

用于3D医学图像分割的空间通道并行网络

谌潼, 谢勤岚*

(中南民族大学 生物医学工程学院, 武汉 430074)

摘要 获取更深层次和更多维度的图像特征有助于3D医学图像分割网络更精确、更稳定地分割图像。现有的3D医学图像分割方法通过提取图像空间特征和通道特征来获取非局部的图像特征的能力有限。为了提取更全面的图像特征,提出了SC-UNet模型。该模型通过外部空间通道注意力块不仅可以捕获图像的空间特征和通道特征,还可以提取图像空间特征与通道特征之间的强相关性。通过并行卷积注意力块能从空间和通道的角度自动学习和获取空间和通道特征。将外部空间通道注意力块和并行卷积注意力块使用拼接的方式进行结合。实验使用MSD数据集进行评估,结果表明SC-UNet模型的DSC为85.51%,比基线模型UNETR++高0.85%,比nn-UNet高1.41%。

关键词 3D医学图像分割;空间通道注意力;外部注意力;混合结构

中图分类号 TP391.4 文献标志码 A 文章编号 1672-4321(2026)01-0069-08

doi:10.20056/j.cnki.ZNMDZK.20250827

A parallel spatial-channel network for 3D medical image segmentation

CHEN Tong, XIE Qinlan*

(College of Biomedical Engineering, South-Central Minzu University, Wuhan 430074, China)

Abstract Obtaining deeper, more multi-dimensional image features facilitates more precise and stable segmentation in 3D medical image segmentation networks. Existing 3D medical image segmentation methods exhibit limited capability in extracting non-local image features through spatial and channel feature extraction. To acquire more comprehensive image features, the SC-UNet model is proposed. This model employs an External Spatial Channel Attention (ESCA) to not only capture spatial and channel features but also extract strong correlations between spatial and channel characteristics. Through the Parallel Convolutional Attention Block (PCAB), the network automatically learns and acquires spatial-channel features from dual perspectives. The ESCA and PCAB are integrated through concatenation. Experimental evaluations on the Medical Segmentation Decathlon (MSD) dataset demonstrate that the proposed SC-UNet achieves a Dice Similarity Coefficient (DSC) of 85.51%, surpassing the baseline model by 0.85% and outperforms nn-UNet by 1.41%.

Keywords 3D medical image segmentation; spatial-channel attention; external attention; hybrid architecture

3D医学图像相对于2D医学图像具有更丰富的图像信息,同时可训练图像少、学习难度高,具有大范围的对象和尺度、多类与不平衡标签、小对象和边界模糊等难点。为了克服以上难点,获取3D医学图像中丰富的特征信息,ÇICEK等在U-Net^[1]的基础上提出了3D U-Net^[2],该网络沿用编码器-解码器的U型结构,使用跳跃连接将编码器和解码器联系在一起。但由于卷积算子本身的局限性,基于CNN的

分割方法擅长提取图像的局部特征,且随着网络的加深,卷积核逐渐关注整个图像中的一个子区域,使得其聚合全局信息的能力不断下降,无法建立远距离依赖关系。

目前有两种主流方法解决以上限制。第一种方法是将卷积层堆叠,进行下采样,或者使用更大的卷积核获得更大的视野范围^[3-4];第二种方法是使用注意力模块来获取全局特征并建立远距离依赖关

收稿日期 2024-12-31

*通信作者 谢勤岚(1968-),男,教授,研究方向:医学图像处理,E-mail:xieqinlan@126.com

基金项目 湖北省自然科学基金资助项目(2016CFB489);2024年中南民族大学研究生教研资助项目“人工智能时代生物医学工程类专业学位硕士培养改革研究”

系^[4-5],如 MedT^[6]、Swin-UNet^[7]、DS-TransUNet^[8]、D-Former^[9]等.

最近的一些研究中,许多方法使用 CNN 和注意力机制相结合的策略,既可以提取局部特征又可以提取全局特征,如 TransUNet^[10]、UNETR^[11]、CT-Net^[12]、HmsU-Net^[13]等.这种策略逐渐成为 3D 医学图像分割的热门方法之一.但是这些工作主要侧重于提高分割精度而忽略了模型规模的大小.DAE-former^[14]、SSCFormer^[15]在空间特征和通道特征两个方面来提取 3D 医学图像的特征,增加了从 3D 医学图像中获取到的信息量,实现更好的性能效果.但这些方法忽略了空间特征与通道特征之间的相关性.

为了解决以上的问题,本文受到 UNETR++^[16]和 CBAM^[17]的启发,提出了 SC-UNet,一种能提取空间特征与通道特征之间相关依赖性的 3D 医学图像分割网络.该网络具有编码器-解码器的 U 型结构,通过空间通道联合注意力模块捕获图像的空间与通道之间的强依赖关系,学习整个数据集的空间和通道特征,从而提高模型的泛化能力.

1 网络模型介绍

首先概述 SC-UNet 的总体结构,之后对其关键模块结构进行详细的介绍.

1.1 总体结构

SC-UNet 是编码器-解码器结构的 U 型网络,总体结构如图 1 所示.编码器分为 4 个阶段,每一阶段都使用了空间通道联合注意力模块(Space Channel Union Attention, SCUA).在第一个阶段使用嵌入层对输入的 3D 医学图像 $X \in R^{H \times W \times D \times C}$ 进行切割,分割的数量为 $N = (H/h) \times (W/w) \times (D/d)$,之后图像进入 SCUA 中.其余阶段每一层都先通过下采样层使图像分辨率降低为原来的一半,通道扩大为原来的两倍,之后再进入 SCUA 中.通过外部空间通道注意力(External Spatial Channel Attention, ESCA)进行外部注意力计算,获取图像全局特征,并捕获通道和空间之间的强依赖关系.通过并行卷积注意力块(Parallel Convolutional Attention Block, PCAB)进行卷积注意力计算,自适应地学习图像的空间和通道特征,获取图像的局部特征.最后将两者提取到的图像特征相互融合,通过跳跃连接与解码器相对应的阶段相连,使图像恢复由下采样造成的信息损失.编码器也分为 4 个阶段,先经过上采样后与编码器的图像特征进行融合,然后进入 SCUA 中,其处理步骤与编码器类似.

1.2 空间通道联合注意力

空间通道联合注意力(SCUA)的结构如图 2 所示,它的主要作用是提取图像的空间和通道特征信

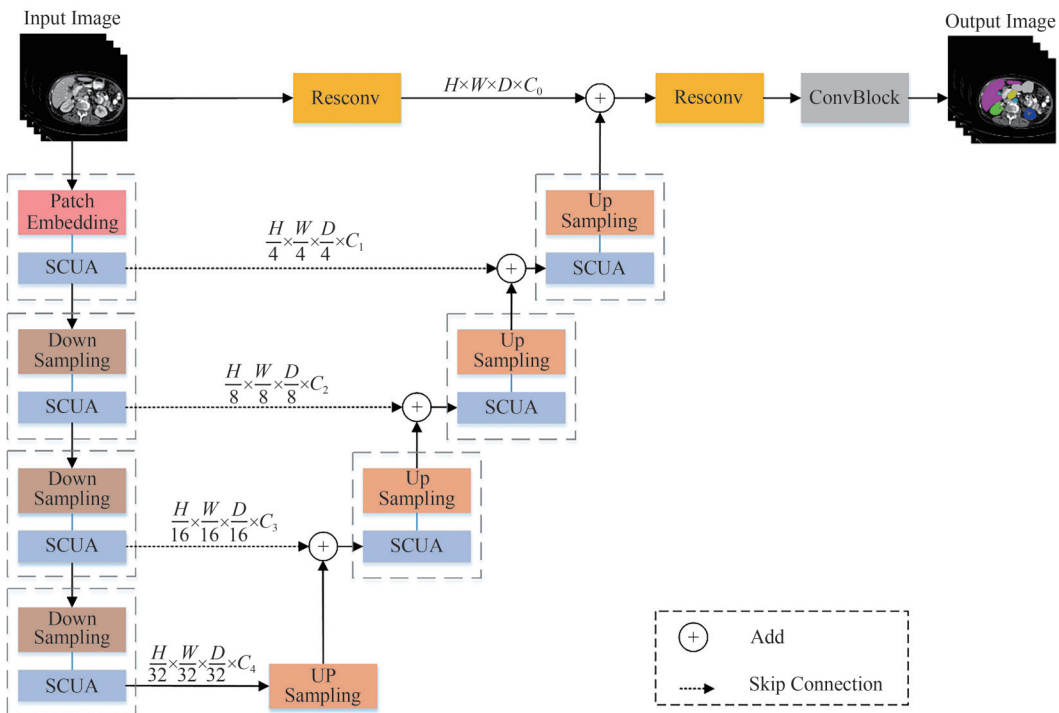


图 1 SC-UNet 结构图

Fig. 1 The structure of SC-UNet

息、捕获空间与通道之间的强依赖关系以及融合特征信息. SCUA 主要由外部空间通道注意力 (ESCA) 和并行卷积注意力 (PCAB) 组成. ESCA 使用点积注意力的方法, 通过查询矩阵和键记忆单元的权重在空间分支和通道分支之间共享, 捕获单张图片中空间与通道之间的相关性, 其中参与运算的记忆单元

可以捕获数据集中每张图片的空间和通道特征信息, 并随着训练过程的进行建模整个数据集不同样本间的联系. PCAB 使用缩放注意力的方法, 对图像分别进行空间卷积注意力计算和通道卷积注意力计算, 学习使用全局信息来选择性地强调信息特征并抑制不重要的特征.

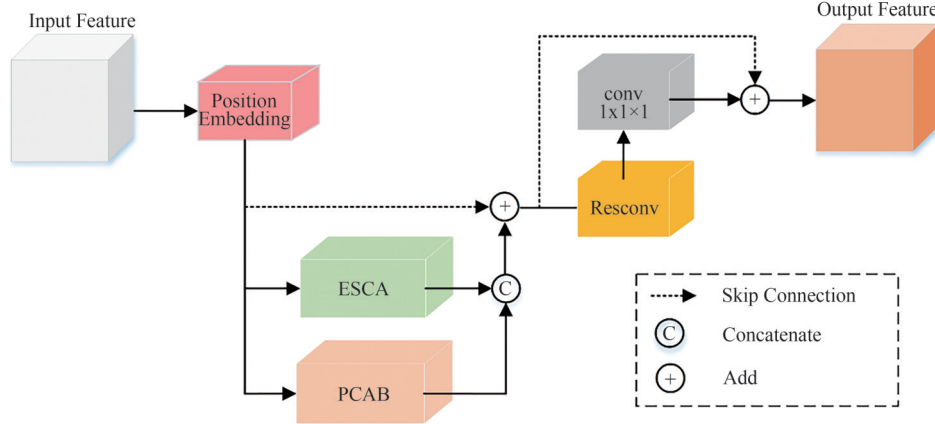


图2 SCUA 结构图

Fig. 2 The structure of SCUA

缩放注意力与点积注意力是两组不同的技术且提取特征目标也不相同^[18]. 缩放注意力是通过对各特征点分配相对应的权重来表示不同特征点的重要性; 点积注意力侧重于关注同个图像内的各个特征点, 计算每个特征点之间的相互重要性, 通过相互重要性得到注意力权重. 由于两种注意力方法的原理不相同, 获得的图像特征信息也不同, 因此 ESCA 和 PCAB 捕获的图像特征信息可以相互补充与融合.

假设输入图像为 X_i , 图像进入 SCUA 后, 首先通过位置嵌入层添加位置信息, 然后将图像分别送入 ESCA 和 PCAB, 经过它们处理后将两者融合, 最后通过残差卷积后得到 X_o . 具体公式如下:

$$X'_i = \text{Position}(X_i), \quad (1)$$

$$X'_f = \text{ESCA}(X'_i) + \text{PCAB}(X'_i), \quad (2)$$

$$F = X'_f + X'_i, \quad (3)$$

$$X_o = \text{Conv}(\text{Res}(F)) + F, \quad (4)$$

其中, X_i 为 SCUA 的输入图像, X_o 为 SCUA 的输出图像. ESCA(\cdot), PCAB(\cdot) 和 Position(\cdot) 分别代表 ESCA、PCAB 和位置嵌入层. Res(\cdot) 为使用 $3 \times 3 \times 3$ 卷积核的残差卷积块. Conv(\cdot) 为使用 $1 \times 1 \times 1$ 卷积核的卷积块.

1.2.1 外部空间通道注意力

外部空间通道注意力 (ESCA) 的结构如图 3 所示. ESCA 通过共享 Q 、 K 获取通道和空间相互依赖关

系. 在通道分支中, Q 与键记忆单元 M_k 计算出相似度 A_c , 再由 A_c 与通道值记忆单元 M_{cv} 计算出图像的通道特征; 在空间分支中, 相似度计算方法与通道部分基本一致, 区别是 M_k 需要经线性投影层后与 Q 计算出相似度 A_s . 之后 A_s 与空间值记忆单元 M_{sv} 计算出图像的空间特征信息. 由于 M_{cv} 、 M_{sv} 和 M_k 是独立于输入的可学习参数, 可记忆每张经过处理的图像的特征, 所以会随着训练过程的进行逐渐记忆整个数据集中每张图像的空间特征和通道特征. 结合以上方法, 实现捕获数据集中每一张图像的空间与通道的相互依赖, 从而提升模型性能.

具体来说, 当图像进入 ESCA 后, 首先通过线性投影生成共享的查询 $Q \in R^{B \times h \times N \times d}$. 然后 Q 与键记忆单元 M_k 计算出注意力图 A_c . 之后与通道值记忆单元 M_{cv} 计算出图像的通道特征. M_k 经过线性映射后得到 M'_k , Q 与 M'_k 计算出注意力图 A_s , 之后与空间值记忆单元 M_{sv} 计算出图像的空间特征. 最后将空间特征和通道特征进行融合, 得到最终输出的图像. 具体公式如下:

$$Q = \text{Linear}(x_{ie}), \quad (5)$$

$$A_c = (\alpha_c)_{i,j} = \text{Norm}(Q^T M_k), \quad (6)$$

$$M'_k = \text{Linear}(M_k), \quad (7)$$

$$A_s = (\alpha_s)_{i,j} = \text{Norm}(Q M'^T_k), \quad (8)$$

$$x_{oe} = A_s M_{sv} + A_c M_{cv}, \quad (9)$$

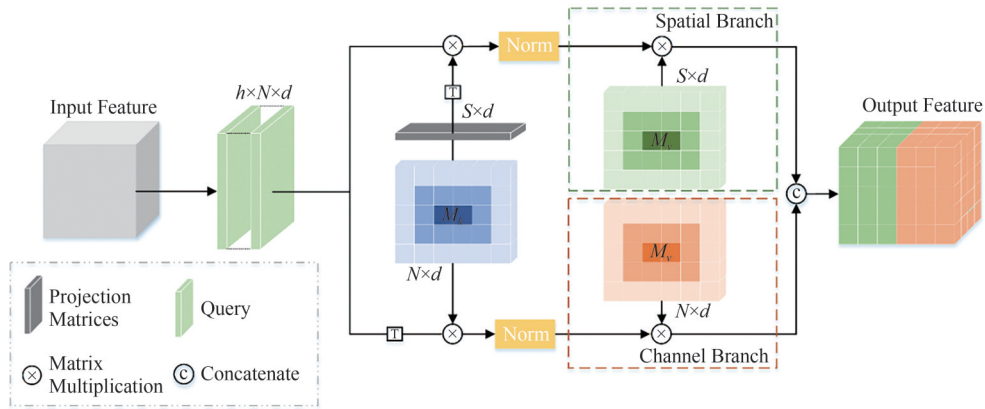


图3 ESCA结构图

Fig. 3 The structure of ESCA

其中, x_{ic} 表示 ESCA 的输入图像, x_{oc} 表示 ESCA 的输出图像. M_k 和 M_v 分别表示键记忆单元和值记忆单元, M'_k 由 M_k 映射产生. $(\alpha)_{i,j}$ 表示是第 i 个特征与 $M \in R^{d \times s}$ 的第 j 行之间的相似度. A 是通道注意力图, A_s 是空间注意力图.

1.2.2 并行卷积注意力块

并行卷积注意力(PCAB)块的结构如图4(a)所示,它主要由通道注意力和空间注意力两个分支组成.通道注意力分支结构如图4(b)所示,通过平均

池化层将图像空间信息压缩为1,由两个不同通道参数的卷积层组成的压缩和激发块提取图像空间特征信息.最后归一化给各个通道特征分配对应的权重来表示不同通道特征的重要性,为每个通道的特征图乘上对应的权重来获得不同的关注度.空间注意力分支结构如图4(c)所示,该分支由一层通道为1的卷积层组成.通过卷积层将图像通道压缩为1,通过归一化给各个空间特征分配对应的权重来表示不同空间特征的重要性.

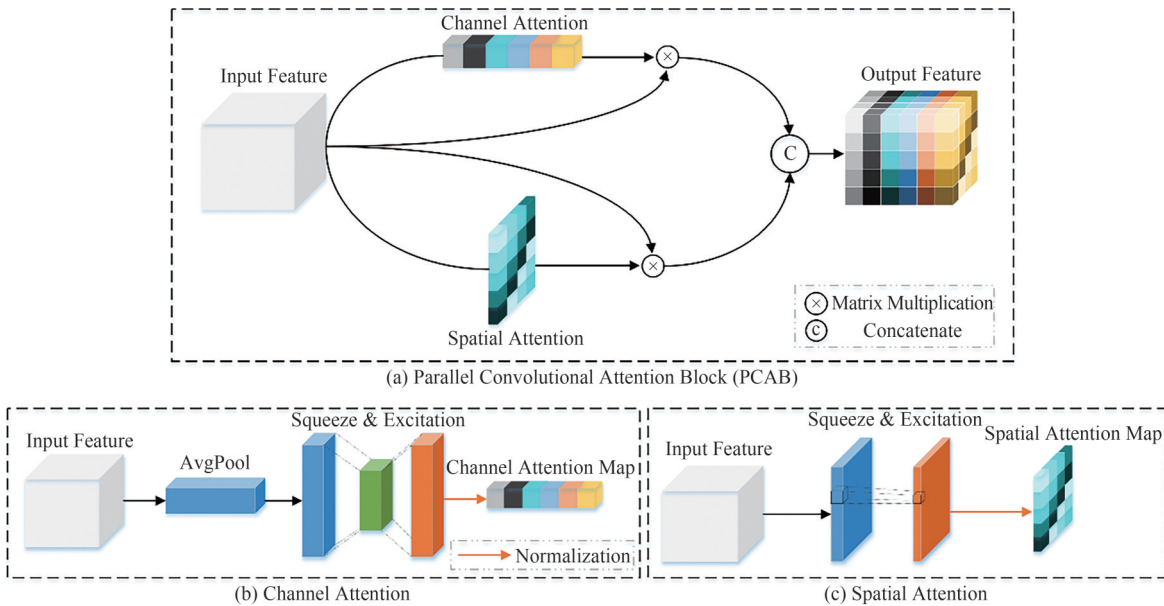


图4 PCAB结构图

Fig. 4 The structure of PCAB

具体来说,当图像进入PCAB后,先将输入图像 $x_{ip} \in R^{B \times H \times W \times D \times C}$ 转化为 $x_{ip} \in R^{B \times N \times C}$, 其中 $N = H \times W \times D$. 之后图像分别进入空间和通道两分支中进行处理.在通道分支中,图像先经过平均池化层处理,将图像空间压缩为1,即 $x_c \in R^{B \times 1 \times C}$,再由两个 1×1 卷积核的卷积层进行压缩和激发操作.第一个

卷积层将通道压缩为 $C/2$,第二个卷积层将通道恢复为 C .之后,进行归一化处理,对通道特征进行权重分配得到图像的通道注意力图.最后将原图像与权重相乘得出相应的图像通道特征.在空间分支中,图像进入 1×1 卷积核的卷积层将图像通道压缩为1,即 $x_s \in R^{B \times N \times 1}$,再通过归一化处理,对空间

特征进行权重分配得到图像的空间注意力图.最后将原图像与权重相乘得出对应的图像空间特征.

具体公式如下:

$$F_s = x_{ip} \cdot \text{Norm}\left(\text{Conv}_{1}\left(x_{ip}\right)\right), \quad (10)$$

$$F_c = x_{ip} \cdot \text{Norm}\left(\text{Conv}_c\left(\text{Conv}_{c/2}\left(\text{AvgPool}\left(x_{ip}\right)\right)\right)\right), \quad (11)$$

$$x_{op} = F_c + F_s, \quad (12)$$

其中, F_s 和 F_c 分别表示空间特征和通道特征. x_{ip} 和 x_{op} 分别表示 PCAB 的输入图像和输出图像. $\text{Conv}_1(\cdot)$ 、 $\text{Conv}_{c/2}(\cdot)$ 和 $\text{Conv}_c(\cdot)$ 分别表示输出通道为 1、 $C/2$ 和 C 的卷积层. $\text{AvgPool}(\cdot)$ 表示平均池化层.

2 实验

2.1 实验数据集

本文使用了医学分割十项全能数据集(The Medical Segmentation Decathlon, MSD)^[19]进行实验.该数据集包括来自不同器官和成像模式的 10 项分割任务,这些任务旨在针对医学图像中的常见困难,如不平衡类、小训练集和小对象等.MSD 提供了在多场景下衡量分割算法的普遍金标准.实验中选取了其中的脑肿瘤分割任务作为实验数据集.

脑肿瘤分割任务的数据集由 484 个多模态多部位并带有真实标签的 MRI 图像组成.每张图像由 4 种脑部 MRI 扫描模态构成,即 native T1-weighted (T1w), post-contrast T1-weighted (T1ce), T2-weighted (T2w) 和 Fluid Attenuated Inversion Recovery (FLAIR).每个模态的体积为 $240 \times 240 \times 155$.体素的空间分辨率为 $1.0 \times 1.0 \times 1.0 \text{ mm}^3$.所有样本按 16:3:1 的比例随机分为训练集、验证集和测试集.脑肿瘤任务可以归纳为具有 4 通道输入的 4 类分割任务.

2.2 评价指标

实验测试基于两个指标进行: Dice Similarity Coefficient (DSC) 和 95% 的 Hausdorff Distance (HD_{95}). DSC 是一种集合相似度的度量指标,通常用于计算两个样本之间的相似度,值的范围为 0~1,具体公式如下:

$$\text{DSC}(T, P) = \frac{2 \sum_{i=1}^l T_i P_i}{\sum_{i=1}^l T_i + \sum_{i=1}^l P_i}, \quad (13)$$

其中, T 和 P 分别表示体素的真实值和预测值.

HD_{95} 是衡量边界重合程度的指标,预测分割区域边界与真实区域边界之间的最大距离.其值

越小代表边界分割误差越小,分割的质量越好.设 X 和 Y 是度量空间的两个非空子集,则定义它们的 HD_{95} 为:

$$\text{HD}_{95}(X, Y) = \max\left\{\max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y)\right\} \times 95\%. \quad (14)$$

为了保持整体数值的稳定性,因此滤除 5% 的离群点,以免造成不合理的距离.一般选择从小到大排名前 95% 的距离作为实际 Hausdorff Distance,计算值越小,表示两个集合的相似度越高.

2.3 实验环境

实验在 Python 3.7、Pytorch 1.7.1 和 Windows10 系统环境中进行.使用带有 24GB RAM 的 NVIDIA TITAN RTX GPU 对模型进行训练.为了与 UNETR++^[16] 和 nn-UNet^[20] 等高性能网络进行对比,本文使用相同的预测策略和数据增强方法,并且不进行预训练.初始学习率设置为 0.01,并使用“poly”策略衰减学习率.使用 SGD 优化器对模型进行优化,将动量和衰减权重分别设置为 0.99 和 0.00003,进行 1000 轮训练.

2.4 损失函数

损失函数是由 dice 损失函数与交叉熵损失函数之和组成,以同时利用这两种损失函数的优势,其表达式为:

$$L_{\text{dice}} = 1 - \frac{\sum_{i=1}^I 2 \times \sum_{v=1}^V Y_{v,i} \cdot P_{v,i}}{\sum_{v=1}^V Y_{v,i}^2 + \sum_{v=1}^V P_{v,i}^2}, \quad (15)$$

$$L_{\text{ce}} = - \sum_{i=1}^I \sum_{v=1}^V Y_{v,i} \log P_{v,i}, \quad (16)$$

$$L_{\text{all}} = L_{\text{dice}} + L_{\text{ce}} \quad (17)$$

其中, L_{all} 、 L_{dice} 、 L_{ce} 分别表示总损失、dice 损失和交叉熵损失; I 表示类别; V 表示像素数量; $Y_{v,i}$ 表示像素 v 是否属于 i 类; $P_{v,i}$ 表示像素 v 属于 i 类的概率.

3 实验结果及分析

本节首先展示 SC-UNet 在 Tumor 数据集上的分割结果,之后对模型的参数量进行对比和分析,并通过消融实验验证 SC-UNet 的有效性.

3.1 Tumor 数据集上的分割结果

表 1 展示了 Tumor 数据集上的实验结果,最好的结果用蓝色表示.SC-UNet 的平均 DSC 值为 85.51%, HD_{95} 的值为 3.59mm.在该两项数值上, DSC 值比

UNETR++^[16]高 0.85%, HD₉₅ 仅仅比 UNETR++^[16]高 0.03mm. 值得注意的是, SC-UNet 在所有的项目中都取得了最高的 DSC 值.

从图 5 中可以直观地看出 Tumor 数据集中的分割目标的边界存在不规则、模糊的问题. 得益于 SC-

UNet 可以捕获整个数据集的空间与通道之间相互依赖的方法, 大大提升了处理此类问题时的性能; 另外还能发现, nnU-Net^[20]对 Tumor 图像分割时出现大量的添加、遗漏等错误. UNETR++^[16]对复杂边界无法做出准确的分割.

表 1 在 Tumor 数据集中分割性能的定量比较

Tab. 1 Quantitative comparison of segmentation performance in the Tumor dataset

方法	平均值		WT		ET		TC	
	DSC/%	HD ₉₅ /mm	DSC/%	HD ₉₅ /mm	DSC/%	HD ₉₅ /mm	DSC/%	HD ₉₅ /mm
TransUNet ^[10]	64.40	13.00	70.60	14.00	54.20	10.40	68.40	14.50
Unet ^[1]	66.40	10.20	76.60	9.20	56.10	11.10	66.50	10.20
Att-UNet ^[21]	66.50	10.00	76.70	9.00	54.30	10.50	68.30	10.50
CoTr ^[22]	68.30	9.70	74.60	9.20	55.70	9.50	74.80	10.50
TransBTS ^[23]	69.60	9.70	77.90	10.00	57.40	10.00	73.50	9.00
UNETR ^[11]	71.10	8.80	78.90	8.30	58.50	9.40	76.10	8.90
Swin-UNet ^[7]	82.40	8.20	88.60	8.20	76.90	6.70	81.60	9.70
nn-UNet ^[20]	84.10	3.70	91.70	3.60	78.40	3.20	82.10	4.30
UNETR++ ^[16]	84.66	3.56	91.36	4.50	78.23	2.39	84.40	3.79
SC-UNet	85.51	3.59	91.74	3.59	80.14	3.35	84.64	3.82

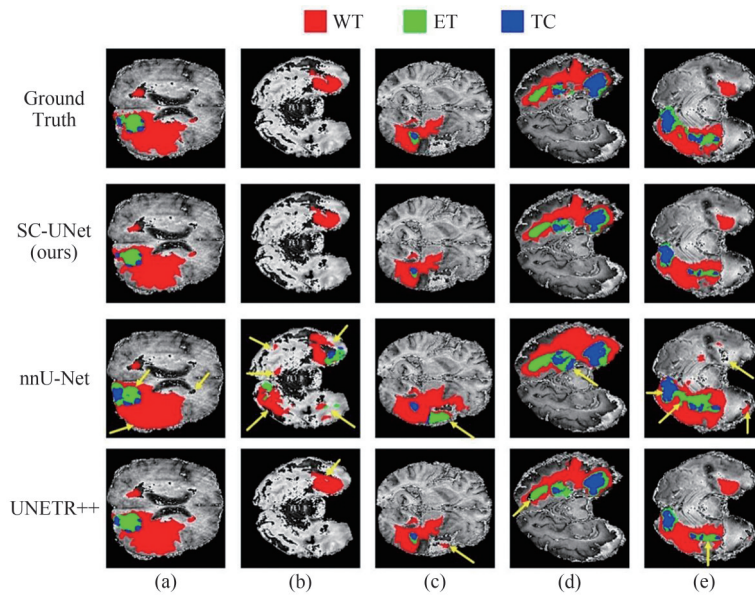


图 5 在 Tumor 数据集中分割性能的定性比较

Fig. 5 Qualitative comparison of segmentation performance in the Tumor dataset

3.2 模型参数量对比

为了公平起见, 实验中使用的样本 Patch 大小均为 $160 \times 160 \times 16$. 表 2 中展示了不同的基于自注意力机制方法的网络和不同的模块的参数, 最好的结果用蓝色表示. 从表中可以看出, SC-UNet 的参数量相对于 UNETR++^[16]减少了 9.07 M. 这些结果表明了 SC-UNet 在计算参数方面具有明显的优势. 同时, 从只使用 ESCA 的参数量相对于 UNETR++ 减少了 9.09 M, 这表示 SC-UNet 主要是通过 ESCA 达到减少模型参数的目的.

ESCA 参数更少的主要原因是 ESCA 的参数主要是由 3 个记忆单元、1 个生成 Q 的线性层和一个映射层构成. 记忆单元的参数量总量为 $2Nd + Sd$, 线性层和映射层的参数量为 $NS + C^2$, 总参数量为 $NS + 2Nd + Sd + C^2$ ($C = h \times d$, h 是头的数量). UNETR++^[16]中使用的 EPA 模块的参数主要由 4 个分别生成 Q 、 K 、 V_s 和 V_c 的线性层和两个映射层构成. 线性层的参数量为 $4C^2$, 映射层的参数量为 $2NS$, 总参数量为 $4C^2 + 2NS$. 因为 NS 远大于 $2Nd$, $3C^2$ 大于 Sd , 所以 EPA 的总参数量远大于 ESCA 的参数量.

表 2 不同网络和模块的参数量对比

Tab. 2 Comparison of parameters among different networks and modules

网络	参数量/M	模块	参数量/M
TransUNet ^[10]	105.28	UNETR++(EPA) ^[16]	66.80
UNETR ^[11]	92.69	EPA+PCAB	66.82
UNETR++ ^[16]	66.80	ESCA	57.71
SC-UNet(Ours)	57.73	SC-UNet(ESCA+PCAB)	57.73

3.3 消融实验

为了验证模型的各个基本组件的有效性,本文在 Tumor 数据集上进行消融实验,设置了 4 组实验, DSC 和 HD₉₅ 作为评估指标,第 1 组实验是基线方法

(UNETR++^[16]),第 2 组实验使用 ESCA,第 3 组实验是在 UNETR++^[16] 中加入 PCAB,第 4 组实验是同时使用 ESCA 和 PCAB.实验结果如表 3 所示,最好的结果用蓝色表示.

表 3 ESCA 和 PCAB 对分割性能的影响

Tab. 3 The impact of ESCA and PCAB on segmentation performance

方法	平均值		WT		ET		TC	
	DSC/%	HD ₉₅ /mm	DSC/%	HD ₉₅ /mm	DSC/%	HD ₉₅ /mm	DSC/%	HD ₉₅ /mm
UNETR++ ^[16]	84.66	3.56	91.36	4.5	78.23	2.39	84.40	3.79
ESCA	84.80	3.59	91.62	3.44	78.44	3.52	84.35	3.83
PCAB	85.00	3.86	91.84	3.69	78.81	2.85	84.36	5.03
ESCA+PCAB	85.51	3.59	91.74	3.59	80.14	3.35	84.64	3.82

在第 2 组实验中,平均 HD₉₅ 结果要略好于第 3 组实验,而平均 DSC 比第 1 组实验提升 0.14%,这可能是由于 ESCA 更擅长处理单模态信息;第 3 组实验的平均 DSC 提升了 0.34%,而且 WT、ET 的 DSC 都高于第 1 组实验.对比第 2 组实验,PCAB 通过自适应学习方法提取图像的空间特征和通道特征,在多模态数据集中表现出更好的稳定性和鲁棒性;在第 4 组实验中,平均 DSC 达到了 85.51%,比第 2 组实验高 0.71%,比第 3 组实验高 0.51%.实验结果表明,在 ESCA 和 PCAB 的共同作用下,平均 DSC 得到了提升.

3.4 讨论

本文从捕获 3D 医学图像的空间与通道之间的相互依赖关系入手,提出了 SC-UNet 网络.其关键组件 SCUA 由 ESCA 和 PCAB 两部分组成,ESCA 实现了捕获图像的空间和通道之间的相互依赖,同时还降低了计算的复杂度,PCAB 捕获图像的空间特征和通道特征,提取重要的信息,抑制不重要的信息.在 Tumor 数据集中,SC-UNet 的 DSC 比基线模型 UNETR++^[16] 提升了 0.85%,同时参数量降低了 9.07 M.

本文提出的分割方法还有进一步的提升空间,虽然 SC-UNet 可以通过捕获数据集的空间特征和通道特征之间的相互依赖来提升分割性能,但在实验中可以看出,SC-UNet 的 HD₉₅ 并没有比 UNETR++^[16] 好,对多模态图像和组织间分界线模糊的图像表现出了边界分割适应性不足的现象.从表 3 中可以看出,ESCA 处理分割边界不规则、模糊和杂糅的图像

时 DSC 提升很有限,这是因为 ESCA 处理带有干扰的图像能力较弱.PCAB 擅长滤除干扰信息,但会导致 HD₉₅ 指标偏高.当 ESCA 与 PCAB 结合后 DSC 和 HD₉₅ 都有稳定的提升,说明两者可以有效结合;同时,医学图像的空间特征和通道特征不仅仅局限于 3D 医学图像,而本文并没有在其他类型的医学图像中验证该方法.

4 结语

SC-UNet 是一种可以捕获整个数据集的空间特征、通道特征以及空间特征与通道特征之间的相互依赖信息的新型 3D 医学图像分割网络.该网络使用 CNN 与注意力相结合的方法,其主要结构 SCUA 由 ESCA 和 PCAB 组成,用于捕获整个数据集的空间信息和通道信息.ESCA 通过记忆模块进行空间和通道的注意力计算,并通过记忆模块学习整个数据集的相关信息,降低了参数量并提升了对 3D 医学图像的分割性能;PCAB 可以学习有效信息并抑制无效信息.消融实验证明了 ESCA 和 PCAB 可以相辅相成.由 ESCA 与 PCAB 构成的 SCUA 使 SC-UNet 拥有了出色的 3D 医学图像分割能力.

参 考 文 献

- [1] RONNEBERGER O, FISCHER P, BROX T. U-Net:

- Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer, 2015: 234-241.
- [2] ÇIÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016. Cham: Springer, 2016: 424-432.
- [3] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. ArXiv:2017, 1706.05587.
- [4] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 2818-2826.
- [5] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 9992-10002.
- [6] VALANARASU J M J, OZA P, HACIHALILOGLU I, et al. Medical transformer: Gated axial-attention for medical image segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer, 2021: 36-46.
- [7] CAO H, WANG Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for Medical image segmentation [C]//Computer Vision-ECCV 2022 Workshops. Cham: Springer, 2023: 205-218.
- [8] LIN A, CHEN B, XU J, et al. DS-TransUNet: Dual swin transformer U-Net for medical image segmentation [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 4005615.
- [9] WU Y, LIAO K, CHEN J, et al. D-former: A U-shaped Dilated Transformer for 3D medical image segmentation[J]. Neural Computing and Applications, 2023, 35(2): 1931-1944.
- [10] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers make strong encoders for medical image segmentation[J]. ArXiv: 2021, 2102.04306.
- [11] HATAMIZADEH A, TANG Y, NATH V, et al. UNETR: Transformers for 3D medical image segmentation [C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2022: 1748-1758.
- [12] ZHANG N, YU L, ZHANG D, et al. CT-Net: Asymmetric compound branch Transformer for medical image segmentation[J]. Neural Networks, 2024, 170: 298-311.
- [13] FU B, PENG Y, HE J, et al. HmsU-Net: A hybrid multi-scale U-Net based on a CNN and transformer for medical image segmentation [J]. Computers in Biology and Medicine, 2024, 170: 108013.
- [14] AZAD R, ARIMOND R, AGHDAM E K, et al. DAE-former: Dual attention-guided efficient transformer for Medical image segmentation [C]//Predictive Intelligence in Medicine. Cham: Springer, 2023: 83-95.
- [15] XIE Q, CHEN Y, LIU S, et al. SSCFormer: Revisiting ConvNet-transformer hybrid framework from scale-wise and spatial-channel-aware perspectives for volumetric medical image segmentation [J]. IEEE Journal of Biomedical and Health Informatics, 2024, 28(8): 4830-4841.
- [16] SHAKER A, MAAZ M, RASHEED H, et al. UNETR++: Delving into efficient and accurate 3D medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2024, 43(9): 3377-3390.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block attention module [M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [18] SHEN Z, ZHANG M, ZHAO H, et al. Efficient attention: Attention with linear complexities [C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2021: 3530-3538.
- [19] ANTONELLI M, REINKE A, BAKAS S, et al. The medical segmentation decathlon [J]. Nature Communications, 2022, 13: 4128.
- [20] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation [J]. Nature Methods, 2021, 18(2): 203-211.
- [21] OKTAY O, SCHLEMPER J, LE FOLGOC L, et al. Attention U-Net: Learning where to look for the pancreas [J]. ArXiv:2018, 1804.03999.
- [22] XIE Y, ZHANG J, SHEN C, et al. CoTr: Efficiently bridging CNN and transformer for 3D medical image segmentation [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021. Cham: Springer, 2021: 171-180.
- [23] WEN X W, CHEN C, MENG D, et al. Transbts: Multimodal brain tumor segmentation using transformer [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Strasbourg: Springer, 2021: 109-119.