

SAU-Net: 融合压缩注意力机制的多器官图像分割

曹国刚^{1*}, 毛红东¹, 张术¹, 陈颖¹, 戴翠霞²

¹上海应用技术大学计算机科学与信息工程学院, 上海 201418;

²上海应用技术大学理学院, 上海 201418

摘要 为了对头颈部多器官 CT 图像实现精准分割, 减少放射治疗对人体正常组织的损伤, 本文提出一种基于卷积神经网络的图像分割方法——SAU-Net 算法, 该算法基于加入残差连接的 3D U-Net 实现。针对器官尺寸差异较大而引起的分割精度不准确问题, 引入压缩注意力模块, 通过非局部的空间注意力机制增加对全局特征的编码能力, 聚合多尺度上下文信息, 实现同一器官的体素分组。此外, 该算法减少了卷积核数量及参数量, 避免因额外的卷积运算造成堆叠局部信息过多而影响模型性能。结果表明, 以 Dice 系数为评估指标, 与 3D U-Net 和 3D ResU-Net 算法相比, SAU-Net 的分割精度分别提高了 13.7% 和 8.2%, 推理速度比 FocusNetv2 提升 73%。SAU-Net 算法显著提高了头颈部器官图像的分割精度和速度, 能够快速准确实现全自动分割任务。

关键词 图像处理; 卷积神经网络; 3D U-Net; 残差连接; 压缩注意力

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202259.0417003

SAU-Net: Multiorgan Image Segmentation Method Improved Using Squeeze Attention Mechanism

Cao Guogang^{1*}, Mao Hongdong¹, Zhang Shu¹, Chen Ying¹, Dai Cuixia²

¹School of Computer Science and Information Engineering, Shanghai

Institute of Technology, Shanghai 201418, China;

²College of Sciences, Shanghai Institute of Technology, Shanghai 201418, China

Abstract Radiotherapy is a common treatment for tumors. To accurately control the radiation dose distribution and reduce the damage caused by radiation to normal tissues and organs in radiotherapy, organs at risk must be delineated precisely. In this study, a novel automatic segmentation method for organs in head and neck, named SAU-Net, was proposed, the architecture of which is based on the three-dimensional (3D) U-Net with residual connections. Nonlocal spatial attention implemented using the squeeze and attention (SA) module was introduced to solve the problem of unbalanced segmentation accuracy caused by massive differences in organs' volumes. This introduction increased the ability to aggregate multiscale contextual information by encoding global features. To avoid the stacking of excess local information by extra convolution operations and reduce the number of parameters, the model reduced the number of convolution kernels. The performance of the model was evaluated using the dice score, and SAU-Net achieved 13.7% and 8.2% higher segmentation accuracy than 3D U-Net and 3D residual U-Net (ResU-Net), respectively. Moreover, the proposed model achieved of an inference time 73% faster than that achieved by FocusNetv2. Thus, SAU-Net delineates organs at risk in the head and neck faster than AnatomyNet and more accurately than FocusNetv2.

Key words image processing; convolution neural network; 3D U-Net; residual connection; squeeze and attention

收稿日期: 2021-07-18; 修回日期: 2021-08-18; 录用日期: 2021-09-13

基金项目: 国家自然科学基金项目(61976140, 61675134, 81827807, 62175156)、上海市科委科技创新行动计划(19441905800)、温州医科大学重点实验室开放项目(K181002)

通信作者: *guogangcao@163.com

1 引言

据统计,全球每年大约有 64.5 万新发头颈部癌变病例^[1],放射治疗是此类疾病的主要治疗手段之一。放射治疗中,需要对危及器官(OARs)的 CT 图像实现精确分割,以控制射线剂量和辐射区域,从而减少射线对人体正常组织的损伤。CT 图像的传统分割大多通过专业医生手动勾画实现,不仅效率低下,且易受到医生主观经验的影响;后续出现的基于图像处理的分割算法,如分层顶点回归算法^[2]、分水岭算法^[3]、区域生长算法^[4]、水平集算法^[5]等,分割精度低,泛化性能差,且需要大量的人工交互;因此出现了基于卷积神经网络(CNN)的方法,如全卷积神经网络^[6]、U-Net^[7]、3D U-Net^[8]等,分割精度高,泛化性能好,人工交互过程少,在肺部分割^[9]、脑肿瘤分割^[10]、细胞图像分割^[11]和视网膜血管分割^[12]等任务中得到了广泛的应用。

基于 3D CNN 的分割模型,能够以端到端的方式实现头颈部多器官 CT 图像的自动分割^[13],如 Ren^[14]提出的交错 3D CNN 模型。然而,经典 3D CNN 模型难以解决因多器官体积差异过大而引起的分割精度不平衡的问题,因此,研究人员通过多种方法改进 CNN 模型来解决这一问题。在 AnatomyNet 模型^[17]中,作者通过在 3D U-Net 模型中引入注意力压缩和激发(SE)^[15]模块以增强模型的特征提取能力,并结合 Dice 损失函数^[6]和 Focal 损失函数^[16]训练模型,平衡了大器官与小器官的分割精度。单个模型的性能有限,而多个模型或结构的组合能够有效提高模型对头颈部器官的分割精度,Tappeiner 等^[18]利用先定位后分割思想,将两个 3D U-Net 级联,通过粗分割与精细分割的结合,提高了器官的分割精度;Gao 等^[19]提出的 FocusNet 模型,通过模拟医生勾画危及器官的过程,组合大器官分割网络、小器官定位网络和小器官分割网络实现了头颈部器官的精准勾画。在模型的训练过程中,引入头颈部器官的形状先验知识可以增加模型对器官形状特征的学习能力,如 Tappeiner 等^[20]在头颈部器官分割网络的自动编码器中引入头颈部器官形状先验信息,Gao 等^[21]提出的 FocusNetv2 模型通过对抗形状约束方法引入小器官形状先验信息,都有效地提高了模型的分割精度。通过不同模态图像之间的互补信息训练 CNN 模型,可以增强模型的泛化性能,Dai 等^[22]通过 CT 图像和核磁共振图像

(MRI)之间的互补信息,显著提高了头颈部器官的分割精度。

上述基于 CNN 的方法在头颈部多器官的分割任务中得到了广泛应用,但大多从逐体素预测的角度实现头颈部多器官分割。本文结合 3D U-Net 和残差连接,提出在残差连接中引入压缩注意力(SA)机制^[23]的分割模型——SAU-Net,通过引入级联的 SA 模块,聚合多尺度上下文信息,通过同一类的非局部体素分组,提高模型对多尺度器官的分割精度。此外,SAU-Net 模型减少了卷积核的数量和模型参数,解决了训练数据较少的问题,同时提高了模型的推理速度。

2 方法

2.1 SAU-Net 结构

本文针对头颈部多器官分割问题,构建基于 3D U-Net 结构和残差连接的 SAU-Net 模型。SAU-Net 的结构如图 1 所示,图中 X 表示模型的输入数据, X' 表示模型最终的分割结果, X_{in} , X_{out} 分别表示 SA 模块的输入特征图、输出特征图, X_{att} 为注意力通道的输出特征。通过两个卷积核为 $3 \times 3 \times 3$,步长为 2 的 3D 卷积前向运算提取局部特征后,其输出特征图用 X_{res} 表示。模型分为编码器和解码器两个主要结构,在编码器中,使用了 3 个带有下采样层的 3D SA 模块,下采样层通过最大池化运算实现;解码器主要结构也是由 3 个 3D SA 模块组成,其上采样运算通过步长为 2 的反卷积运算实现。为了能够实现深层特征和浅层特征的融合,基于 U-Net 的设计思想,通过跳跃连接的方式将编码器的浅层特征和解码器的深层语义特征相融合。在 SAU-Net 的最后一层输出中,通过 10 个通道的卷积层和 softmax 激活函数实现多器官体素分类。

SAU-Net 的编码器由 3 个带有下采样层的 SA 模块构成,SA 模块通过注意力通道编码非局部特征。SAU-Net 模型残差连接的主通道和注意力通道中,卷积运算是由步长为 1,卷积核大小为 $3 \times 3 \times 3$ 的卷积实现,并通过 batchnorm 层实现归一化,激活函数为非线性整流单元(ReLU)。为了提取深层特征,编码器的通道数不断增加,分别是 32, 36, 40。每次经过下采样运算后,特征图尺寸减小为原来的一半,编码器层输出特征图的尺寸缩小为原来的 $1/8$ 。

SAU-Net 的解码器与编码器相对应,由 3 个 SA

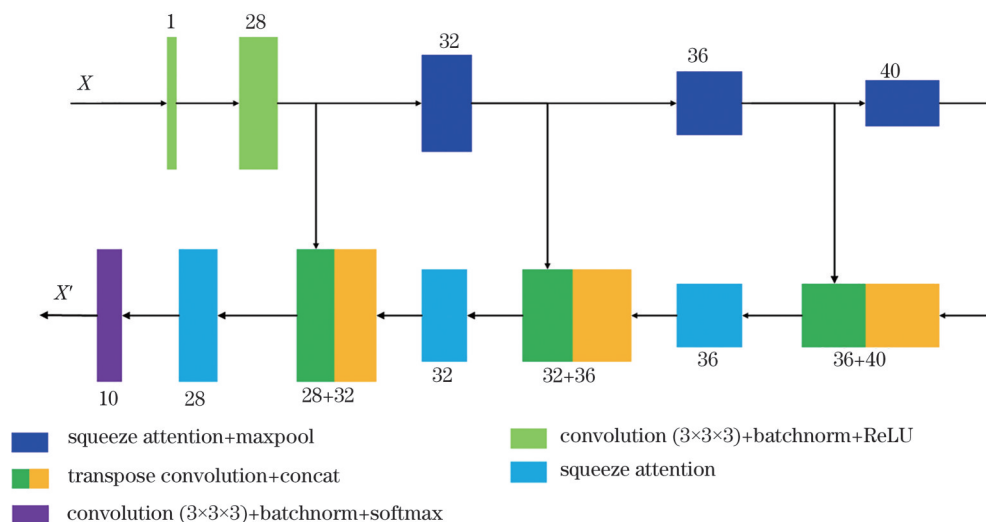


图 1 SAU-Net 结构图

Fig. 1 Architecture of SAU-Net

模块和反卷积运算构成的上采样层实现, 解码器的 SA 模块的通道数不断减小。基于 U-Net 的设计思想, SAU-Net 通过编码器浅层特征和解码器的深层特征利用快捷连接的方法实现特征融合。解码器的上采样层通过卷积核大小为 $3 \times 3 \times 3$, 步长为 2 的反卷积实现, 每次上采样运算后, 特征图的大小变成原来的 2 倍, 为了实现浅层特征与深层特征的融合, 上采样得到的特征图与浅层编码器特征图在通道维度上拼接, 拼接后特征图的通道数量分别是 76, 68, 60, 拼接后的特征图再通过 SA 模块聚合上下文多尺度信息, 其通道数量分别是 36, 32 和 28。

2.2 压缩注意力模块

基于卷积神经网络的语义分割任务中, 通常是

通过逐像素预测的机制实现像素级别的分类, 进而完成图像分割任务。多尺度的语义分割任务 (如头颈部多器官的分割), 通过压缩空间注意力机制, 鼓励非局部的同一类信息分组, 聚合非局部特征, 从而提高多尺度语义分割任务的精度。SA 模块与 SE 模块结构相似, 不同之处在于, SA 模块用平均池化层代替 SE 模块中的全局池化层, 保留了特征图通道信息的同时还保留了空间信息。此外, 注意力通道中, 以卷积层代替 SE 模块中的全连接层, 实现了体素级别的分类, 进而鼓励同一类体素分组。SA 模块的结构如图 2 所示。

SA 模块的残差连接中, 其注意力通道, 通过步长为 2 的平均池化层, 扩大了模型的感受野, 再经过

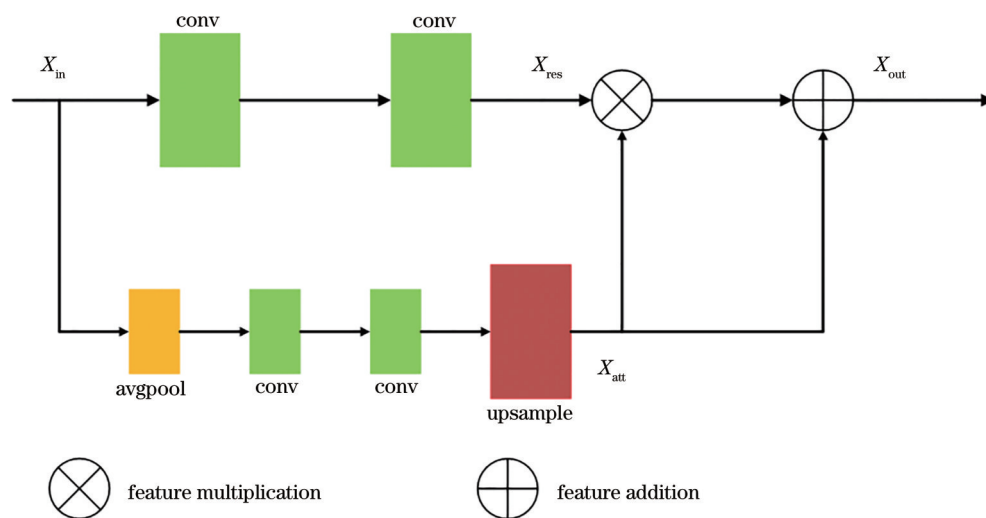


图 2 压缩注意力结构图

Fig. 2 Architecture of squeeze attention module

卷积核大小为 $3 \times 3 \times 3$, 步长为 1 的两个连续的卷积运算, 实现逐体素特征学习, 而平均池化层实现了下采样运算, 使得模型更加关注非局部特征, 通过注意力通道和主通道上的特征值相乘, 实现全局特征的学习。此外, SA 模块还利用特征相加的方式, 融合多尺度特征, 增强了模型对多尺度特征的学习能力。最后通过上采样运算, 将残差连接中所求得的特征图尺寸恢复到原来的尺寸。SA 模块输入与输出的关系如(1)~(3)式所示。

$$X_{\text{res}} = F_{\text{conv}}(X_{\text{in}}), \quad (1)$$

$$X_{\text{att}} = U \{ F_{\text{conv}} [F_{\text{avg}}(X_{\text{in}})] \}, \quad (2)$$

$$X_{\text{out}} = X_{\text{res}} \times X_{\text{att}} + X_{\text{att}}, \quad (3)$$

式中: F_{conv} 表示连续两次卷积核大小为 $3 \times 3 \times 3$, 步长为 1 的 3D 卷积运算, F_{avg} 表示步长为 2 的平均池化运算, U 表示步长为 2 的上采样运算。

2.3 损失函数

头颈部多器官 CT 图像的分割, 其本质是基于体素(像素)的分类问题, 而器官的大小, 反映在 CT 图像中的特征即是各个器官体素数量的多寡。脑干、视交叉、下颌骨、左视神经(Optic. L)、右视神经(Optic. R)、左腮腺(Paro. L)、右腮腺(Paro. R)、左下颌骨(Subm. L)和右腮腺(Subm. R)的体素数量如图 3 所示, 横坐标表示器官类型, 纵坐标表示每个器官的平均体素数量。在同样的分辨率中, 体积最大的下颌骨与体积较小的视神经的体素数量差异超过 60 倍以上。体素数量的巨大差异带来严重的分割精度不平衡的问题, 具体表现在体积较大的器官分割精度较高, 而体积较小的器官分割精度较低。

在模型的训练过程中, 基于 Tversky 系数^[24]的

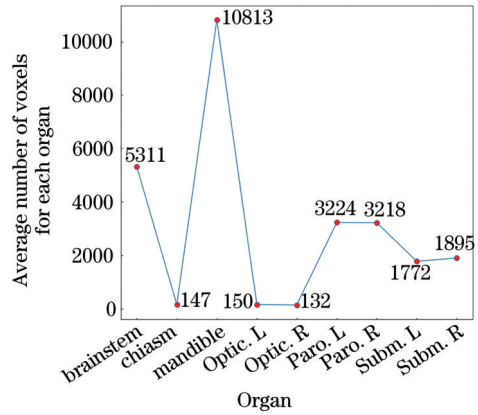


图 3 头颈部各器官 CT 图像体素数量
Fig. 3 Number of voxels in each organ of head and neck CT images

损失函数, 通过 α 和 β 参数调节假阴性(FN)和假阳性(FP)的比例, 可以有效解决因体素数量差异太大而引起的分割精度不平衡的问题, Tversky 系数的损失函数为

$$T(P, G, \alpha, \beta) = \frac{|PG|}{|PG| + \alpha |P/G| + \beta |G/P|}, \quad (4)$$

式中: P 表示模型对当前体素的预测值, 即模型 softmax 激活函数的输出, G 表示当前体素真实标签, $|PG|$ 表示真阳性, $|P/G|$ 表示假阳性, $|G/P|$ 表示假阴性, 通过调节系数 α 和 β 可以对 FP 和 FN 加以调节, 有效解决头颈部多器官分割精度不平衡的问题。实验验证, 当 $\alpha = 0.3, \beta = 0.7$ 时, 模型性能最佳。

针对多类分割和训练样本中头颈部器官标签数量不一致的问题, 定义 Tversky 损失函数 L_{Tversky} 为

$$L_{\text{Tversky}} = \sum_{i=1}^C m(i) - \frac{\sum_{i=1}^C m(i) \sum_{j=1}^N p_{0j}(i) g_{0j}(i)}{\sum_{i=1}^C \sum_{j=1}^N p_{0j}(i) g_{0j}(i) + \alpha \sum_{i=1}^C \sum_{j=1}^N p_{1j}(i) g_{1j}(i) + \beta \sum_{i=1}^C \sum_{j=1}^N p_{1j}(i) g_{0j}(i)}, \quad (5)$$

式中: C 表示头颈部器官的类别数目, 此处取 $C = 9$; N 表示体素的数量; $m(i)$ 表示训练样本器官是否被标注, 如果存在, 则为 1, 否则为 0; $p_{j0}(i)$ 表示第 j 个体素没有被预测为第 i 个器官的概率, 与之相反, $p_{1j}(i)$ 表示第 j 个体素被预测为第 i 个器官的概率; $g_{0j}(i)$ 表示第 j 个体素不属于第 i 个器官的概率, $g_{1j}(i)$ 表示第 j 个体素属于第 i 个器官的概率。

3 实验结果与分析

3.1 数据

对于有监督学习的分割算法, 数据集的多寡与处理方式有直接关系。本文原始数据集为来源于 MICCAI Head and Neck Auto Segmentation Challenge 2015 (以下简称 MICCAI 2015) 的 PDDCA 数据集^[13], 包含 38 个训练样本和 10 个测试样本。为了提高模型性能, 本文采用 Zhu 等^[17]提供

的数据处理方法,用 Head-Neck Cetuximab collection^[25](数据集可以在 The Cancer Imaging Archive 下载)的 CT 图像数据集和四个不同的加拿大魁北克研究机构公开的头颈部 CT 图像数据集^[26]

(数据集可以在 The Cancer Imaging Archive 下载)作为扩充训练集。最终,训练数据集包含 261 个样本,测试集包含 10 个样本,数据集的来源分布如表 1 所示。

表 1 数据集来源与分布

Table 1 Source and distribution of dataset

Type	Source	Number
Training dataset	MICCAI 2015 training dataset	38
	Head and Neck Cetuximab collections	46
	Public dataset of the Quebec Institute of Canada	177
Test dataset	MICCAI 2015 test dataset	10

每个样本包含了头颈部 CT 图像及其分割图像的二值标签。对原始 CT 图像进行裁剪,以节省硬件资源,提高模型的推理速度和分割精度。为了保证扩充后的训练数据与原始数据的一致性,通过重采样运算,将 CT 图像横断面、冠状面和矢状面的体素间隔分别设置为 3.0,1.12,1.12 mm。

PDDCA 数据集包含 9 个头颈部器官图像的标注数据,而扩充训练集所用到的部分数据并没有完全包含这 9 个器官图像的标注数据,以横坐标表示器官的类型,纵坐标表示训练集中器官的标注数量,261 例训练数据集标注情况如图 4 所示。

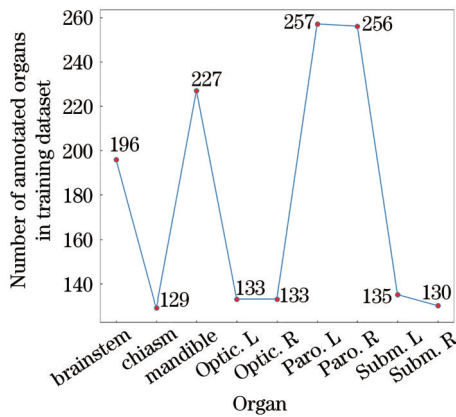


图 4 训练集中器官标注数据数量分布

Fig. 4 Number of annotated organs in training dataset

训练集中标注数据数量不一致,这给模型训练带来了困难,为了解决这一问题,本文引入一个由 0 和 1 构成的向量,如果训练样本的第 c 个标注器官存在,则该向量的第 c 个元素为 1,否则为 0。

3.2 评估指标与实验设置

本文以 Dice 相似系数(DSC) E_{dice} 、过分割率(FPR) E_{fpr} 和欠分割率(FNR) E_{fnr} 作为分割模型的评价指标,其定义为

$$E_{dice}(P, G) = \frac{2|P \cap G|}{|P| + |G|}, \quad (6)$$

$$E_{fpr} = \frac{|P/G|}{|PG| + |P/G| + |G/P|}, \quad (7)$$

$$E_{fnr} = \frac{|G/P|}{|PG| + |P/G| + |G/P|}, \quad (8)$$

式中: P 表示模型的预测值, G 表示体素真实标签的体素数量, $|P \cap G|$ 表示模型预测正确的体素个数, $|P| + |G|$ 表示模型预测值和真实标签的体素数量之和。

实验采用 NVIDIA RTX 2080Ti GPU 和 Inter i7-10700 CPU 搭建的硬件平台及 SAU-Net 模型来处理数据。为了节省硬件资源,提高模型训练速度,利用 NVIDIA 的 APEX 实现在 Pytorch 中混合精度计算,通过 RMSProp 算法优化训练过程,学习率设置为 0.001,扩充后的训练集每张 CT 图像的维度不一致,因此设置 BatchSize 为 1,损失函数采用 Tverkey 损失函数。本文以 3D U-Net 作为 baseline 模型,对比了加入残差连接的 baseline 模型 (baseline+Res)、加入 SE 模块的 baseline 模型 (baseline+SE)、与 SAU-Net 模型的损失函数迭代曲线,如图 5 所示。经过 200 次迭代以后,模型损失趋于收敛,通过对比,发现 SAU-Net 在训练过程中具有更好的稳定性。

3.3 消融实验

本文以 Dice 系数为评估指标,验证 SA 模块对实验结果的影响,如表 2 所示。结果表明,SAU-Net 模型的 Dice 平均精度要比 baseline 模型、baseline+Res 模型、baseline+Res 模型分别高出 13.7%、8.2% 和 2.8%,说明 SA 模块能够显著提高模型性能,增加模型对多尺度 OARs 的分割精度。

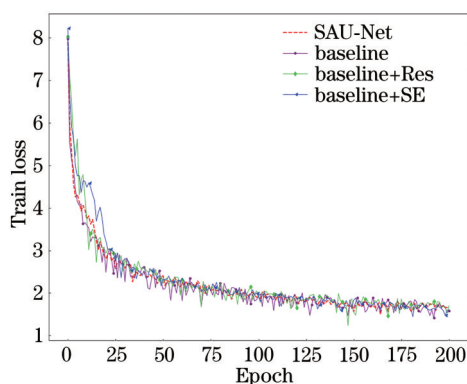


图5 不同模型的损失值曲线对比

Fig. 5 Comparison of loss value curves of different models

以 FPR 和 FNR 指标作为模型分割结果的体素在前景目标与背景信息误差率的评估标准,如表 3 所示。相比其他模型,SAU-Net 具有更小的 FPR 和 FNR,且模型的 FPR 普遍小于 FNR。这是因为基于深度学习体素分类的分割任务中,器官标注数量

表 2 不同模型的 Dice 精度对比

Table 2 Comparison of DSC with different models

Organ	Baseline	Baseline+Res	Baseline+SE	SAU-Net
Brainstem	0.773	0.834	0.857	0.884
Chiasm	0.484	0.495	0.522	0.554
Mandible	0.830	0.858	0.917	0.932
Optic. L	0.662	0.692	0.712	0.738
Optic. R	0.628	0.683	0.709	0.729
Paro. L	0.787	0.818	0.863	0.887
Paro. R	0.768	0.824	0.859	0.881
Subm. L	0.723	0.743	0.798	0.812
Subm. R	0.732	0.738	0.805	0.820
Average	0.707	0.743	0.782	0.804

和体素数量限制了模型对特征的学习能力,且器官的形状也会限制模型的分割精度,模型难以学习 ground truth 的全部形状特征,导致模型的 FPR 普遍小于 FNR。

表 3 不同网络结构模型的 FPR 和 FNR 对比

Table 3 Comparison of FPR and FNR with different models

Organ	Baseline		Baseline+Res		Baseline+SE		SAU-Net	
	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR
Brainstem	0.117	0.138	0.110	0.124	0.072	0.101	0.063	0.091
Chiasm	0.105	0.461	0.091	0.457	0.081	0.347	0.074	0.292
Mandible	0.127	0.119	0.127	0.093	0.097	0.058	0.106	0.047
Optic. L	0.118	0.428	0.106	0.415	0.092	0.327	0.034	0.295
Optic. R	0.121	0.437	0.126	0.397	0.114	0.294	0.108	0.313
Paro. L	0.137	0.154	0.120	0.139	0.088	0.131	0.076	0.092
Paro. R	0.141	0.161	0.131	0.148	0.112	0.129	0.090	0.123
Subm. L	0.115	0.146	0.109	0.139	0.092	0.122	0.084	0.117
Subm. R	0.129	0.152	0.116	0.142	0.098	0.118	0.101	0.071
Average	0.123	0.244	0.115	0.228	0.094	0.181	0.082	0.160

3.4 与其他方法实验结果对比

头颈部多器官 CT 图像分割方法,国内外已提出很多。DSC 作为标签与模型分割结果的内部填充的相似度评价指标,比较不同方法的 DSC,结果如表 4 所示。这些方法包括 MICCAI 2015 中最优结果的分割方法^[13]、Ren 等^[14]提出的头颈部小器官 CT 图像分割方法、Wang 等人^[2]提出的分层顶点回归分割方法、FocusNet^[17]、AnatomyNet^[19]、FocusNetv2^[21]方法。结果表明,FocusNetv2 的 DSC 比 SAU-Net 提高 5%,主要是因为有限的数据集下,模型难以学习器官复杂的结构,如图 4 所示,小尺寸器官的标注数量只有大尺寸器官的一半,极大限制了模型对小尺寸器官的分割精度;此

外,FocusNetv2 的训练集包含 1044 例私有 CT 图像样本和 38 例 PDDCA 训练样本,样本数量是 SAU-Netv2 的 4 倍,能够保证大型网络的分割精度。

DSC 对模型分割的内部填充较为敏感,而 95% Hausdorff 距离 (95HD)^[18]对分割的边界比较敏感,以 95HD 作为模型预测结果与真实标签的边界相似度评估标准,SAU-Net 的 95HD 比起 FocusNetv2 约降低 8%,如表 5 所示。小器官的 95HD 精度较差的主要原因在有限的数据集中,模型难以学习器官的全部形状特征,而 FocusNetv2 引入了形状先验知识解决这一问题。

模型的参数量和计算速度是衡量模型性能的重要指标,模型的参数量越少,推理速度越快,且越

表 4 不同方法 DSC 精度对比
Table 4 Comparison of DSC score of different methods

Organs	Ref. [13]	Ref. [14]	Ref. [2]	AnatomyNet ^[17]	FocusNet ^[19]	FocusNetv2 ^[21]	SAU-Net
Brainstem	0.880	—	0.903	0.867	0.875	0.882	0.884
Chiasm	0.557	0.580	—	0.532	0.596	0.713	0.554
Mandible	0.930	—	0.944	0.925	0.935	0.947	0.932
Optic. L	0.634	0.720	—	0.721	0.735	0.790	0.738
Optic. R	0.639	0.700	—	0.706	0.744	0.817	0.729
Paro. L	0.827	—	0.823	0.881	0.863	0.898	0.887
Paro. R	0.814	—	0.829	0.874	0.879	0.881	0.881
Subm. L	0.723	—	—	0.814	0.798	0.840	0.812
Subm. R	0.723	—	—	0.813	0.801	0.838	0.820
Average	0.749	—	—	0.793	0.803	0.845	0.804

表 5 不同方法的 95HD 对比
Table 5 Comparison of 95HD score of different methods unit:mm

Organs	Ref. [13]	Ref. [14]	AnatomyNet ^[17]	FocusNet ^[19]	FocusNetv2 ^[21]	SAU-Net
Brainstem	4.59	—	6.42±2.38	2.14±0.6	2.32±0.7	2.02±0.8
Chiasm	2.78	2.81±1.6	5.76±2.49	3.16±1.3	2.25±0.8	2.65±1.3
Mandible	1.97	—	6.28±2.21	1.18±0.3	1.08±0.4	2.12±0.7
Optic. L	2.26	2.33±0.8	4.85±2.32	3.76±2.9	1.92±0.8	2.14±1.8
Optic. R	3.15	2.13±1.0	4.77±4.27	2.65±1.5	2.17±0.7	2.32±1.4
Paro. L	5.11	—	9.31±3.32	2.52±1.0	1.81±0.4	2.54±1.1
Paro. R	6.13	—	10.08±5.09	2.07±0.8	2.43±2.0	2.32±1.7
Subm. L	5.35	—	7.01±4.44	2.67±1.3	2.84±1.2	2.27±1.6
Subm. R	5.44	—	6.02±1.08	3.41±1.4	2.74±1.2	2.92±1.8
Average	4.14	—	6.30	2.62	2.17	2.37

容易在小微型计算平台上快速部署,以 NVIDIA RTX 2080Ti GPU 和 Inter i7-10700 CPU 搭建的硬件平台作为计算平台,对比不同模型的参数量和计算速度,如表 6 所示。AnatomyNet 模型的参数量比 SAU-Net 减少 8.2%,而 SAU-Net 的平均推理速度比 AnatomyNet 提高 33%,原因在于 AnatomyNet 只采用了一次下采样,而 SAU-Net 采用了三次下采样,模型深层特征图更小,运算速度更快。SAU-Net 的参数量比 FocusNetv2 减少 60%,推理速度提高 73%,原因为 FocusNetv2 是由三个子网络构成

表 6 不同方法的参数量和计算速度对比

Table 6 Comparison of parameters and inference time of different methods

Models	Metrics	
	Parameters /million	Time /s
AnatomyNet	0.73	0.68
FocusNetv2	2.02	1.88
SAU-Net	0.79	0.51

(主分割网络、小器官定位网络和小器官分割网络),模型参数量更大,且对比单分支网络,FocusNetv2 需要单独训练三个子网络,最后通过参数微调,实现所有子网络的融合。

3.5 可视化

SAU-Net 模型的预测结果与 ground truth 的可视化对比如图 6 所示,不同颜色标记了不同的器官,图 6(a)、(b)、(c)、(d) 分别代表 SAU-Net 的预测结果、ground truth、预测结果与 ground truth 的横断面视叠加图、预测结果与 ground truth 的三维叠加图。从图中可以看出,SAU-Net 模型对大器官(如下颌骨、左右腮腺)实现了高精度分割,而对于小器官(如左右视神经、视交叉)而言,SAU-Net 预测结果与 ground truth 在体积方面差异较小,而结构相差较为明显,如视交叉器官的交叉性结构,右视神经的横断面结构为非连续结构,而模型预测结果为连续结构。

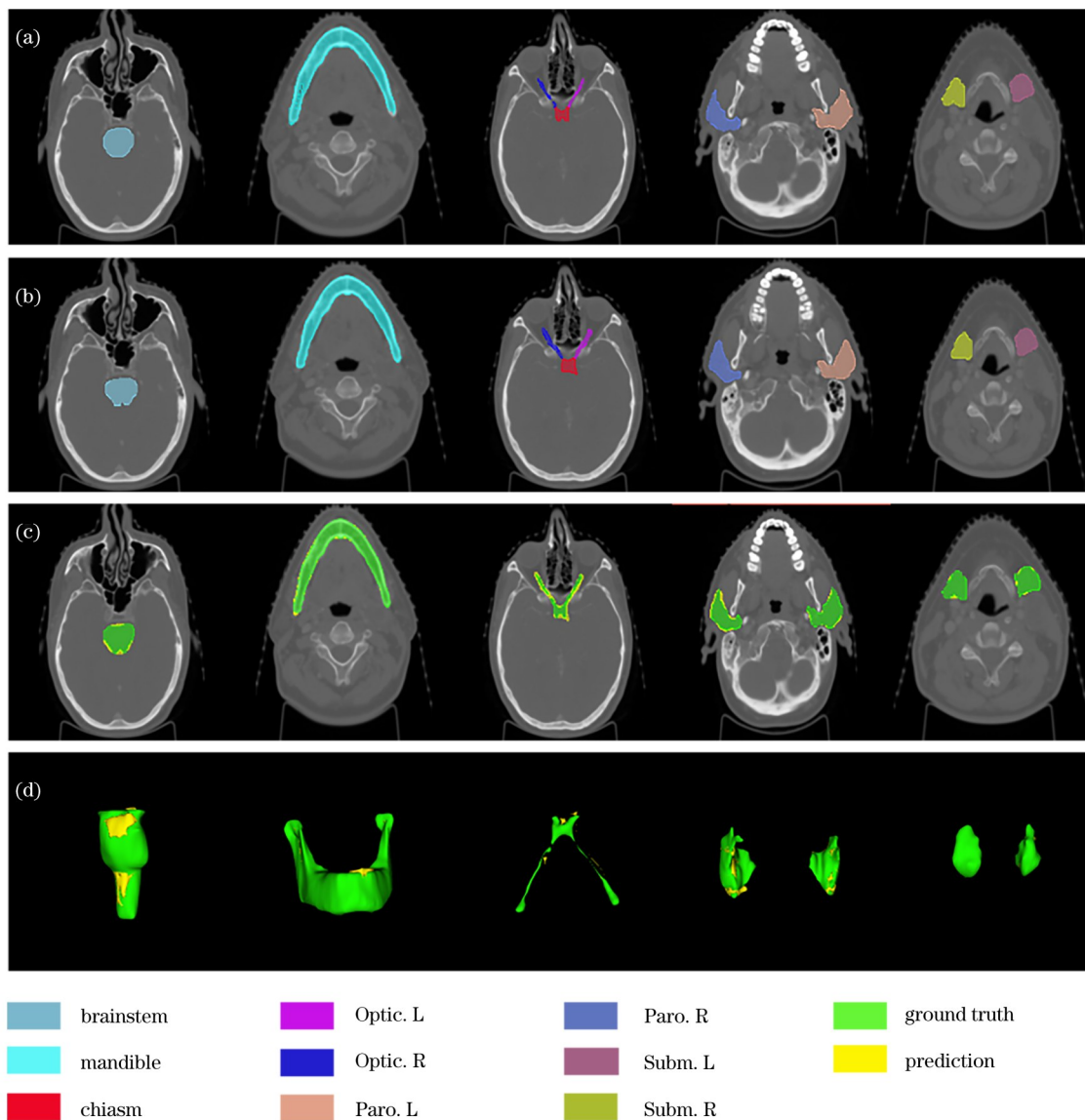


图6 SAU-Net分割结果可视化对比。(a)预测结果的横截面视图;(b)真实标签的横截面视图;(c)预测结果与真实标签的横截面叠加视图;(d)预测结果与真实标签的三维叠加视图

Fig. 6 Comparison of visualization for SAU-Net segmentation results. (a) The cross-sectional view of prediction; (b) the cross-sectional view of ground truth; (c) the cross-sectional view of overlap between prediction and ground truth; (d) the 3D view of overlap between prediction and ground truth

4 结 论

本文提出了一种基于卷积神经网络的头颈部多器官CT图像自动分割方法——SAU-Net。为了提高精度,在模型的残差连接中,引入了压缩注意力机制,在实现体素分类的同时,聚合非局部多尺度上下文信息,提高多尺度器官分割精度。为了解决小尺寸器官和大尺寸器官因体积差异巨大而引起的分割精度不平衡的问题,本文以Tversky损失函数训练模型。实验结果表明,SAU-Net模型能够

显著地提高头颈部多器官的分割精度,解决了头颈部器官尺寸差异过大而引起的分割精度不平衡的问题。此外,SAU-Net减少了卷积核的数量及模型参数量,提升了模型的推理速度,也避免了过多的卷积运算堆叠太多的局部信息而影响危及器官的分割精度。相对自然图像而言,人体头颈部器官存在较为固定的形状且图像空间结构较为稳定,实验结果表明,具有对称结构的器官,如左视神经和右视神经,虽然与视交叉器官体素数量相差不大,但分割精度更高。

参 考 文 献

- [1] Long Y, Mi K, Fan J C, et al. Prognosis and related influencing factors of HPV-positive head and neck squamous cell carcinoma[J]. *Journal of Cancer Control and Treatment*, 2011, 24(4): 196-200.
龙宇, 米昆, 樊晋川, 等. HPV 阳性头颈鳞癌预后及相关影响因素[J]. *肿瘤预防与治疗*, 2011, 24(4): 196-200.
- [2] Wang Z S, Wei L F, Wang L, et al. Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning[J]. *IEEE Transactions on Image Processing*, 2018, 27(2): 923-937.
- [3] Lin G, Adiga U, Olson K, et al. A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks[J]. *Cytometry Part A*, 2003, 56(1): 23-36.
- [4] Lu J F, Lin H, Pan Z G. Adaptive region growing algorithm in medical images segmentation[J]. *Journal of Computer Aided Design & Computer Graphics*, 2005, 17(10): 2168-2173.
陆剑锋, 林海, 潘志庚. 自适应区域生长算法在医学图像分割中的应用[J]. *计算机辅助设计与图形学学报*, 2005, 17(10): 2168-2173.
- [5] Lei X L, Yu X S, Chi J N, et al. Brain tumor segmentation in MR images using a sparse constrained level set algorithm[J]. *Expert Systems With Applications*, 2021, 168: 114262.
- [6] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39(4): 640-651.
- [7] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. *Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science*. Cham: Springer, 2015, 9351: 234-241.
- [8] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-net: learning dense volumetric segmentation from sparse annotation[M]//Ourselin S, Joskowicz L, Sabuncu M R, et al. *Medical image computing and computer-assisted intervention-MICCAI 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9901: 424-432.
- [9] Yi S L, Wang T W, Yang X L, et al. Improved lung field segmentation method of u-net network[J]. *Laser & Optoelectronics Progress*, 2022, 59(02): 0210010.
易三莉, 王天伟, 杨雪莲, 余芙蓉, 贺建峰. 改进 u-net 网络的肺野分割方法[J]. *激光与光电子学进展*, 2022, 59(02): 0210010.
- [10] Feng B W, Lü X Q, Gu Y, et al. Three-dimensional parallel convolution neural network brain tumor segmentation based on dilated convolution[J]. *Laser & Optoelectronics Progress*, 2020, 57(14): 141009.
冯博文, 吕晓琪, 谷宇, 等. 基于空洞卷积的三维并行卷积神经网络脑肿瘤分割[J]. *激光与光电子学进展*, 2020, 57(14): 141009.
- [11] Zhang W X, Zhu Z C, Zhang Y H, et al. Cell image segmentation method based on residual block and attention mechanism[J]. *Acta Optica Sinica*, 2020, 40(17): 1710001.
张文秀, 朱振才, 张永合, 等. 基于残差块和注意力机制的细胞图像分割方法[J]. *光学学报*, 2020, 40(17): 1710001.
- [12] Li D X, Zhang Z. Improved U-net segmentation algorithm for the retinal blood vessel images[J]. *Acta Optica Sinica*, 2020, 40(10): 1010001.
李大湘, 张振. 基于改进 U-Net 视网膜血管图像分割算法[J]. *光学学报*, 2020, 40(10): 1010001.
- [13] Raudaschl P F, Zaffino P, Sharp G C, et al. Evaluation of segmentation methods on head and neck CT: auto-segmentation challenge 2015[J]. *Medical Physics*, 2017, 44(5): 2020-2036.
- [14] Ren X, Xiang L, Nie D, et al. Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images[J]. *Medical Physics*, 2018, 45(5): 2063-2075.
- [15] 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.
- [16] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327.
- [17] Zhu W, Huang Y, Zeng L, et al. AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy[J]. *Medical Physics*, 2019, 46(2): 576-589.
- [18] Tappeiner E, Pröll S, Hönig M, et al. Multi-organ segmentation of the head and neck area: an efficient

- hierarchical neural networks approach[J]. *International Journal of Computer Assisted Radiology and Surgery*, 2019, 14(5): 745-754.
- [19] Gao Y H, Huang R, Chen M, et al. FocusNet: imbalanced large and small organ segmentation with an end-to-end deep neural network for head and neck CT images[EB/OL]. (2019-07-28) [2021-06-17]. <https://arxiv.org/abs/1907.12056>.
- [20] Tappeiner E, Pröll S, Fritscher K, et al. Training of head and neck segmentation networks with shape prior on small datasets[J]. *International Journal of Computer Assisted Radiology and Surgery*, 2020, 15(9): 1417-1425.
- [21] Gao Y H, Huang R, Yang Y W, et al. FocusNetv2: imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images[J]. *Medical Image Analysis*, 2021, 67: 101831.
- [22] Dai X J, Lei Y, Wang T H, et al. Head-and-neck organs-at-risk auto-delineation using dual pyramid networks for CBCT-guided adaptive radiotherapy[J]. *Physics in Medicine & Biology*, 2021, 66(4): 045021.
- [23] Zhong Z L, Lin Z Q, Bidart R, et al. Squeeze-and-attention networks for semantic segmentation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13062-13071.
- [24] Salehi S S M, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks[M]// Wang Q, Shi Y H, Suk H I. *Machine learning in medical imaging*. Lecture notes in computer science. Cham: Springer, 2017, 10541: 379-387.
- [25] Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository[J]. *Journal of Digital Imaging*, 2013, 26(6): 1045-1057.
- [26] Vallières M, Kay-Rivest E, Perrin L J, et al. Radiomics strategies for risk assessment of tumor failure in head-and-neck cancer[J]. *Scientific Reports*, 2017, 7(1): 10117.