

# SAU-Net: 基于 U-Net 和自注意力机制的 医学图像分割方法

张淑军, 彭 中, 李 辉

(青岛科技大学信息科学技术学院, 山东青岛 266061)

**摘 要:** 基于深度学习的生物医学图像分割由于其精度的提高, 可以更好地辅助医生做精确的诊断. 目前主流的基于 U-Net 的分割模型通过多层卷积进行局部特征的提取, 缺失了全局信息, 使分割过于局部化而产生误差. 本文通过自注意力机制和分解卷积策略对 U-Net 模型进行改进, 提出一种新的深度分割网络 SAU-Net, 使用自注意力模块增加全局信息, 将原 U-Net 中的级联结构改为逐像素相加, 减小维度, 降低计算量; 提出一种快速简洁的分解卷积方法, 将传统卷积分解为两路一维卷积, 并加入残差连接强化上下文信息. 在 BRATS 和 Kaggle 两个脑肿瘤数据集上进行的实验结果表明, SAU-Net 在参数量和 Dice 系数上都有更优的性能.

**关键词:** 自注意力; 分解卷积; 医学图像分割; 深度学习; U-Net

中图分类号: TP301.6

文献标识码: A

文章编号: 0372-2112(2022)10-2433-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20200984

## SAU-Net: Medical Image Segmentation Method Based on U-Net and Self-Attention

ZHANG Shu-jun, PENG Zhong, LI Hui

(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China)

**Abstract:** Biomedical image segmentation based on deep learning can better help doctors make an accurate diagnosis due to its enhanced accuracy. At present, the U-Net-based mainstream segmentation model extracts local features through multi-layer convolutions, which lacks global information and leads to over-localized results with errors. This paper improves the U-Net model through the self-attention mechanism and decomposition convolution and proposes a new deep segmentation network called SAU-Net. The model uses the self-attention module to increase global information, and changes the cascade structure in the original U-Net to pixel-by-pixel addition in order to reduce the dimension and cut down the calculation cost. A fast and concise decomposition convolution method is proposed which integrates the traditional convolution into a two-way one-dimensional convolution, and the residual connection is added to enhance the context information. The experimental results conducted on the two brain tumor datasets of BRATS and Kaggle show that SAU-Net has better performance in terms of parameters and the Dice coefficients.

**Key words:** self-attention; decomposition convolution; medical image segmentation; deep learning; U-Net

### 1 引言

医学图像分割是医学图像处理与分析领域复杂而关键的步骤, 其目的是将医学图像中具有某些特殊含义的区域如器官、病灶等分割出来, 并提取相关特征, 为临床诊疗和病理学研究提供可靠的依据, 辅助医生做出更为准确、快速的诊断, 从而减轻医生的工作量, 提高计算机辅助诊断的智能化和效率. 基于深度学习的方法通过

海量数据的学习和训练取得更优的准确率, 在 CT 图像、X 光图像、超声图像、PET 图像和 MRI 图像分割等领域均有良好的表现<sup>[1]</sup>. 比较经典的算法是 2015 年 Ronneberger 等人<sup>[2]</sup>提出的 U-Net 分割网络, 其采用编解码对称结构设计, 形式简单, 通过跳接传递高级语义信息, 采用 U 型结构提取底层特征, 但其连续的卷积和池化层会造成全局空间信息的损失, 影响了其普适性与扩展性. 针

收稿日期: 2020-09-06; 修回日期: 2021-12-04; 责任编辑: 孙瑶

基金项目: 山东省高等学校青创人才引育计划“人工智能与医学影像分析创新团队”建设项目

对医学图像分割中多种类图像的差异性、病灶形状及尺度的多样性等问题,本文基于自注意力机制和分解卷积对经典的 U-Net 模型进行改进,提出了一种新型的网络架构 SAU-Net,设计逐层深入的自注意力模块(Self-Attention Block, SABlock)连接对应的编解码结构,充分挖掘病变区域的上下文信息,通过分解卷积降低不必要的参数量,使网络更为紧凑,有效提高分割精度。

## 2 相关工作

### 2.1 U-Net 结合全局信息分割

U-Net 和类似的体系结构通常需要更多的全局信息来执行准确的分割。2018 年,Zhou 等人<sup>[3]</sup>对 U-Net 中使用的长连接进行重排,将传统的 U-Net 结构进行剪枝和叠加,网络对大小不一的目标对象的敏感度不同,可以更好地提取不同层次的特征,并将它们通过特征叠加的方式整合,达到良好的分割效果。Kamnitsas 等人<sup>[4]</sup>和 Ghafoorian 等人<sup>[5]</sup>提出使用具有不同比例或不均匀采样斑块的多流网络来增强全局信息。Hu 等人<sup>[6]</sup>提出了一种选择机制,将特征图首先使用全局平均池化进行聚合,然后缩减为单个通道描述符,使用激活门来突出显示最具区别性的特征。

循环神经网络结构也是获取全局信息的重要方法之一,可以增强分割效果,提高分割精度。2016 年 Xie 等人<sup>[7]</sup>使用空间发条 RNN (spatial Clock Work RNN, CW-RNN) 分割 H&E 组织病理学图像中的肌束膜(perimysium)。该网络结合了来自当前补丁的行列信息以及先前的信息。为了合并来自左(上)和右(下)邻域的双向信息,RNN 在不同的方向上应用了 4 次,最终将结果连接起来并送入全连接层,产生单个补丁输出。CW-RNN 可以端到端训练,并且能够将全局上下文信息编码为每个局部图像的补丁特征,并将结构化输出用于每个局部图像补丁,使其对训练和测试均有较好的分割效果。kU-Net<sup>[8]</sup>结合了 U-Net 和 RNN,可以捕获沿轴向不同切片之间的依赖性。利用多模式 MRI 图像并应用长期短期记忆网络(Long Short-Term Memory, LSTM),可以对 kU-Net 进一步改善<sup>[9]</sup>。但是,RNN 和 LSTM 本质上仅限于短距离信息获取<sup>[10]</sup>,不能真正捕获远程的特征信息。本文引入新的卷积模块将局部信息和全局信息充分融合,以达到更好的分割精度。

### 2.2 基于自注意力机制的分割

非局部神经网络<sup>[11]</sup>首先采用自注意力机制作为计算机视觉任务、视频分类、对象检测和实例分割的子模块。自注意力机制也称作内部注意力(intra-attention),2019 年 Zhang 等人<sup>[12]</sup>将自注意力思想用于图像生成,该模块与非局部神经网络用了相似的结构,其目的都是获

取到更多的图像远端信息。CCNet<sup>[13]</sup>通过堆叠两个串行纵横交错的注意力模块来收集所有特征的上下文信息。DANet<sup>[14]</sup>采用和自注意力类似的空间和通道注意力模块来从所有像素生成信息,这要比非局部运算<sup>[11]</sup>花费更多的计算时间和 GPU 内存。ACFNet<sup>[15]</sup>提出了一种基于注意力类别特征模块的粗细分网络,将粗分割网络与自注意力所得到的类别特征进行融合,对粗分割结果进行优化。Sinha 等人<sup>[16]</sup>使用多个 ResNet 网络及多层注意力架构进行 MRI 图像的腹部器官分割,得到了较好的分割效果,但其网络复杂度较高,训练过程缓慢。上述工作都是将自注意力思想用在自然场景的分割中。

近年来有学者将自注意力机制运用在医学图像分割领域,2018 年 Oktay 等人<sup>[17]</sup>提出了一种基于栅格的注意力门(Attention Gate, AG),将 U-Net 下采样过程的输出结果经过 AG 过滤,再与上采样结果进行连接,以消除跳接过程中的噪声和无关信息,减少模型计算开销。Li 等人<sup>[18]</sup>受 Attention U-Net 的启发,提出了连接敏感注意 U-Net 用于视网膜血管的准确分割。2019 年 Qi 等人<sup>[19]</sup>将 Self-Attention 模块加入编解码网络,并在常规卷积中加入残差连接,并对脑卒中数据集进行分割。Zhang 等人<sup>[20]</sup>将 Self-Attention 用作 3 个不同的解剖面,并在 3DU-Net 的分割框架下对脑肿瘤图像进行分割。3D 卷积网络更加依赖计算设备的 GPU 显存,且在数据量不足时训练效果较差。

U-Net 编码器特征提取是采用常规卷积操作,使用小卷积核扫描全图像而得来的,这在减小特征图像空间尺寸的同时也牺牲了大量的远距离信息,不能有效地获取上下文信息。常规的卷积如膨胀卷积,通过扩大卷积感受野、获取更多的复杂信息来提升分割模型效果。Gao 等人<sup>[21]</sup>提出了 channel-wise 深度可分离卷积,在 channel 维度上进行滑动,并基于该结构替换网络最后的“全连接层+全局池化”的操作,一定程度上解决了卷积操作中输入输出的复杂全连接特性。

受上述工作的启发,为进一步加强网络中全局信息的传递,尽可能准确地捕获到病变区域的边界,同时减少网络参数量,提高网络对不同规模和特点的医学图像数据集的分割效果,本文提出两个创新点:(1)将自注意力机制引入 U-Net 的下采样与对应的上采样层连接部分,保留长跳接结构,使网络能充分挖掘全局信息,对病灶细节进行更为准确的提取;(2)设计了分解卷积模块,改变了传统的卷积方法,减少网络参数量,将近端信息与远端信息更有效地结合。因此,SAU-Net 可以提取图像中更丰富的特征信息,对医学图像做到更有效的自动分割。

### 3 SAU-Net分割网络

SAU-Net由3部分组成:U-Net基础网络结构、自注意力模块(SABlock)以及分解卷积模块(图1).

图1中,SAU-Net网络包含一个Encoder、一个Decoder以及中间的自注意力模块SABlock,图像经过编码卷积模块,提取底层特征,再依次输入两个下采样块用于减小空间大小并获得高级特征.经过每个下采样块之后,通道数量增加了一倍.然后,下采样末端送入自注意力模块汇总全局信息并产生编码器的输出.编码模块由 $3 \times 3$ 卷积和分解卷积串联而成, $3 \times 3$ 卷积用于提取局部信息,分解卷积模块将更关注上下文信息,并

减少卷积的参数量.相应地,解码器使用3个上采样块来恢复其对应的特征图像.通过自注意力模块实现下采样到上采样的跳接结构,将特征从编码器复制到解码器.由于低层特征已经包含较多的全局信息,可以直接跳接到解码器端,高级特征全局信息较少,使用SABlock可以更有效地提取全局信息.因此在网络编码器的后两层加入SABlock.在解码器端,将原U-Net结构的特征图级联操作改为逐像素相加,如图1所示.解码模块仍由分解卷积和 $3 \times 3$ 卷积组成,先通过分解卷积恢复原始信息,再输入 $3 \times 3$ 卷积输出解码特征,将上采样后的特征与SABlock的输出相加,传递到下一阶段.

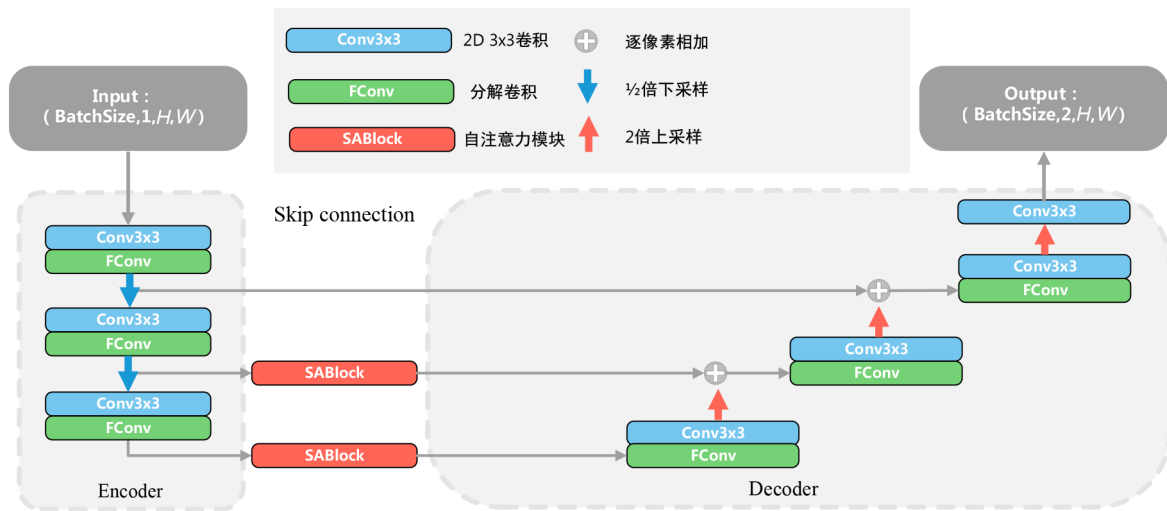


图1 SAU-Net网络结构

在SAU-Net中,编码器得到的高层特征通过SABlock模块对特征图处理后,再与解码特征图求和,形成图1中的跳接结构.使用求和来连接SABlock和解码特征具有两个优势:首先,求和不会增加特征图大小和通道的数量,能够减少下一层中可训练参数的数量;其次,具有自注意力机制的求和结构的跳接可被视为远程残差连接,能够提高模型的分割精度.

#### 3.1 Self-Attention 结构

Self-Attention结构用来加强跳接结构中全局信息的获取.U-Net本身的解码器构建与编码器类似,反卷积等流行的上采样运算符也是局部运算符<sup>[22]</sup>,没有充分挖掘有效的上下文信息对特征进行深入提取,使得网络在处理远层依存关系方面效率较低.本文采用非局部模型将自注意机制引入图像分割框架中,将编码器得到的高层特征进行进一步的优化处理,Self-Attention的处理机制如图2所示.

图2所示的结构可视为网络特征提取的SABlock模块,可以有效提取全局信息. $B$ 代表一次送入网络的特征图数量(BatchSize), $C$ 代表输入的图像的通道数, $H$

和 $W$ 分别代表输入特征图的长和宽.

Self-Attention结构将特征图进行3次 $1 \times 1$ 的卷积操作,图2中输入部分是图像的特征图 $x$ ,且 $x \in \mathbb{R}^{C \times H \times W}$ ,通过 $1 \times 1$ 卷积核得到3个输出 $F(x), G(x), Q(x)$ . $F(x)$ 和 $G(x)$ 的输出通道数是 $C/k$ ,其中 $k$ 为正整数; $Q(x)$ 不改变通道数,输出的通道数仍为 $C$ .故得到 $F(x)=[B, C/k, H, W]$ ,  $G(x)=[B, C/k, H, W]$ ,  $Q(x)=[B, C, H, W]$ .分别对 $F(x)$ 和 $G(x)$ 进行维度和形状变化,得到

$$\begin{aligned} \text{Reshape}(F(x)) &= [B, C/k, HW] \\ \text{Reshape}(G(x)) &= [B, C/k, HW] \end{aligned} \quad (1)$$

如图2中的Step1所示,对 $B$ 中的每一个特征图,将变形后的 $F(x)$ 进行转置,与 $G(x)$ 相乘,并经过Softmax得到关系图 $\beta(x_i, x_j)$ 为

$$\beta(x_i, x_j) = \frac{\exp(M_{ij})}{\sum_{i=1}^N \exp(M_{ij})}, \quad M_{ij} = F(x_i)^T G(x_j) \quad (2)$$

上述的 $\beta(x_i, x_j) \in \mathbb{R}^{N \times N}$ 表示在图像特征图 $x$ 的 $j$ 位置对 $i$ 位置的注意力程度, $N$ 表示输入特征图中像素的位置数量,即 $N=HW$ .将得到的关系图 $\beta(x_i, x_j)$ 和 $H(x)$ 做矩阵

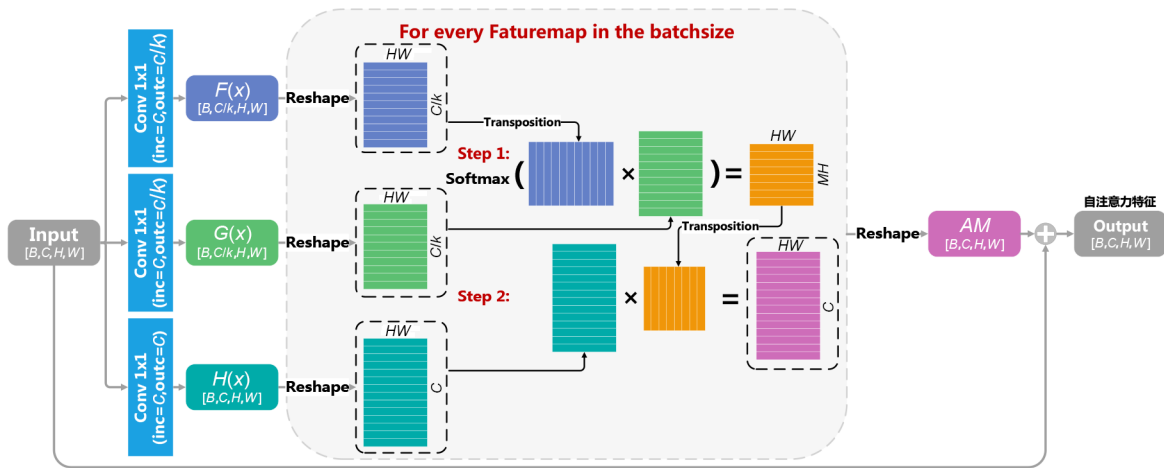


图2 Self-Attention 结构处理机制

相乘运算, 得到注意力图  $AM$ . 其输出为  $AM = (AM_1, AM_2, \dots, AM_j, \dots, AM_N) \in \mathbb{R}^{C \times N}$ , 其中

$$AM_j = \sum_{i=1}^N \beta(x_i, x_j) H(x_i) \quad (3)$$

在式(3)中, 得到的  $AM = [B, C, HW]$ . 实验发现在将  $F(x)$  和  $G(x)$  进行矩阵相乘时, 通道数的大小并不会影响最终分割准确性, 本文选取  $k=8$  来提高计算效率<sup>[12]</sup>. 此外, 将输出的注意力图  $AM$  进行变形得到  $\text{Reshape}(AM) = [B, C, H, W]$ , 结果与输入的特征图相加, 这样可以保留更广的上下文信息和全局特征. 最终注意力特征输出为

$$y = x + AM \quad (4)$$

### 3.2 分解卷积结构

本文使用改进的分解卷积对局部特征表示进行丰富的上下文相关性建模, 不仅可以获取远距离的复杂特征, 还可以考虑近距离的简单特征, 同时可实现轻量化分割. 将  $3 \times 3$  卷积核分解为  $3 \times 1$  和  $1 \times 3$  卷积核, 使原始卷积具有相同的接受域 (receptive field) 和较少的参数 (可以看作  $3 \times 3$  卷积内核的 1 阶近似操作). 假设不同的空间语义区域具有不同的特征, 则当原始卷积核的膨胀率是 1 或较小时, 卷积核可能不会跨越多个不同的空间语义区域, 获取特征的信息量较少且简单, 因此  $3 \times 3$  卷积核的 1 阶近似值会更有效, 分解过程如图 3 所示.

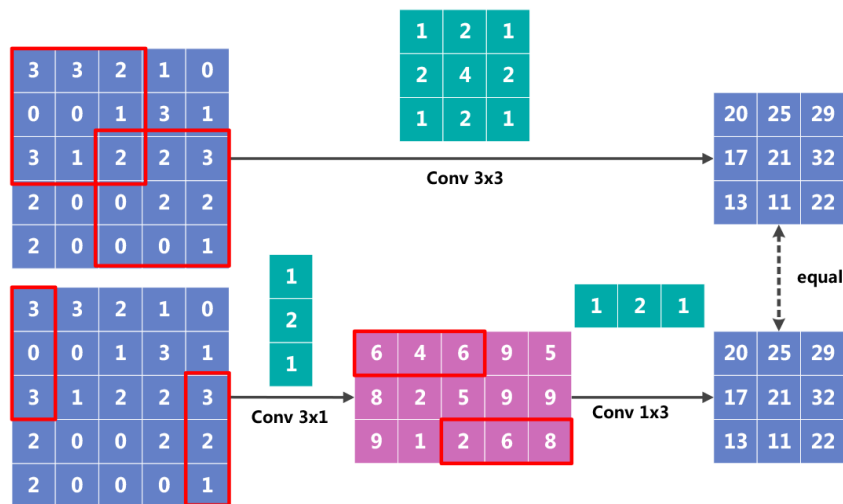


图3 卷积核  $3 \times 3$  分解为  $3 \times 1$  和  $1 \times 3$  卷积核图示

下面计算单路分解卷积的计算量. 单路分解卷积被称为空间可分离卷积 (Spatial Separable Convolution), 设在卷积核为  $n \times n$ 、卷积步长为 1、填充为 0 的  $N \times N$  图像上做卷积. 传统的卷积需要进行  $(N-2)(N-2)n^2$  次

乘法运算, 而分解卷积只需要进行  $N(N-2)n + (N-2)(N-2)n = (2N-2)(N-2)n$  次乘法运算.

由于在进行分解卷积前对通道进行了分解, 假设输入通道数是  $C$ , 则单路分解卷积的通道数是  $C/2$ , 常规

卷积的计算量为  $C(N-2)(N-2)n^2$ , 而分解卷积的计算量为  $(C/2 \times 2)(2N-2)(N-2)n$ , 因此分解卷积与原始卷积的计算成本之比为

$$\frac{2}{n} + \frac{2}{n(N-2)} \quad (5)$$

当卷积核的大小为  $n=3$  时, 由于图片的边长  $N$  较大, 计算量可近似为原始卷积的  $2/3$ , 减少了卷积的计算量. 大卷积核可以在空间中接收复杂的或具有空间信息的远距离特征, 并且在空间中需要更多参数. 同时, 较小的卷积核将在空间中接收简单或较少信息的短距离特征, 并且空间中的参数较少.

如图 4 所示<sup>[23-25]</sup>, 本文提出的 FConv 模块首先在 2

个完全分开的通道中通过一维分解卷积处理短距离和空间较少的信息特征, 大大减少了参数和计算量. 通过单路卷积后将两个通道的特征级联, 再通过  $1 \times 1$  的卷积进行降维, 与输入特征相加, 最后设置了通道混洗 Shuffle 操作将卷积后的信息进行相连, 用到的通道混洗操作和 ShuffleNet<sup>[24]</sup> 中的一致. FConv 在空间上对短距离特征使用低阶近似, 对远距离特征使用残差相连, 从而实现了更轻便、高效且强大的特征提取.

分解卷积之后, 参数降低至原始卷积的  $2/3$ , 通过两路分解卷积中卷积核以及卷积维度的变化, 增加图像远距离和近距离的信息, 与单纯的原始卷积层处理相比, 能够在提取特征过程中更多考虑到上下文信息的影响.

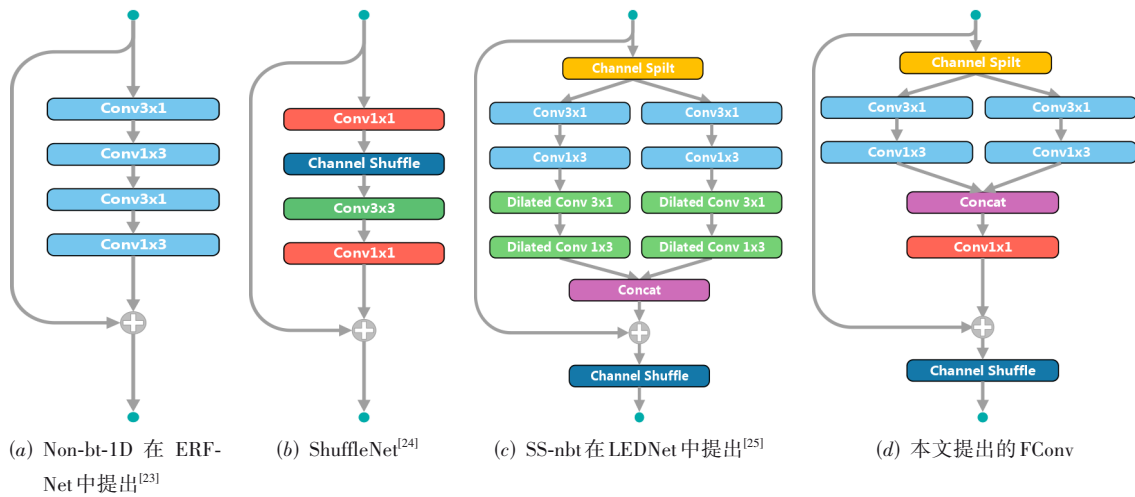


图 4 比较不同的分解卷积模块

## 4 实验结果与分析

### 4.1 数据集与实验配置

为验证上述方法的有效性, 本文选用 2 种目前公知的医学图像数据集进行实验, 如图 5 所示.

(1) 脑部肿瘤数据集 BRATS<sup>[26]</sup>. 实验使用 BRATS

2017 数据集. 其使用“.nii”格式存储, 训练集包括 210 例高级别胶质瘤 (High Grade Gliomas, HGG) 数据和 75 例低级别胶质瘤 (Low Grade Gliomas, LGG) 数据. 采用多像存储, 包括 Flair, T1, T2, T1c, 每个像包括 155 张横断面解剖图. 手动标注结果包含 4 个标签: 标签 1 为坏

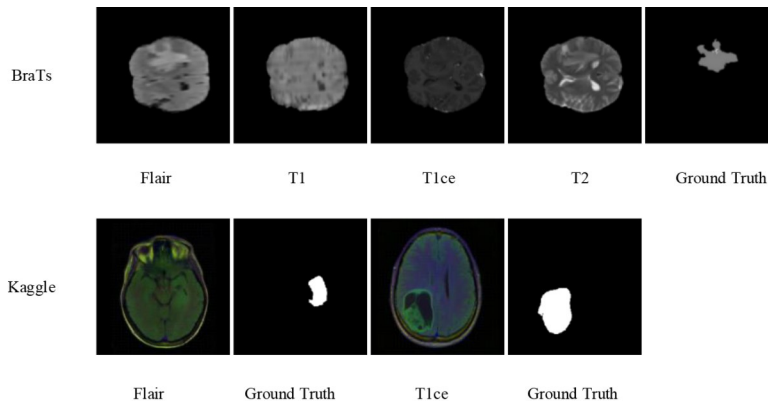


图 5 数据集示例

死核(Necrotic Core), 标签2为水肿(Edema), 标签4为增强核(Enhancing Core), 标签0为其他部分. 使用标签4是由于BRATS 2017将BRATS 2015中的标签3和标签1结合为新的标签1, 但其他标签仍保留了原有的命名顺序. 为了减小不同数据之间亮度不同所带来的差异, 本实验将对上述BRATS数据集采用N4BiasFieldCorrection进行校正(数据集地址为<http://braintumorsegmentation.org/>).

(2) Kaggle 脑部低级别胶质瘤数据集. 图像从TCIA(The Cancer Imaging Archive)获得. 数据集只有脑部低级别胶质瘤MR图的Flair像以及带有经杜克大学董事会认证的放射线医生批准的分割掩膜. 其包括110个训练样本, 每个案例横断面图像数量不同, 且采用“.tiff”格式存储, 数据已做过预处理, 不对其进行校正(数据集地址为<https://www.kaggle.com/mateuszbudalgg-mri-segmentation>).

医学图像分割的一个局限性是数据稀缺, 通常会过拟合. 针对此问题, 实验中进行了数据增强, 通过对数据本身与其对应的分割图像变换合成新的图像, 包括缩放、旋转、平移、水平翻转和镜像反转操作来保证训练的数据量.

实验配置如下: GPU为2块TITAN Xp 12G, 内存128G, 编码框架PyTorch 0.4.1, CUDA10.0. 本文将以U-Net为基准进行对比, 并对提出的模型进行消融实验.

## 4.2 BRATS数据集实验结果及分析

SABlock提升分割效果对比实验: 分别在HGG小数据(50例)和HGG全数据(210例)上进行, 以U-Net作为基线, SAU-Net为对比, 分别在数据量、Dice loss以及交并比(Intersection over Union, IoU)3个指标方面进行对比分析. 实验使用BRATS HGG中的全像数据, 模型训练采用Adam优化器, batchsize为10, 学习率为0.00001, 迭代epoch为100次, 终止次数为迭代次数. 实验结果分别如表1和图6所示.

表1 不同数据量下分割结果对比

数据量	方法	Train Dice loss	Train IoU	Test Dice loss	Test IoU
50例	U-Net	0.094 8	0.839 0	0.236 6	0.737 9
	SAU-Net	0.050 1	0.865 7	0.126 1	0.850 3
210例	U-Net	0.058 9	0.857 4	0.185 5	0.790 1
	SAU-Net	0.052 0	0.866 8	0.171 5	0.802 3

从表1可见, 通过加入单个SABlock对U-Net进行改进之后, 小数据量和全数据量分割的Dice loss和IoU性能都有提升. 训练时间均在12h左右, 没有明显增加.

从图6可以看出, 加入SABlock对分割效果有优化作用: 小数据量时比U-Net更能获取病灶信息, 对非病灶区域有较好的过滤作用; 随着数据量的增加, 对病灶区域有更好的优化效果, 细化病灶边界, 与Ground Truth更贴合.

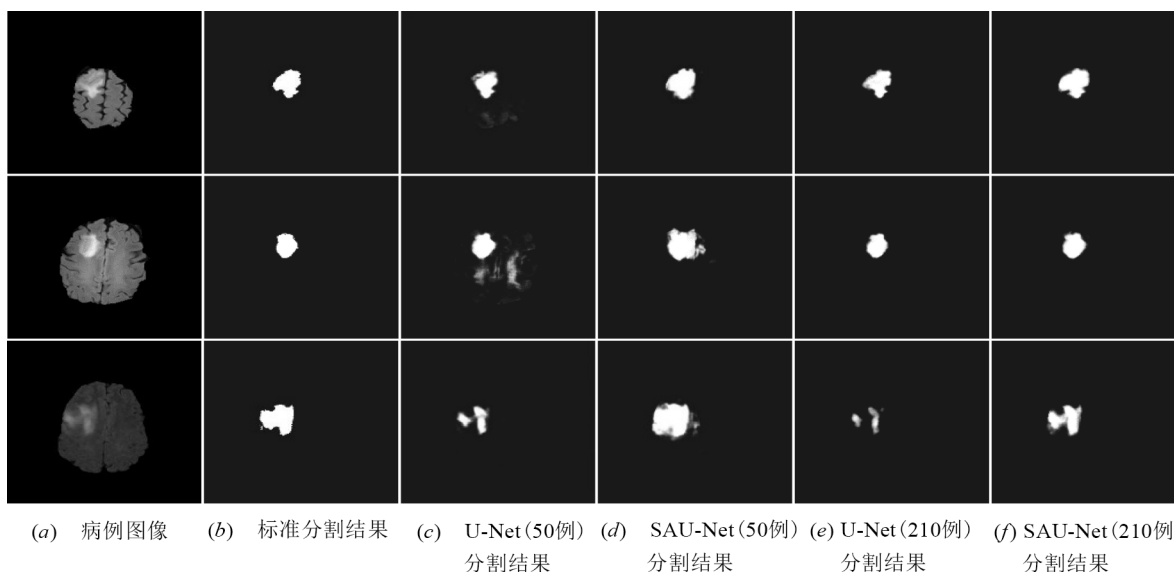


图6 不同数据量下的分割结果

单像Flair分割消融实验: Flair像便于鉴别脑室内/周围高信号病灶(如多发性硬化、脑室旁梗塞灶)以及与脑脊液信号难于鉴别的蛛网膜下腔出血、肿瘤及肿

瘤周围水肿等. 为了进一步验证SABlock对模型的影响, 将模型在Flair单像下进行消融实验, 实验结果如表2和图7所示. 在表2中, 表左侧代表不同的分割方

法,命名方式解释:如SAU-Net-FCB-1,其中FCB模块代表加入分解卷积模块,1代表加入的SABlock模块的个数,Concat表示级联方式,Add表示逐像素相加.

表2中,比较SAU-Net-1和SAU-Net-FCB-1(Concat)模型可以发现,在同样的SABlock数量下,加入FCB模块后测试的效果更好,模型不容易过拟合.比较SAU-Net-FCB-1(Add)和SAU-Net-FCB-1(Concat),融合方式改为Add在测试数据集上表现更好,可以更好地平衡训练和测试的Dice Loss.对比加入不同的SABlock可知,在SABlock的数量为1时,测试的Dice Loss可以保证最小,且训练过程更稳定,训练和测试的Dice Loss达到的平衡最优.

表2 BRATS 2017 Flair单像分割消融实验

分割方法	Train Dice Loss (均值±标准差)	Test Dice Loss (均值±标准差)
U-Net	0.396 11±0.000 23	0.441 89±0.000 25
SAU-Net-1	0.362 89±0.000 29	0.458 29±0.000 43
SAU-Net-FCB-1(Concat)	0.402 72±0.000 17	0.450 73±0.000 27
SAU-Net-FCB-1(Add)	0.406 21±0.000 20	0.427 71±0.000 30
SAU-Net-FCB-2(Concat)	0.362 81±0.000 32	0.469 14±0.000 40
SAU-Net-FCB-2(Add)	0.370 14±0.000 20	0.467 17±0.000 39

BRATS 2017 Flair单像训练与测试过程中Loss函数曲线对比如图7所示.

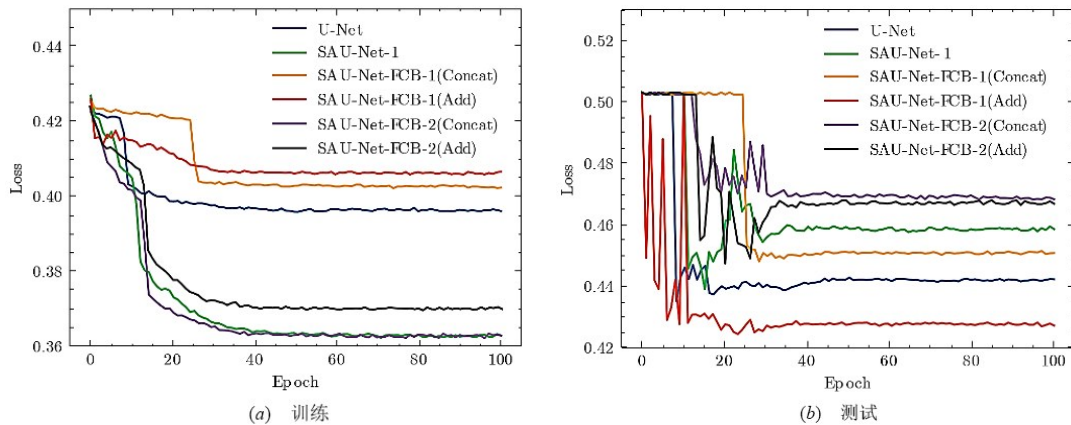


图7 BRATS 2017 Flair单像训练与测试曲线

从图7可以看到,训练过程中SAU-Net-FCB-2(Concat)模型Loss函数收敛速度最快,但在测试过程中收敛较慢,而SAU-Net-FCB-1(Concat)模型也比SAU-Net-FCB-1(Add)模型收敛要慢,这反映了Concat连接方式比较容易对训练数据产生过拟合,而Add方式在测试过程中效果更佳.从模型使用SABlock的个数来看,SAU-Net-FCB-1模型比SAU-Net-FCB-2模型在测试曲线上性能更优,其中,SAU-Net-FCB-1(Add)测试结果最好.

训练时Batchsize为25,学习率为0.000 1,迭代epoch为100次,终止次数为迭代次数,采用Adam优化器.比较各模型的参数量大小,如表3所示.

选用参数量较少,训练和测试效果相对较好的模

表3 各模型参数量大小

模型	参数量
U-Net	39, 401, 026
SAU-Net	40, 699, 523
SAU-Net-FCB-1(Concat)	11, 087, 555
SAU-Net-FCB-1(Add)	8, 273, 091
SAU-Net-FCB-2(Concat)	11, 169, 796
SAU-Net-FCB-2(Add)	8, 355, 332

型SAU-Net-FCB-1(Add),作为最终的分割模型.

BRATS 2017数据集有3项分割任务,分别是全肿瘤分割(Whole Tumor, WT)、肿瘤核分割(Tumor Core, TC)、增强肿瘤核分割(Enhancing Tumor core, ET).其中WT使用标签1, 2, 4;TC使用标签1, 4;ET使用标签4.使用SAU-Net-FCB-1(Add)模型进行3项任务分割,实验结果如表4及图8所示.从表4可见,在BRATS 2017的3项分割任务中,本文提出的模型分割Loss较低,标准差更小.

图8是将分割后得到的标签和原Flair像进行叠加.以图8(a)为依据,由图8(b)和(c)这2种模型的结果对比可知,在病灶的不同形状和尺度下,本文提出的方法在精度上比U-Net都有显著的提高,漏分割和过分割现象明显减少.

### 4.3 Kaggle脑部低级别胶质瘤数据集

对于Kaggle LGG数据,同样使用3种模型对其进行对比,训练学习率为0.001,训练epoch为100次,结束训练时间为epoch结束时间.为了优化分割效果,在此次训练中使用二值交叉熵损失和Dice损失相结合,交叉熵损失保持在0.5左右.实验结果分别如表5和图9、

表 4 BRATS 2017 的 3 项分割任务的 Dice Loss

模型	Dice Loss		
	WT(均值±标准差)	TC(均值±标准差)	ET(均值±标准差)
U-Net	0.069 57±0.006 7	0.086 11±0.010 6	0.081 59±0.009 2
SAU-Net-FCB-1(Add)	0.068 53±0.006 2	0.054 78±0.008 0	0.056 27±0.010 5

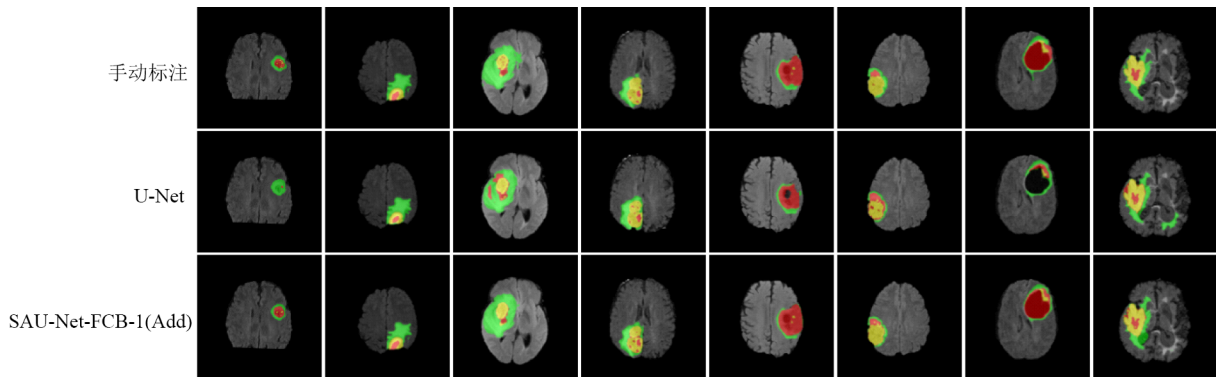


图 8 BRATS 2017 数据集下的分割结果

注:其中红色代表坏死核,绿色代表水肿,黄色代表增强核

表 5 Kaggle LGG 数据分割 Dice Loss

分割方法	Train Dice Loss (均值±标准差)	Test Dice Loss (均值±标准差)
U-Net	0.265 63±0.007 60	0.274 03±0.011 77
SAU-Net-1	0.239 88±0.007 58	0.276 76±0.008 41
SAU-Net-FCB-1(Add)	0.238 56±0.008 87	0.272 58±0.011 19

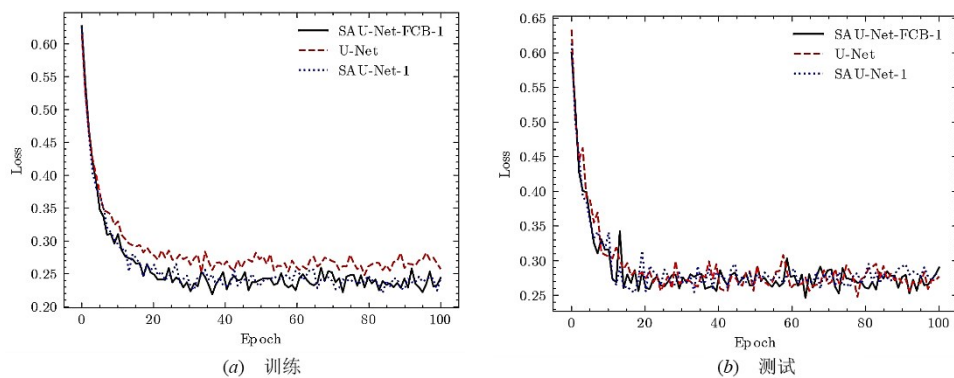


图 9 Kaggle 数据训练和测试 Dice Loss

图 10 所示. 由此可以看出 SAU-Net-FCB-1(Add) 模型分割的结果 Loss 值更小且分割的结果相对更好, 验证了本文提出的结合 SABlock 和分解卷积的方法可以获得更准确的分割结果.

## 5 结论

针对经典 U-Net 全局信息提取不充分、易因局部

化而导致分割误差的问题, 本文提出了一种结合自注意力机制和分解卷积的改进的医学图像分割方法 SAU-Net, 以获得更具代表性的图像底层和高层特征, 提升图像分割的精度. 在两个医学图像数据集上的实验表明, 本文方法较其他方法有更强的移植性, 且在降低了网络参数的情况下得到了更高的精度. 后续工作中我们将对模型进行多尺度的进一步优化.

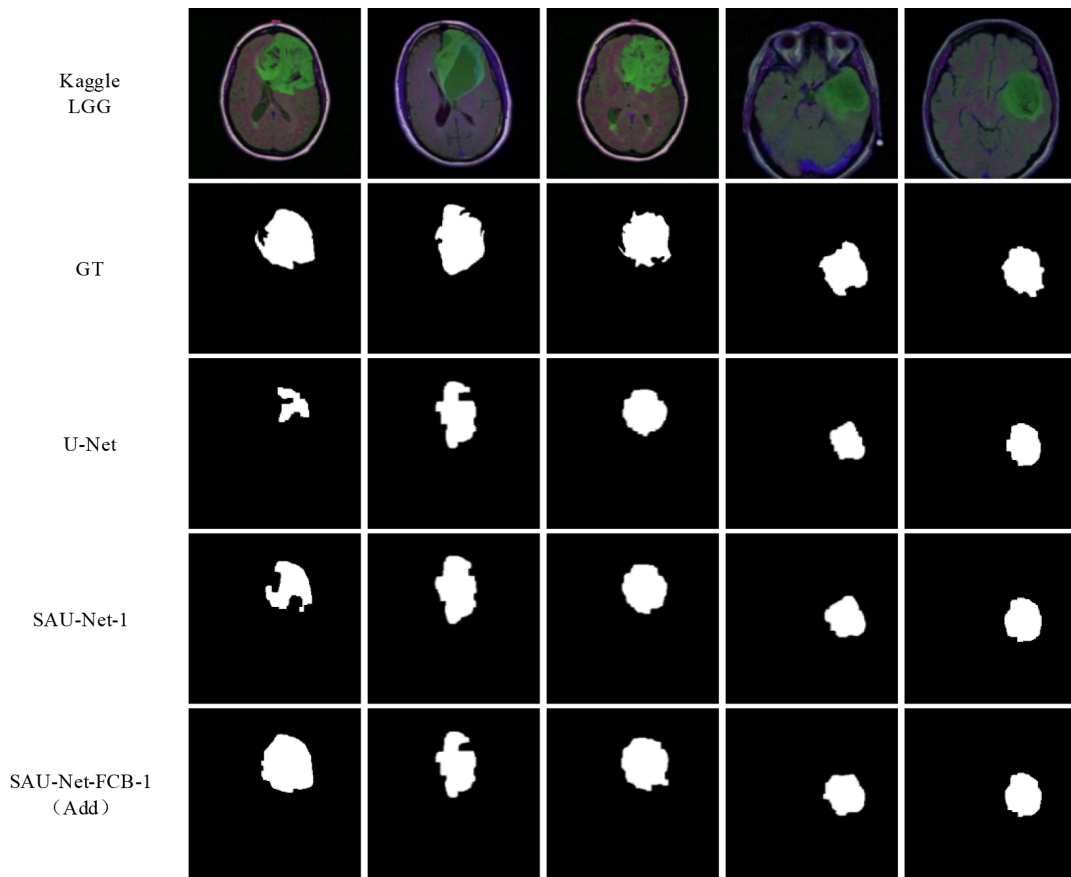


图 10 Kaggle数据集分割结果

## 参考文献

- [1] ALTAF F, ISLAM S M S, AKHTAR N, et al. Going deep in medical image analysis: Concepts, methods, challenges, and future directions[J]. IEEE Access, 2019, 7: 99540-99572.
- [2] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015: 234-241.
- [3] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: A nested U-net architecture for medical image segmentation[J]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, 11045: 3-11.
- [4] KAMNITSAS K, LEDIG C, NEWCOMBE V F J, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation[J]. Medical Image Analysis, 2017, 36: 61-78.
- [5] GHAFORIAN M, KARSSEMEIJER N, HESKES T, et al. Non-uniform patch sampling with deep convolutional neural networks for white matter hyperintensity segmentation[C]//2016 IEEE 13th International Symposium on Biomedical Imaging. Prague: IEEE, 2016: 1414-1417.
- [6] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [7] XIE Y P, ZHANG Z Z, SAPKOTA M, et al. Spatial clockwork recurrent neural network for muscle perimysium segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2016: 185-193.
- [8] CHEN J X, YANG L, ZHANG Y Z, et al. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation[C]//NIPS' 16: Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016: 3044-3052.
- [9] TSENG K L, LIN Y L, HSU W, et al. Joint sequence learning and cross-modality convolution for 3D biomedical segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 3739-3746.
- [10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]//NIPS' 17: Proceedings of the 31st

- International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
- [11] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7794-7803.
- [12] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019: 12744-12753.
- [13] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 603-612.
- [14] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach: IEEE, 2019: 3141-3149.
- [15] ZHANG F, CHEN Y Q, LI Z H, et al. ACFNet: Attentional class feature network for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV). Seoul: IEEE, 2019: 6797-6806.
- [16] SINHA A, DOLZ J. Multi-scale self-guided attention for medical image segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(1): 121-130.
- [17] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[C]//1st Conference on Medical Imaging with Deep Learning (MIDL 2018). Amsterdam: MIDL, 2018: 4861068.
- [18] LI R R, LI M M, LI J C, et al. Connection Sensitive Attention U-NET for Accurate Retinal Vessel Segmentation [EB/OL]. (2019)[2020]. <https://arxiv.org/abs/1903.05558>.
- [19] QI K H, YANG H, LI C, et al. X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2019: 247-255.
- [20] ZHANG H, ZHANG J W, ZHANG Q H, et al. RSANet: Recurrent slice-wise attention network for multiple sclerosis lesion segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2019: 411-419.
- [21] GAO H Y, WANG Z Y, CAI L, et al. ChannelNets: Compact and efficient convolutional neural networks via channel-wise convolutions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2570-2581.
- [22] GAO H Y, YUAN H, WANG Z Y, et al. Pixel transposed convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(5): 1218-1227.
- [23] ROMERA E, ÁLVAREZ J M, BERGASA L M, et al. ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 263-272.
- [24] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6848-6856.
- [25] WANG Y, ZHOU Q, LIU J, et al. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation[C]//2019 IEEE International Conference on Image Processing. Taipei: IEEE, 2019: 1860-1864.
- [26] MENZE B H, JAKAB A, BAUER S, et al. The multimodal brain tumor image segmentation benchmark(BRATS) [J]. IEEE Transactions on Medical Imaging, 2015, 34(10): 1993-2024.

#### 作者简介



张淑军 女, 1980年10月出生, 山东泰安人. 现为青岛科技大学信息科学技术学院副教授. 主要研究方向为计算机视觉、人工智能、虚拟现实技术等.

E-mail: zhangsj@qust.edu.cn



彭中 男, 1994年12月出生, 河南南阳人. 现为青岛科技大学信息科学技术学院在读硕士研究生. 主要研究方向为计算机视觉.

E-mail: 964451354@qq.com



李辉 男, 1984年3月出生, 河南平顶山人. 现为青岛科技大学信息科学技术学院副教授. 主要研究方向为计算机视觉、人工智能等.

E-mail: lipelin1984xyz@163.com