Segmentation [C]// Machine Learning in Medical Imaging. 2021:267-276.



LIU Chenhong, born in 2000, postgraduate, is a member of CCF (No. Z5771G). Her main research interest is medical image segmentation.



LI Fenglian, born in 1972, Ph.D, professor, Ph.D supervisor, is a member of CCF (No. 65512S). Her main research interest is medical signal processing and analysis.

(责任编辑:柯颖)