

文章编号: 1007-5321(2025)05-0151-08

DOI: 10.13190/j.jbupt.2024-169

基于 Transformer 的轻量级脑肿瘤图像分割算法

侯蓓蓓, 关赛宗, 王亚敏

(河南理工大学 计算机科学与技术学院, 焦作 454000)

摘要: 脑肿瘤分割是医学图像分析领域的关键任务,且脑肿瘤及其亚区域具有复杂和不规则的边界结构。针对现有脑肿瘤分割算法在建模长距离依赖和资源开销方面的不足,提出了一种轻量级的脑肿瘤图像分割算法。具体而言,设计了一种融合卷积神经网络(CNN)与Transformer的混合轻量级编码器。在浅层特征提取阶段引入深度可分离卷积,以降低计算复杂度;在深层特征提取阶段,提出局部-全局双分支特征提取模块(SFB),结合Transformer与ShuffleNet v2的优势,实现局部与全局上下文信息的融合。此外,在Transformer模块中引入轻量级全局注意力模块和局部感知模块,分别用于建模长距离依赖关系和增强局部特征感知能力。最后,在BraTS 2019数据集上进行了充分的实验。实验结果表明,该模型在仅使用0.98 M参数数量和54.60 G每秒浮点运算次数的条件下,分别在整体肿瘤(WT)、肿瘤核心(TC)和增强肿瘤(ET)区域达到了93.1%、92.2%和91.2%的戴斯相似性系数。在保持极低计算成本的同时,模型在各关键区域的分割精度整体优于现有主流方法,实现了在效率与性能之间的最佳权衡。

关键词: 轻量级; Transformer; 脑肿瘤; 图像分割

中图分类号: TP391.4

文献标志码: A

Lightweight Brain Tumor Image Segmentation Algorithm Based on Transformer

HOU Beibei, GUAN Saizong, WANG Yamin

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000)

Abstract: Brain tumor segmentation is a key task in medical image analysis due to the heterogeneous and irregular nature of tumor regions. To address the limitations of existing methods in modeling long-range dependencies and reducing resource consumption, we propose a lightweight segmentation model based on a hybrid convolutional neural network (CNN) and Transformer encoder. Depthwise separable convolutions are employed in shallow layers to reduce computation, while the proposed shuffle former block (SFB) integrates Transformer and ShuffleNet v2 to effectively capture both global and local context. Furthermore, lightweight attention modules are introduced to model long-range dependencies and enhance local perception. Experimental results on the BraTS 2019 dataset demonstrate that our model achieves Dice scores of 93.1% in whole tumor (WT), 92.2% in tumor core (TC), and 91.2% in enhancing tumor (ET), with only 0.98 M parameters and 54.60 G floating point operations per second, achieving a superior balance between segmentation accuracy and computational efficiency for deployment in resource-constrained clinical settings.

Key words: lightweight; Transformer; brain tumors; image segmentation

收稿日期: 2024-08-18

基金项目: 河南省科技攻关项目(242102211042); 河南理工大学博士基金项目(B2022-14)

作者简介: 侯蓓蓓(1992—), 女, 讲师, 硕士生导师, 邮箱: houbeibei0120@hpu.edu.cn.

在医学图像分析的各种任务中,脑肿瘤分割引起了研究界的广泛关注^[1]。脑胶质瘤由于其症状隐蔽,且脑组织的特殊性质限制了手术和治疗方法,导致高发病率和致死率^[2]。核磁共振成像(MRI, magnetic resonance imaging)在脑部肿瘤的诊断和治疗方面发挥着重要作用。临床上,常用的 MRI 序列有 t1, t2, t1ce, flair, 这些序列提供了丰富的互补信息。通常,脑胶质瘤包含水肿区域、坏死核心与非增强肿瘤核心区域、增强肿瘤核心区域^[3]。通过脑部多模态 3 维磁共振自动、准确地分割这些恶性肿瘤,对于临床诊断具有重要意义^[4]。

近年来,卷积神经网络(CNN, convolutional neural networks)已广泛应用于图像分割领域^[5]。U 型网络(U-Net, u network)^[6]是在全卷积神经网络^[7]的基础上采用对称的编-解码结构,显著提高医学图像分割精度。许多 U-Net 的变体,如 U-Net++^[8], Attention U-Net^[9], Res-UNet^[10] 等网络,均取得了良好的性能。但由于 CNN 的局部感受野,模型的全局上下文信息提取能力有限。基于此,分割 Transformer (SETR, segmentation transformer)^[11]、Segmenter^[12]、检测 Transformer (DETR, detection transformer)^[13] 等模型使用 Transformer 结构代替堆叠卷积层,捕获长距离依赖关系。然而在分割中,局部特征和全局特征都至关重要^[14]。因此,学者们提出了结合 CNN 和 Transformer^[15] 的方法。TransUNet^[16]、TransFuse^[17]、基于 Transformer 的脑肿瘤分割网络(TransBTS, transformer for brain tumor segmentation)^[18]、Swin-UNet^[19]、收缩 3 重注意力(STA-Former, shrinkage triplet attention in a hybrid CNN-transformer model)^[20] 等模型通过 3 维 CNN 在体积空间上提取局部特征,将每个体素作为 Token 输入 Transformer 进行全局特征建模,实现对 3 维图像的全分辨率分割。但是,由于模型参数庞大、计算成本高、内存占用量大等问题,限制了它们在移动医疗应用等场景中的可行性。在此背景下, MobileNet^[21]、ShuffleNet^[22]、Mobile-Former^[23]、SeaFormer^[24] 等轻量化模型通过各种策略实现语义分割任务中精度和推理延迟之间的有效权衡。然而 Transformer 中的自注意力机制在处理长序列时存在较高的时间和空间复杂度。

针对上述问题,提出了一种基于 Transformer 的轻量化脑肿瘤 3 维分割算法。算法采用轻量级的编码器,分为浅层特征提取和深层特征提取 2 部分。

在浅层特征提取阶段,采用了深度可分离卷积。在深层特征提取阶段中引入了局部-全局双分支特征提取模块(SFB, shuffle former block),将 Transformer 与 ShuffleNet v2 相结合。此外,在 Transformer 模块中引入轻量级的注意力模块和局部感知模块,分别用来捕获长距离依赖关系和强化局部特征感知。进一步提升了在实际应用中的效果和性能。所提算法模型架构如图 1 所示。

1 算法描述

1.1 局部-全局双分支特征提取模块

在图 1 中,编码器逐渐将输入图像编码为高级特征,输入图像的尺寸(D, W, H)逐渐缩小为($D/16, W/16, H/16$),通道维度 $C=4$ 最终变为 $C=128$ 。编码器分为浅层特征提取、深层特征提取 2 部分。为减少模型计算复杂度,在浅层特征提取阶段,模型使用 $3 \times 3 \times 3$ (步长为 1) 的深度可分离卷积(DWConv, depthwise separable convolution)对多模态融合数据进行初步特征提取,可表示为

$$\mathbf{X} = f_0(X_0) \quad (1)$$

对于给定的输入图像 X_0 ,通过浅层特征提取操作 $f_0(\cdot)$,得到图像的初始特征矩阵 $\mathbf{X} \in R^{C \times D \times W \times H}$ 。为了更好地集成局部和全局语义信息,在深层特征提取阶段,所提模型使用了局部-全局双分支特征提取模块 SFB。通过改进 ShuffleNet v2 下采样架构,使用 MetaFormer^[25] 架构替代左分支的卷积,目的是在获得局部信息的同时,也能捕获全局上下文信息。MetaFormer 作为从 Transformer 中抽象出来的通用架构,通过不指定具体的 Token 混合器,为研究者提供了一种新的思路。

SFB 右分支使用了 $1 \times 1 \times 1$ Conv, $3 \times 3 \times 3$ DWConv, $1 \times 1 \times 1$ Conv 来获取输入特征的局部信息。左分支使用块嵌入和 Transformer 模块来获取全局上下文信息。定义经过 2 条分支得到的特征矩阵分别是 $\mathbf{X}, \mathbf{Y} \in R^{C \times D \times W \times H}$, SFB 使用拼接融合了 2 条分支的局部信息和全局上下文信息:

$$\mathbf{Z} = S_E(S_{\text{shuffle}}(C_{\text{oncat}}(\mathbf{X}, \mathbf{Y}))) \in R^{C \times D \times W \times H} \quad (2)$$

其中: S_{shuffle} 表示通道混洗操作, C_{oncat} 表示拼接操作,随后通过 S_E 注意力机制调整每个通道的权重,最终获得特征 \mathbf{Z} 。

最后,使用 $1 \times 1 \times 1$ 卷积来降低输入特征图的深度,减小计算复杂度,降低网络参数量,得到最终特征矩阵 \mathbf{Z}' 为

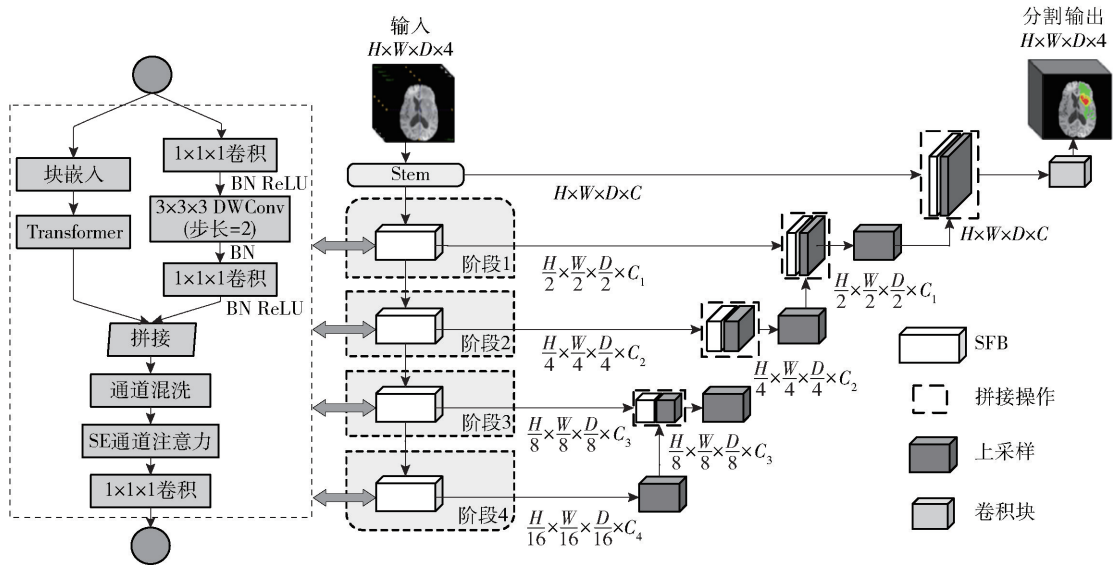


图 1 所提算法模型整体架构图

$$Z' = C_{\text{ov}}(Z, 1 \times 1 \times 1) \quad (3)$$

1.2 Transformer 模块

为了在 SFB 中构建轻量且快速的 Transformer 分支,所提模型基于 MetaFormer 架构提出了一种新的 Transformer 模块。如图 2 所示,Transformer 模块由局部感知模块(LPM, local perception module)和长距离注意力模块(LAB, long-distance attention block)堆叠构成。

给定输入特征 I , 首先经过块嵌入层, 通过使用 3 维卷积操作, 使用了 $2 \times 2 \times 2$ 的卷积核, 步幅为 2, 映射为嵌入向量 $X \in R^{C \times D \times W \times H}$:

$$X = E_{\text{mb}}(I) \quad (4)$$

其中 E_{mb} 表示块嵌入。

LAB 保留了 MetaFormer 的部署优势, 在保证模型轻量的同时, 也获得了 Transformer 块的优秀性能。LAB 部分是为了建模长距离依赖关系而设计

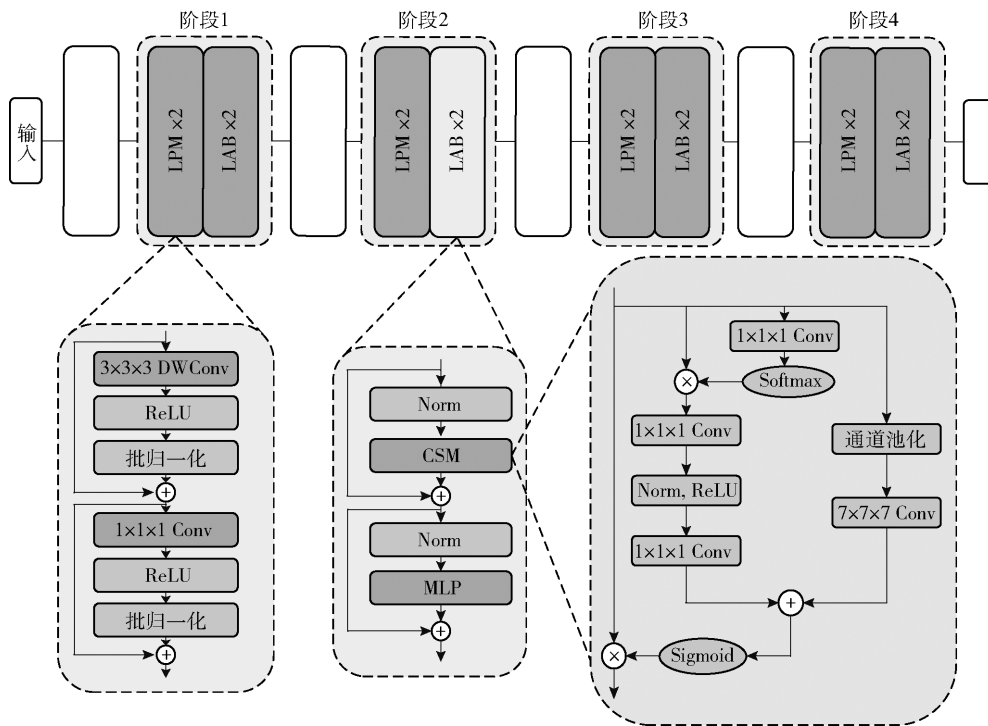


图 2 Transformer 分支的整体分层结构图

的,它引入了一种通道与空间注意力联合模块(CSM, channel-spatial module),将特征提取在通道和空间上分开进行。这种设计使得模型能够更好地理解数据中的全局上下文信息,并且能够关注对当前任务更为重要的区域。LAB 可以表述为

$$Y = C_{SM}(L_N(X)) + X \quad (5)$$

$$Z = M_{LP}(L_N(Y)) + Y \quad (6)$$

其中: L_N 表示归一化层,LAB 遵循 MetaFormer 的通用架构, C_{SM} 是 MetaFormer 架构中基于注意力的令牌混合器模块,用数学表达式表示为

$$L(X) = W_{v2} \text{ReLU}(L_N(W_{v1}(W_k(X)))) \quad (7)$$

$$g(X) = C_{ovv}(A_p(X, 7 \times 7 \times 7)) \quad (8)$$

$$X' = X \otimes \sigma(L(X)) \oplus g(X) \quad (9)$$

其中: $L(X)$ 表示上下文分支, W_k 表示上下文建模, W_{v1}, W_{v2} 表示 $1 \times 1 \times 1$ 卷积运算, $g(X)$ 表示空间分支,其中 A_p 表示平均池化。

为了解决 Transformer 容易忽视局部信息的问题,提出 LPM 模块。LPM 使用残差结构和深度可分离卷积操作来捕捉图像中的局部信息。给定输入特征矩阵 X ,输出特征矩阵 Z :

$$Y = B_N(\text{ReLU}(D_w(X))) + X \quad (10)$$

$$Z = B_N(\text{ReLU}(P_w(Y))) + Y \quad (11)$$

其中: D_w 表示分组卷积,分组数等于通道数, P_w 使用 $1 \times 1 \times 1$ 的卷积核对深度上的每个位置进行卷积。总之,LPM 是在保持计算效率的同时,提高模型对输入数据的局部特征提取能力,确保图像细节不会在层层传递中被模糊或丢失。

最后,采用了 $LPM \times 2 + LAB \times 2$ 的堆叠策略,实现了局部和全局信息的融合。

2 实验及分析

2.1 实验环境

所有实验均在 NVIDIA GeForce RTX 3090 图形处理器(GPU, graphics processing unit)硬件平台上进行,并采用 PyTorch 1.12.1 深度学习框架以构建和评估所提出的算法模型。鉴于 3 维医学数据的复杂性,原始数据被裁剪至 $192 \times 192 \times 140$ 。在训练阶段,通过实施数据增强策略(包括随机裁剪、翻转、强度偏移、缩放等)提升模型的鲁棒性。模型的输入数据和网络输出的尺寸均统一设置为 $128 \times 128 \times 128$ 。模型的分割结果通过 3 线性插值方法恢复到裁剪前的原始尺寸($192 \times 192 \times 140$),并与相同尺寸的标签图像($192 \times 192 \times 140$)进行对比,进

而计算各项性能指标。由于受到 GPU 内存限制,实验选择了较小的批大小,设定为 2 进行模型训练,总的训练周期为 300。优化器选用随机梯度下降优化器,其初始学习率设置为 0.001。

2.2 数据集和评估指标

实验使用官方的 BraTS 2019、BraTS 2020 和 BraTS 2021 的数据集进行训练和验证,标签结果为整体肿瘤(WT, whole tumor)、肿瘤中心(TC, tumor core)和增强肿瘤(ET, enhancing tumor)。

由于 BraTS 2019 的训练集在 BraTS 2018 的基础上增加了 50 例数据,实验将 BraTS 2018 数据集随机分为训练集(256)、验证集(29),并将新增的 50 例数据作为测试集。另外,使用了 BraTS 2020 数据集进行训练,从 BraTS 2021 数据集中随机抽取了 50 个病例作为独立验证集。

对于医学图像,戴斯相似性系数(D_{dice} , Dice similarity coefficient)对 mask 的内部填充比较敏感,而 95% 豪斯多夫距离(D_{hd95} , 95% Hausdorff distance)对分割出的边界比较敏感。因此,实验同时使用 D_{dice} 和 D_{hd95} 来评价分割精度。

D_{dice} 是一种常用的评估图像分割算法精度的指标,特别是在医学图像分割中,其表达式为

$$D_{dice} = (2|P \cap T|) / (|P| + |T|) \quad (12)$$

其中: P 和 T 分别表示预测和真实的肿瘤区域, $|P|$ 和 $|T|$ 分别表示预测和真实肿瘤区域的像素数。 D_{dice} 越接近 1,表示预测和真实的结果越相似。

D_{hd95} 则是另一种评价图像分割算法精度的指标,它衡量的是 2 个非空点集之间的最远距离。在图像分割中,其表达式为

$$D_{hd95} = \max(h_{95}(P, T), h_{95}(T, P)) \quad (13)$$

其中 $h_{95}(P, T)$ 表示从预测肿瘤区域 P 到真实肿瘤区域 T 的豪斯多夫距离的 95th 百分位数:

$$h_{95}(P, T) = \max_{p \in P} \min_{t \in T} d(p, t) \quad (14)$$

其中 $d(p, t)$ 表示点 p 和点 t 之间的欧氏距离。

表 1 在 BraTS 2019 数据集上的消融实验

基线模型	LAB	LPM	SFB	D_{dice}	D_{hd95}
✓	-	-	-	0.892	4.285
✓	✓	-	-	0.908	3.750
✓	-	-	✓	0.905	4.206
✓	✓	-	✓	0.911	3.345
✓	✓	✓	-	0.914	3.381
✓	✓	✓	✓	0.922	2.876

2.3 消融实验结果

为了验证所提算法对类别不平衡的脑肿瘤图像分割的有效性,在 BraTS 2019 数据集上进行消融实验,评估了不同模块对模型性能的影响。基线模型在 MetaFormer 的基础上进行了改进和扩展,采用了 MetaFormer 作为编码器,并增加了相应的解码器模块。

由表 1 可知,基线模型的 D_{dice} 为 0.892, D_{hd95} 为 4.285。基线模型加入 SFB 架构后, D_{dice} 升至 0.905, D_{hd95} 降至 4.206。在基线模型中引入 LAB 模块, D_{dice} 升至 0.908, D_{hd95} 降至 3.750。

当同时加入 LAB 和 SFB 模块时,模型性能进一步提升, D_{dice} 为 0.911, D_{hd95} 为 3.345。同时,引入 LAB 和 LPM 模块, D_{dice} 为 0.914, D_{hd95} 为 3.381。最后,当加入所有模块时,模型性能达到最佳, D_{dice} 为 0.922, D_{hd95} 降至 2.876。以上结果表明,所提算法的每个模块均对提高模型性能具有显著贡献。

2.4 对比实验结果

为了验证所提算法的有效性,所提算法与 V 型网络 (V-Net, v-shaped network)^[26]、TransBTS、NestFormer^[27]、U 型 Transformer (UNETR, u-net transformer)^[28]、伴注意力的扩张分层解耦卷积网络 (ADHDC-Net, dilated hierarchical decoupled convolution network with attention)^[29]、3 维轻量化卷积网络 (3D UX-Net, 3 dimensional lightweight convnet)^[30] 和 UNETR ++^[31] 算法在相同的数据集和实验环境下进行实验对比,结果如表 2 ~ 表 5 所示。

表 2 不同算法在 BraTS 2019 数据集上的 D_{dice} 结果

方法	Param/ M	FLOPs/ G	D_{dice}			
			WT	TC	ET	均值
V-Net	9.450	100.760	0.874	0.883	0.870	0.875
NestFormer	10.290	209.020	0.922	0.926	0.895	0.914
UNETR	148.560	41.190	0.915	0.895	0.896	0.902
UNETR ++	19.979	137.520	0.926	0.916	0.900	0.914
ADHDC-Net	0.300	25.800	0.785	0.860	0.834	0.826
TransBTS	30.630	177.740	0.827	0.841	0.839	0.836
3D UX-Net	10.310	389.700	0.928	0.919	0.898	0.915
笔者方法	0.980	54.600	0.931	0.922	0.912	0.922

由表 2 可知,所提算法在 WT, TC, ET 上分别取得了 93.1%, 92.2% 和 91.2% 的 D_{dice} , 平均 D_{dice} 为

表 3 不同算法在 BraTS 2019 数据集上的 D_{hd95} 结果

方法	D_{hd95}			
	WT	TC	ET	均值
V-Net	6.453	4.500	2.490	4.481
NestFormer	3.266	3.262	2.038	2.855
UNETR	5.322	5.075	2.626	4.341
UNETR ++	4.411	4.464	3.118	3.998
ADHDC-Net	12.703	10.072	4.798	9.191
TransBTS	12.584	16.852	13.083	14.173
3D UX-Net	3.731	2.781	2.062	2.858
笔者方法	4.258	2.670	1.700	2.876

表 4 不同算法在 BraTS 2020 数据集上的 D_{dice} 结果

方法	D_{dice}			
	WT	TC	ET	均值
V-Net	0.895	0.879	0.869	0.881
NestFormer	0.900	0.882	0.859	0.880
ADHDC-Net	0.845	0.861	0.823	0.843
TransBTS	0.823	0.838	0.818	0.827
笔者方法	0.905	0.900	0.882	0.896

表 5 不同算法在 BraTS 2020 数据集上的 D_{hd95} 结果

方法	D_{hd95}			
	WT	TC	ET	均值
V-Net	9.208	5.668	4.182	6.353
NestFormer	7.298	5.492	3.387	5.392
ADHDC-Net	9.724	9.467	7.630	8.940
TransBTS	12.534	14.142	9.586	12.087
笔者方法	6.196	4.353	3.364	4.638

92.2%, 总体优于其他几种算法。平均 D_{dice} 比 V-Net, TransBTS, NestFormer, UNETR, ADHDC-Net, 3D UX-Net 和 UNETR ++ 算法分别领先了 4.7%, 8.6%, 0.8%, 2.0%, 9.6%, 0.7% 和 0.8%。在 TC 区域 NestFormer 表现最佳, 为 91.4%。在模型复杂度方面, 虽然 ADHDC-Net 仅使用 0.30 M 参数 25.80 G 浮点运算, 但是由于 ADHDC-Net 只关注局部特征信息, 导致分割性能较低。相比而言, 所提算法模型的参数数量 (Param, parameters) 和每秒浮点操作数 (FLOPs, floating-point operations per second) 在相差不大的同时, 仅使用了 0.98 M 参数和 54.60 G

FLOPs 达到了更高的性能指标。

由表 3 可知,所提算法在 3 个区域上的 D_{hd95} 分别为 4.258, 2.670 和 1.700。其中,平均 D_{hd95} 为 2.876,与最优的 NestFormer 相差 0.021。这可能因为 NestFormer 使用了嵌套模态感知特征聚合模块来实现多模态融合。但是,这加大了模型的复杂度,相比而言,所提算法具有较少的参数数量和计算复杂度,在提高分割性能的同时,保持了较高的效率。

由表 4 和表 5 可知,所提算法在 BraTS 2020 数据集进行训练,在 BraTS 2021 数据集随机抽取的 50 例验证表明,所提算法在 3 个区域的 D_{dice} 和 D_{hd95} 均表现最优。实验说明,所提算法在脑肿瘤分割中的性能更为鲁棒,更为准确。

如图 3 所示,可以清楚地观察到所提算法在 BraTS 2019 测试集中的 50 个样本在 WT, TC, ET 上的 D_{dice} 。其中,3 个区域的中位数分别达到 94%, 96%, 93%。外部包裹的就是核密度图,区域对应图形面积越大,该区域值附近分布的概率越大,可以看出,模型的结果具有较高的可靠性,不太容易出现偶然性的波动。

2.5 可视化结果分析

如图 4 所示,使用可视化软件 ITK-SNAP 对比所提算法与 V-Net, TransBTS, NestFormer, UNETR 这 4 种算法在 BraTS 2019 数据集上的分割掩膜与标签掩膜。图 4 中绿色为浮肿区域(ED, peritumoral edema)、黄色为 ET、红色为坏疽(NET, necrotic and non-enhancing tumor)。可以发现,所提算法在脑肿

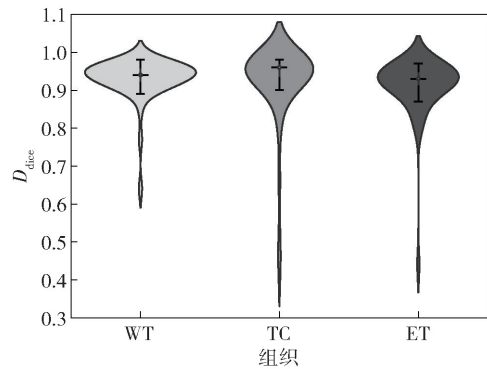


图 3 在 BraTS 2019 测试集上的戴斯系数小提琴图

瘤及瘤周水肿区域分割上更接近标签。通过引入 LAB 和 LPM 模块,模型有效捕获长距离依赖与局部特征,结合 SFB 融合局部与全局语义信息,从而提升分割精度与结果质量。

如图 5 所示,所提算法与 TransBTS, UNETR 等算法的分割结果在随机 2 维切片 60 层和 70 层中进行差异热力图对比,蓝色区域集中显示了模型分割结果与标签不一致的部分。对于其他算法,蓝色区域较为明显且范围更广,特别是在肿瘤边界复杂或形状不规则的区域,分割误差较大。而所提算法的热力图中蓝色部分显著减少,说明模型在这些区域的分割更精确,特别是在边界和关键区域对标签的拟合度更高。

3 结束语

针对脑肿瘤图像分割的需求,提出了一种基于

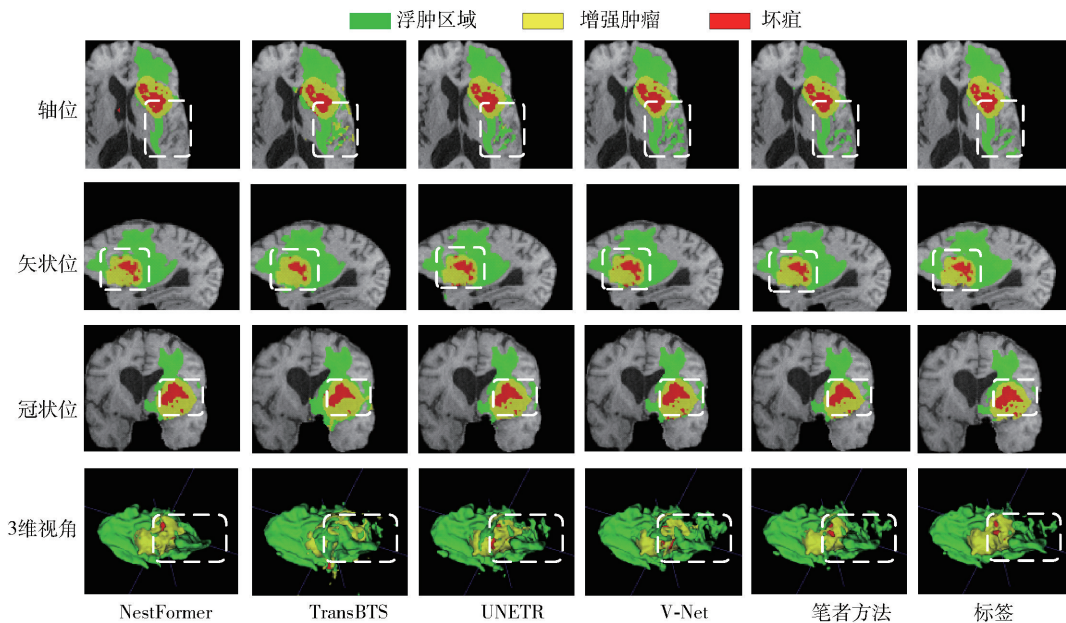


图 4 各算法分割掩膜与标签掩膜的对比图

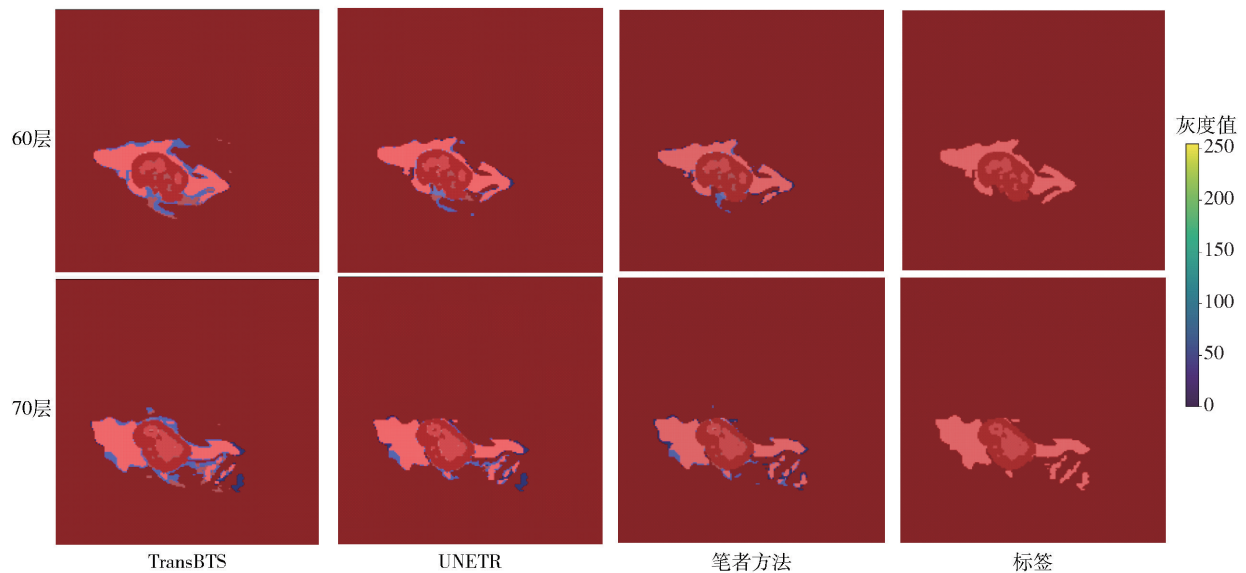


图 5 各算法在 BraTS 2019 测试集上的 2 维切片差异热力图

Transformer 的轻量级 3 维医学图像分割算法, 通过结合 CNN 和 Transformer 的优点, 实现了快速、准确的分割。模型仅使用 0.98 M 参数和 54.60 G 浮点运算, 在 BraTS 2019 数据集上 WT、TC 和 ET 的戴斯系数分别达到 93.1%、92.2% 和 91.2%。编码器采用 2 阶段设计: 在浅层特征提取阶段, 采用深度可分离卷积进行初步特征提取; 在深层特征提取阶段, 使用 SFB 同时获取图像的局部信息和全局上下文信息。总体而言, 所提算法在保持分割性能的同时, 有效降低了计算成本, 适用于脑肿瘤分割的移动端应用场景。下一步将探索将 Mamba 机制引入模型中, 以增强模型对长程依赖的建模能力。

参考文献:

- [1] LIU Z, TONG L, CHEN L, et al. Deep learning based brain tumor segmentation: A survey [J]. *Complex and Intelligent Systems*, 2023, 9(1): 1001-1026.
- [2] MILLER K D, OSTROM Q T, KRUCHKO C, et al. Brain and other central nervous system tumor statistics [J]. *CA: A Cancer Journal for Clinicians*, 2021, 71(5): 381-406.
- [3] 褚晶辉, 李晓川, 张佳琪, 等. 一种基于级联卷积网络的三维脑肿瘤精细分割[J]. *激光与光电子学进展*, 2019, 56(10): 67-76.
CHU J H, LI X C, ZHANG J Q, et al. A fine segmentation method for 3D brain tumors based on cascaded convolutional networks [J]. *Laser and Optoelectronics Progress*, 2019, 56(10): 67-76.
- [4] ZHU Z, SUN M, QI G, et al. Sparse dynamic volume TransUNet with multi-level edge fusion for brain tumor segmentation [J]. *Computers in Biology and Medicine*, 2024: 108284.
- [5] ISM A, DIREKOĞLU C, ŞAH M. Review of MRI-based brain tumor image segmentation using deep learning methods [J]. *Procedia Computer Science*, 2016, 102: 317-324.
- [6] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C] // *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, Oct 5-9, 2015, Proceedings, Part III 18*, Springer International Publishing, 2015: 234-241.
- [7] LONG J, SHEHAMEER E, DARRELL T. Fully convolutional networks for semantic segmentation [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3431-3440.
- [8] ZHOU Z, RAHMAN S M M, TAJBAKHS N, et al. U-Net ++: A nested U-Net architecture for medical image segmentation [C] // *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, Springer International Publishing, 2018: 3-11.
- [9] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas [J]. *arXiv preprint arXiv: 1804. 03999*, 2018.

- [10] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-Net[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(5): 749-753.
- [11] ZHENG S, LU J, ZHAO H, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 6881-6890.
- [12] STRUDEL R, GARCIA R, LAPTEV I, et al. Segformer: Transformer for semantic segmentation[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 7262-7272.
- [13] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//*European Conference on Computer Vision*. Cham: Springer International Publishing, 2020: 213-229.
- [14] HEIDARI M, KAZEROONI A, SOLTANY M, et al. Hiformer: Hierarchical multi-scale representations using transformers for medical image segmentation[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023: 6202-6212.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. *arXiv preprint arXiv: 2010.11929*, 2020.
- [16] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. *arXiv preprint arXiv: 2102.04306*, 2021.
- [17] ZHANG Y, LIU H, HU Q. Transfuse: Fusing transformers and CNNs for medical image segmentation[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference, Strasbourg, France, Sep 27-Oct 1, 2021, Proceedings, Part I 24*, Springer International Publishing, 2021: 14-24.
- [18] WENXUAN W, CHEN C, et al. Transbts: Multimodal brain tumor segmentation using transformer[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021: 109-119.
- [19] CAO H, WANG Y, CHEN J, et al. Swin-UNet: U-Net-like pure transformer for medical image segmentation[C]//*European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 205-218.
- [20] LIU Y, HAN L, YAO B, et al. STA-former: Enhancing medical image segmentation with shrinkage triplet attention in a hybrid CNN-transformer model[J]. *Signal, Image and Video Processing*, 2024, 18(2): 1901-1910.
- [21] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. *arXiv preprint arXiv: 1704.04861*, 2017.
- [22] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 6848-6856.
- [23] CHEN Y, DAI X, CHEN D, et al. Mobile-former: Bridging mobilenet and transformer[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 5270-5279.
- [24] WAN Q, HUANG Z, LU J, et al. Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation[J]. *arXiv preprint arXiv: 2301.13156*, 2023.
- [25] YU W, LUO M, ZHOU P, et al. Metaformer is actually what you need for vision[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 10819-10829.
- [26] MILLETARI F, NAVAB N, AHMADI S A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation[C]//*2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, 2016: 565-571.
- [27] XING Z, YU L, WAN L, et al. NestedFormer: Nested modality-aware transformer for brain tumor segmentation[C]//*International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2022: 140-150.
- [28] HATAMIZADEH A, TANG Y, NATH V, et al. UNETR: Transformers for 3D medical image segmentation[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022: 574-584.
- [29] LIU H, HUO G, LI Q, et al. Multiscale lightweight 3D segmentation algorithm with attention mechanism: Brain tumor image segmentation[J]. *Expert Systems with Applications*, 2023, 214: 119166.
- [30] LEE H H, BAO S, HUO Y, et al. 3D UX-Net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation[J]. *arXiv preprint arXiv: 2209.15076*, 2022.
- [31] SHAKER A M, MAAZ M, RASHEED H, et al. UNETR++: Delving into efficient and accurate 3D medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2024, 43(9): 3377-3390.