

# 基于非对称 U 型卷积神经网络的脑肿瘤图像分割<sup>①</sup>



刘盼盼, 安典龙, 丰 艳

(青岛科技大学 信息科学技术学院, 青岛 266042)  
通信作者: 丰 艳, E-mail: [fywmh@163.com](mailto:fywmh@163.com)

**摘 要:** 在计算机视觉分割任务中, 基于 Transformer 的图像分割模型需要大量的图像数据才能达到最好的性能, 医学图像相对于自然图像, 数据量非常稀少, 而卷积本身具有更高的感应偏差, 使得它更适合医学图像方面的应用. 为了将 Transformer 的远程表征学习与 CNN 的感应偏差相结合, 本文设计了残差 ConvNeXt 模块来模拟 Transformer 的设计结构, 采用深度卷积和逐点卷积组成的残差 ConvNeXt 模块来提取特征信息, 极大地降低了参数量. 并对感受野和特征通道进行了有效的缩放和扩展, 丰富了特征信息. 此外, 本文提出了一个非对称 3D U 型网络 ASUNet 用于脑肿瘤图像的分割. 在非对称 U 型结构中, 采用残差连接, 将最后两个编码器的输出特征进行连接来扩大通道数. 最后, 在上采样的过程中采用深度监督, 促进了上采样过程中语义信息的恢复. 在 BraTS 2020 和 FeTS 2021 数据集上的实验结果表明, ET、WT 和 TC 的骰子分数分别达到了 77.08%、90.83%、83.41% 和 75.63%、90.45%、84.21%. 并且通过对比实验, ASUNet 在准确性方面完全可以与 Transformer 构建的模型竞争, 同时保持了标准卷积神经网络的简单性和高效性.

**关键词:** 非对称 U 型结构; 倒置瓶颈; 深度监督; ASUNet

引用格式: 刘盼盼, 安典龙, 丰艳. 基于非对称 U 型卷积神经网络的脑肿瘤图像分割. 计算机系统应用, 2024, 33(8): 196–204. <http://www.c-s-a.org.cn/1003-3254/9613.html>

## Brain Tumor Image Segmentation Based on Asymmetric U-shaped Convolutional Neural Network

LIU Pan-Pan, AN Dian-Long, FENG Yan

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266042, China)

**Abstract:** In computer vision segmentation, the Transformer-based image segmentation model needs a large amount of image data to achieve the best performance. However, the data volume of medical images is very scarce compared with natural images. Convolution, with its higher inductive bias, is more suitable for medical images. To combine the long-range representation learning of Transformer with the inductive bias of CNN, a residual ConvNeXt module is designed to simulate the design structure of Transformer in this research. The module, composed of deep convolution and point wise convolution, is used to extract feature information, which greatly reduces the number of parameters. The receptive field and feature channel are effectively scaled and expanded to enrich the feature information. In addition, an asymmetric 3D U-shaped network called ASUNet is proposed for the segmentation of brain tumor images. In the asymmetric U-shaped structure, the output features of the last two encoders are connected by residual connection to expand the number of channels. Finally, deep supervision is used in the process of upsampling, which promotes the recovery of semantic information. Experimental results on the BraTS 2020 and FeTS 2021 datasets show that the dice scores of ET, WT, and TC reach 77.08%, 90.83%, 83.41%, and 75.63%, 90.45%, 84.21%, respectively. Comparative experiments show that ASUNet can fully compete with Transformer-based models in terms of accuracy while maintaining the simplicity and

① 基金项目: 国家自然科学基金 (62172248); 山东省自然科学基金 (ZR2021MF098)

收稿时间: 2024-01-23; 修改时间: 2024-03-05, 2024-04-19; 采用时间: 2024-04-23; csa 在线出版时间: 2024-07-03

CNKI 网络首发时间: 2024-07-08

efficiency of standard convolutional neural networks.

**Key words:** asymmetric U-shaped structure; inverted bottleneck; deep supervision; ASUNet

脑肿瘤的发病率越来越高,严重危害人类的健康生活.核磁共振成像(MRI)作为一种典型的非侵入式成像技术,为脑肿瘤的诊断和治疗做出了突出的贡献.卷积神经网络(CNN)被广泛地应用于医学图像中主要是因为固有的感应偏差,可以实现特征的自动学习,获取局部感受野等优点<sup>[1]</sup>.在语义分割中实现模型精度跨度比较大的是FCN<sup>[2]</sup>,它全部由卷积组成.基于FCN,Ronneberger等人<sup>[3]</sup>提出了一种新颖的模型架构,成为后来很多优秀模型的基准模型,即U-Net.它是由编码器和解码器,以及编码器和解码器之间的跳跃连接组成,跳跃连接实现了不同层次特征的融合,促进上采样特征图细节特征的恢复.ResNet<sup>[4]</sup>中的残差结构,将输出特征与原始特征相结合,解决了深层网络降级的问题.研究者们致力于找到效率高,计算量和参数量小的分割方法.群卷积、逐点卷积和深度可分离卷积在轻量级模型中起到主要的作用.为了在有限的计算资源上实现高效的计算,ShuffleNet<sup>[5]</sup>利用逐点卷积和通道混洗的操作大大减少了计算量,这对小规模网络精度的提升具有重大意义.ResNeXt<sup>[6]</sup>是一个多分支的结构,通过简化分组卷积来降低模型的计算量.MobileNet<sup>[7]</sup>将普通的卷积替换为深度可分离卷积,在精度不变的情况下,使得模型的参数量降低.

因此,训练一个参数少、计算量低并且简单高效的卷积神经网络是主流的发展方向.

在医学图像领域利用Transformer<sup>[8]</sup>来处理图像数据,已经取得了巨大的成功,这主要归功于Transformer对数据建模的远程依赖性,能够实现特征的可扩展性,高效性等.从Vision Transformer(ViT)<sup>[9]</sup>和Swin Transformer<sup>[10]</sup>提出以来,在医学图像中应用比较多的CNN和Transformer的混合模型<sup>[11-14]</sup>,将Transformer的变体与CNN相结合.但是这些混合网络相对于CNN,计算复杂度要高得多,并且需要将图像转化为图像块的序列,破坏了图像的空间结构,为了获得感应偏差,需要大量的数据进行训练,这些缺点对于医学图像数据稀少的特点,以及医学图像密集分割的性质来说是不利的.

为了在利用Transformer结构进行改进的同时保持卷积固有的感应偏差,ConvNeXt<sup>[15]</sup>重新考虑卷积神

经网络在图像方面的作用,用卷积神经网络来模拟ViT和Swin Transformer的结构特点,利用大的感受野和特征缩放行为来改进卷积神经网络,并且在更简明的操作方法下,实现和Transformer最先进的模型相当甚至更高的水平.但是ConvNeXt只是应用在自然图像这些2D图像中.显然,自然图像与脑肿瘤图像无论是数据维度还是数据量方面都存在很大的不同.MedNeXt<sup>[1]</sup>将ConvNeXt块扩展到上采样和下采样的过程中,构建了第1个完全ConvNeXt的U型结构.在本文中,我们设计了残差ConvNeXt模块,将Transformer的远程表征学习与CNN的感应偏差相结合.并且改变了传统的U型结构,提出了一个非对称3D U型卷积神经网络用于MRI脑肿瘤分割.

## 1 方法

本节详细介绍所提出的非对称U型卷积神经网络ASUNet.首先,介绍总体的架构组成.其次,介绍用于下采样的残差ConvNeXt模块.最后介绍解码器架构.

### 1.1 提出的ASUNet

ASUNet是一个具有编码器解码器结构的轻量级模型.如图1所示,不同颜色的长方体表示不同的功能模块,每个长方体上面的信息表示特征图的尺寸和通道数.ASUNet是由收缩路径,跳跃连接和扩展路径组成.从图中可以看出,首先,ASUNet是一个不对称的U型结构,编码器的数量要多于解码器的数量.为了防止在下采样的过程中特征图尺寸缩小的幅度过大导致特征细节信息严重丢失,保持最后一个下采样特征图的尺寸不变.此外,将最后两个编码器的输出在通道级别上进行连接,通道数量的增加有助于特征更详细的表示,这种连接方式可以保存上一阶段网络的一部分信息从而保证信息的有效性.其次,与传统的U型网络在下采样过程中使用步长为2的残差连接进行特征提取不同的是,我们在下采样部分采用残差ConvNeXt模块进行特征提取,对特征图在通道和感受野两个方面进行缩放和扩展,减少特征信息在下采样过程中不必要的丢失问题.然后,在上采样之后,采用 $3 \times 3 \times 3$ 的卷积对上采样特征图进行细化,有助于特征的突出表示进而促进上采样语义信息的恢复.最后,ASUNet在

上采样的过程中采用深度监督, 通过三次线性插值将每一级别的上采样特征图尺寸转变为  $128 \times 128 \times 128$ ,

并采用  $1 \times 1 \times 1$  的卷积将特征图的通道数转变为 3, 进而对每一个上采样特征图计算损失, 实现对网络的监督.

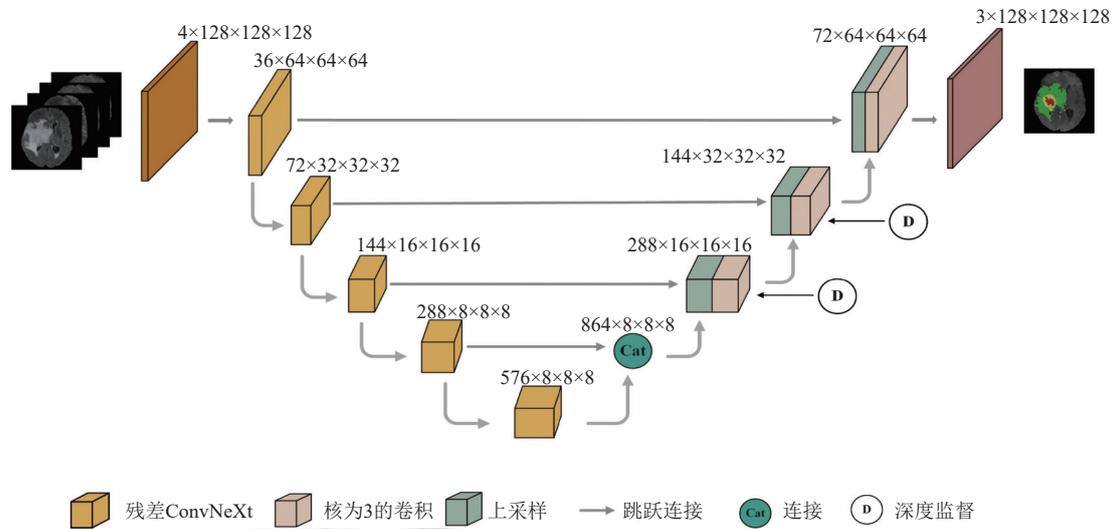


图1 提出的ASUNet架构

## 1.2 残差 ConvNeXt 模块

在下采样的过程中我们设计了残差 ConvNeXt 模块, 丰富了下采样过程中的语义特征信息, 有利于密集语义分割任务. 我们对传统的瓶颈结构进行改进, 传统的瓶颈结构组成如图 2(b) 所示. 除此之外, 如图 2(c), MedNeXt 在下采样的过程中也采用了具有残差结构的 MedNeXt 模块, 但是采用将主路径与捷径连接直接相加的方式, 并且在捷径连接中采用卷积核大小为 1. 我们所提出的残差 ConvNeXt 根据 ResNet 的设计特点, 将主路径与捷径连接相加, 然后送到激活层. 为了将不同感受野所提取的特征进行融合, 在捷径连接中采用了卷积核大小为 3 的卷积. 通过第 2.6.1 节的实验, 发现 ResNet 的残差连接方法性能更好, 更适合于医学图像分割. 改进后的残差 ConvNeXt 模块如图 2(d) 所示, 主要分为 3 部分: 深度卷积层、扩张层、收缩层. 从图 2 中可以看出, 第 1 层深度卷积层是由核大小为  $K$  的深度卷积 (DW) 组成的, 采用深度卷积层, 有效地降低了参数的数量. 如图 2(a) 所示, 在 Swin Transformer 的多头自注意力 (MSA) 中, 所采用的滑动窗口大小至少为 7, 并且在 ConvNeXt 模块中采用核大小为 7 的卷积. 通过第 2.6.1 节的实验表明, 当  $K=7$  时分割效果是最好的, 因此在残差 ConvNeXt 模块中, 采用核大小为 7, 即  $K=7$ , 输出通道数不变, 仍然为  $C$ . 深度卷积在每一个通道内利用大尺寸的卷积核提取特征, 模拟了 Swin Transformer 的注意力窗口. 由于 MSA 块在 MLP 块之

前, 因此, 我们将深度卷积层放在了残差 ConvNeXt 的第 1 层, 在深度卷积层之后, 特征图的空间分辨率发生变化, 在其后加入了归一化层来稳定训练的过程. 随后的扩张层是由  $1 \times 1 \times 1$  卷积组成, 输出通道是  $CR$ , 表示将输入通道扩大  $R$  倍, 由于 Transformer 中 MLP 隐藏层的通道数是输入层的 4 倍. 因此在本次实验中我们将  $R$  设置为 4, 这是在宽度上的扩展, 而采用  $1 \times 1 \times 1$  卷积核不仅限制了计算量而且在感受野上进行了缩放. 收缩层采用  $1 \times 1 \times 1$  的卷积主要是执行特征通道的压缩, 由于在下采样的过程中, 特征图的通道数逐渐倍增, 所以在收缩层将特征通道数减少为  $2C$ , 这是在宽度上的压缩. 最后, 按照图 2(b) 残差连接的方式得到的特征图通过激活函数 GELU, 进而得到最终的输出.

总之, 采用深度卷积是在空间维度上混合信息,  $1 \times 1 \times 1$  卷积是在通道维度上混合信息, 在空间上和通道上混合信息的方式, 模拟了 Swin Transformer 的结构特点. 此外, 为了更好地实现梯度的反向传播, 我们使用了残差结构. 残差 ConvNeXt 模拟了 Swin Transformer 倒置瓶颈的优势, 在宽度上收缩和扩展以及在感受野上的收缩, 使得在下采样过程中的语义特征信息变得丰富, 不易丢失. 此外, 采用残差 ConvNeXt 来模拟 Swin Transformer 的特性和训练特点, 同时结合卷积网络固有的感应偏差, 在医学数据集稀少的情况下, 能以相对较低的参数量在下采样的过程中实现准确的特征提取.

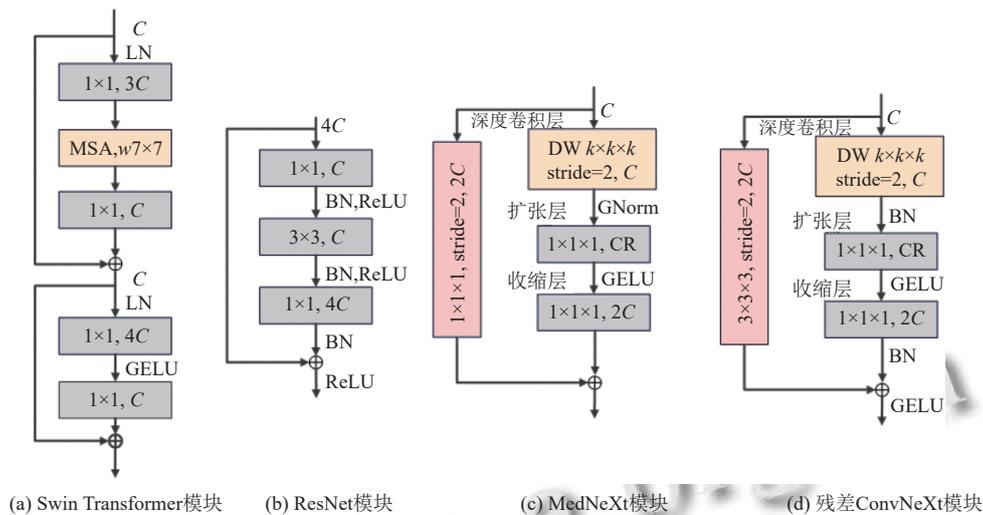


图2 4种模块详细结构

### 1.3 解码器结构

解码器中的上采样模块主要包括转置卷积、批量归一化 (batch norm, BN) 和 ReLU 激活函数, 在上采样的过程中, 首先通过转置卷积逐渐恢复特征图的分辨率, 转置卷积减少了网络扩展情况下的特征损失, 同时实现了降维, 并且逐步降低特征图的通道数. 在转置卷积之后, 采用 BN 和 ReLU 激活函数, 加快模型的收敛. 在上采样模块之后, 采用  $3 \times 3 \times 3$  的卷积对特征图进一步细化, 除了上采样的最后一层外, 我们在  $3 \times 3 \times 3$  的卷积之后也采用了 BN 和 ReLU 激活函数, 经过细化之后的特征图与相同尺寸的下采样特征图在通道级别上相连, 连接之后特征图通道的数目是之前的两倍. 通道数从 864 开始, 然后逐步进行减少, 依次为 288、144、72 和 3, 最终得到了通道数为 3 的脑肿瘤预测分割图.

## 2 实验

### 2.1 数据集

在本实验中, 使用 BraTS 2020 和 FeTS 2021 数据集进行实验, 这两个数据集都是由 MICCAI 脑肿瘤分割挑战赛提供的完全注释的公开数据集, 从不同患者中利用不同的 MRI 扫描仪所获得的, 这些 MRI 图像都是严格对齐的, 重采样至  $1 \times 1 \times 1$  mm 的各向同性分辨率, 并剥离颅骨<sup>[16]</sup>. MRI 数据集通常包括 4 个序列: T1 加权序列 (T1)、T2 加权序列 (T2)、T1 加权增强序列 (T1ce) 和 FLAIR 序列 (FLAIR). 脑肿瘤组织在不同序列中表现不同, 每一个序列的作用都是为了突出观察脑肿瘤内某一个组织的位置, 大小和形状, 并提供脑

肿瘤的互补信息, 便于分割. BraTS 2020 由训练集和验证集组成, 其中训练集是由 369 个病例组成, 验证集包括 125 个病例. FeTS 2021 训练集包括 341 个病例, 验证集有 111 个病例. 每一个病例的 MRI 图像都是由 4 个模态组成. 训练集提供事实标签, 主要包含 4 类: 背景 (标签 0)、坏死 (标签 1)、瘤周水肿 (标签 2) 和增强的肿瘤 (标签 4). 而我们的评估指标是针对这 4 类标签组合而成的 3 个子区域进行评估, 分别是: 标签 4 组成的增强肿瘤区域 (ET)、由标签 1 和 4 组成的肿瘤核心区 (TC)、除了背景区域外的整个肿瘤区域 (WT). 验证集并没有提供真实标签, 我们将训练好的模型对验证集进行分割, 然后将分割结果上传至 CBICA 在线评估平台, 通过这个官方平台来获得预测分割图的准确程度.

### 2.2 实现细节

本文的模型使用 PyTorch 实现, 在 3 台 NVIDIA RTX A5000 上训练了 2 000 个周期, 将批处理大小设置为 6, 并使用初始学习率为 0.000 6 的 AdamW 优化器来优化我们的模型. 使用了相邻体素重叠为 0.7 的滑动窗口方法来进行推论. 将 3D 原始脑肿瘤 MRI 图像随机裁剪成  $128 \times 128 \times 128$  的尺寸作为模型的输入尺寸. 训练过程使用了 50 个周期的线性预热的方法, 先从一个较小的学习率训练 50 个周期, 然后再使用初始学习率来训练模型, 同时使用余弦退火的方法来调整学习速率.

#### 2.2.1 预处理和后处理

为了避免由于脑肿瘤的数据量比较稀少而导致

的过拟合问题,首先本次实验将MRI脑肿瘤图像进行了随机裁剪,将图像尺寸由 $240 \times 240 \times 155$ 裁剪为 $128 \times 128 \times 128$ 的输入数据来增加图像的数量。然后,我们对数据应用概率为0.5的随机轴镜像翻转,最后,在每一个通道上进行数据增强,应用(0.9, 1.1)的随机强度比例,在范围(-0.1, 0.1)中应用随机强度平移。

由于脑肿瘤在整个MRI图像中占比非常小,在预测的过程中出现了很多评分为0的数值,这是假阳性导致的结果,所以在预测的过程中,使用测试时间增强的技术(TTA),将验证集输入图像进行水平,垂直,对角等多个角度进行7次翻转,然后分别送入训练好的模型进行测试,测试完成的7次分割结果再分别进行相对应的角度翻转,翻转回原来的角度。我们将翻转后的7次分割结果与原始图像的分割结果求取平均值得到最终的分割图。通过使用TTA技术,可以对验证数据集进行数量以及特征的增强,很大程度上解决了假阳性的问题,并且模型精度得到了提升。

### 2.2.2 损失函数

损失函数是为了测量真实分割与预测分割之间的偏差,在实验中使用软骰子损失函数直接优化重叠度量<sup>[17]</sup>,产生了比通常使用的加权损失函数更好的精度,这个值在(0, 1)之间,我们的目标是最小化这个损失函数。

$$L(G, Y) = 1 - \frac{2}{J} \sum_{j=1}^J \frac{\sum_{i=1}^I G_{i,j} Y_{i,j}}{\sum_{i=1}^I G_{i,j}^2 + \sum_{i=1}^I Y_{i,j}^2} \quad (1)$$

其中,  $G_{i,j}$ 表示在体素 $i$ 上的关于类别 $j$ 的one-hot编码真实值,  $Y_{i,j}$ 表示在体素 $i$ 上的关于类别 $j$ 的预测概率,  $I$ 表示体素的数量,  $J$ 表示脑肿瘤类别总数。

### 2.3 评价指标

我们采用骰子相似系数(Dice)和豪斯多夫距离(Hausdorff)两个评价指标来测试模型的分割精度,这些指标是针对WT, TC和ET来计算的,所以我们不根据提供的事实标签(0, 1, 2, 4)进行训练,有助于针对性的增加模型的分割准确度。其中骰子分数目的是评估我们的预测分割结果和真实分割图之间的相似程度,这个数值越大说明模型的分割效果越好。骰子分数被定义为:

$$Dice = \frac{2TP}{FN + FP + 2TP} \quad (2)$$

豪斯多夫距离目的是测量预测的分割结果和真实分割图之间的距离差距,这个数值越小说明预测结果图越接近真实分割图。豪斯多夫距离被定义为:

$$Hausdorff95 =$$

$$\max \left( \max_{p \in P} (95\%) \max_{t \in T} \|p - t\|, \max_{t \in T} (95\%) \max_{p \in P} \|t - p\| \right) \quad (3)$$

其中,  $TP$ 、 $FN$ 和 $FP$ 分别表示真阳性、假阴性和假阳性。 $t$ 和 $p$ 表示地面真实值和预测分割结果图的体素。

### 2.4 在BraTS 2020上的实验结果展示

本节将展示ASUNet在BraTS 2020验证集上与其他先进的模型在骰子分数和豪斯多夫距离上的比较。在BraTS 2020验证集上,ASUNet通过实验在ET、WT和TC上分别取得了77.08%、90.83%、83.41%和24.13 mm、4.92 mm、12.28 mm的Dice分数和豪斯多夫距离。表1中的数据对比表明,ASUNet在WT和TC上表现出了较大的优势。此外,为了验证ASUNet和真实分割图(GT)直观上的差异,我们将训练集随机划分为4:1,将1/5的训练集抽出GT作为测试集,剩余的数据集用于训练,将训练好的模型来预测测试集的数据。由于测试集是训练集的一部分,所以GT是可以获得的,进而比较预测分割结果图与GT之间视觉分割的差异。可视化结果如图3所示,其中前4列表示MRI脑肿瘤的4个模态,GT表示脑肿瘤的真实分割图,最后一列是所提出模型的预测分割图。从图3可以看出,ASUNet的分割预测结果与GT是相近的,进一步证明ASUNet模型的有效性。

#### 2.4.1 与基于Transformer和CNN的混合模型进行比较

在基于Transformer和CNN的混合模型中,TransBTS<sup>[12]</sup>结合CNN来提取局部特征信息,然后将提取的局部特征信息送入Transformer中实现全局信息建模,相比于TransBTS,ASUNet在ET的分割上低了1.65%,这主要是由于TransBTS同时在深度维度和空间维度上有效的建模局部和全局特征,促进特征信息的充分表示。但TransBTS的参数数量高达32.99M,比ASUNet的参数高出2.7倍。UNETR<sup>[18]</sup>方法采用ViT作为编解码器。ASUNet是一个完全基于卷积的网络,在捕获多尺度上下文信息方面存在不足,导致对具有不同形状和尺度的脑结构的次优分割。在TC的分割上,豪斯多夫距离相比于UNETR差了1.65。SwinBTS<sup>[19]</sup>将Swin Transformer作为编解码器并使用CNN进行上下采样来获得最终的分割结果。Peiris等人<sup>[20]</sup>利用自我注意和交叉注意设计了一个新颖的Transformer架构。

表 1 在 BraTS 2020 数据集上将所提出的模型与先进的 BraTS 方法进行比较

类型	方法	Dice (%)			豪斯多夫距离 (mm)		
		ET	WT	TC	ET	WT	TC
基于卷积的模型方法	CH-UNet <sup>[21]</sup>	78	90	82	32.78	7.41	12.82
	Point-UNet <sup>[22]</sup>	78.98	89.71	82.75	—	—	—
	Akbar等人 <sup>[23]</sup>	72.91	88.58	80.19	31.97	10.26	13.58
	DR-UNet104 <sup>[24]</sup>	67.56	88.62	67.21	47.62	12.11	15.74
	3D-UNet <sup>[25]</sup>	72.00	83.00	77.00	37.42	12.34	13.11
	Attention-UNet <sup>[26]</sup>	<b>82.07</b>	88.93	81.36	<b>7.89</b>	10.47	<b>9.14</b>
Transformer + 卷积的混合模型	TransBTS <sup>[12]</sup>	78.73	90.09	81.73	16.72	6.47	10.47
	UNETR <sup>[18]</sup>	71.18	88.3	75.85	34.46	8.18	10.63
	SwinBTS <sup>[19]</sup>	77.36	89.06	80.30	26.84	8.56	15.78
	Peiris等人 <sup>[20]</sup>	76.45	88.73	80.39	28.99	9.54	14.76
	ASUNet	77.08	<b>90.83</b>	<b>83.41</b>	24.13	<b>4.92</b>	12.28

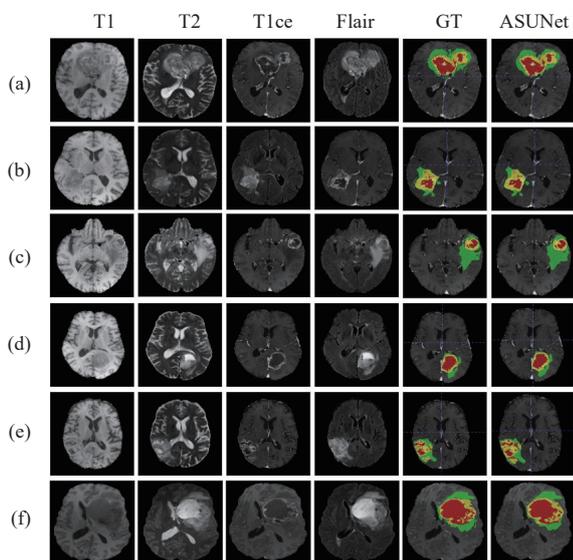


图 3 在 BraTS 2020 数据集上的视觉分割结果

#### 2.4.2 与基于 CNN 的模型进行比较

在基于 CNN 的方法中, CH-UNet<sup>[21]</sup>提出了角注意模块分别在 3D MRI 脑肿瘤的 3 个轴方向上提取信息. Point-UNet<sup>[22]</sup>设计了一种高效的上下文感知点采样机制; Point-UNet 在 ET 分割上的 *Dice* 得分, 相比于 ASUNet 高出了 1.9%. 但是在 WT 和 TC 上分别比 ASUNet 低了 1.12% 和 0.66%, 并且模型复杂度要高于 ASUNet. Akbar 等人<sup>[23]</sup>提出将 *atrous* 卷积和注意力进行有效的结合. DR-UNet104<sup>[24]</sup>使用 104 层的卷积层和瓶颈残差单元, 但是使用的 2D 卷积, 对于 3D 脑肿瘤图像的结构细节信息并未完全获取, 导致在 ET 和 TC 的较小区域分割上效果较差, 仅为 67.56% 和 67.21%. Attention-UNet<sup>[26]</sup>将注意力与 U-Net 进行结合. 该注意力模块产生的门控信号有效地抑制了输入图像的不相关信息同

时突出特定局部区域的相关特征. ASUNet 通过跳跃连接直接将低级语义信息和高级语义信息相连接的方式, 并不能充分发挥低级语义信息在上采样的作用, 这就导致了 ASUNet 模型在 ET 小区域图像的分割性能较差, 在 ET 的 *Dice* 分数上低了 4.99%. 但是 ASUNet 对远程依赖关系进行有效的建模, 并且结合卷积固有的感应偏差, 在 WT 和 TC 上获得了比较好的分割效果.

#### 2.5 在 FeTS 2021 上的实验结果展示

本节在 FeTS 2021 数据集上进行实验来验证该网络的泛化性能. 如表 2 所示, 展示了 ASUNet 与其他先进模型在骰子分数和豪斯多夫距离上的比较. 其中, Khan 等人<sup>[27]</sup>提出两种权重聚合方法, 在 FeTS 2021 挑战赛中获得了第 2 名. 观察表 2, 可以发现该方法在 ET 上获得了最佳的豪斯多夫距离, 但它的骰子分数在 3 个脑肿瘤子区域中是最低的. Hu 等人<sup>[28]</sup>提出了一种基于集成学习的方法, 但它可能导致模型过拟合. 从表 2 可以看出, 该方法在 ET 上的骰子得分很差, 比 ASUNet 低了 4.1%. Pawar 等人<sup>[29]</sup>介绍了一种两阶段的方法用于脑肿瘤精确分割, 细化了脑肿瘤区域的边界位置信息. 与其他最先进的模型相比, 实现了最佳性能, 但需要训练 2 次, 这就导致 2 倍的参数量和计算量, 而且相比 ASUNet, 在 ET、WT 和 TC 上分别低 0.5%、1.2% 和 0.7%. 观察表 2, 可以明显发现, ASUNet 模型在所有脑肿瘤子区域上的 *Dice* 分数明显优于其他最先进的办法. 虽然 ET 的豪斯多夫距离不是最好的, 但 WT 和 TC 的豪斯多夫距离明显优于其他最先进的办法. 综上所述, 在 BraTS 2020 和 FeTS 2021 数据集上的实验结果表明, ASUNet 表现出较强的泛化能力.

表2 在 FeTS 2020 数据集上将所提出的模型与先进的 BraTS 方法进行比较

方法	Dice (%)			豪斯多夫距离 (mm)		
	ET	WT	TC	ET	WT	TC
Khan等人 <sup>[27]</sup>	70.9	85.0	70.4	<b>15.2</b>	13.2	16.3
Hu等人 <sup>[28]</sup>	71.5	89.0	76.3	—	—	—
Pawar等人 <sup>[29]</sup>	75.1	89.2	83.5	32.9	7.2	7.1
Sahil等人 <sup>[30]</sup>	72.1	87.4	77.3	22.1	22.2	37.0
ASUNet	<b>75.6</b>	<b>90.4</b>	<b>84.2</b>	27.1	<b>4.9</b>	<b>6.8</b>

为了直观地说明 ASUNet 和 GT 预测分割结果之间的差异, 使用与 BraTS 2020 相同的数据集划分方法来预测 ASUNet 的分割结果. 图4显示了 ASUNet 预测分割图与真实分割图之间的视觉比较. 从图4可以发现, ASUNet 的分割结果与 GT 非常相似, 同时进一步表明了其对 FeTS 2021 数据集的适用性.

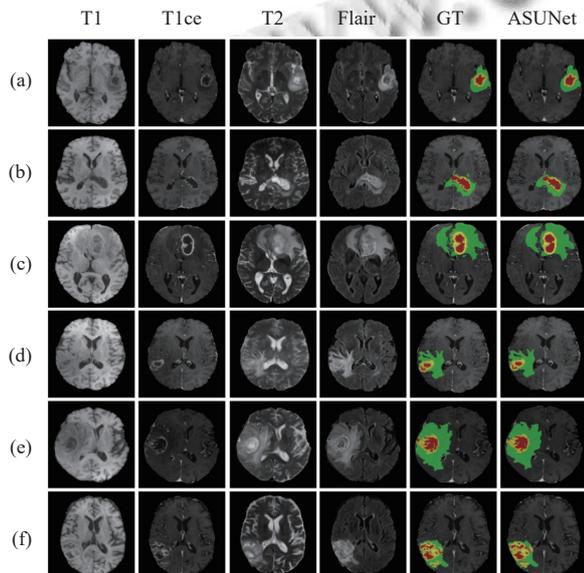


图4 在 FeTS 2021 数据集上的视觉分割结果

## 2.6 消融实验

为了更好地说明 ASUNet 中的每一个模块发挥的重要作用, 本节将利用 BraTS 2020 数据集进行消融实验来验证. 我们分别评估使用残差 ConvNeXt 模块进行下采样的有效性, 对比 U 型结构不同设计对实验结果的影响, 以及有无后处理、有无深度监督对实验结果的影响.

### 2.6.1 不同下采样策略对模型的影响

本节通过采用不同的下采样策略来验证残差 ConvNeXt 模块的有效性, 所对比的下采样方法是步长为 2 的残差结构, 配置如图5所示, 其中 C 表示通道

数, Stride 表示步长. 在下采样中, 采用步长为 2 的残差结构, 以一种常规卷积的方式降低特征图的尺寸, 增加特征图的通道数, 但是忽略了下采样过程中特征细节丢失的问题.

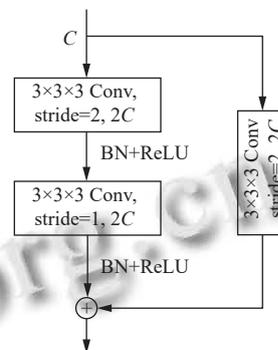


图5 用于下采样步长为 2 的残差结构

从表3中的实验数据可以明显发现, 当采用步长为 2 的残差连接时, ET 和 TC 的 Dice 得分非常低, 分别为 71.41%、81.00%. 并且从参数的数量可以发现, 残差 ConvNeXt 模块极大地降低了参数的数量, 大约是传统残差结构的 1/2. 残差 ConvNeXt 模块利用特征尺寸和通道的缩放和扩展, 不仅保持了下采样过程中语义信息丰富性, 而且以一种计算代价小的方式建立了远程依赖关系, 保持了卷积固有的感应偏差.

表3 在 BraTS 2020 验证集上对不同的下采样策略进行 Dice 值的对比

方法	Dice (%)			参数量 (M)
	ET	WT	TC	
残差结构	71.41	90.22	81.00	24.32
MedNeXt模块	74.93	89.79	80.62	6.54
残差ConvNeXt (K=3)	77.46	90.03	82.52	<b>12.1</b>
残差ConvNeXt (K=5)	<b>77.88</b>	90.49	82.86	12.16
残差ConvNeXt (K=7)	77.08	<b>90.83</b>	<b>83.41</b>	12.27

表3也显示, 使用残差 ConvNeXt 进行下采样在 ET、WT 和 TC 的 Dice 分数上提升了 5.67%、0.61% 和 2.41%. 在残差 ConvNeXt 模块的深度卷积层, 不同的 K 值产生了不同的分割结果, 从表3还可以看出, 虽然残差 ConvNeXt 模块采用 K=5 时, 在 ET 上获得了最高的 Dice 分数, 但是当采用 K=7 时, 在 WT 和 TC 上的分割效果是最好的, 因此在深度卷积层, 我们将 K 设置为 7.

### 2.6.2 U 型结构的不同设计对模型的影响

为了在编码过程中提取详细的语义特征信息, 我们采用非对称的 U 型网络架构, 在非对称的 U 型结构

中,底层采用了残差连接的方式,扩充了U型结构底层的通道数,促进特征图更详细的表示.在不改变网络的其他功能模块的前提下,我们将特征图按照对称U型结构进行实验,在下采样的过程中将特征图的尺寸减少到 $4 \times 4 \times 4$ ,然后通过转置卷积和细化操作得到最终的预测分割图.得到的骰子分数如表4所示.从表4中的实验数据,我们可以发现,使用不对称的U型结构比对称U型结构在ET, WT和TC上分别提升了0.09%, 0.38%和1.15%,效果比较明显,说明了不对称U型结构对脑肿瘤的分割具有显著的效果.

表4 在BraTS 2020验证集上进行不对称U型结构和对称U型结构的骰子分数Dice值对比(%)

方法	ET	WT	TC
对称U型结构	76.99	90.45	82.26
ASUNet	<b>77.08</b>	<b>90.83</b>	<b>83.41</b>

### 2.6.3 后处理对模型的影响

本文应用TTA技术来提高模型的稳健性,由于脑肿瘤MRI图像类别的不均衡性,其中有一些MRI图像的标签为0,这样就会导致假阳性的出现,为了避免出现这种情况,通过使用TTA技术来扩充验证集的数量.实验结果如表5所示,从中可以发现,通过使用TTA技术,ASUNet在ET, TC和WT上的Dice分数值分别增长了0.23%, 0.06%和0.57%,此外TTA作为后处理技术同样也适用于其他类型的医学图像分割以及用于其他模型的精度提升.

表5 在BraTS 2020验证集上使用TTA与不使用TTA的Dice值对比(%)

方法	ET	WT	TC
不使用TTA	76.85	90.77	82.84
使用TTA	<b>77.08</b>	<b>90.83</b>	<b>83.41</b>

### 2.6.4 深度监督对模型的影响

本节中对U型结构有无深度监督进行了比较,在上采样过程中加入深度监督,通过使用不同的权重,对每一个上采样特征图根据不同的权重求和并计算损失,在没有深度监督的U型结构中,保持其他结构不变.如表6所示,我们可以发现,加入深度监督在ET, WT和TC上分别提升了4.59%、0.41%和0.26%.

表6 在BraTS 2020验证集上使用深度监督与不使用深度监督的Dice值对比(%)

方法	ET	WT	TC
不使用深度监督	72.49	90.42	83.15
使用深度监督	<b>77.08</b>	<b>90.83</b>	<b>83.41</b>

## 3 结论与展望

本文提出了一个新型的非对称U型卷积神经网络ASUNet,用于MRI脑肿瘤图像的精确分割.首先,为了减少特征语义信息丢失,我们设计了一个非对称的U型框架结构,在U型结构的底层建立残差连接,保存了之前阶段的特征信息,解决了网络训练过程中的梯度问题.其次,在下采样中使用残差ConvNeXt模块,结合深度可分离卷积和逐点卷积,避免了在下采样过程中丢失过多的细节信息.最后,我们在上采样的过程中采用了深度监督算法,通过对每一层的上采样特征图进行监督,有助于最终的预测分割图更接近于真实分割图.在BraTS 2020数据集的实验中,与多个先进的模型进行对比,本文的方法取得了较好的结果.此外,本文还对不同的下采样方法、不同U型结构的设计以及TTA和深度监督对模型效果的影响进行了消融实验,实验数据表明网络在设计 and 实现上也有独特的思路.

通过一系列实验证明,ASUNet具有良好的分割性能.但也有一定的局限性:ASUNet只在多模式公共数据集上进行实验.目前,我们正在与医院谈判合作,以获得更多临床数据.在未来的工作中,我们将使用临床数据,以便真正将其应用于医学临床研究.

### 参考文献

- Roy S, Koehler G, Ulrich C, *et al.* MedNeXt: Transformer-driven scaling of convNets for medical image segmentation. Proceedings of the 26th International Conference on Medical Image Computing and Computer-assisted Intervention. Vancouver: Springer, 2023. 405-415.
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431-3440.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-assisted Intervention. Munich: Springer, 2015. 234-241.
- He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770-778.
- Zhang XY, Zhou XY, Lin MX, *et al.* ShuffleNet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake: IEEE,

2018. 6848–6856.
- 6 Xie SN, Girshick R, Dollár P, *et al.* Aggregated residual transformations for deep neural networks. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5987–5995.
  - 7 Howard AG, Zhu ML, Chen B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
  - 8 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017 6000–6010.
  - 9 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2020.
  - 10 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 9992–10002.
  - 11 Chen JN, Lu YY, Yu QH, *et al.* TransUNet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306, 2021.
  - 12 Wang WX, Chen C, Ding M, *et al.* TransBTS: Multimodal brain tumor segmentation using Transformer. Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention. Strasbourg: Springer, 2021. 109–119.
  - 13 Li JY, Wang WX, Chen C, *et al.* TransBTSV2: Towards better and more efficient volumetric segmentation of medical images. arXiv:2201.12785, 2022.
  - 14 Xie YT, Zhang JP, Shen CH, *et al.* CoTr: Efficiently bridging CNN and Transformer for 3D medical image segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg: Springer, 2021. 171–180.
  - 15 Liu Z, Mao HZ, Wu CY, *et al.* A convNet for the 2020s. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 11966–11976.
  - 16 Hatamizadeh A, Nath V, Tang YC, *et al.* Swin UNETR: Swin Transformers for semantic segmentation of brain tumors in MRI images. Proceedings of the 7th International MICCAI Brainlesion Workshop. Springer, 2021. 272–284.
  - 17 Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proceedings of the 4th International Conference on 3D Vision (3DV). Stanford: IEEE, 2016. 565–571.
  - 18 Hatamizadeh A, Tang YC, Nath V, *et al.* UNETR: Transformers for 3D medical image segmentation. Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. New Orleans: IEEE, 2022. 1748–1758.
  - 19 Jiang Y, Zhang Y, Lin X, *et al.* SwinBTS: A method for 3D multimodal brain tumor segmentation using Swin Transformer. Brain Sciences, 2022, 12(6): 797. [doi: 10.3390/brainsci12060797]
  - 20 Peiris H, Hayat M, Chen ZL, *et al.* A robust volumetric Transformer for accurate 3D tumor segmentation. Proceedings of the 25th International Conference on Medical Image Computing and Computer-assisted Intervention. Singapore: Springer, 2022. 162–172.
  - 21 Xu WJ, Yang HH, Zhang MY, *et al.* Brain tumor segmentation with corner attention and high-dimensional perceptual loss. Biomedical Signal Processing and Control, 2022, 73: 103438. [doi: 10.1016/j.bspc.2021.103438]
  - 22 Ho NV, Nguyen T, Diep GH, *et al.* Point-UNet: A context-aware point-based neural network for volumetric segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention. Strasbourg: Springer, 2021. 644–655.
  - 23 Akbar AS, Faticah C, Suciati N. Single level UNet3D with multipath residual attention block for brain tumor segmentation. Journal of King Saud University—Computer and Information Sciences, 2022, 34(6): 3247–3258. [doi: 10.1016/j.jksuci.2022.03.022]
  - 24 Colman J, Zhang L, Duan WT, *et al.* DR-Unet104 for Multimodal MRI brain tumor segmentation. Proceedings of the 6th International MICCAI Brainlesion Workshop. Lima: Springer, 2021. 410–419.
  - 25 Çiçek Ö, Abdulkadir A, Lienkamp SS, *et al.* 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Proceedings of the 19th International Conference on Medical Image Computing and Computer-assisted Intervention. Athens: Springer, 2016. 424–432.
  - 26 Oktay O, Schlemper J, Le Folgoc L, *et al.* Attention U-Net: Learning where to look for the pancreas. arXiv:1804.03999, 2018.
  - 27 Khan MI, Jafaritadi M, Alhoniemi E, *et al.* Adaptive weight aggregation in federated learning for brain tumor segmentation. Proceedings of the 7th International MICCAI Brainlesion Workshop. Springer, 2021. 455–469.
  - 28 Hu JY, Gu XJ, Wang ZQ, *et al.* Mixture of calibrated networks for domain generalization in brain tumor segmentation. Knowledge-based Systems, 2023, 270: 110520. [doi: 10.1016/j.knsys.2023.110520]
  - 29 Pawar K, Zhong SJ, Chen ZL, *et al.* Brain tumor segmentation using two-stage convolutional neural network for federated evaluation. Proceedings of the 2021 International MICCAI Brainlesion Workshop. Springer, 2021. 494–505.
  - 30 Nalawade S, Ganesh C, Wagner B, *et al.* Federated learning for brain tumor segmentation using MRI and Transformers. Proceedings of the 2021 International MICCAI Brainlesion Workshop. Springer, 2021. 444–454.

(校对责编: 张重毅)