

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2023)04-1157-16

论文引用格式: Sun J K, Zhang R, Guo L J and Wang J H. 2023. Multi-scale feature fusion and additive attention guide brain tumor MR image segmentation. Journal of Image and Graphics, 28(04): 1157-1172(孙家阔, 张荣, 郭立君, 汪建华. 2023. 多尺度特征融合与加性注意力指导脑肿瘤MR图像分割. 中国图象图形学报, 28(04): 1157-1172)[DOI:10.11834/jig.211073]

多尺度特征融合与加性注意力指导脑肿瘤MR图像分割

孙家阔¹, 张荣^{1*}, 郭立君¹, 汪建华²

1. 宁波大学信息科学与工程学院, 宁波 315211; 2. 宁波大学医学院附属医院, 宁波 315211

摘要: **目的** U-Net是医学图像分割领域中应用最为广泛的基础分割网络,然而U-Net及其各种增强网络在跳跃连接时仅利用相同尺度特征,忽略了具有互补信息的多尺度特征对当前尺度特征的指导作用。同时,跳跃连接时编码器特征和解码器特征所处的网络深度不同,二者直接串联会产生语义特征差距。针对这两个问题,提出了一种新型分割网络,以改进现有网络存在的不足。**方法** 首先,将编码器不同层级具有不同尺度感受野的特征进行融合,并在融合特征与编码器各层级特征间引入加性注意力对编码器特征进行指导,以增强编码器特征的判别性;其次,在编码器特征和解码器特征间采用加性注意力来自适应地学习跳跃连接特征中的重要特征信息,以降低二者间的语义特征差距。**结果** 在多模态脑肿瘤数据集BraTS2020(multimodal brain tumor segmentation challenge 2020)上评估了所提出的网络模型,并进行了消融实验和对比实验。实验结果表明,所提出的网络在BraTS2020验证数据集上关于整个肿瘤、肿瘤核心和增强肿瘤的平均Dice分别为0.887 5、0.719 4和0.706 4,优于2D网络DR-Unet104(deep residual Unet with 104 convolutional layers)的分割结果,其中肿瘤核心和增强肿瘤的分割结果分别高出后者4.73%和3.08%。**结论** 所提出的分割网络模型,通过将编码器中具有互补信息的多尺度特征进行融合,然后对当前尺度特征进行加性注意力指导,同时在编码器和解码器特征间采用加性注意力机制来降低跳跃连接时二者间的语义特征差距,能更精准地分割MR(magnetic resonance)图像中脑肿瘤子区域。

关键词: 医学图像分割;脑肿瘤;磁共振(MR)图像;U-Net;多尺度特征融合;加性注意力

Multi-scale feature fusion and additive attention guide brain tumor MR image segmentation

Sun Jiakuo¹, Zhang Rong^{1*}, Guo Lijun¹, Wang Jianhua²

1. Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China;

2. Affiliated Hospital of Medicine School of Ningbo University, Ningbo University, Ningbo 315211, China

Abstract: **Objective** U-Net can be as the basic network in medical image segmentation. For U-Net and its various augmented networks, the encoder can extract features from input images in terms of a series of convolution and down-sampling operations. With the convolution and down-sampling operations at each layer of the encoder, the feature map sizes are decreased and the receptive field sizes can be remained to increase. For the network training, each level of the encoder can

收稿日期: 2021-11-24; 修回日期: 2022-04-25; 预印本日期: 2022-05-02

* 通信作者: 张荣 zhangrong@nbu.edu.cn

基金项目: 浙江省公益技术研究项目(LGF21F020008); 浙江大学CAD&CG国家重点实验室开放项目(A2119)

Supported by: Zhejiang Provincial Public Welfare Technology Research Project (LGF21F020008); Open Project Program of the State Key Laboratory of CAD&CG, Zhejiang University (A2119)

learn discriminative feature information at the current scale. To improve its feature utilization, the augmented U-Net schemes can melt skip connections between the encoder features and the decoder features into feature information-reused of shallow layers. However, the same scale are concatenated the features only via the skip-connected channel, and the role of multi-scale features with complementary information can be ignored. In addition, encoder features are oriented at a relatively shallow position in the overall network structure, while decoder features are based on a relatively deep position. As a result, a semantic feature gap is required to be bridged between encoder features and decoder features when skip connections are made. To optimize the U-Net and its augmented networks model, a novel segmentation network model is developed. **Method** We construct a segmentation network in terms of multi-scale feature fusion and additive attention mechanism. First, the features are fused in relevant to multi-scale receptive fields at different levels of the encoder. To guide the encoder features and enhance their discrimination ability, additive attention is introduced between the fused features and the encoder features at each level of the encoder. Second, to bridge the gap between the two semantic features, encoder and decoder features-between additive attention is used to learn important feature information in skip connections features adaptively. Experiments are carried out based on five-fold cross-validation. Multimodal magnetic resonance (MR) images of 234 high-grade glioma (HGG) samples and 59 low-grade glioma (LGG) samples in the BraTS2020 training dataset are used as the training data. MR images of 59 HGG samples and 15 LGG samples from the BraTS2020 training dataset are regarded as the validation data. The validation dataset of BraTS2020 is used as the final test data. The images of each modality are normalized using the Z-Score approach on the basis of the original data. The loss function is used in terms of the categorical cross-entropy loss function. Our model proposed is equipped with Ubuntu 18.04 operating system using Pycharm based on Keras, and the network model is trained and predicted on a workstation with a 16 GB graphics memory NVIDIA Quadro P5000 GPU. An ADAM optimizer is used in terms of a learning rate of 0.0001 and the parameters in the network are initialized using the he_normal parameter initialization method. Our batch size for training the network is set to 12 and the model took 3 days to train after 150 iterations. **Result** To evaluate the performance of the proposed model, the Dice coefficient and the 95% Hausdorff distance (HD95) are used as evaluation metrics for the segmented regions of whole tumor (WT), tumor core (TC) and enhancing tumor (ET). To obtain quantitative evaluation results for these evaluation metrics, network-based segmentation results are uploaded to the BraTS2020 online evaluation platform. First, the segmentation effectiveness of the proposed network is verified on the BraTS2020 validation dataset. The experimental results show that the average Dice of the proposed network in relevant to ET, WT and TC are 0.7064, 0.8875 and 0.7194 of each. Then, the proposed network is investigated in ablation experiments to validate the effectiveness of the proposed multi-scale feature fusion module, the fused feature additive attention module, and the encoder-decoder additive attention concatenate module. The results of the ablation experiments show that the addition of the proposed multi-scale feature fusion module to the backbone network improves the average Dice of the network of ET, WT and TC by 2.23%, 2.13% and 0.97%, respectively. In addition, the average Dice values of the network about ET, WT and TC are increased by 1.54%, 0.58% and 1.45% more after adding the proposed multi-scale feature fusion and fused feature additive attention modules to the network. The average Dice values of the network related to ET, WT and TC are increased by 2.46%, 0.82% and 3.51% further after the proposed encoder-decoder additive attention concatenate module is added to the network. Finally, our optimal network is compared to U-Net and popular augmented networks, as well as other non-U-Net segmentation networks. The proposed network to the 2D network DR-Unet104 is optimized by 4.73%, 3.08% and 0.13% for TC, ET and WT. Furthermore, the visualization results show that the proposed network can segment the boundaries of different tumor regions more accurately and achieve a better overall segmentation effect. **Conclusion** To segment brain tumor sub-regions in MR images more accurately, we develop a novel segmentation network model. It can fuse multi-scale features with complementary information in the encoder, and additive attention guidance can be applied to the features in the current scale. To reduce the gap between the two semantic features, additive attention mechanism is also used between the encoder features and the decoder features when skip connections are made.

Key words: medical image segmentation; brain tumor; magnetic resonance (MR) images; U-Net; multi-scale feature fusion; additive attention

0 引言

胶质瘤是最常见的原发性中枢神经系统脑肿瘤,约占所有中枢神经系统肿瘤的27%,约占恶性中枢神经系统肿瘤的80%(Ostrom等,2015)。磁共振成像技术具有非侵入式、对人体无害且获得的图像非常清晰等优点,广泛应用于临床脑肿瘤诊断(Bauer等,2013)。为了更好地辅助医生对病人进行诊断治疗,需要知道准确的脑肿瘤的大小、形状和位置。但是,医生手动从MR图像中分割脑肿瘤需要花费大量的时间和精力,并且由于脑肿瘤外观的高度复杂性,很容易出现误差。因此,脑肿瘤MR图像的自动分割可以极大提高医生的诊断效率。然而,由于不同患者脑肿瘤的大小、形状和位置存在高度变化性,且在同一患者大脑中,正常软组织与病变组织之间的边界比较模糊(任璐等,2018),准确的自动分割仍然是一项艰巨的任务。

卷积神经网络(convolutional neural network, CNN)在图像分割领域取得了重大突破。Long等人(2015)基于CNN提出了全卷积网络(fully convolutional network, FCN)用于语义分割,通过使用卷积层替换卷积神经网络结尾的全连接层,得到与输入图像相同大小的输出,并取得了较为准确的分割效果。Ronneberger等人(2015)在FCN的基础上提出了U-Net分割网络,主要包括编码器、解码器、桥接模块和跳跃连接等4部分,是目前医学图像分割领域中应用最为广泛的基础网络。为了提出更有效的医学图像分割模型,基于U-Net网络的各种增强版本应运而生。Xiao等人(2018)基于注意力机制和残差连接(He等,2016)提出了加权Res-UNet网络(weighted Res-UNet),并用于视网膜图像的分割。该网络通过在U-Net网络的各个组件中引入残差连接,进一步提高了特征的利用率。Colman等人(2020)通过将U-Net编码器中各个层级的卷积块替换为堆叠的bottleneck残差块,提出了DR-Unet104网络并用于脑肿瘤分割任务,取得了2020年脑肿瘤2D分割任务中SOTA(state-of-the-art)的分割效果。Ibtehaz和Rahman(2020)基于扩展的多分辨率块和改进的残差跳跃连接,提出了MultiResUNet网络用于皮肤癌分割任务,将U-Net网络中的卷积块替换为扩展的多分辨率块。另外,在编码器特征与解码

器对应层级特征拼接之前,对编码器特征进行残差卷积操作,以降低两者间的语义特征差距。Oktay等人(2018)基于加性注意力机制提出了attention U-Net分割网络,通过在U-Net中引入编码器和解码器特征间的门控注意力机制,以突出跳跃连接特征中的重要特征信息。

在U-Net及其增强网络中,编码器通过一系列的卷积和下采样操作逐渐对输入图像进行特征提取。随着卷积和下采样操作的进行,编码器中各层级特征图的大小在逐渐减小,而感受野在逐渐增大。随着网络的训练,编码器各个层级特征学习到了当前尺度下关键性的判别特征信息。U-Net中的跳跃连接,直接向解码器中引入了网络浅层的编码器特征信息,进一步提高了特征的利用率。然而,观察U-Net及其增强网络发现,跳跃连接时仅利用了相同尺度的特征按通道进行拼接,忽略了编码器中具有不同感受野的多尺度特征间的互补信息。因此,本文提出将编码器中具有不同感受野的各尺度特征进行融合,融合特征中兼具输入图像的局部细节特征信息和全局高级语义特征信息。然后,通过在融合特征与各尺度特征间引入加性注意力机制对各尺度特征进行指导。融合特征来自于各尺度特征,又通过加性注意力机制反作用于各尺度特征,二者相辅相成。相较于原始编码器特征,被指导后的编码器特征更具判别性。此外,编码器特征在整个网络结构中处于相对较浅的位置,而解码器特征处于相对较深的位置。因此,跳跃连接时,编码器特征和解码器特征间存在语义特征差距。为了降低跳跃连接时二者由于所处网络深度不同而产生的语义特征差距,本文在经过指导的编码器特征和解码器特征间引入加性注意力机制。通过加性注意力机制来自适应的学习跳跃连接特征中的重要特征信息。

本文的贡献主要分为以下3部分。1)向U-Net网络中引入了多尺度特征融合模块(multi-scale feature fusion, MSFF)。该模块将编码器中具有不同感受野的各尺度特征进行融合,融合特征中兼具局部细节特征信息和全局高级语义特征信息。然后,基于MSFF模块,提出了融合特征加性注意力模块(fusion feature additive attention, FFAA)。通过在融合特征与编码器各尺度特征间引入加性注意力机制对各尺度特征进行指导。相较于原始编码器特征,被指导后的编码器特征更具判别性。2)提出了编码

器—解码器加性注意力拼接模块(encoder-decoder additive attention concatenate, E-DAAC)。该模块在经过融合特征指导的编码器特征和解码器特征间采用加性注意力机制,以降低跳跃连接时二者由于所处网络深度不同而产生的语义特征差距,即通过加性注意力机制来自适应地学习跳跃连接特征中的重要特征信息。3)基于多模态脑肿瘤分割挑战赛 BraTS2020 验证数据集的实验结果表明,在相同的实验环境下,本文提出的2D网络取得了最好的分割效果。关于整个肿瘤(whole tumor, WT)、肿瘤核心(tumor core, TC)和增强肿瘤(enhancing tumor, ET)的平均Dice分别为0.887 5, 0.719 4, 0.706 4。

1 相关工作

1.1 脑肿瘤图像分割

基于CNN的脑肿瘤分割方法取得了显著成就,尤其是基于U-Net的分割网络。何康辉和肖志勇(2021)提出了一种轻量级的脑肿瘤快速语义分割网络(light-weighted rapid semantic segmentation network, LRUNet),通过缩减U-Net网络中的特征通道数,并引入深度可分离卷积,以降低参数量、提高计算效率。Awasthi等人(2021)提出了一种基于Attention U-Net的多阈值分割模型MTAU(multi-threshold attention U-Net),采用多路径分割,为不同的脑肿瘤子区域分别建立独立的子模型。Xu等人(2021)提出了一种用于脑肿瘤分割的深监督模型U-Attention Net(deep supervised U-attention net),向U-Net中引入了注意力机制和深监督,即使对于较小的肿瘤区域,依然能够实现高低分辨率的特征表示。夏峰等人(2022)向U-Net网络中引入了跨阶段局部网络结构来提取特征,并改进了损失函数以消除训练样本类别不均衡的影响,从而提高脑肿瘤的分割精度。

1.2 U-Net及其改进

U-Net网络是目前医学图像分割领域中应用最为广泛的基础网络。U-Net及其各种增强版本应用于医学图像处理的众多领域。Ronneberger等人(2015)提出并利用U-Net赢得了2015年生物医学成像国际研讨会(International Symposium on Biomedical Imaging, ISBI)细胞追踪挑战赛的冠军。Çiçek等人(2016)提出使用3D卷积操作来替换U-Net中所有的2D卷积操作,所提出的3D U-Net网络更适用于

医学图像体素处理任务。但是3D卷积具有巨大的参数量,训练耗时且对于硬件的要求较高。受到残差连接(He等,2016)的启发,Ibtehaz和Rahman(2020)提出了多分辨率块和残差跳跃连接,与U-Net结合后得到了MultiResUNet模型并用于皮肤癌图像的分割。Zhou等人(2018)基于密集连接(Huang等,2017)和U-Net提出了带有深监督模块的嵌套型网络UNet++(a nested U-Net),为了降低跳跃连接时编码器和解码器特征间的语义差距,提出在编码器特征和解码器特征间使用嵌套的、密集的跳跃路径进行连接。

1.3 多尺度特征分割网络

目前,也存在一些分割方法通过改进多尺度特征来提高网络的分割性能。如金字塔场景解析网络PSPNet(pyramid scene parsing network)中通过采用多尺度特征池化模块,从同一特征引入具有不同尺度感受野的池化特征,以增强局部特征和全局特征的互补作用(Zhao等,2017)。Wang等人(2018b)通过利用卷积神经网络不同层中编码的互补信息,提出了一种配备有深度注意力特征模块的新型深度神经网络,以实现经直肠超声TRUS(transrectal ultrasound)中更好的前列腺分割。Sinha和Dolz(2021)在融合提取特征时得到的多尺度特征信息的基础上,尝试在融合特征与局部特征间使用引导的自注意力机制,以捕获更丰富的上下文依赖关系。该方法能够将局部特征与其相应的全局依赖关系进行集成,并以自适应方式突出显示相互依赖的通道图。Shi等人(2020)在分割小目标海马体时将多分辨率图像输入网络以获取多分辨率的特征信息,并利用密集连接将编码器和解码器的各分辨率特征同时利用以加强特征传播,从而改善编码器和解码器之间的信息交互。

1.4 注意力机制

注意力机制广泛应用于图像处理领域中的目标检测、分类和分割。自注意力机制(Jetley等,2018; Wang等,2018a)被提出用于消除对外部信息的依赖性。Hu等人(2018)通过采用通道注意力机制来建模特征通道之间的相互依赖关系,从而以自适应的方式调整特征中各通道的重要性程度。Woo等人(2018)在通道注意力机制的基础上,提出采用空间注意力来对特征图上各个位置间的依赖关系进行建模,可以得到各位置间的相关性系数矩阵。基于自注意力机制的Transformer(Valanarasu等,2021; Liu

等,2021)用来捕获特征间的长范围依赖。加性注意力机制最初用于句子到句子的翻译(Bahdanau 等, 2014; Shen 等, 2018)领域中,最近才应用于图像领域。加性注意力机制通过将两个输入特征进行线性变换后的特征相加,再分别经过激活函数和线性变换。将变换后的特征通过 sigmoid 函数进行激活,用来表征两个特征间的相似性关系。Oktay 等人(2018)在加性注意力的基础上,提出在 U-Net 中引入编码器和解码器特征间的门控注意力机制,进而用来突出跳跃连接中重要的特征信息。注意力机制的引入,使得神经网络对重要特征分配更多的权重,进一步提高了网络的分割性能。

通过分析以上的领域知识发现,U-Net 及其增强网络在跳跃连接时仅利用了相同尺度的特征进行拼接,忽略了编码器中多尺度特征间的互补特征信息。因此,本文提出将编码器中具有不同感受野的各尺度特征进行融合,融合特征同时包含输入图像的细节特征信息和高级语义特征信息。然后,通过在融合特征与各尺度特征间引入加性注意力机制对各尺度特征进行指导,经过全局特征信息指导后的编码

器特征更具判别性。此外,由于经过全局特征信息指导后的编码器特征和解码器特征所处的网络深度不同,因此两者之间仍存在语义特征差距。为了降低二者由于所处的网络深度不同而产生的语义特征差距,本文在经过指导的编码器特征和解码器特征间采用加性注意力机制来自适应的学习跳跃连接特征中的重要特征信息。本文阐明了所提出模块的实现方式,并基于脑肿瘤医学图像数据集 BraTS2020 来验证所提出网络模型的有效性。

2 基于多尺度特征融合和加性注意力的脑肿瘤 MR 图像分割网络

本文网络模型结构主要由 6 部分构成。除了编码器模块、解码器模块和桥接模块,还有提出的融合编码器各尺度特征信息的多尺度特征融合模块、用于指导编码器特征的融合特征加性注意力模块和用于降低语义特征差距的编码器—解码器加性注意力拼接模块。具体结构如图 1 所示,图中蓝色的矩形表示特征图,矩形左下角的数字表示特征图的尺寸

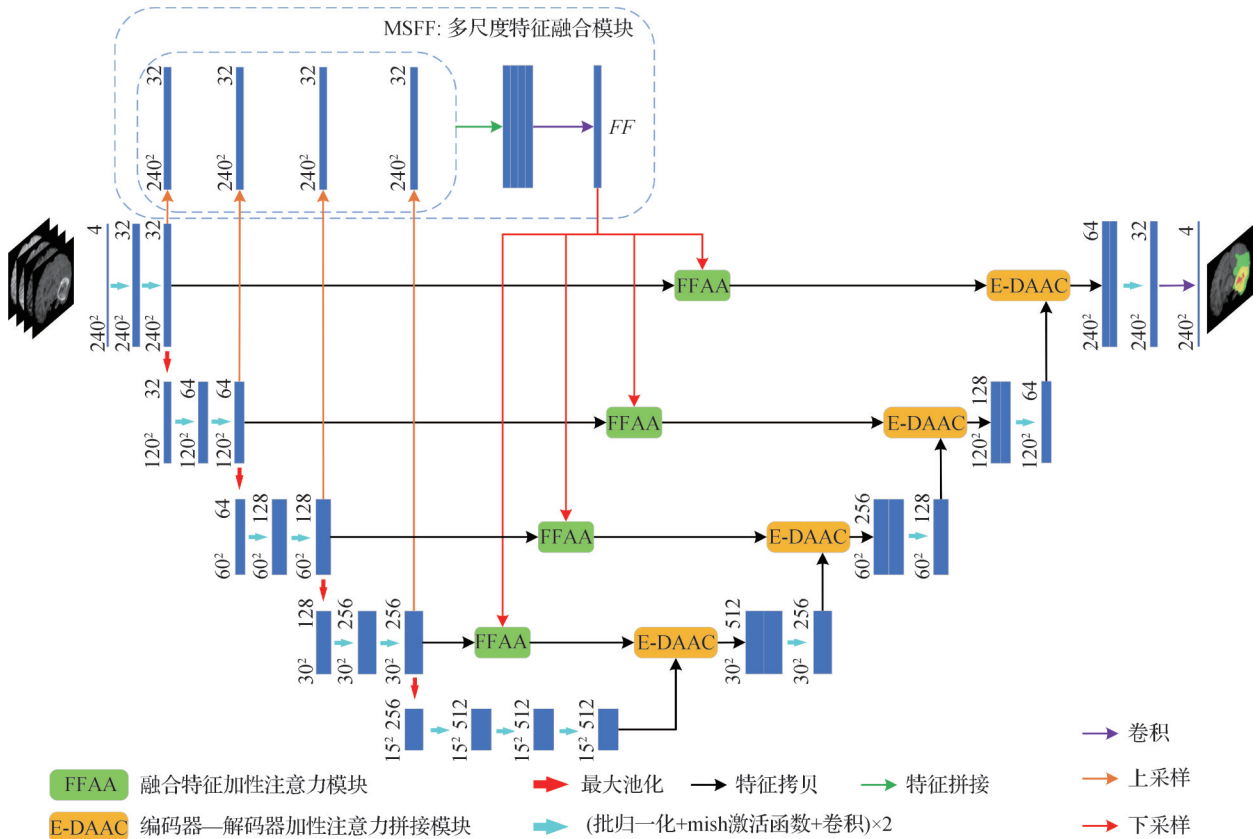


图1 本文网络结构图

Fig. 1 Structure of the proposed network

大小,左上角的数字表示特征图的通道数,不同颜色的箭头表示不同的操作。

2.1 编码器模块

编码器模块通过一系列的卷积和下采样操作对输入网络的整个图像进行特征编码,用于提取输入多模态MR图像中脑肿瘤的特征信息。编码器模块中包含有4层降采样。在这4次降采样过程中,采用最大池化操作来降低特征图的尺寸大小。最大池化可以加快网络模型的训练速度,提高运行效率。另外,通过卷积操作将特征图的通道数增加1倍以获得不同维度的特征信息。本文对编码器模块中每一层级上的特征均使用了两次叠加的批归一化(Ioffe和Szeqedy, 2015)、mish激活函数(Misra, 2019)和卷积操作。

2.2 多尺度特征融合模块

由于编码器模块经过一定数量的卷积和下采样操作,因此编码器模块中各个层级特征图的尺寸在逐渐缩小,而特征图的感受野在逐渐增大。最终,随着网络模型的训练,浅层编码器特征捕获脑肿瘤的颜色和纹理等关键的局部细节判别特征信息,而深层编码器特征捕获脑肿瘤的全局高级语义判别特征信息。本文通过将编码器模块中的各尺度特征信息上采样到编码器中特征的最大分辨率后进行拼接。然后,通过卷积操作以融合多尺度特征信息。具体细节如图1中MSFF所示。因此,多尺度特征通过卷积得到的融合特征FF中既包含了局部的细节特征信息,又包含了全局的高级语义特征信息。

2.3 融合特征加性注意力模块

U-Net及其各种增强网络模型中,编码器特征处于整个网络结构中的浅层,而解码器特征处于深层。因此,编码器与解码器相同层级特征间的跳跃连接在一定程度上利用了浅层特征信息。但是,仅利用相同尺度特征进行跳跃连接,会忽略编码器模块中各层级特征图间互补的特征信息对当前尺度特征的指导作用。因此,本文提出在多尺度融合特征和编码器特征间引入加性注意力。通过兼具局部和全局信息的多尺度融合特征对编码器模块中各尺度特征进行加性注意力指导。其中,融合特征来源于各尺度特征,又通过注意力机制反作用于各尺度特征。相较于原始编码器特征,经过融合特征加性注意力指导后的编码器特征更具判别性。在学习判别性特征的过程中,加性注意力机制起到了关键性作用。

缩放后的融合特征FF'和编码器各尺度特征间加性注意力操作的具体实现方式如图2所示。图中FF'表示融合特征FF下采样后得到的特征。由于融合特征FF中既包含脑肿瘤的纹理、颜色和边界等局部细节特征信息,又包含脑肿瘤的全局高级语义特征信息。那么FF'同样包含局部和全局特征信息。首先,将缩放后的融合特征FF'与相同尺度的编码器特征分别进行一次1×1卷积操作,以获得相同的特征通道数。随后,将得到的两个维度相同的特征相加求和,再分别经过mish激活函数、1×1卷积操作和sigmoid激活函数。此时,经过上述操作后得到的特征图表示缩放后的融合特征FF'关于当前尺度编码器特征的加性注意力权重。然后,将该加性注意力权重作用于编码器特征上。最后,将经过加性注意力加权的编码器特征和输入编码器特征相加后输入到下一个模块中。

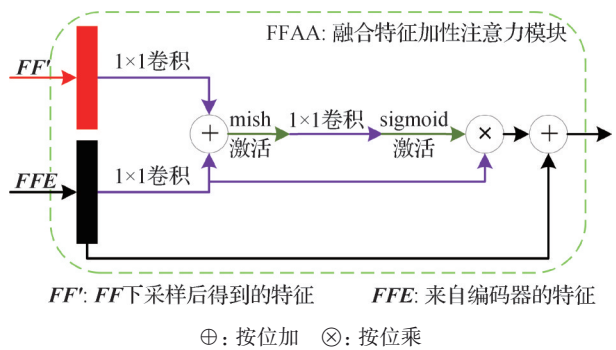


图2 融合特征加性注意力模块框图
Fig. 2 Flow chart of the fusion feature additive attention module

2.4 编码器—解码器加性注意力拼接模块

经过融合特征加性注意力模块指导后的编码器特征更具判别性。但是,由于指导后的编码器特征和解码器特征在整个网络结构中所处的深度不同,因此跳跃连接时两者之间仍存在语义特征差距。为了降低由于网络深度不同所产生的语义特征差距,本文在经过指导的编码器特征和解码器特征间采用加性注意力机制。通过加性注意力机制来自适应地学习跳跃连接特征中的重要特征信息。

如图3所示,E-DAAC模块的输入分别是经过融合特征指导的编码器特征FFM,以及来自解码器模块中下一尺度的特征FFD。E-DAAC模块中加性注意力的操作与FFAA模块类似。首先,将输入的解码器特征FFD进行上采样,再与指导编码器特征FFM分别经过1×1卷积操作后进行相加。接着,将

得到的求和特征再分别经过 mish 激活函数、 1×1 卷积和 sigmoid 激活函数。此时, 经过上述操作后得到的特征图表示解码器特征关于当前跳跃连接的指导编码器特征的加性注意力权重。然后, 将该加性注意力权重作用于指导编码器特征上。最后, 将经过注意力加权的指导编码器特征与上采样后的解码器特征进行拼接作为下一个操作的输入。与 FFAA 模块中的加性注意力机制建模融合特征对于单尺度特征进行指导不同, E-DAAC 模块中的加性注意力机制用于建模解码器特征和指导编码器特征间的相关性程度。通过该加性注意力机制的指导, 用于降低经过指导的编码器特征和解码器特征在跳跃连接时由于所处网络深度不同而产生的语义特征差距。

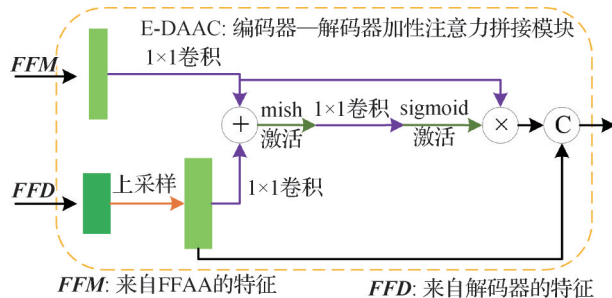


图3 编码器-解码器加性注意力拼接模块框图

Fig. 3 Flow chart of the encoder-decoder additive attention concatenate module

值得注意的是, E-DAAC 模块中所采用的操作和 Attention U-Net (Oktay 等, 2018) 中的门控注意力有所不同。如图 4 所示, 在门控注意力中, 首先, 将输入编码器特征进行下采样后再与解码器特征进行加性注意力计算。然后, 将得到的注意力权重上采样至输入编码器特征的分辨率后再作用于编码器特征。最后, 将经过注意力加权的编码器特征与经过上采样的解码器特征进行拼接作为下一个操作的输入。而在 E-DAAC 中, 直接将解码器特征上采样后与编码器特征进行加性注意力操作。简言之, E-DAAC 简化了门控注意力中复杂的操作。

2.5 解码器模块

在解码器模块中, 逐步采用卷积并插值上采样操作, 以恢复特征图大小至网络输入图像的原始分辨率大小, 并在解码器结尾进行最终脑肿瘤 MR 图像的逐像素多标签分类。与编码器模块类似, 解码器模块中具有 4 个上采样层, 每个上采样层将特征

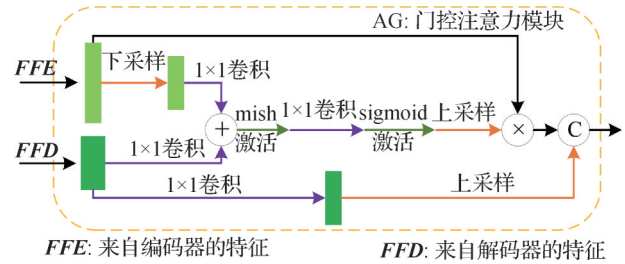


图4 Attention U-Net中门控注意力模块框图

Fig. 4 Flow chart of attention gate module in Attention U-Net

图的大小加倍而通道数减少为输入的一半。值得注意的是, 与编码器模块中采用的操作略有不同。解码器模块中关于各个尺度特征的操作仅使用了 1 次批归一化、mish 激活函数和卷积操作。解码器中每个层级的输入来自于编码器-解码器加性注意力拼接模块的输出特征。

在所提出网络的桥接模块中, 本文采用了 3 次叠加的批归一化、mish 激活函数和卷积操作。在本文网络结构的最后, 使用 1×1 卷积对解码器模块的最后一层输出特征进行卷积和 sigmoid 激活, 以达到将解码器模块的输出特征映射到期望分割出的类别数。通过 1×1 卷积生成按像素分类的分割结果, 分别表示 3 种脑肿瘤和背景的像素分类类别, 该类别数即为网络模型最后输出的通道数。另外, 在编码器模块中每个尺度的输出之后还使用了 dropout (Hinton 等, 2012), 用于防止训练网络模型时出现过拟合, 从而提高泛化能力。所提网络模型的输入是一个 4 通道的 2D 便携式网络图形 (portable network graphics, PNG) 图像, 其中每个通道代表每个样本相同轴向位置的不同模态 MR 图像切片。

3 实验及结果分析

3.1 实验数据集

实验数据采用多模态脑肿瘤分割数据集 BraTS2020 (Menze 等, 2015; Bakas 等, 2017, 2018)。该数据集是通过不同的临床协议, 从 19 个机构的 MR 扫描仪中获取的。所有的图像数据都经过了相同的解剖模板配准、头骨剥离和插值等预处理操作。经过预处理后, 所有图像的分辨率均为 1 mm^3 , 尺寸均为 $240 \times 240 \times 155$ 。

BraTS2020 训练集中每个样本包括 FLAIR、T1、T1ce 和 T2 等 4 个模态的 3 维 MR 图像, 以及一个分

割掩码图像(ground truth, GT)。其中,分割掩码图像是由1~4个专业标记人员按照相同的标注协议手动分割得到的,并通过了具有丰富经验的神经放射科医生的认证。标注示例如图5所示,包括GD增强肿瘤(GD-enhancing tumor, ET)、肿瘤周围水肿(peritumoral edema, ED)、坏死(necrotic, NCR)和非增强肿瘤核心(non enhancing tumor core, NET)等

4个区域。为了方便评估,对标注结果进行合并可得到3个脑肿瘤子区域,分别是整个肿瘤(WT),肿瘤核心(TC)和增强肿瘤(ET)。在GT分割掩码图像中,NCR和NET(绿色和红色)用标签1标记,ED(黄色)用标签2标记,ET(蓝色)用标签4标记,其他区域用标签0标记。与训练集不同,验证集和测试集中的每个样本不包括分割掩码图像。

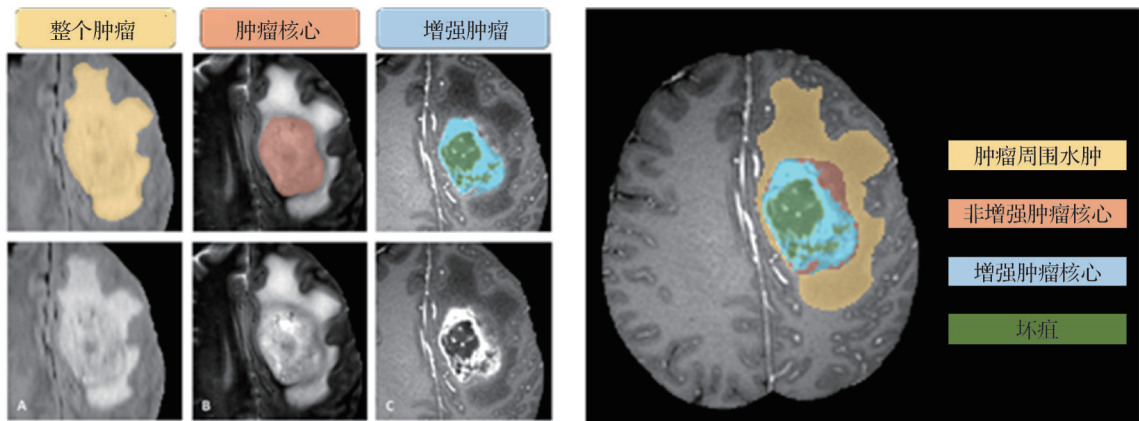


图5 标注示例
Fig. 5 Annotation example

由于测试集仅对参赛人员开放,因此本文使用训练集和验证集进行网络模型的训练、验证和测试。实验采用5折交叉验证避免偏差。数据具体划分如下,将BraTS2020训练集中234个高级别胶质瘤(high-grade glioma, HGG)样本和59个低级别胶质瘤(low-grade glioma, LGG)样本作为训练数据。将BraTS2020训练集中59个HGG样本和15个LGG样本作为验证数据。将BraTS2020的验证集作为最终的测试数据。

由于数据集中4种模态的MR图像具有不同的对比度,因此需要对各模态图像分别进行Z-Score标准化处理,将其标准化为零均值和单位标准差。由于GT是多标签二值掩码图像,因此不需要进行标准化处理。具体操作如下,首先计算每个样本的各个模态MR图像中体素值大于零区域的均值和标准差;然后再对每个模态图像进行标准化计算。经过Z-Score标准化处理后的MR图像数据符合标准正态分布。其计算为

$$v' = \begin{cases} 254 & v > \mu + 3 \times \sigma \\ \frac{v - \mu}{3 \times \frac{\sigma}{128}} + 127 & \text{其他} \\ 0 & v < \mu - 3 \times \sigma \end{cases} \quad (1)$$

式中, v 和 v' 分别表示初始模态以及标准化后的3D MR图像体素值, μ 和 σ 分别表示初始模态的3D MR图像中体素值大于零区域的均值和标准差。

3.2 评价指标

为了评估所提出的网络模型,本文对WT、TC和ET的分割区域使用Dice系数和95%的Hausdorff距离作为评价指标。将网络产生的分割结果上传到BraTS2020在线评估平台,以获得关于这些评价指标的定量评估结果。

1) Dice系数(Dice coefficient)用于计算两个样本集之间的重叠程度,是一种几何相似度度量的指标。其范围在0~1之间。Dice系数越高,表示分割的精度越好。其计算为

$$Dice(S, L) = \frac{2|S \cap L|}{|S| + |L|} \times 100\% \quad (2)$$

式中, S 和 L 分别表示实际分割预测结果和真实分割标签的集合。

2) 豪斯多夫距离(Hausdorff distance, HD)用来表示两个样本集之间的最大不匹配程度。HD值越小,表示分割的效果越好。Hausdorff 95(HD95)表示将计算得到的HD值乘以95%,目的是消除离群值对计算结果的影响。其计算为

$$HD(S, L) = \max \left(\max_{s \in S} \min_{l \in L} \|s - l\|, \max_{l \in L} \min_{s \in S} \|l - s\| \right) \quad (3)$$

式中, S 和 L 分别表示实际分割预测结果和真实分割标签的集合。

3.3 实现细节

本文网络采用多模态 MR 图像转化的 2D PNG 图像作为输入, 并输出像素级分类的语义分割结果。具体操作是将 MR 图像进行 Z-Score 规范化和转换, 以得到 2D 的 PNG 图像切片。另外, 网络模型训练过程中采用的图像增强方式是在左右和上下方向随机翻转图像, 概率为 50%。在测试时, 对测试数据进行了与训练数据相同的预处理, 但是没有进行数据增强。在定量评估时, 将网络模型输出的分割结果转换为 3D 文件后上传至在线评估平台获取定量评估结果。

本文网络模型是在 Ubuntu18.04 操作系统上使用 Pycharm 基于 Keras 实现的, 并且在带有 16 GB 显存 NVIDIA Quadro P5000 GPU 的工作站上训练、验证和测试。本文使用学习率为 0.000 1 的 ADAM 优化器, 并使用 he_normal(He 等, 2015) 参数初始化方法对网络模型进行初始化。由于受到显存大小的限制, 训练时所采用的批处理大小设置为 12, 并且经过 150 次迭代训练, 最终保存的结果是在验证数据上表现最好的网络模型参数。此外, 在编码器模块中使用了 dropout 为 0.2 的正则化, 可以避免训练过程中出现过拟合, 并提高网络模型的泛化能力。

3.4 损失函数

本文使用的损失函数是分类交叉熵损失函数 (categorical cross entropy, CE), 其计算为

$$CE = -\frac{1}{N} \sum_{c=1}^C \sum_{n=1}^N Y_{true_n}^c \times \log(Y_{pred_n}^c) \quad (4)$$

式中, N 表示的是输入图像中像素的总个数, C 表示的是分割类别数, $Y_{true_n}^c$ 和 $Y_{pred_n}^c$ 分别表示第 n 个像素关于第 c 个类别的真实标签值和预测结果值。

3.5 结果和讨论

本节基于 BraTS2020 验证数据集, 首先展示了所提网络模型的分割效果; 然后, 进行了消融实验, 以验证所提出模块的有效性; 最后, 与其他分割模型进行了对比实验, 以验证所提出网络的优越性。

3.5.1 实验结果

本文在 BraTS2020 验证数据集上验证了所提网络的分割效果。以下定量评价指标是将模型的分割

结果提交至官方网站后得到的。该网站计算并评估提交的分割结果, 并输出 Dice 和 HD95 等分割评价指标。

所提网络在 BraTS2020 验证集上的分割结果如表 1 所示。其中关于 ET、WT 和 TC 的平均 Dice 分别为 0.706 4、0.887 5 和 0.719 4。可以看出, 提出的网络模型对脑肿瘤的 WT 区域有更好的分割效果, 且明显优于 ET 和 TC 子区域的平均分割结果。出现这种情况的原因是 WT 的区域范围最大, 更容易分割。此外, 神经外科医生、神经放射科医生和放射科医生报告称, ET 和 TC 区域标签之间的边界具有高度的不确定性(Bakas 等, 2017)。这也是 ET 和 TC 的分割精度低于 WT 的原因。

表 1 所提网络的分割结果

Table 1 Segmentation results of the proposed network

统计量	Dice			HD95		
	ET	WT	TC	ET	WT	TC
Mean	0.706 4	0.887 5	0.719 4	41.9	14.13	17.42
StdDev	0.304 8	0.082 2	0.308 7	109.14	22.42	49.32
Median	0.832 7	0.913 4	0.878	2.45	4.24	4.24
25quantile	0.711 3	0.867 3	0.594 5	1.41	2.45	2
75quantile	0.890 4	0.936 1	0.928	8.93	11.75	14.04

注: Mean, StdDev, Median, 25quantile, 75quantile 分别表示验证集样本分割结果的平均值、标准差、中位数、第 25 个百分点值以及第 75 个百分点值。

3.5.2 消融实验

消融实验在 BraTS2020 验证数据集上进行, 主要包括 6 组实验。a 组实验用于验证基础网络关于 ET、WT 和 TC 的分割效果。b 组实验用于验证向基础网络中添加多尺度特征融合模块 MSFF 后的分割效果。值得注意的是, 这里本文先将融合特征下采样, 再经过 1×1 卷积后直接加到跳跃连接的编码器特征上, 即此时没有加性注意力的相关操作。c 组实验用于验证同时向基础网络中加入 MSFF 和 FFAA 模块后的分割效果。d 组实验用于验证编码器—解码器特征加性注意力拼接模块 E-DAAC 的作用。e 组实验用于验证向基础网络中添加 Attention U-Net 中的门控注意力模块后的分割效果。f 组实验用于验证同时增加本文提出的 3 个模块后网络所取得的分割效果。

消融实验结果如表 2 所示,通过逐渐向基础 U-Net 中添加模块至提出的网络模型。在表 2 中, a 组表示基础网络, b 组表示向基础网络中添加提出的多尺度特征融合模块。对比 a 组和 b 组的分割结果发现, MSFF 模块的引入提升了网络的分割精度。其中 ET、WT 和 TC 的平均 Dice 分别提高了 2. 23%、2. 13% 和 0. 97%。 c 组表示在 b 组实验基础上, 添加融合特征加性注意力模块。对比 b 组和 c 组的分割结果发现, 添加加性注意力操作后, 模型关于 3 个区域分割结果的平均 Dice 进一步提高。这表明编码器各尺度特征经过全局融合特征加性注意力的指导后更具判别性。对比 c 组和 a 组, 观察到关于 ET、WT 和 TC 的平均 Dice 值分别提高了 3. 77%、2. 71% 和 2. 42%。这得益于多尺度特征融合和加性注意力共同作用的结果。 d 组表示向基础网络中添加编码器—解码器加性注意力拼接模块。对比 a 组和 d 组分割结果发现, 向基础网络中添加 E-DAAC 模块后, 关于 ET、WT 和 TC 的平均 Dice 分别提高了 2. 94%、2. 21% 和 2. 25%。这说明通过在编码器和解码器特

征间引入加性注意力机制可以降低二者由于所处网络深度不同而导致的语义特征差距。并且, 对比 c 组和 d 组实验结果发现, MSFF 和 FFAA 模块的引入对于网络分割性能的提升优于 E-DAAC 模块。 e 组表示在 c 组实验的基础上, 向网络中添加 Attention U-Net 中的门控注意力模块。对比 e 组和 c 组发现, 增加门控注意力模块后, 网络的参数增加了约 2 M, 但是模型的分割结果仅有略微提升。 f 组表示同时向基础网络中添加本文提出的 3 个模块, 即提出的网络模型结构。对比 a、b、c、d 和 f 五组实验结果发现, 所提网络在 3 个肿瘤区域上取得了最好的分割效果。相较于基础网络, 本文网络模型关于 ET、WT 和 TC 的平均 Dice 分别提高了 6. 23%、3. 53% 和 5. 93%。此外, 对比 e 组和 f 组发现, f 组模型的参数比 e 组少约 1. 6 M, 但是分割结果远高于后者。这表明提出的 E-DAAC 模块与门控注意力模块相比, 具有更少的参数且对于分割结果的提升具有更好的效果。这些结果表明了提出的 MSFF、FFAA 和 E-DAAC 模块对于脑肿瘤 MR 图像分割问题的有效性。

表 2 消融实验结果
Table 2 Results of ablation experiments

实验组	方法					Dice			HD95			Params/ M	FLOPs/ M
	U-Net	MSFF	FFAA	E-DAAC	AG	ET	WT	TC	ET	WT	TC		
a	√					0.644 1	0.852 2	0.660 1	58.21	31.54	21.95	23.96	47.88
b	√	√				0.666 4	0.873 5	0.669 8	55.67	23.72	25.44	26.55	53.05
c	√	√	√			0.681 8	0.879 3	0.684 3	47.61	16.79	26.81	26.64	53.22
d	√			√		0.673 5	0.874 3	0.682 6	50.42	20.10	27.82	24.23	48.40
e	√	√	√		√	0.685 5	0.880 8	0.688 6	46.35	22.51	24.99	28.56	61.71
f(本文)	√	√	√	√		0.706 4	0.887 5	0.719 4	41.90	14.13	17.42	26.91	54.20

注:加粗字体表示各列最优结果。MSFF表示多尺度特征融合模块,FFAA表示融合特征加性注意力模块,E-DAAC表示编码器—解码器特征加性注意力拼接模块,AG表示门控注意力模块。

此外,观察表 2 最后一列发现,提出的网络模型在增加少量计算量的前提下,可以显著提升分割效果。对比 a、b 和 c 三组实验以及 d 和 a 两组实验,发现 c 组实验相对于 b 组实验计算量增加了 0. 17 M, d 组实验相对于 a 组实验计算量增加了 0. 52 M,但是关于 ET、WT 和 TC 的分割结果均有不同程度的提升。另外,对比 f、e 和 c 三组实验发现,相较于门控注意力模块,提出的编码器—解码器加性注意力拼接模块降低了约 7. 51 M 的计算量,却得到了更好的分割效果,这体现了提出模块的优越性。

3. 5. 3 对比实验

为了证明所提出网络模型的优越性,本文与 U-Net 及其经典的、最新提出的改进模型,如 Res-UNet (Xiao 等, 2018)、Attention U-Net (Oktay 等, 2018)、UNet++ (Zhou 等, 2018)、MultiResUNet (Iltis-haz 和 Rahman, 2020)、DR-Unet104 (Colman 等, 2020)、MTAU (Awasthi 等, 2021)、U-Attention Net (Xu 等, 2021)、MVP U-Net (multi-view pointwise U-Net) (Zhao 等, 2021) 等,以及非基于 U-Net 的优秀分割模型 HRNet (high-resolution network) (Wang 等, 2021) 和 DeepLab V3+ (Chen 等, 2018) 进行了比较。

对比实验结果如表 3 所示。表 3 中各模型的分割指标结果,除了 DR-Unet104、DeepLab V3+、MTAU、U-Attention Net 和 MVP U-Net 模型是来自算法原论文中的实验数据外,其他各网络关于整个肿瘤、肿瘤核心和增强肿瘤的分割指标结果均是从原论文指定网站下载模型源码后,重新训练所得。之所以需要重新训练是因为暂时没有这些网络在脑肿瘤数据集 BraTS2020 上的验证结果。所有模型的训练均采用 5 折交叉验证避免偏差。在训练过程中所采用的软、硬件配置和实验细节中所描述的完全相同。最终保存的结果是在验证数据上表现最好的模型参数。最后将测试数据输入到训练好的模型中,得到输出结果后,上传到官方评价网站以获取分割结果的各项定量评价指标,最终得到如表 3 中所展示的分割结果。观察表 3 发现,本文模型关于 ET、WT 和 TC 取得了显著的分割效果。在 Dice 指标上,所提网络关于 ET、WT 和 TC 的分割结果均优于其他网络。在 HD95 指标上,所提网络关于 ET、WT 和 TC 的分割结果在所有网络中取得了第 2、第 4 和第 3,分别比第 1 低了 1.29、6.19 和 1.68。这证明了所提出网络的优越性。值得注意的是,所提出模型关于 3 个分割区域的平均 Dice 值均高于 DR-Unet104 的分割结果。其中关于 TC、ET 和 WT 分割结果的平均 Dice 值分别

超出 DR-Unet104 有 4.73%、3.08% 和 0.13%。这些实验结果表明所提出的网络具有优越性。

为了进一步证明所提出网络整体分割效果的优越性。在图 6 中展示了各种模型在 BraTS2020 验证数据集上关于 3 个区域分割结果的 Dice 值箱形图统

表 3 对比实验结果
Table 3 Comparative experimental results

模型	Dice			HD95		
	ET	WT	TC	ET	WT	TC
U-Net	0.644 1	0.852 2	0.660 1	58.21	31.54	21.95
Res-U-Net	0.645 4	0.848 1	0.691 7	60.58	33.64	23.87
Attention U-Net	0.670 5	0.857 9	0.635 5	45.90	27.12	36.72
UNet++	0.646 6	0.863 7	0.673 6	56.09	21.90	20.08
MultiResU-Net	0.650 9	0.841 3	0.699 3	58.83	37.11	27.37
MTAU	0.570 0	0.730 0	0.610 0	47.22	24.03	31.53
U-Attention Net	0.673 6	0.860 8	0.704 2	40.61	7.94	15.75
MVP U-Net	0.671 0	0.862 0	0.623 0	47.33	12.58	50.15
HRNet	0.654 1	0.871 9	0.684 9	53.09	22.88	27.11
DeepLab V3+	0.652 4	0.877 1	0.681 3	49.10	14.87	17.96
DRU-net104	0.675 6	0.886 2	0.672 1	47.62	12.11	15.74
本文	0.706 4	0.887 5	0.719 4	41.90	14.13	17.42

注:加粗字体表示各列最优结果。

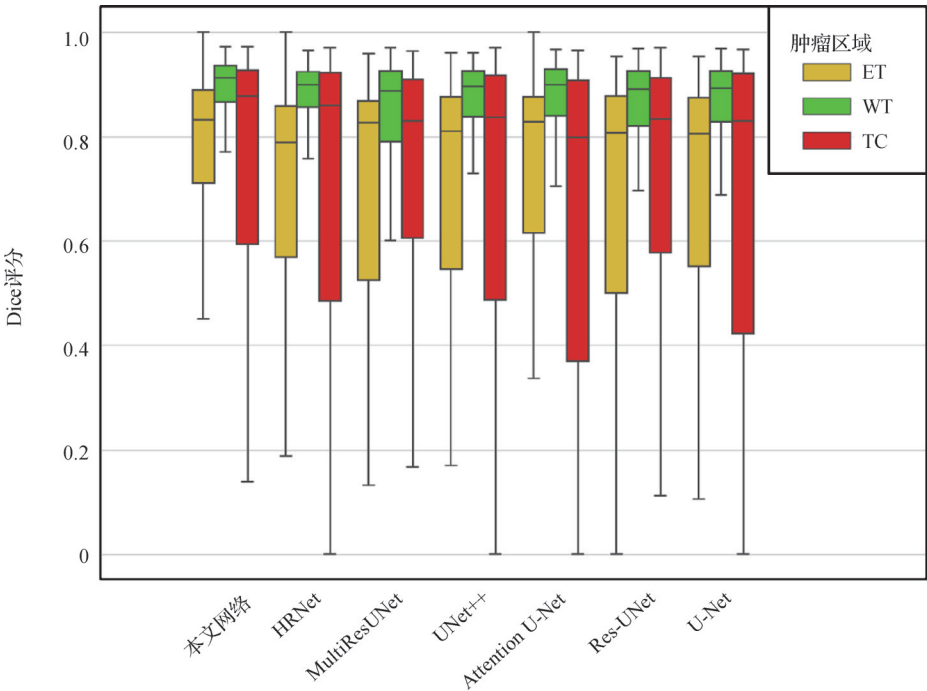


图 6 各种模型在 BraTS2020 数据集上关于 ET、WT 和 TC 的 Dice 值箱形统计图

Fig. 6 Box statistics of Dice values of various models for ET, WT and TC on BraTS2020 validation dataset

计,其中箱体中间的水平线代表一组数据的中位数,箱体上下两端分别表示这组数据的上四分位数和下四分位数。最上面和最下面的两个短横线分别代表上界和下界。

从图6可以看出,本文网络关于ET、WT和TC分割结果的Dice值波动非常小。在所有网络的分割

结果中,本文模型的Dice值集中在较高值处。这表明所提网络的整体分割效果优于其他分割模型。

为了更加直观地评估本文模型分割效果的优越性,本文对各网络的分割结果进行可视化,结果如图7所示。图7上方的 BraTS20_Valid_P_Q 表示验证数据中样本ID为P的第Q幅切片。第1行表示网

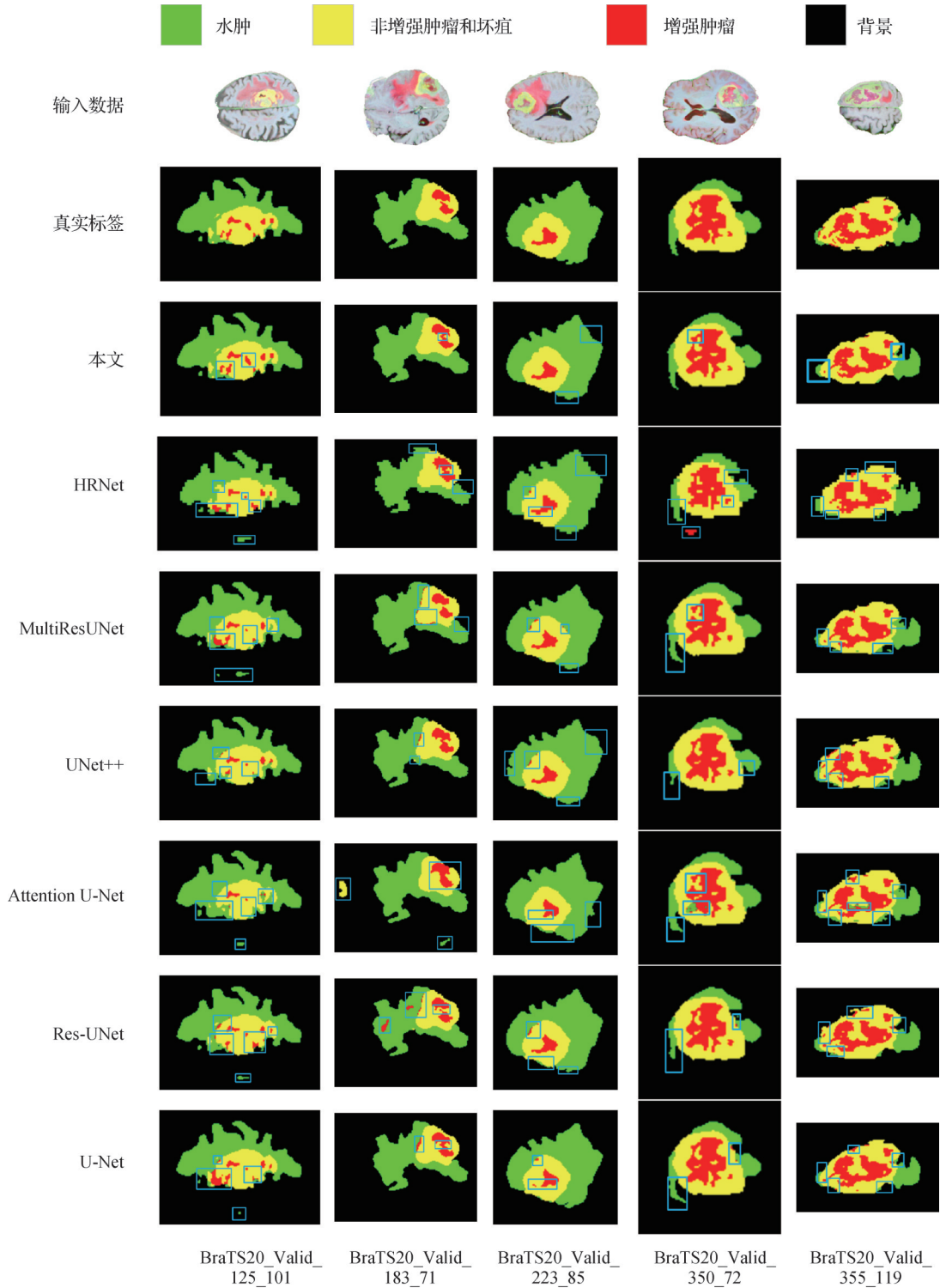


图7 各网络分割结果可视化

Fig. 7 Visualization of the segmentation results of each network

络的输入,它是经过Z-Score预处理并且合并后的PNG图像,是一个4通道的RGBA图像。第2行表示对应输入的真实分割标签,红色区域表示增强肿瘤区域,黄色和红色区域共同表示肿瘤核心区域,绿色、黄色和红色区域共同表示整个肿瘤区域。其他图像表示各网络的实际分割预测结果。

图7中的蓝色矩形框表示各模型分割结果和真实标签存在显著差异的区域。观察发现,与基于U-Net的分割网络进行比较,非基于U-Net的分割网络HRNet关于ET、WT和TC的分割结果具有明显的锯齿状边缘,这表明在医学图像分割领域中,基于U-Net的分割网络更具优势。对比图中基于U-Net的分割示例发现,U-Net、Res-UNet和Attention U-Net等网络在分割不同区域时易出现混淆,与真实标签相比存在显著差异。UNet++和MultiResUNet等网络的分割结果非常接近真实标签,但不同肿瘤区域边界容易出现错误分割。在此基础上,本文网络可以较为准确地分割不同肿瘤区域的边界,达到更好的整体分割效果。这表明本文网络可以准确地分割病变子区域,证明了本模型在脑肿瘤MR图像分割方面的潜力。

4 结 论

本文提出了用于多模态脑肿瘤MR图像自动分割的多尺度特征融合加性注意力网络。首先,将U-Net编码器中具有互补信息的多尺度特征进行融合,然后对各尺度特征进行加性注意力指导,以增强编码器特征的判别性。此外,本文提出在编码器特征和解码器特征间采用加性注意力机制,以降低跳跃连接时编码器特征和解码器特征由于所处网络深度不同而产生的语义特征差距。最后,在BraTS2020数据集上进行了训练、验证和测试。实验结果表明,所提网络模块可以显著提升分割精度,所提网络模型优于经典的、最新提出的U-Net增强网络,以及非基于U-Net的优秀分割网络模型,能更精准地分割MR图像中脑肿瘤子区域。

虽然本文网络模型取得了显著的分割效果,但是仍存在问题。与整个肿瘤相比,肿瘤核心和增强肿瘤的分割精度相对较低。这是由于整个肿瘤区域范围最大、更容易分割,且肿瘤核心和增强肿瘤区域之间的边界具有高度不确定性。因此,后续将对损

失函数进行改进,引导网络模型更加关注于对肿瘤核心和增强肿瘤特征的学习,从而达到与整个肿瘤相近的分割效果。

参考文献(References)

- Awasthi N, Pardasani R and Gupta S. 2021. Multi-threshold attention U-Net (MTAU) based model for multimodal brain tumor segmentation in MRI scans//Proceedings of the 6th International MICCAI Brainlesion Workshop. Cham, Switzerland: Springer: 168-178 [DOI: 10.1007/978-3-030-72087-2_15]
- Bahdanau D, Cho K and Bengio Y. 2014. Neural machine translation by jointly learning to align and translate [EB/OL]. [2022-04-23]. <https://arxiv.org/pdf/1409.0473.pdf>
- Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J S, Freymann J B, Farahani K and Davatzikos C. 2017. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4 (1): #170117 [DOI: 10.1038/sdata.2017.117]
- Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara R T, Berger C, Ha S M, Rozycki M, Prastawa M, Alberts E, Lipkova J, Freymann J, Kirby J, Bilello M, Fathallah-Shaykh H, Wiest R, Kirschke J, Wiestler B, Colen R, Kotrotsou A, Lamon-tagne P, Marcus D, Milchenko M, Nazeri A, Weber M A, Mahajan A, Baid U, Gerstner E, Kwon D, Acharya G, Agarwal M, Alam M, Albiol A, Albiol A, Albiol F J, Alex V, Allinson N, Amorim P H A, Amrutkar A, Anand G, Andermatt S, Arbel T, Arbelaez P, Avery A, Azmat M, Pranjal B, BAI W J, Banerjee S, Barth B, Batchelder T, Batmanghelich K, Battistella E, Beers A, Belyaev M, Bendszus M, Benson E, Bernal J, Bharath H N, Biros G, Bisdas S, Brown J, Cabezas M, Gao S L, Cardoso J M, Carver E N, Casamitjana A, Castillo L S, Catà M, Cattin P, Cerigues A, Chagas V S, Chandra S, Chang Y J, Chang S Y, Chang K, Chazalon J, Chen S C, Chen W, Chen J W, Chen Z L, Cheng K, Choudhury A R, Chylla R, Clérigues A, Coleman S, Rodriguez Colmeiro R G, Combalia M, Costa A, Cui X M, Dai Z Z, Dai L T, Daza L A, Deutsch E, Ding C X, Dong C, Dong S D, Dudzik W, Eaton-Rosen Z, Egan G, Escudero G, Estienne T, Everson R, Fabrizio J, Fan Y, Fang L W, Feng X, Ferrante E, Fidon L, Fischer M, French A P, Fridman N, Fu H, Fuentes D, Gao Y Z, Gates E, Gering D, Gholami A, Gierke W, Glocker B, Gong M M, González-Villá S, Groses T, Guan Y F, Guo S, Gupta S, Han W S, Han I S, Harmuth K, He H G, Hernández-Sabaté A, Herrmann E, Himthani N, Hsu W, Hsu C, Hu X J, Hu X B, Hu Y T, Hu Y F, Hua R, Huang T Y, Huang W L, Van Huffel S, Huo Q, Vivek H V, Iftikharuddin K M, Isensee F, Islam M, Jackson A S, Jambawalikar S R, Jesson A, Jian W J, Jin P, Jose V J M, Jungo A, Kainz B, Kamnitsas K, Kao P Y, Karnawat A, Kell-

- ermeier T, Kermi A, Keutzer K, Khadir M T, Khened M, Kickingereder P, Kim G, King N, Knapp H, Knecht U, Kohli L, Kong D R, Kong X M, Koppers S, Kori A, Krishnamurthi G, Krivov E, Kumar P, Kushibar K, Lachinov D, Lambrou T, Lee J, Lee C, Lee Y, Lee M, Lefkovits S, Lefkovits L, Levitt J, Li T F, Li H W, Li W Q, Li H Y, Li X C, Li Y X, Li H, Li Z Y, Li X Y, Li Z J, Li X G, Li W Q, Lin Z S, Lin F M, Lio P, Liu C, Liu B Q, Liu X, Liu M Y, Liu J, Liu L Y, Llado X, Lopez M M, Lorenzo P R, Lu Z T, Luo L, Luo Z G, Ma J, Ma K, Mackie T, Madabushi A, Mahmoudi I, Maier-Hein K H, Maji P, Mammen C P, Mang A, Manjunath B S, Marcinkiewicz M, McDonagh S, McKenna S, McKinley R, Mehl M, Mehta S, Mehta R, Meier R, Meinel C, Merhof D, Meyer C, Miller R, Mitra S, Moiyadi A, Molina-Garcia D, Monteiro M A B, Mrukwa G, Myronenko A, Nalepa J, Ngo T, Nie D, Ning H, Niu C, Nuechterlein N K, Oermann E, Oliveira A, Oliveira D D C, Oliver A, Osman A F I, Ou Y N, Ourselin S, Paragios N, Park M S, Paschke B, Pauloski J G, Pawar K, Pawlowski N, Pei L M, Peng S T, Pereira S M, Perez-Beteta J, Perez-Garcia V M, Pezold S, Pham B, Phophalia A, Piella G, Pillai G N, Piraud M, Pisov M, Popli A, Pound M P, Pourreza R, Prasanna P, Prkowska V, Pridmore T P, Puch S, Puybareau É, Qian B Y, Qiao X, Rajchl M, Rane S, Rebsamen M, Ren H L, Ren X H, Revanuru K, Rezaei M, Rippel O, Rivera L C, Robert C, Rosen B, Rueckert D, Safwan M, Salem M, Salvi J, Sanchez I, Sánchez I, Santos H M, Sartor E, Schellingerhout D, Scheufele K, Scott M R, Scussell A A, Sedlar S, Serrano-Rubio J P, Shah N J, Shah N, Shaikh M, Shankar B U, Shboul Z, Shen H P, Shen D G, Shen L L, Shen H C, Shenoy V, Shi F, Shin H E, Shu H, Sima D, Sinclair M, Smedby O, Snyder J M, Soltaninejad M, Song G D, Soni M, Stawiaski J, Subramanian S, Sun L, Sun R, Sun J W, Sun K, Sun Y, Sun G X, Sun S, Suter Y R, Szilagyi L, Talbar S, Tao D C, Tao D C, Teng Z Z, Thakur S, Thakur M H, Tharakan S, Tiwari P, Tochon G, Tran T, Tsai Y M, Tseng K L, Tuan T A, Turlapov V, Tustison N, Vakalopoulou M, Valverde S, Vanguri R, Vasiliev E, Ventura J, Vera L, Vercauteren T, Verrastro C A, Vidyaratne L, Vilaplana V, Vivekanandan A, Wang G T, Wang Q, Wang C J, Wang W, Wang D, Wang R X, Wang Y Y, Wang C L, Wang G T, Wen N, Xin Wen X, Weninger L, Wick W, Wu S C, Wu Q, Wu Y H, Xia Y, Xu Y W, Xu X W, Xu P Y, Yang T L, Yang X P, Yang H Y, Yang J L, Yang H J, Yang G, Yao H D, Ye X J, Yin C C, Young-Moxon B, Yu J H, Yue X Y, Zhang S T, Zhang A, Zhang K, Zhang X J, Zhang L C, Zhang X Y, Zhang Y Z, Zhang L, Zhang J G, Zhang X, Zhang T H, Zhao S C, Zhao Y, Zhao X M, Zhao L, Zheng Y F, Zhong L M, Zhou C H, Zhou X B, Zhou F, Zhu H T, Zhu J, Ying Z G, Zong W W, Kalpathy-Cramer J, Farahani K, Davatzikos C, Van Leemput K and Menze B. 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge [EB/OL]. [2021-11-09]. <https://arxiv.org/pdf/1811.02629.pdf>
- Bauer S, Wiest R, Nolte L P and Reyes M. 2013. A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine and Biology*, 58(13): R97-R129 [DOI: 10.1088/0031-9155/58/13/R97]
- Çiçek Ö, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O. 2016. 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, Greece: Springer: 424-432 [DOI: 10.1007/978-3-319-46723-8_49]
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]
- Colman J, Zhang L, Duan W T and Ye X J. 2020. DR-Unet104 for multimodal MRI brain tumor segmentation//Proceedings of the 6th International MICCAI Brainlesion Workshop. Lima, Peru: Springer: 410-419 [DOI: 10.1007/978-3-030-72087-2_36]
- He K H and Xiao Z Y. 2021. LRUNet: a lightweight rapid semantic segmentation network for brain tumors. *Journal of Image and Graphics*, 26(9): 2233-2242 (何康辉, 肖志勇. 2021. LRUNet: 轻量级脑肿瘤快速语义分割网络. *中国图象图形学报*, 26(9): 2233-2242) [DOI: 10.11834/jig.200436]
- He K M, Zhang X Y, Ren S Q and Sun J. 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1026-1034 [DOI: 10.1109/ICCV.2015.123]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R. 2012. Improving neural networks by preventing co-adaptation of feature detectors [EB/OL]. [2021-11-09]. <https://arxiv.org/pdf/1207.0580.pdf>
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Ibtehaz N and Rahman M S. 2020. MultiResUNet: rethinking the U-Net

- architecture for multimodal biomedical image segmentation. *Neural Networks*, 121: 74-87 [DOI: 10.1016/j.neu-net.2019.08.025]
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift//*Proceedings of the 32nd International Conference on International Conference on Machine Learning*. Lille, France: JMLR.org: 448-456
- Jetley S, Lord N A, Lee N and Torr P H S. 2018. Learn to pay attention [EB/OL]. [2021-11-09]. <https://arxiv.org/pdf/1804.02391.pdf>
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin transformer: hierarchical vision transformer using shifted windows//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Menze B H, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber M A, Arbel T, Avants B B, Ayache N, Buendia P, Collins D L, Cordier N, Corso J J, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst C R, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X T, Hamamci A, Iftekharuddin K M, Jena R, John N M, Konukoglu E, Lashkari D, Mariz J A, Meier R, Pereira S, Precup D, Price S J, Raviv T R, Reza S M S, Ryan M, Sarikaya D, Schwartz L, Shin H C, Shotton J, Silva C A, Sousa N, Subbanna N K, Szekely G, Taylor T J, Thomas O M, Tustison N J, Unal G, Vasseur F, Wintermark M, Ye D H, Zhao L, Zhao B S, Zikic D, Prastawa M, Reyes M and Leemput K V. 2015. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10): 1993-2024 [DOI: 10.1109/TMI.2014.2377694]
- Misra D. 2019. Mish: a self regularized non-monotonic neural activation function [EB/OL]. [2021-11-09]. <https://arxiv.org/pdf/1908.08681v1.pdf>
- Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N Y, Kainz B, Glocker B and Rueckert D. 2018. Attention U-Net: learning where to look for the pancreas [EB/OL]. [2021-11-09]. <https://arxiv.org/pdf/1804.03999.pdf>
- Ostrom Q T, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, Wolinsky Y, Kruchko C and Barnholtz-Sloan J S. 2015. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. *Neurooncology*, 17(S4): iv1-iv62 [DOI: 10.1093/neuonc/nov189]
- Ren L, Li Q, Guan X and Ma J. 2018. Three-dimensional segmentation of brain tumors in magnetic resonance imaging based on improved continuous max-flow. *Laser and Optoelectronics Progress*, 55(11): #111011 (任璐, 李镔, 关欣, 马杰. 2018. 改进的连续型最大流算法脑肿瘤磁核共振成像三维分割. *激光与光电子学进展*, 55(11): #111011 [DOI: 10.3788/LOP55.111011])
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Shen T, Jiang J, Zhou T Y, Pan S R, Long G D and Zhang C Q. 2018. DiSAN: directional self-attention network for RNN/CNN-free language understanding//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence and the 13th Innovative Applications of Artificial Intelligence Conference and the 18th AAAI Symposium on Educational Advances in Artificial Intelligence*. Louisiana, USA: AAAI Press: #668
- Shi J L, Zhang R, Guo L J, Gao L L, Li Y Q, Ma H F and Wang J H. 2020. Dual dense context-aware network for hippocampal segmentation. *Biomedical Signal Processing and Control*, 61: #102038 [DOI: 10.1016/j.bspc.2020.102038]
- Sinha A and Dolz J. 2021. Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 25(1): 121-130 [DOI: 10.1109/JBHI.2020.2986926]
- Valanarasu J M J, Oza P, Hacihaliloglu I and Patel V M. 2021. Medical transformer: gated axial-attention for medical image segmentation [EB/OL]. [2022-04-23]. <https://arxiv.org/pdf/2102.10662v2.pdf>
- Wang J D, Sun K, Cheng T H, Jiang B R, Deng C R, Zhao Y, Liu D, Mu Y D, Tan M K, Wang X G, Liu W Y and Xiao B. 2021. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10): 3349-3364 [DOI: 10.1109/TPAMI.2020.2983686]
- Wang X L, Girshick R, Gupta A and He K M. 2018a. Non-local neural networks//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- Wang Y, Deng Z J, Hu X W, Zhu L, Yang X, Xu X M, Heng P A and Ni D. 2018b. Deep attentional features for prostate segmentation in ultrasound//*Proceedings of the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention*. Granada, Spain: Springer: 523-530 [DOI: 10.1007/978-3-030-00937-3_60]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xia F, Shao H J and Deng X. 2022. Cross-stage deep-learning-based MRI fused images of human brain tumor segmentation. *Journal of Image and Graphics*, 27(3): 873-884 (夏峰, 邵海见, 邓星. 2022. 融合跨阶段深度学习的脑肿瘤MRI图像分割. *中国图象图形学报*, 27(3): 873-884 [DOI: 10.11834/jig.210330])
- Xiao X, Lian S, Luo Z M and Li S Z. 2018. Weighted Res-UNet for

- high-quality retina vessel segmentation//Proceedings of the 9th International Conference on Information Technology in Medicine and Education. Hangzhou, China: IEEE: 327-331 [DOI: 10.1109/ITME.2018.00080]
- Xu J H, Teng W P K, Wang X J and Nürnberger A. 2021. A deep supervised U-attention net for pixel-wise brain tumor segmentation//Proceedings of the 6th International MICCAI Brainlesion Workshop. Lima, Peru: Springer: 278-289 [DOI: 10.1007/978-3-030-72087-2_24]
- Zhao C C, Zhao Z M, Zeng Q R and Feng Y J. 2021. MVP U-Net: multi-view pointwise U-Net for brain tumor segmentation//Proceedings of the 6th International MICCAI Brainlesion Workshop. Lima, Peru: Springer: 93-103 [DOI: 10.1007/978-3-030-72087-2_9]
- Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

- Zhou Z W, Siddiquee M M R, Tajbakhsh N and Liang J M. 2018. UNet++: a nested U-Net architecture for medical image segmentation//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Granada, Spain: Springer: 3-11 [DOI: 10.1007/978-3-030-00889-5_1]

作者简介

孙家阔,男,硕士研究生,主要研究方向为医学图像处理、计算机视觉与模式识别。E-mail: sunjiakuo@foxmail.com

张荣,通信作者,女,副教授,主要研究方向为数字图像取证、计算机视觉和医学成像分析。E-mail: zhangrong@nbu.edu.cn

郭立君,男,教授,主要研究方向为计算机视觉与模式识别、智能科学、医学影像分析。E-mail: guolijun@nbu.edu.cn

汪建华,男,主任医生,主要研究方向为医学影像分析。
E-mail: woxingw@sina.com