



## 融合CNN与Transformer的MRI脑肿瘤图像分割

刘万军, 姜岚, 曲海成, 王晓娜, 崔衡

引用本文:

刘万军, 姜岚, 曲海成, 王晓娜, 崔衡. 融合CNN与Transformer的MRI脑肿瘤图像分割[J]. 智能系统学报, 2024, 19(4): 1007-1015.

LIU Wanjun, JIANG Lan, QU Haicheng, et al. MRI brain tumor image segmentation by fusing CNN and Transformer[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(4): 1007-1015.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202301016>

## 您可能感兴趣的其他文章

### 用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction

智能系统学报. 2021, 16(3): 518-527 <https://dx.doi.org/10.11992/tis.202008036>

### 混合神经网络和条件随机场相结合的文本情感分析

Text sentiment analysis combining hybrid neural network and conditional random field

智能系统学报. 2021, 16(2): 202-209 <https://dx.doi.org/10.11992/tis.201907041>

### 融合整体与局部信息的武夷岩茶叶片分类方法

Classification of Wuyi rock tealeaves by integrating global and local information

智能系统学报. 2020, 15(5): 919-924 <https://dx.doi.org/10.11992/tis.202003018>

### 基于注意力融合的图像描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740-749 <https://dx.doi.org/10.11992/tis.201910039>

### 基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956-963 <https://dx.doi.org/10.11992/tis.201903001>

### 基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network

智能系统学报. 2019, 14(6): 1152-1162 <https://dx.doi.org/10.11992/tis.201812003>

DOI: 10.11992/tis.202301016

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240306.1400.002>

# 融合 CNN 与 Transformer 的 MRI 脑肿瘤图像分割

刘万军, 姜岚, 曲海成, 王晓娜, 崔衡  
(辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105)

**摘要:** 为解决卷积神经网络 (convolutional neural network, CNN) 在学习全局上下文信息和边缘细节方面受到很大限制的问题, 提出一种同时学习局语义信息和局部空间细节的级联神经网络用于脑肿瘤医学图像分割。首先将输入体素分别送入 CNN 和 Transformer 分支, 在编码阶段结束后, 采用一种双分支融合模块将 2 个分支学习到的特征有效地结合起来以实现全局信息与局部信息的融合。双分支融合模块利用哈达玛积对双分支特征之间的细粒度交互进行建模, 同时使用多重注意力机制充分提取特征图通道和空间信息并抑制无效的噪声信息。在 BraTS 竞赛官网评估了本文方法, 在 BraTS2019 验证集上增强型肿瘤区、全肿瘤区和肿瘤核心区的 Dice 分数分别为 77.92%, 89.20% 和 81.20%。相较于其他先进的三维医学图像分割方法, 本文方法表现出了更好的分割性能, 为临床医生做出准确的脑肿瘤细胞评估和治疗方案提供了可靠依据。

**关键词:** 医学图像分割; 脑肿瘤; 级联神经网络; 卷积神经网络; Transformer; 特征融合; 多重注意力; 残差学习中图分类号: TP391 文献标志码: A 文章编号: 1673-4785(2024)04-1007-09

中文引用格式: 刘万军, 姜岚, 曲海成, 等. 融合 CNN 与 Transformer 的 MRI 脑肿瘤图像分割 [J]. 智能系统学报, 2024, 19(4): 1007-1015.

英文引用格式: LIU Wanjun, JIANG Lan, QU Haicheng, et al. MRI brain tumor image segmentation by fusing CNN and Transformer[J]. CAAI transactions on intelligent systems, 2024, 19(4): 1007-1015.

## MRI brain tumor image segmentation by fusing CNN and Transformer

LIU Wanjun, JIANG Lan, QU Haicheng, WANG Xiaona, CUI Heng

(School of Software, Liaoning Technical University, Huludao 125105, China)

**Abstract:** This study presents a cascaded neural network that learns both global semantic information and local spatial details for medical image segmentation of brain tumors, solving the problem that convolutional neural networks(CNN) are greatly restricted in learning global contextual information and edge details. First, the input voxels are fed into the CNN and Transformer branches separately. After the encoding phase, a two-branch fusion module is used to effectively combine the features learned in both branches to achieve the fusion of global and local information. The two-branch fusion module uses Hadamard products to model the fine-grained interactions between the two-branch features, while using multiple attention mechanisms to fully extract the feature map channels and spatial information and suppress the invalid noise information. The method of this paper has been evaluated on the BraTS competition website, with Dice scores of 77.92%, 89.20% and 81.20% for the enhanced tumor region, full tumor region and tumor core region on the BraTS2019 validation set, respectively. Compared with other advanced 3D medical image segmentation methods, this method shows better segmentation performance, which provides a reliable basis for clinicians to make accurate brain tumor cell assessment and treatment plans.

**Keywords:** medical image segmentation; brain tumor; cascaded neural network; convolutional neural networks; Transformer; feature fusion; multiple attention; residual learning

收稿日期: 2023-01-16. 网络出版日期: 2024-03-08.

基金项目: 辽宁省高等学校基本科研课题项目 (LJKMZ20220699);  
辽宁工程技术大学学科创新团队项目 (LNTU20T-D-23).

通信作者: 刘万军. E-mail: [liuwanjun@lntu.edu.cn](mailto:liuwanjun@lntu.edu.cn).

©《智能系统学报》编辑部版权所有

胶质瘤是成年患者中最常见的原发性脑肿瘤, 具有不同程度的侵袭性, 占脑肿瘤的 81%<sup>[1]</sup>。目前, 脑肿瘤的检查和分析主要应用核磁共振成像 (magnetic resonance imaging, MRI) 技术。然而,

由于胶质瘤在多模态磁共振成像中的形状和外观多变,导致脑肿瘤的精确分割成为了一项极具挑战性的医学图像分析任务。通常脑肿瘤及其子区域的精确描绘由经验丰富的神经放射科医生手动完成,这种方式成本高、耗时长、可重复性差,容易出现人为的错误导致不正确的预后和治疗<sup>[2]</sup>。因此,临床对于自动分割脑肿瘤的方法提出了较高的要求。

随着深度学习发展至今,“U型”编解码结构已在各种医学图像分割任务中展现出卓越的性能<sup>[3-5]</sup>。对于典型架构 U-Net<sup>[6]</sup>,编码器通过下采样操作捕捉图像的底层细节特征,解码器通过上采样对编码器得到的底层特征进行还原,恢复到原始输入尺寸来实现图像的精确分割。同时,利用跳跃连接将编解码器同一层级的输出结合起来以捕捉图像的多尺度信息。因此出现了许多基于 U-Net 及其变体网络应用于脑肿瘤分割。Myronenko<sup>[7]</sup>将 3D U-Net 与附加的变分解码器分支相结合,为编码器分支提供附加的监督和正则化,并对原始输入的脑肿瘤图像进行重构,提高了脑肿瘤分割模型的泛化性能。Chen 等<sup>[8]</sup>提出了一种新的 3D 多纤维空洞卷积网络 DMFNet,使用空洞卷积扩大输入特征图的感受野并通过将卷积通道分割为多组来减少特征图与内核之间的连接,在不过多降低脑肿瘤分割精度的情况下,显著提升了脑肿瘤的分割效率。张睦卿等<sup>[9]</sup>提出了一种多尺度伪影生成对抗网络的脑肿瘤图像分割方法,解决了肿瘤空间信息变化大和精细样本少的问题。Valanarasu 等<sup>[10]</sup>设计了一种过完备卷积结构 Kite-Net,将输入图像映射到更高维度,限制了感受野在网络深层的增加,同时通过交叉残差块与 U-Net 结合来学习互补特征,实现了对脑肿瘤边缘细节以及小肿瘤目标的精确分割。Isensee 等<sup>[11]</sup>应用 nnU-Net 框架对数据后处理、基于区域的训练、数据增强和 nnU-Net 管道的修改进行了特定的 BraTS 设计,大幅提高了分割性能。尽管这些网络表现出了较好的分割性能,但它们在捕获远距离依赖方面存在局限性,这可能严重影响图像分割的性能。

最近,基于自注意力的 Transformer<sup>[12]</sup>被用于建模长距离依赖关系来捕获序列到序列任务中任意位置间的相关性并且在脑肿瘤分割任务上达到了最先进的性能。Hatamizadeh 等<sup>[13]</sup>提出了 UNETR 模型,使用 Transformer 作为编码器来捕获远程依赖关系,同时采用类似 U-Net 网络中的跳跃连接以有效地融合脑肿瘤图像的多尺度信息。Valanarasu 等<sup>[14]</sup>提出了一个门控的轴向注意力模型(即 medical Transformer, MedT),该模型通过在

自我注意力中引入一个总结性的控制机制,能够精确控制输入脑肿瘤图像的位置嵌入,扩展了现有的卷积神经网络架构。Jiang 等<sup>[15]</sup>设计了一个 ETrans 模块增强对脑肿瘤细节特征的提取。Cao 等<sup>[16]</sup>基于 Swin-Transformer 块构建了一个具有跳跃连接的对称编解码结构 Swin-UNet,它利用块扩展层在不使用卷积或插值操作的情况下实现了上采样和特征维数的增加,取得了较好的分割性能。Lin 等<sup>[17]</sup>在 Swin-UNet 的基础上提出了 DS-TransUNet,首次将分层 Swin-Transformer 结合到 U 型编解码结构中,同时设计了一种融合模块,利用自注意力机制有效捕获了不同尺度特征之间的相关性。同样,虽然 Transformer 能很好地学习全局上下文信息,但它不能很好地捕捉输入图像的局部信息,进而导致不能很好地分割小目标肿瘤。

为了解决上述脑肿瘤分割方法存在的问题,提出了一个双分支级联神经网络,将学习局部高级特征的卷积神经网络(convolutional neural network, CNN)与捕获全局多尺度信息的 Transformer 通过基于多重注意力机制的双分支融合模块结合在一起,实现了全局与局部信息的融合。相比于其他融合 CNN 与 Transformer 的 3D MRI 脑肿瘤图像分割方法,本文采用将 Transformer 作为单独的分支来捕获全局多尺度信息,避免了输入体素在进行卷积操作和不断的下采样过程中导致重要信息丢失。实验结果证明,提出的网络实现了全局信息与局部信息的有效融合,提高了脑肿瘤的分割准确率。

## 1 本文方法

本文提出的方法旨在将 CNN 和 Transformer 结合在一起学习更有效的医学图像分割表示来实现脑肿瘤的精确分割。网络整体结构如图 1 所示,它由用于捕捉空间和深层特征的 CNN 编码器、学习输入体的序列表示并有效地捕获全局多尺度信息的 Transformer 编码器以及用于分割的解码器组成。

### 1.1 CNN 编码器

CNN 编码器是一种广泛用于对局部特征进行  $\mathbf{R}^{C \times H \times W \times D}$  的 MRI 扫描图像,其中  $H$ 、 $W$  为扫描图像的高度和宽度,  $D$  为扫描的二维图像切片数,  $C$  为通道数。通过一个 3D 卷积层和 dropout<sup>[18]</sup> 层对输入体素进行初始化,然后是 4 个阶段的 3D 残差块,数量分别为 1、2、2、4。下采样操作采用步长为 2 的卷积代替池化操作,最终得到  $128 \times 16 \times 16 \times 16$  大小的输出特征图。



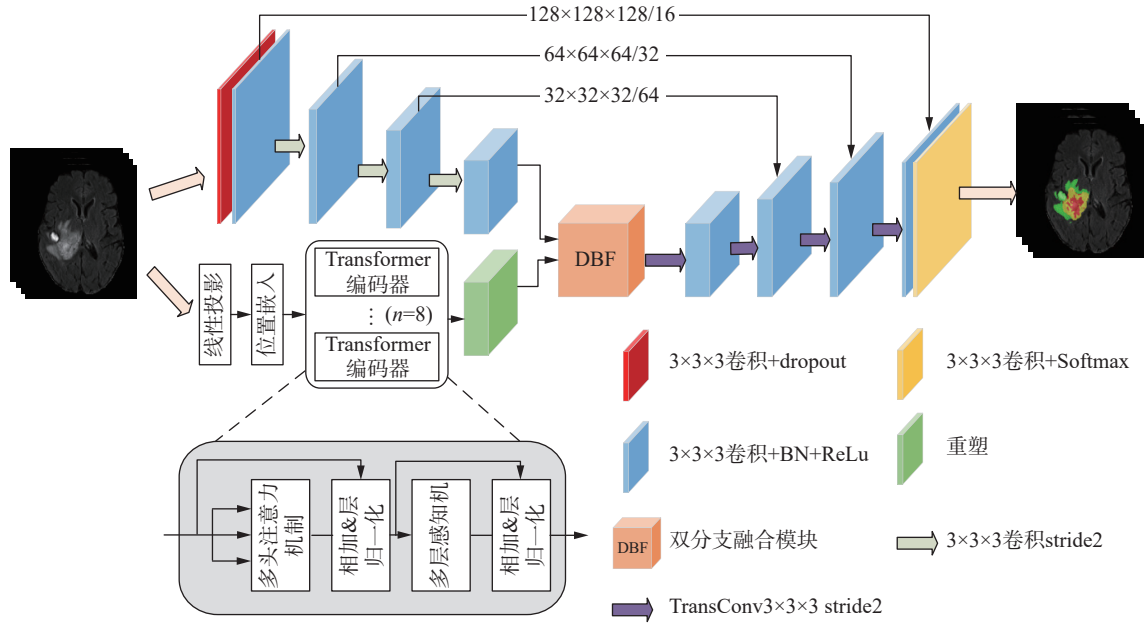


图 1 本文模型整体结构

Fig. 1 Overall structure of the model in this paper

## 1.2 Transformer 编码器

CNN 编码器由于自身的局限性不能有效地捕获输入体素的远距离相关性。为此, 使用 Transformer 编码器进行全局上下文建模。首先, 将输入体素  $X \in \mathbf{R}^{C \times H \times W \times D}$  重塑为均匀不重叠的块  $X \in \mathbf{R}^{N \times (P^3 \cdot C)}$ , 其中  $(P, P, P)$  表示每个块的分辨率,  $N = (H \times W \times D) / P^3$  为输入序列的长度。使用线性层将块映射到  $K$  维嵌入空间中, 其在整个 Transformer 层中保持不变。为了编码空间位置信息, 在块嵌入中添加了一维可学习的位置嵌入中来保留位置信息。公式可以表示为

$$Z_0 = [x^1 E; x^2 E; \dots; x^N E] + E_{\text{pos}}$$

式中:  $E \in \mathbf{R}^{(P^3 \cdot C) \times K}$  为块嵌入投影,  $E_{\text{pos}} \in \mathbf{R}^{N \times K}$  为位置嵌入。

最终, Transformer 编码器由  $L$  层多头自注意力 (multi-head attention, MHA) 和多层感知器 (multi-layer perceptron, MLP) 模块组成, 每个子层中采用加性跳跃连接策略来避免梯度消失。因此, 第  $i$  层的输出可以表示成:

$$Z'_i = T_{\text{mha}}(T_{\text{ln}}(Z_i - 1)) + Z_{i-1}, \quad i = 1, 2, \dots, L$$

$$Z_i = T_{\text{mlp}}(T_{\text{ln}}(Z'_i)) + Z'_i, \quad i = 1, 2, \dots, L$$

式中:  $T_{\text{ln}(\cdot)}$  为层归一化,  $T_{\text{mha}}$  为多头自注意力,  $T_{\text{mlp}}$  由具有 GELU<sup>[19]</sup> 激活函数的 2 个线性层组成,  $i$  为中间块标识符,  $L$  为 Transformer 的层数。

Transformer 编码得到输出特征图经过一个步长为 2 的反卷积层后得到与 CNN 编码后相同大小的输出特征图, 通过 DBF 模块将 Transformer 和 CNN 编码阶段学习到的全局信息与局部信息

融合。

## 1.3 DBF 模块设计

为了有效地结合 CNN 和 Transformer 分支的编码特征, 提出了一种新的融合模块——DBF (dual branch fusion) 模块, 其结构如图 2 所示。将编码阶段得到的特征图  $F(x)$  和  $S(x)$  分别送入空间注意力 (spatial attention, SA) 模块和通道注意力 (channel attention, CA) 模块得到输出特征  $\hat{F}(x)$  和  $\hat{S}(x)$ 。为了对特征之间的细粒度交互进行建模,  $F(x)$  和  $S(x)$  在  $1 \times 1 \times 1$  卷积和批归一化后通过哈达玛积使对应位置特征进行加权, 然后经过  $3 \times 3 \times 3$  卷积、批归一化和 ReLU 激活得到交互特征  $G(x)$ 。将得到的特征  $\hat{F}(x)$ 、 $\hat{S}(x)$  和  $G(x)$  沿通道维度拼接起来, 使用  $1 \times 1 \times 1$  卷积进行特征降维后通过瓶颈残差模块得到最终的特征输出  $Z(x)$ 。DBF 模块实现了全局信息与局部信息的融合, 弥补了 CNN 与 Transformer 只能分别关注单一特征的不足, 并通过注意力机制实现了目标特征增强和噪声抑制的目的。本文设计改进的 SA 模块和 CA 模块结构如图 3 所示。

1) SA 模块。空间注意力表示图像中不同区域像素点之间的权重关系。由于 CNN 低层特征存在较多噪声并且在不断的下采样过程中导致空间位置信息丢失, CNN 分支利用空间注意力来增强局部特征并抑制不相关的区域, 提高小目标肿瘤的分割结果。首先使特征图  $F(x)$  沿通道维度应用最大池化和平均池化, 并将它们拼接成一个有效的特征图。接着使用  $7 \times 7 \times 7$  卷积进行通道降维

并通过 Sigmoid 激活函数得到空间注意力的特征权重, 最后与输入特征相乘得到输出特征  $\hat{F}(x)$ , 公式可以表示为

$$\hat{F}(x) = F(x) \cdot \sigma(f^7([Z_{\text{gmp}}(F(x)); Z_{\text{gap}}(F(x))]))$$

式中:  $\sigma(\cdot)$  为 Sigmoid 函数,  $f^7$  为  $7 \times 7 \times 7$  卷积,  $Z_{\text{gmp}}$  为全局最大池化,  $Z_{\text{gap}}$  为全局平均池化。

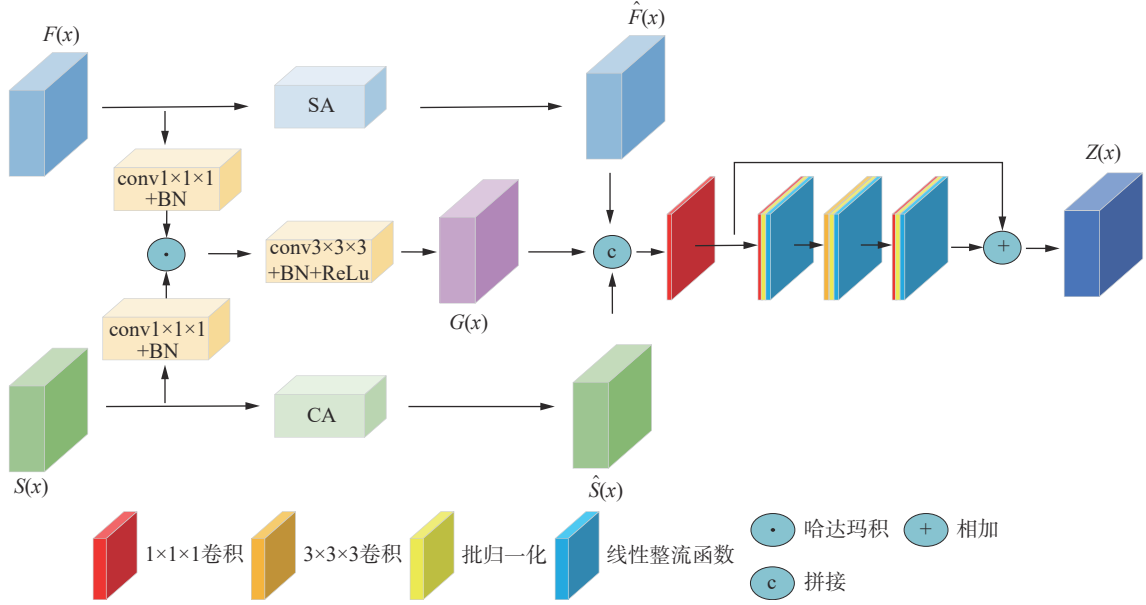


图 2 DBF 模块结构

Fig. 2 DBF module structure

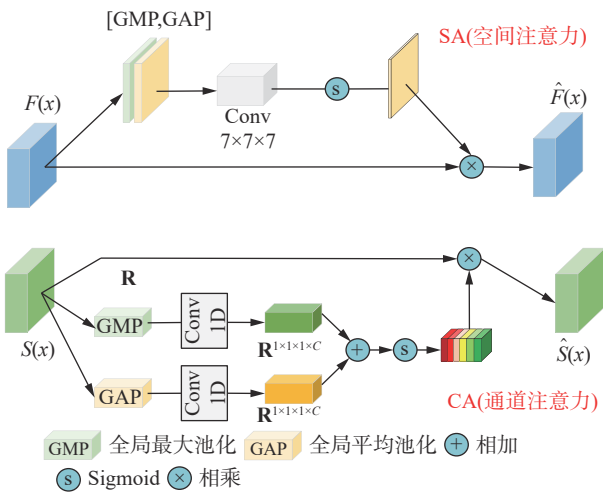


图 3 SA 模块和 CA 模块结构

Fig. 3 SA module and CA module structure

2) CA 模块。通道注意力表示不同通道之间特征的重要程度。Transformer 分支通过每个特征通道的权重大小对当前脑肿瘤分割任务相关或无关的特征进行增强或抑制, 从而提高去噪和获取准确信息的能力。首先对输入特征  $S(x)$  同时使用全局最大池化和全局平均池化操作, 为了避免降维对于通道注意力特征学习的影响, 因此使用 1 维卷积 (Conv1D) 替代经典通道注意力中的全连接层降低模型复杂度, 同时增强跨通道信息获取能力<sup>[20]</sup>, 公式可以表示为

$$S_{\text{max}} = \text{Conv1D}(Z_{\text{gmp}}(S(x)))$$

$$S_{\text{avg}} = \text{Conv1D}(Z_{\text{gap}}(S(x)))$$

式中  $S_{\text{max}}$  和  $S_{\text{avg}}$  分别为输入特征应用最大池化和平均池化后进行 1 维卷积后的特征。

将大小为  $R^{1 \times 1 \times 1 \times C}$  的权重向量  $S_{\text{max}}$  和  $S_{\text{avg}}$  进行对应像素相加操作并通过 Sigmoid 激活函数得到通道注意力的特征权重, 最后与输入特征相乘得到输出特征  $\hat{S}(x)$ , 公式可以表示为

$$\hat{S}(x) = \sigma(S_{\text{max}} + S_{\text{avg}}) \cdot S(x)$$

#### 1.4 解码器

解码器是纯 CNN 架构, 使用转置卷积将特征图逐步上采样到输入分辨率 ( $H \times W \times D$ ), 并且在上采样过程中使用 3D 残差块细化特征图。最后, 使用  $1 \times 1 \times 1$  卷积和 softmax 激活函数将特征映射为概率分割结果。此外, 编码器和解码器之间使用跳跃连接以保留更多的底层信息, 提高脑肿瘤的分割精度。最后, 本文方法采用多标签损失<sup>[21]</sup>作为训练过程中的损失函数来解决正负样本之间的极端不平衡问题。具体可以表示为

$$\text{ML}(G, P) = 1 - \frac{2}{C} \sum_{j=1}^C \frac{\sum_{i=1}^N G_{i,j} P_{i,j} + s}{\sum_{i=1}^N G_{i,j}^2 + \sum_{i=1}^N Y_{i,j}^2 + s}$$

式中:  $C$  为类的数量,  $N$  为输入体素的数量,  $P_{i,j}$  和  $G_{i,j}$  为表示输入体素  $i$  在类  $j$  的预测值输出和真实标

签,  $s$ 值设置为  $1 \times 10^{-5}$ 。

## 2 实验及结果分析

### 2.1 数据和评估指标

本文采用多模态脑肿瘤分割挑战赛 (Multimodal Brain Tumor Segmentation Challenge 2019, BraTS 2019) 公开数据集 BraTS 2019(<https://ipp.cbica.upenn.edu/>) 来训练和评估所提出的方法。训练数据包含 335 例患者的 MRI 图像及真实分割标签, 包括 259 例高级别胶质瘤 (high-grade glioma, HGG) 患者和 76 例低级别胶质瘤 (low-grade glioma, LGG) 患者。验证集为 125 例无标签患者的 MRI 图像。由官网在线评估服务器评估在验证集上的分割结果, 以此验证所提出的方法的有效性。每个样本有 T1、T1ce、T2 和 FLAIR 共 4 种模态, 所有模态体积均为  $240 \times 240 \times 155$ 。数据集标签共有背景 (标签 0)、坏疽 (标签 1)、水肿区 (标签 2) 和增强肿瘤区 (标签 4) 4 个标签, 通过 Dice 分数和豪斯多夫距离来评估整体肿瘤区 (WT, 标签 1、2、4)、肿瘤核心区 (TC, 标签 1、4) 和增强肿瘤区 (ET, 标签 1、2、4)。

### 2.2 预处理

鉴于每个病例的 4 种模态的 MRI 图像大小均为  $240 \times 240 \times 155$ , 且共享分割标签, 因此将 4 种模态堆叠成一个 4D 图像 ( $H \times W \times D \times C$ ,  $C=4$ )。然后裁剪每个体积边界上的冗余背景体素 (体素值为零), 因为它们不提供任何有用信息, 可以被神经网络忽略。随后, 对于每张图像的灰度区域进行  $z$ -score 标准化处理, 背景区域体素始终保持为零。

对于数据增强采用以下技术: 1) 对 MRI 图像进行尺寸从  $240 \times 240 \times 155$  体素到  $128 \times 128 \times 128$  体素的随机裁剪; 2) 以 0.5 的概率在轴向、冠状面和矢状面上进行随机翻转; 3) 随机旋转角度范围为  $[-10^\circ, +10^\circ]$ ; 4) 随机强度在  $[-0.1, 0.1]$  偏移, 尺度在  $[0.9, 1.1]$ 。在测试阶段, 将 MRI 图像从  $240 \times 240 \times 155$  填充到  $240 \times 240 \times 160$ , 并使用测试时间增强技术进一步提高模型的性能, 该技术已得到验证<sup>[22]</sup>。

### 2.3 实验环境配置

实验使用的计算机操作系统为 Ubuntu 18.04, 硬件环境为 NVIDIA GeForce RTX 3090 显卡, Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz 处理器, 360 GB 内存。

实验过程中使用深度学习框架 PyTorch 进行分布式训练, 采用 Adam 优化器, 批大小为 8, 初始学习率为  $4 \times 10^{-4}$  并随每次迭代以 0.9 的幂次进

行衰减, 同时使用权重衰减率为  $1 \times 10^{-5}$  的  $L_2$  范数进行正则化, 模型的最大迭代轮数为 1000。

## 2.4 实验结果与分析

### 2.4.1 实验结果

为了评估所提出模型的分割性能, 在 BraTS-2019 验证集上与 4 种模型进行对比:

1) 3D U-Net。对称编解码结构, 同一维度添加了跳跃连接。

2) Attention U-Net。在跳跃连接处增加了门控注意力。

3) UNETR。遵循 U 型编解码设计的纯基于 Transformer 的结构。

4) TransBTS。CNN 编码器最后一层级应用 Transformer 多头自注意力对全局依赖关系进行建模。

定量结果见表 1, 本文模型在增强肿瘤区、整体肿瘤区和肿瘤核心区的 Dice 分数分别为 77.92%、89.20% 和 81.20%, 豪斯多夫距离分别为 3.214、5.444 和 6.915 mm。与 3D U-Net 和 Attention U-Net 相比, 所提出的方法在 3 个肿瘤区域的分割准确率上都有显著的提升, 这说明通过 DBF 模块将 Transformer 与 CNN 结合在一起有效地实现了全局信息与局部信息的融合, 同时揭示了利用 Transformer 建模全局关系的好处。完全基于 Transformer 的 UNETR 在脑肿瘤分割任务上并没有实现很好的分割效果, 这是由于其对局部信息的捕获能力上不如卷积神经网络。最后, 与同样将 CNN 与 Transformer 结合的 TransBTS 相比, 在相同的实验设置情况下, 所提出方法的分割性能更好。这是由于 TransBTS 只是对 CNN 编码得到的最后一层输出特征图进行局部全局关系建模, 而本文方法则是将 Transformer 作为单独的分支来捕获全局多尺度信息, 避免了输入体素在进行卷积操作和不断的下采样过程中导致重要信息丢失。

表 1 各种模型的实验结果对比

Table 1 Comparison of experimental results of various models

方法	Dice分数↑/%			Hausdorff95↓/mm		
	ET	WT	TC	ET	WT	TC
3D U-Net <sup>[6]</sup>	69.78	87.18	75.75	8.026	6.255	8.896
Attention U-Net <sup>[23]</sup>	76.05	88.82	77.77	4.732	7.656	9.676
UNETR <sup>[11]</sup>	74.61	87.06	78.11	7.297	14.087	10.975
TransBTS <sup>[24]</sup>	76.20	89.08	80.54	<b>3.133</b>	5.879	<b>6.235</b>
本文方法	<b>77.92</b>	<b>89.20</b>	<b>81.20</b>	3.214	<b>5.444</b>	6.915

注: 加粗字体为最优结果, ↑表示越高越好, ↓表示越低越好, 下同。

图 4 给出了 5 种方法在 BraTS2019 无标签在线验证集的分割示例。图 4 中 ET、WT、TC 表示每个病例的 Dice 分数, 其中, 绿色区域表示水肿区, 黄色区域代表增强肿瘤区, 红色区域表示坏疽。可以看出, 与其他方法相比, 本文方法

的分割预测更接近真实的分割结果。这是因为与其他方法相比, 本文方法同时实现了对全局信息与局部信息的获取, 并通过带有注意力机制的 DBF 模块有效结合在一起, 提高了特征提取的质量。

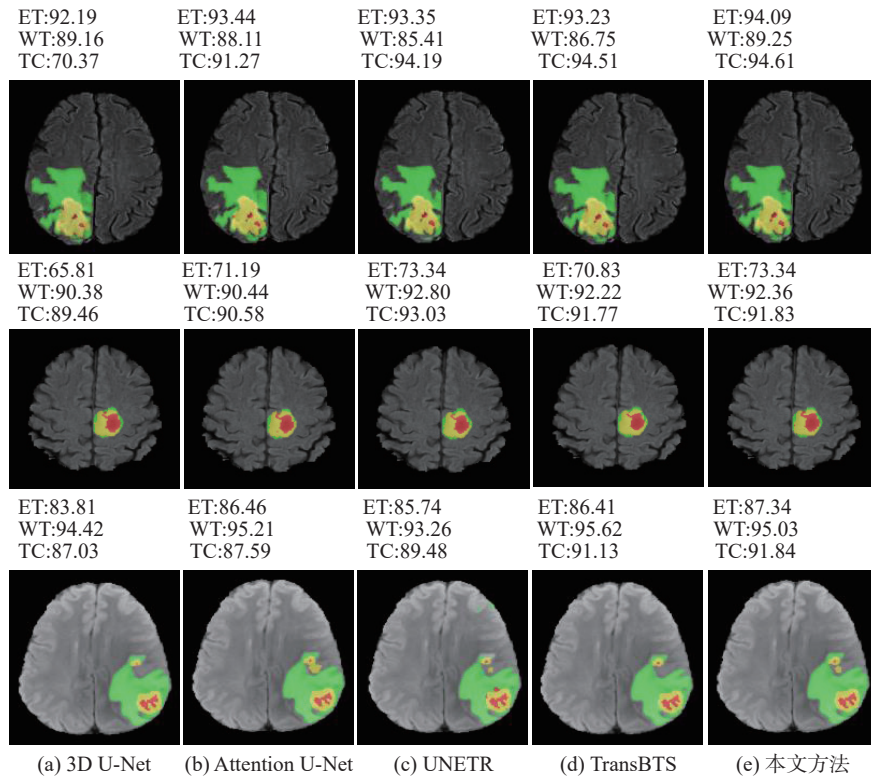


图 4 各模型分割效果对比

Fig. 4 Comparison of segmentation effect of each model

#### 2.4.2 与其他方法比较

为了进一步验证本文方法在小目标脑肿瘤分割任务上的有效性, 将本文方法在 BraTS2019 验证集上的分割结果与其他先进的方法进行比较, 结果见表 2。相比于近两年提出的 3D 脑肿瘤分割模型, 对于整体肿瘤区和肿瘤核心区, 本文取

得了与其他方法相当或更好的分割表现。而对于增强肿瘤区, 本文方法的分割精确度有明显的提升, 取得了最好的分割性能。这说明通过带有注意力机制的 DBF 模块将 CNN 分支与 Transformer 分支有效结合在一起, 增强了模型对重要目标特征的提取, 显著提升了对小目标肿瘤的分割性能。

表 2 与其他先进方法的比较

Table 2 Comparison with other advanced methods

方法	Dice分数↑/%			Hausdorff95↓/mm		
	ET	WT	TC	ET	WT	TC
Valanarasu等 <sup>[10]</sup>	73.21	87.60	73.92	6.323	8.942	9.893
Huang等 <sup>[25]</sup>	73.00	82.70	78.80	6.100	8.500	9.200
Vadacchino等 <sup>[26]</sup>	73.30	89.60	79.10	—	—	—
Li等 <sup>[27]</sup>	74.00	89.50	81.70	—	—	—
Xu等 <sup>[28]</sup>	74.00	<b>90.00</b>	<b>82.00</b>	3.490	<b>4.510</b>	<b>6.250</b>
Akbar等 <sup>[29]</sup>	74.20	88.48	80.98	6.670	10.83	10.25
Liang等 <sup>[30]</sup>	76.64	88.32	81.07	4.968	6.635	7.105
本文方法	<b>77.92</b>	89.20	81.20	<b>3.214</b>	5.444	6.915



## 2.5 消融实验

为了评估所提出的框架并验证在不同条件下的性能,进行了各种消融研究,包括不同模块的选择、补丁分辨率和Transformer规模。

### 2.5.1 不同模块的选择

表3为不同模块对模型的影响结果。表3第1行是由2个分支(BaseLine)直接结合在一起的基线网络,其不包含跳跃连接(skip connection, SK)以及双分支融合模块(dual branch fusion, DBF)。然后依次添加跳跃连接和双分支融合模块以观察不同模块对所提方法的影响。表3给出了对本文方法中不同模块的消融研究结果。与基线网络相比,添加双分支融合模块后,在脑肿瘤3个区域的Dice分数分别提高了1.35%、1.41%和3.56%,说明通过DBF模块将CNN分支与Transformer分支编码得到的特征结合在一起,改善了特征融合的效果,实现了全局信息与局部信息的有效融合。同样添加跳跃连接后Dice分数分别提高了1.22%、1.83%和2.5%,说明跳跃连接使解码器在上采样过程中获得了更多高分辨率信息,更好地保留了原始图像的细节特征。通过表3及以上分析可以得出,脑肿瘤的分割性能会随着网络中每个块的增加而得到一定提升。

表3 不同模块对模型Dice分数的影响结果  
Table 3 Results of the effect of different modules on the model Dice score %

方法	ET	WT	TC
BaseLine	75.86	87.61	76.24
BaseLine+SK	77.08	<b>89.44</b>	78.74
BaseLine+DBF	77.21	88.98	79.80
BaseLine+SK+DBF(本文方法)	<b>77.92</b>	89.20	<b>81.20</b>

### 2.5.2 补丁分辨率

本文研究了补丁分辨率对所提出模型的影响。由于Transformer的序列长度与补丁分辨率的立方成反比,输入的补丁分辨率越小,序列长度则越长,从而产生较高的内存消耗。表4中实验结果表明:减小补丁分辨率(增加序列长度)可以提高模型性能,这是由于Transformer为更长的输入序列编码每个元素之间更复杂的依赖关系。具体而言,在脑肿瘤分割任务中,将补丁分辨率从32降到16,除了整体肿瘤区指标没有明显变化,增强肿瘤区和肿瘤核心区的Dice分数分别提高了0.85%和1.2%。参照ViT<sup>[10]</sup>中的设置,本文中使用了16×16×16作为默认的补丁分辨率大小,同时由于内存限制,没有实验更低的尺寸设置。

表4 补丁分辨率对模型Dice分数的影响结果  
Table 4 Results of the effect of patch resolution on the model Dice score %

补丁分辨率	ET	WT	TC
16	<b>77.92</b>	89.20	<b>81.20</b>
32	77.07	<b>89.29</b>	80.00

### 2.5.3 Transformer规模

Transformer的规模受隐藏层尺寸和Transformer层数的影响。因此,本文通过修改隐藏层尺寸和Transformer层数来验证Transformer规模对分割性能的影响。“基础”模型(Base)的隐藏层尺寸和注意力头数分别设置为512和8,而“大型”(Large)模型的超参数设置为768和12,实验结果见表5。从表5中可以看到较大的模型使脑肿瘤分割性能只得到了略微提升,但这带来额外的计算成本,使得模型训练时间增长。为了提高效率、减小计算成本,本文采用“基本”模型进行所有实验。

表5 Transformer规模对模型Dice分数的影响结果  
Table 5 Results of the effect of Transformer size on the model Dice score %

Transformer规模	ET	WT	TC
基础模型	77.92	89.20	<b>81.20</b>
大模型	<b>78.08</b>	<b>89.34</b>	80.65

## 3 结束语

本文提出了一种双分支级联网络用于脑肿瘤的准确分割。其中,CNN分支通过卷积运算提取输入特征图的局部信息;Transformer分支通过多头自注意力和多层感知器学习输入图像全局上下文信息。在编码阶段结束后,2个分支使用DBF模块结合在一起,通过不断地进行上采样操作最终生成脑肿瘤的分割预测。实验结果表明,所提方法实现了全局信息与局部信息的有效融合,提高了脑肿瘤的分割精度,为临床医生做出准确的脑肿瘤细胞评估和治疗方案提供了可靠依据。但是,本文方法由于使用了Transformer导致模型参数量较大,使得模型训练速度与传统卷积神经网络相比较慢。在后续工作中,将使用更加轻量化的Transformer结构,使得模型在保持分割精度的基础上减少运算量。

## 参考文献:

- [1] OSTROM Q T, BAUCHET L, DAVIS F G, et al. The epidemiology of glioma in adults: a “state of the science”



- review[J]. *Neuro-oncology*, 2014, 16(7): 896–913.
- [2] MENZE B H, JAKAB A, BAUER S, et al. The multimodal brain tumor image segmentation benchmark (BRATS)[J]. *IEEE transactions on medical imaging*, 2014, 34(10): 1993–2024.
  - [3] HUANG Huimin, LIN Lanfen, TONG Ruofeng, et al. Unet 3+: a full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020: 1055–1059.
  - [4] NGUYEN T, HUA B S, LE N. 3D-UCaps: 3D capsules unet for volumetric image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 548–558.
  - [5] LOU A, GUAN S, LOEW M. DC-UNet: rethinking the U-Net architecture with dual channel efficient CNN for medical image segmentation[C]//Medical Imaging 2021: Image Processing. Virtual Online: SPIE, 2021: 758–768.
  - [6] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2016: 424–432.
  - [7] MYRONENKO A. 3D MRI brain tumor segmentation using autoencoder regularization[C]//International MICCAI Brainlesion Workshop. Cham: Springer, 2019: 311–320.
  - [8] CHEN Chen, LIU Xiaopeng, DING Meng, et al. 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2019: 184–192.
  - [9] 张睦卿, 韩雨童, 陈柏年, 等. 基于多尺度伪影生成对抗网络的磁共振成像脑肿瘤分割方法[J]. *光子学报*, 2023, 52(8): 194–205.  
ZHANG Muqing, HAN Yutong, CHEN Bonian, et al. Magnetic resonance imaging brain tumor segmentation using multiscale ghost generative adversarial network[J]. *Acta photonica sinica*, 2023, 52(8): 194–205.
  - [10] VALANARASU J M J, SINDAGI V A, HACIHALILOGLU I, et al. KiU-Net: overcomplete convolutional architectures for biomedical image and volumetric segmentation[J]. *IEEE transactions on medical imaging*, 2021, 41(4): 965–976.
  - [11] ISENSEE F, JÄGER P F, FULL P M, et al. NNU-Net for brain tumor segmentation[C]//International MICCAI Brainlesion Workshop. Cham: Springer, 2021: 118–132.
  - [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. (2021–06–03)[2023–01–16]. <https://arxiv.org/pdf/2010.11929.pdf>.
  - [13] HATAMIZADEH A, TANG Yuchen, NATH V, et al. Unetr: transformers for 3d medical image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2022: 574–584.
  - [14] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: gated axial-attention for medical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 36–46.
  - [15] JIANG Yun, ZHANG Yuan, LIN Xin, et al. SwinBTS: a method for 3D multimodal brain tumor segmentation using swin transformer[J]. *Brain sciences*, 2022, 12(6): 797.
  - [16] CAO Hu, WANG Yueyue, CHEN J, et al. Swin-Unet: Unet-like pure transformer for medical image segmentation[EB/OL]. (2021–05–12)[2023–01–16]. <https://arxiv.org/pdf/2105.05537.pdf>.
  - [17] LIN Ailiang, CHEN Bingzhi, XU Jiayu, et al. Ds-transunet: dual swin transformer u-net for medical image segmentation[J]. *IEEE transactions on instrumentation and measurement*, 2022, 71: 1–15.
  - [18] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. *The journal of machine learning research*, 2014, 15(1): 1929–1958.
  - [19] HENDRYCKS D, GIMPEL K. Gaussian error linear units (GELUs)[EB/OL]. (2020–07–08)[2013–01–16]. <https://arxiv.org/pdf/1606.08415v4.pdf>.
  - [20] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11534–11542.
  - [21] 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. Stanford: IEEE, 2016: 565–571.
  - [22] WANG Guotai, LI Wenqi, OURSELIN S, et al. Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation[C]//International MICCAI Brainlesion Workshop. Cham: Springer, 2019: 61–72.
  - [23] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[EB/OL]. (2018–05–20)[2023–01–16]. <https://arxiv.org/pdf/1804.03999.pdf>.
  - [24] WANG Wenxuan, CHEN Chen, DING Meng, et al. Transbts: multimodal brain tumor segmentation using transformer[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention.

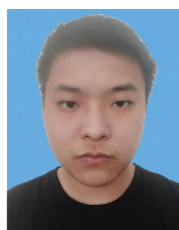
- Cham: Springer, 2021: 109–119.
- [25] HUANG He, YANG Guang, ZHANG Wenbo, et al. A deep multi-task learning framework for brain tumor segmentation[J]. *Frontiers in oncology*, 2021, 11: 690244.
- [26] VADACCHINO S, MEHTA R, SEPAHVAND N M, et al. HAD-Net: a hierarchical adversarial knowledge distillation network for improved enhanced tumour segmentation without post-contrast images[EB/OL]. (2021–03–30)[2023–01–16]. <https://arxiv.org/abs/2103.16617>.
- [27] LI Sahohua, SUI Xiuchao, LUO Xiangde, et al. Medical image segmentation using squeeze-and-expansion transformers[C]//International Joint Conference on Artificial Intelligence. Montreal: IJCAI, 2019: 61–72.
- [28] XU Weijin, YANG Huihua, ZHANG Mingying, et al. Brain tumor segmentation with corner attention and high-dimensional perceptual loss[J]. *Biomedical signal processing and control*, 2022, 73: 103438.
- [29] AKBAR A S, FATICHAH C, SUCIATI N. Single level UNet3D with multipath residual attention block for brain tumor segmentation[J]. *Journal of King Saud University-computer and information sciences*, 2022: 3247–3258.
- [30] LIANG Junjie, YANG Cihui, ZHONG Jingting, et al. BTSwin-Unet: 3D U-shaped symmetrical swin transformer-based network for brain tumor segmentation with

self-supervised pre-training[J]. *Neural processing letters*, 2022, 55(4): 3695–3713.

#### 作者简介:



刘万军, 教授, 博士生导师, 主要研究方向为模式识别与人工智能、计算机视觉与图像处理。主持国家自然科学基金面上项目等各类科研项目 20 余项。发表学术论文 200 余篇。E-mail: liuwanjun@lntu.edu.cn。



姜岚, 硕士研究生, 主要研究方向为计算机视觉与图像处理。E-mail: 13562859231@163.com。



曲海成, 副教授, 博士, 主要研究方向为遥感大数据智能处理和目标识别与跟踪。主持辽宁省自然科学基金 1 项、省教育厅面上项目 2 项。发表学术论文 60 余篇。E-mail: quhaicheng@lntu.edu.cn。