

基于深度学习的肢体骨肉瘤CT医学图像分割

赵林林¹, 王潜¹, 王军², 唐子硕¹, 刘雨¹, 樊卓明¹, 陈继民¹

1. 北京工业大学材料与制造学部激光工程研究院, 北京 100124; 2. 北京大学人民医院骨与软组织肿瘤治疗中心, 北京 100044

【摘要】针对骨肉瘤CT图像自动化分割任务, 本文制作了骨肉瘤CT图像数据集Osteosarcoma, 并提出具有双特征提取结构Double-CNN的D-TransUNet模型。D-TransUNet模型在深度学习TransUNet分割模型的基础上新增特征提取结构。用3×3卷积核进行特征提取, 同时在特征通道上进行缩减和拼接, 最后将双特征提取结构提取的图像信息进行融合。模型整体提取的原始图像信息更加丰富, 进一步提升了分割精度。

【关键词】骨肉瘤; 医学图像分割; 深度学习; TransUNet模型; D-TransUNet模型

【中图分类号】R318; R445.3

【文献标志码】A

【文章编号】1005-202X(2023)10-1204-08

Deep learning based approach for extremity osteosarcoma segmentation in CT image

ZHAO Linlin¹, WANG Qian¹, WANG Jun², TANG Zishuo¹, LIU Yu¹, FAN Zhuoming¹, CHEN Jimin¹

1. Institute of Laser Engineering, Department of Materials and Manufacturing, Beijing University of Technology, Beijing 100124, China; 2. Bone and Soft Tissue Tumor Treatment Center, Peking University People's Hospital, Beijing 100044, China

Abstract: For the task of automatic segmentation of osteosarcoma in CT images, a dataset of CT images (Osteosarcoma) is established, and a D-TransUNet model with Double-CNN feature extraction structure based on TransUNet is proposed. The 3×3 convolution kernel is used for feature extraction, and the feature channel is reduced and spliced. Finally, the image information extracted by the Double-CNN feature extraction structure is fused. The original image information extracted by the proposed model is more abundant, which further improves the segmentation accuracy.

Keywords: osteosarcoma; medical image segmentation; deep learning; TransUNet model; D-TransUNet model

前言

骨肉瘤是最常见的儿童及青少年骨原发恶性肿瘤, 以形成肿瘤样骨基质为病理学特点, 大部分肿瘤位于膝关节周围, 以股骨远端和胫骨近端最为常见^[1-5]。骨肉瘤因具有高度侵袭性和转移性特点, 患者预后较差, 严重危害青少年的生命健康。由于骨肉瘤发病率低, 广大基层医生对此疾病的诊断治疗认识不足, 影像阅片能力和水平参差不齐, 严重制约

了基层医生对该疾病的诊断能力。

医学图像分割技术能帮助基层医生快速确定患者病灶区, 精确诊断病情, 在图像处理和临床应用中具有重要意义。骨肉瘤的影像学诊断是临床医生重要的诊断手段, 而骨肉瘤医学图像的准确分割在前期诊断、手术制定和术后检测中至关重要。目前针对骨肉瘤CT图像分割任务主要以人工勾画为主。但这种人工分割方式耗时长, 分割结果个体差异大且复现性差, 因此极难推广, 极大阻碍了基层医生对该疾病的诊疗能力。

随着计算机图像处理技术的发展, 图像的自动化处理逐渐被应用到医学图像分割中。当前医学图像的自动化分割主要分为传统的医学图像分割算法和基于深度学习的图像分割算法两类。传统分割方法主要有基于阈值的分割算法、基于边缘检测的分割算法、基于聚类的分割算法、基于可变形模型的分割算法等。传统分割算法因适用于分割场景简单的情形, 使其难以大规模应用到实际医学图像分割任务中。基于深度学习的图像分割算法是一种以人工

【收稿日期】2023-05-12

【基金项目】国家自然科学基金(82272947); 北京大学人民医院研究与发展基金(RDG2021-02, RDL2022-14和RZ2023-02); 北京大学临床医学+X青年专项, 中央高校基本科研业务费(PKU2023LCXQ016)

【作者简介】赵林林, 硕士, 研究方向: 医学图像处理, E-mail: zll13022273902@163.com

【通信作者】王潜, 博士, 助理研究员, 研究方向: 激光增材制造、医学图像处理, E-mail: wangqian@bjut.edu.cn; 王军, 博士, 主治医师, 研究方向: 骨与软组织肿瘤的基础和临床研究, E-mail: pkuwj2010@163.com

神经网络为架构,对图像进行表征学习的算法^[6-11]。同时随着卷积神经网络(Convolution Neural Network, CNN)在计算机视觉领域的不断发展^[12],特别是基于端到端的网络分割模型U-Net的提出,使得医学图像分割精度进一步提升^[13]。利用深度学习自动分割在医学图像分割任务中成果斐然,并迅速用于放射治疗领域^[14]。但目前针对肢体的骨肉瘤CT医学图像分割任务仍然极具挑战性。例如:骨肉瘤在组织学中类型多样;肿瘤形态和大小不一;病灶中成骨位置随机;肿瘤影像学图像受造影剂影响较大等。同时与其他分割任务不同,基于深度学习的医学图像分割存在着固有的困难。首先是模型训练数据集缺少;以骨肉瘤CT医学图像为例,目前还没有公开的数据集用于深度学习模型训练。其次,医学图像分割任务中存在类不平衡的问题,即图像中的目标区域,如病变、器官等,占据整幅图像的比例很小,导致背景区域在训练阶段占主导地位,降低模型的优化质量^[15]。

自2020年起,随着BERT^[16]、ViT^[17]、MAE^[18]、Swin Transformer^[19]等深度学习模型的提出,自注意

力机制因其能兼顾全集信息而逐渐被引入到计算机视觉中,在医学图像分割方面将自注意力机制和U-Net的结合模型TransUNet应用到医学图像分割任务中并取得了很好的分割效果^[20]。针对以上骨肉瘤CT图像分割任务中存在的问题,本文构建了骨肉瘤CT图像分割数据集Osteosarcoma。同时以TransUNet网络模型为基础,提出含有双特征提取结构的D-TransUNet模型,进一步提升了分割精度。

1 骨肉瘤CT图像数据集 Osteosarcoma

1.1 骨肉瘤CT图像特点

骨肉瘤CT图像为DICOM文件格式,DICOM格式包含信息冗余,首先应进行图像格式转换。在CT扫描图像中,通常包含冠状面、矢状面、轴位3种视图图像。以骨肉瘤CT图像为例,3种视图如图1所示。通过对比3种视图发现,在骨肉瘤CT图像中轴位视图骨肉瘤特征不明显,对骨肉瘤特征提取作用不大,同时还会增加模型提取骨肉瘤信息的难度,会对冠状面和矢状面的体征提取造成干扰,因此选择对肢体骨肉瘤冠状面和矢状面图像进行标注。

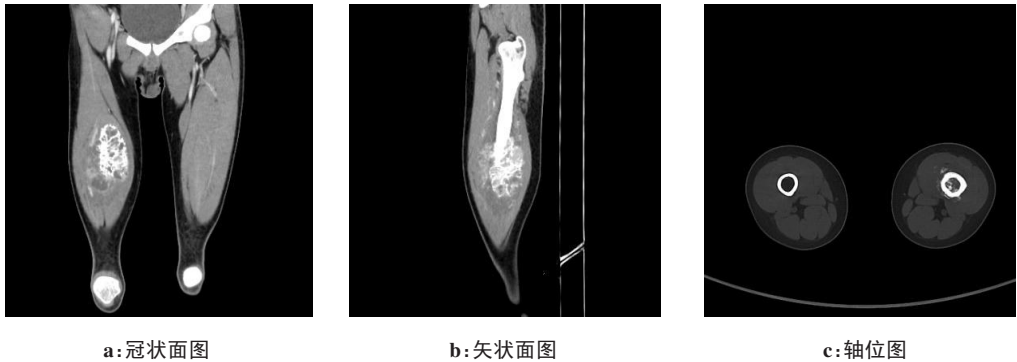


图1 骨肉瘤CT图像3种视图
Figure 1 Three views of the CT image of osteosarcoma

1.2 标注骨肉瘤CT图像数据集 Osteosarcoma

骨肉瘤CT图像数据集以开源标注模型labelMe作为标注工具^[21],labelMe模型的标注界面如图2所示。

标注过程主要分为3个步骤:第一步,首先由北京大学人民医院骨与软组织肿瘤治疗中心临床医生对骨肉瘤CT图像进行标注示范,对骨肉瘤组织形态知识进行讲解;第二步,课题组成员学习骨肉瘤相关医学知识进行试标注,之后针对标注过程存在的疑难点进行汇总,与医院医生进行沟通和开会研讨;第三步,课题组成员对骨肉瘤CT图像数据进行标注,医生对标注结果进行查验和优化完成骨肉瘤CT图

像数据标注任务。本次标注骨肉瘤CT图像数据集以Osteosarcoma命名,共标注1207张图像,分别以I+序号命名(I00001~I01207)。数据集分为训练集(train)1192张、测试集(val)15张。部分患者输入图像(Input images)和真实标签(Ground truth)展示如图3所示。

2 D-TransUNet网络分割模型

2.1 TransUNet网络结构

在基于卷积网络的图像分割任务中,U-Net网络模型因其结构简单、分割精确度高、能将像素级特征与语义级特征相融合实现像素级别语义分割的特

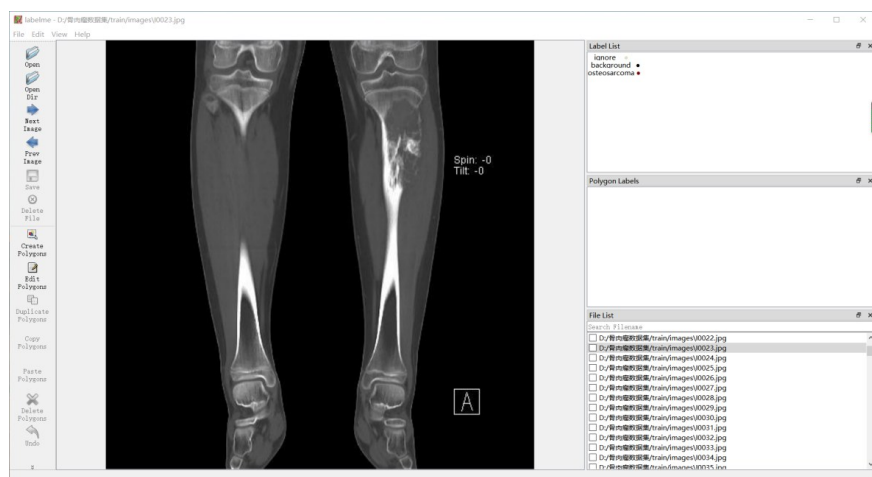


图 2 labelMe 图像标注界面

Figure 2 labelMe image annotation interface

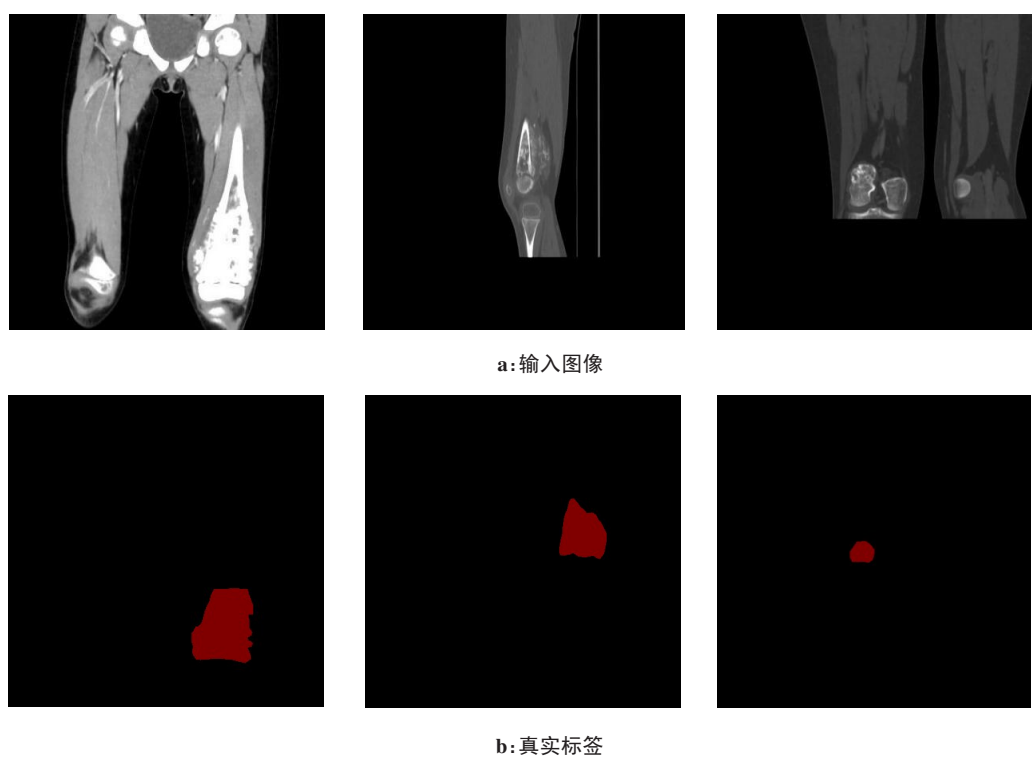


图 3 部分患者输入图像和真实标签

Figure 3 Input images and ground truth

第 1~3 列分别为 3 位不同患者的图像

点,被广泛应用到医学图像分割任务中。但受卷积核感受野的固有局限性,CNN在处理图像细节的同时无法兼顾到全局信息,而基于自注意力机制的Transformer结构在处理图像时将全局信息作为输入能很好地兼顾全局信息^[6,22]。但单纯使用Transformer又会使得低层次细节不足,同时可能导致定位不准确。因此将Transformer与U-Net网络结合起来的TransUNet网络模型既有全局信息的输入又能兼顾到细节信息。TransUNet网络模型整体结

构如图 4 所示。TransUNet网络是一个整体呈“U”形的 Encoder-Decoder 网络模型。“U”形左半边为 Encoder 部分,采用 CNN+Transformers 混合模型,CNN首先用作特征提取器,输入生成特征映射,之后将CNN提取到的特征图编码作为提取全局上下文信息的输入序列输入到Transformer Layer层。Decoder部分对编码后的特征进行上采样,同时与CNN提取到的浅层特征图进行特征拼接以实现精准定位防止浅层特征细节丢失。

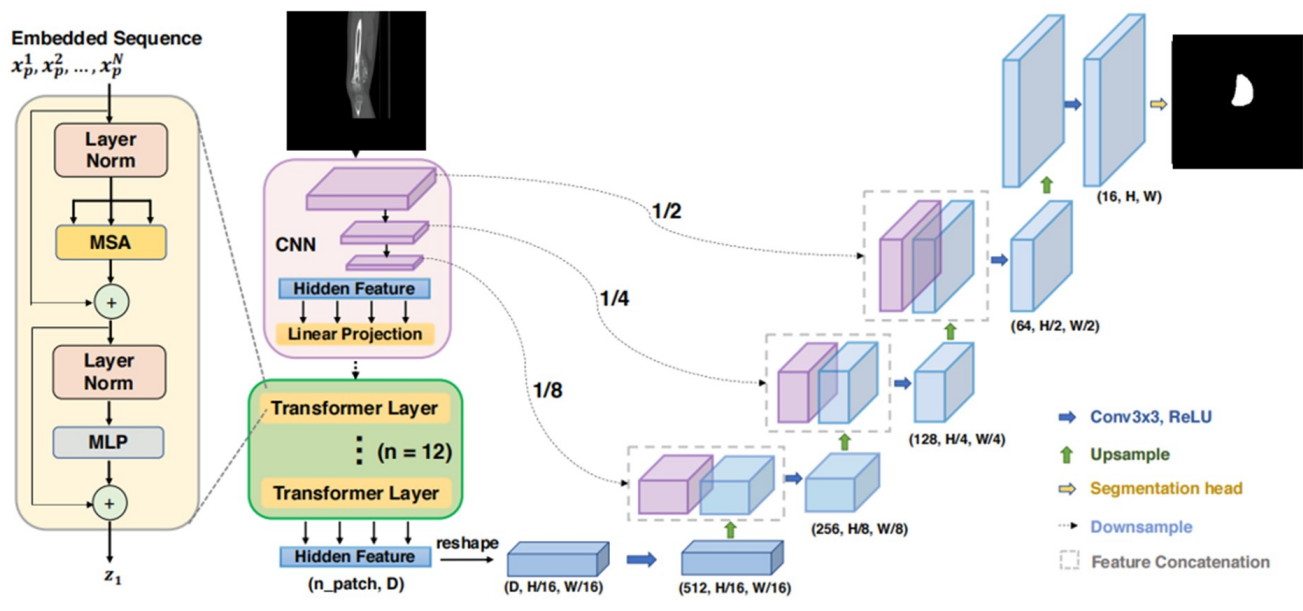


图4 TransUNet网络模型的整体结构
Figure 4 Overall structure of TransUNet network model

2.2 TransUNet网络改进模型 D-TransUNet模型

TransUNet网络模型虽然将自注意力机制与CNN结合起来,但如果在卷积特征提取阶段对图像特征提取不充分,就会造成后续分割结果不准。这使得模型在骨肉瘤CT图像实际分割任务中出现一

些问题,如小目标难分割且分割边界难确定等。TransUNet网络模型实际分割结果如图5所示。针对TransUNet模型在骨肉瘤分割任务中存在的问题,本文设计了D-TransUNet模型。该模型在TransUNet模型中CNN模块设计双特征提取结构形成

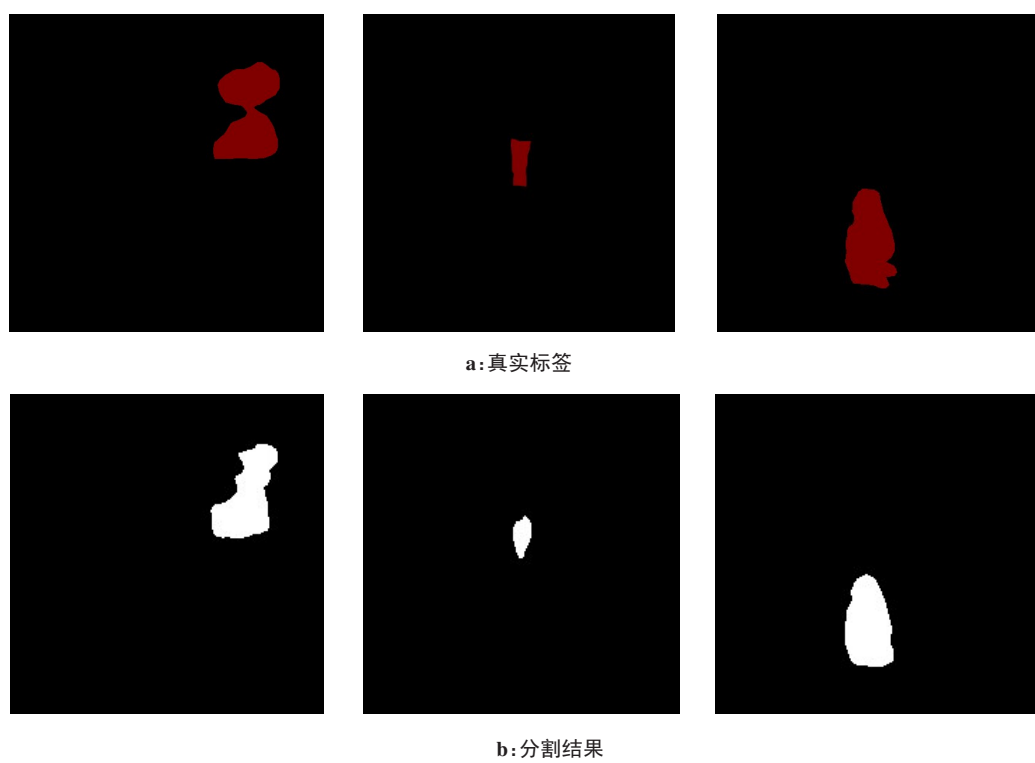


图5 部分真实标签与分割结果
Figure 5 Ground truth and segmentation results

第1~3列分别为3位不同患者的图像

Double-CNN 结构,用来提取特征信息;使得输入 Transformer Layer 层的特征图包含的特征信息更加丰富,从而提升分割效果,因此将其命名为 D-TransUNet 模型。模型的 Double-CNN 结构如图 6 所示。

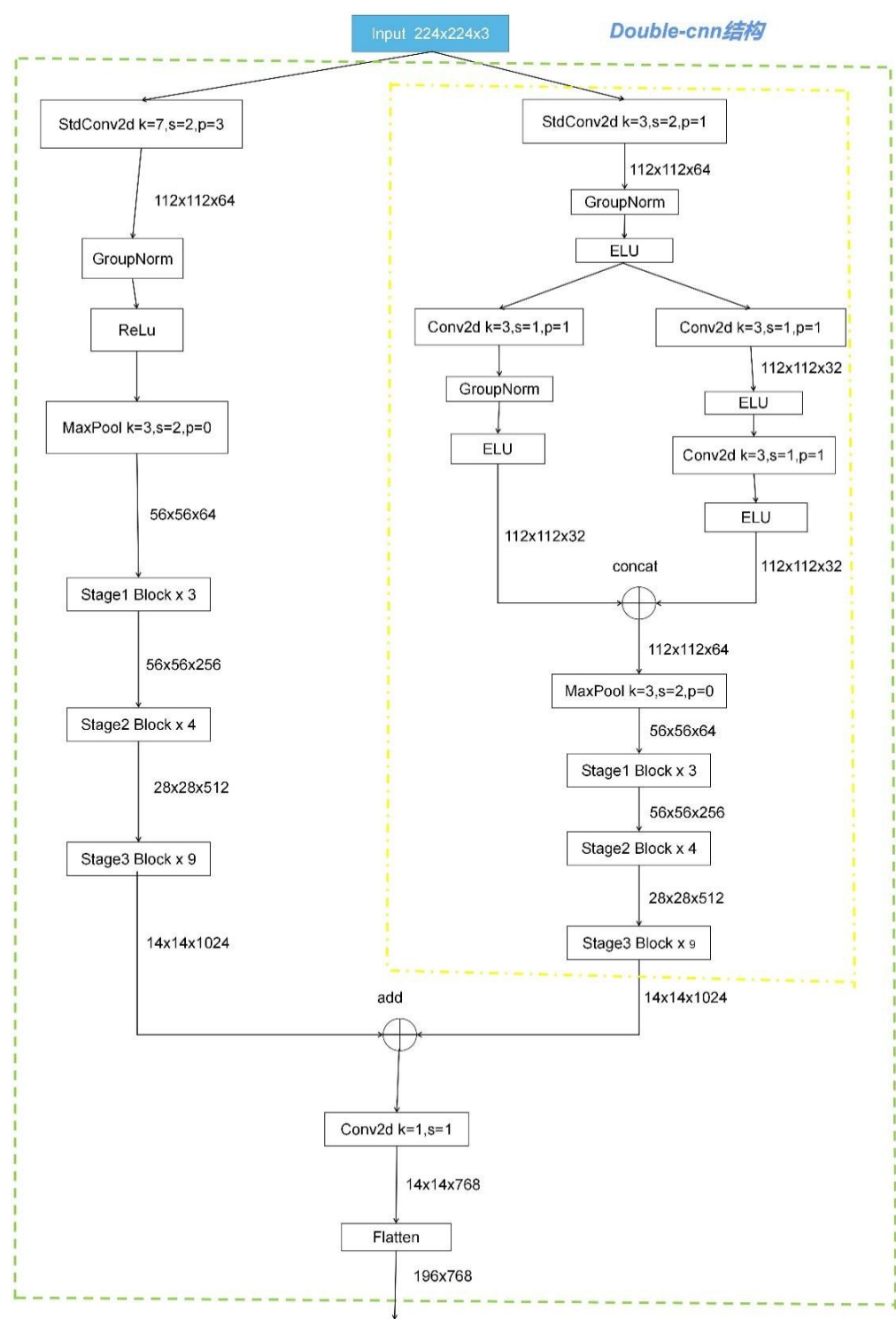


图 6 Double-CNN 结构图

Figure 6 Double-CNN structure diagram

D-TransUNet 模型的 Double-CNN 结构就是在 TransUNet 原有特征提取结构 CNN 的基础上新增加了一个特征提取分支,构成双特征提取结构。新增加的特征提取结构为 ResNet50 残差网络,其具体结

构如下:①新特征提取结构以 3×3 的卷积来代替 7×7 卷积;这样设计的原因是模型引入新的特征提取结构后会极大增加计算量,而以 3×3 的小卷积核来代替 7×7 卷积能降低模型的计算量,同时引入不同大小的

卷积核更能使网络提取不同特征信息。②网络经GN归一化层和激活层后将特征通道减半,再次分为2个不同子分支,将其中1条子分支多进行一次卷积核为3、步长为1、padding为1的卷积操作,之后将2条子分支进行拼接处理,这样做的目的是为了保持特征图的尺寸不发生变化,有利于在后续网络中进行相近特征图的融合,防止产生语义鸿沟,同时又增加了模型不同特征信息提取的能力。最后将双特征提取结构提取的不同特征信息在Transformer前一层进行相加融合。使得D-TransUNet模型在特征提取阶段能提取到更加丰富的图像信息。在新特征提取通道中使用ELU激活函数代替ReLU激活函数,使得模型得到进一步优化^[23-24]。D-TransUNet模型对原始图像信息提取更加丰富,减少了原始特征信息的缺失,从而提高模型的分割精度。

3 实验结果与分析

3.1 分割评价指标

医学图像分割算法通常需要定量评估优劣,对算法的评估方法一般是对分割结果与真实标签进行比较。本文延续采用TransUNet的评价指标对分割结果进行评价,分别为Dice相似性系数(Dice Similarity Coefficient, DSC)、豪斯多夫距离(Hausdorff Distance, HD)。

DSC是医学图像分割常见的指标之一,反映了模型分割结果与真实标签之间的近似程度,取值范

围在[0,1]之间:

$$DSC(X,Y)=\frac{2|X\cap Y|}{|X|+|Y|}\tag{1}$$

其中,|X|和|Y|分别代表分割结果和真实标签中像素点个数,数值越大表示分割值与真实标签越相似,表示分割结果越准确。

HD代表的是模型分割结果与真实标签上点集之间的最大不匹配程度。HD值越小,代表算法分割结果越接近真实值,表示分割结果越准确。95%HD就是HD最后的值乘以95%,目的是为了消除离群值的一个非常小的子集的影响。

$$HD(X_s,Y_s)=\max(h(X_s,Y_s),h(Y_s,X_s))\tag{2}$$

$$h(X_s,Y_s)=\max_{x_i\in X_s}\min_{y_j\in Y_s}\|x_i-y_j\|\tag{3}$$

$$h(Y_s,X_s)=\max_{y_j\in Y_s}\min_{x_i\in X_s}\|y_j-x_i\|\tag{4}$$

其中, X_s 、 Y_s 分别代表分割结果和真实标签的像素点集, x_i 、 y_j 分别代表了分割结果和真实标签中某一像素点。

3.2 实验结果

为验证双特征提取结构Double-CNN在医学图像分割任务上的有效性,在定量评估的角度下,使用TransUNet模型和D-TransUNet模型在Synapse多器官分割数据集和Osteosarcoma数据集上进行比较。实验中网络输入CT图像尺寸为224×224,优化器为SGD优化器,初始学习率设置为0.02,Batch-size设置为48,迭代次数为300。模型分割性能如表1所示。

表1 优化算法在不同数据集上的定量比较
Table 1 Quantitative comparison between optimization algorithms on different datasets

网络结构	Synapse		Osteosarcoma	
	平均DSC/%	平均95%HD/mm	平均DSC/%	平均95%HD/mm
TransUNet	76.820	32.017	67.335	10.151
D-TransUNet	77.352	27.041	79.020	6.879

两种网络模型在Osteosarcoma数据集分割预测结果如图7所示,其中图7e、f中红色曲线圈出部分为模型过分割部分。

由表1 Osteosarcoma数据集测试结果得出D-TransUNet模型较TransUNet模型在平均DSC和95%HD指标上分别提升了11.7%和3.27 mm;从图7e、f模型预测在原图复现结果,可以直观看出D-TransUNet网络预测结果过分割区域要少于TransUNet网络;充分验证具有双特征提取结构Double-CNN的D-TransUNet网络模型更适合用于处理骨肉瘤CT图像分割任务。同时,在多标签医学公

开数据集Synapse上,D-TransUNet模型测试分割指标也优于TransUNet网络,进一步说明具有Double-CNN网络结构的D-TransUNet网络模型在医学分割任务中具有较好的适应能力和优越的分割性能。

4 结论

本文针对骨肉瘤CT图像自动分割任务,标注了骨肉瘤CT图像Osteosarcoma数据集;为进一步提高分割精度,以深度学习TransUNet分割模型为基础,设计了具有双特征提取结构Double-CNN的D-TransUNet模型。同时在医学图像数据集上进行

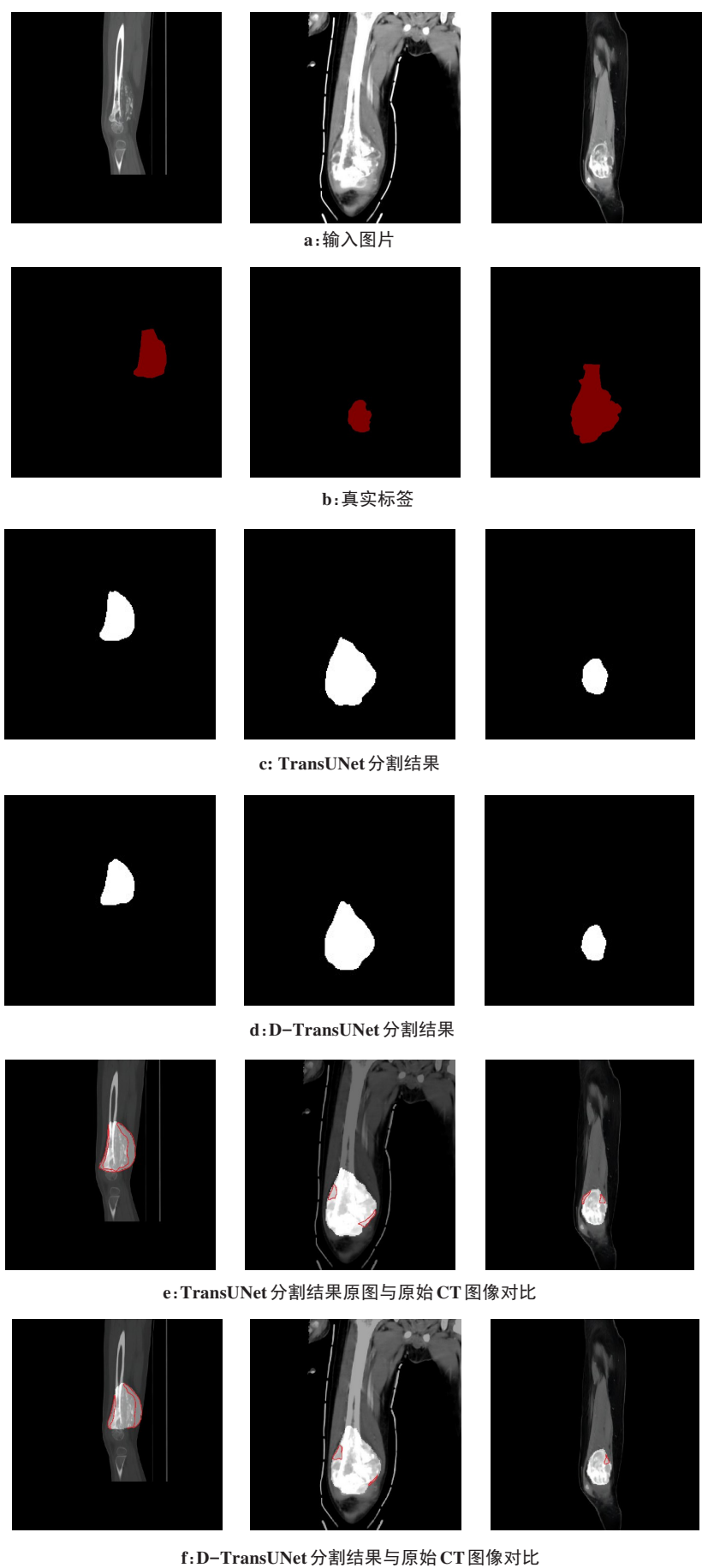


图 7 TransUNet 和 D-TransUNet 模型分割结果对比

Figure 7 Comparison of segmentation results between TransUNet and D-TransUNet models

第 1~3 列分别为 3 位不同患者的图像

了自动分割实验,充分验证了D-TransUNet模型在医学图像分割任务中的优秀分割性能。但在骨肉瘤分割任务中,基于深度学习的图像分割精度仍与医生的手动分割有一定的差距,今后如何进一步提升分割精度并将算法推行到实际应用仍是一个研究重点。

【参考文献】

- [1] 邓磊,邹玉坚,曾雪伟,等. 颅面骨骨肉瘤的CT和MRI诊断及鉴别[J]. 中国医学计算机成像杂志, 2019, 25(1): 53-57.
Deng L, Zou YJ, Zeng XW, et al. Diagnosis and differential diagnosis of craniofacial osteosarcoma by CT and MRI[J]. Chinese Computed Medical Imaging, 2019, 25(1): 53-57.
- [2] 郭卫,牛晓辉,肖建如,等. 骨肉瘤临床循证诊疗指南[J]. 中华骨与关节外科杂志, 2018, 11(4): 288-301.
Guo W, Niu XH, Xiao JR, et al. Guideline for clinical evidence-based diagnosis and treatment of osteosarcoma[J]. Chinese Journal of Bone and Joint Surgery, 2018, 11(4): 288-301.
- [3] Klein MJ, Siegal GP. Osteosarcoma: anatomic and histologic variants[J]. Am J Clin Pathol, 2006, 125(4): 555-581.
- [4] 倪明. 2021版美国国家综合癌症网络(NCCN)《骨肉瘤临床实践指南》更新与解读[J]. 中国修复重建外科杂志, 2021, 35(9): 1186-1191.
Ni M. Update and interpretation of 2021 National Comprehensive Cancer Network (NCCN)"Clinical Practice Guidelines for Bone Tumors"[J]. Chinese Journal of Reparative and Reconstructive Surgery, 2021, 35(9): 1186-1191.
- [5] 吴蔚,景豆豆,曹理,等. 骨肉瘤免疫治疗的现状和前景[J]. 肿瘤防治研究, 2022, 49(7): 721-726.
Wu W, Jing DD, Cao L, et al. Current status and prospects of immunotherapy for osteosarcoma[J]. Cancer Research on Prevention and Treatment, 2022, 49(7): 721-726.
- [6] Deng L. Deep learning: methods and applications[M]. Now Publishers Inc., 2014.
- [7] Bengio Y. Learning deep architectures for AI[J]. Foundations Trends in Machine Learning, 2009, 2(1): 1-127.
- [8] Bengio Y, Courville A, Vincent P, et al. Representation learning: a review and new perspectives[J]. IEEE Trans Pattern Anal Mach Intell, 2013, 35(8): 1798-1828.
- [9] Schmidhuber J. Deep learning in neural networks: an overview[J]. Neural Netw, 2015, 61: 85-117.
- [10] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436.
- [11] 黄天怡,冉小炜,郭增,等. 基于深度学习的轴突三维图像分割与重构[J]. 神经解剖学杂志, 2020, 36(3): 261-269.
Huang TY, Rang XW, Guo Z, et al. Deep-learning based three-dimensional axon image segmentation and reconstruction[J]. Chinese Journal of Neuroanatomy, 2020, 36(3): 261-269.
- [12] Dhruv P, Naskar S. Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): a review[M]. Machine Learning and Information Processing, 2020: 367-381.
- [13] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, 2015.
- [14] 吴青南,王运来,全红,等. 基于有限训练样本的融合网络模型用于盆腔危及器官自动分割的研究[J]. 生物医学工程学杂志, 2020, 37(2): 311-316.
Wu QN, Wang YL, Quan H, et al. A fusion network model based on limited training samples for the automatic segmentation of pelvic endangered organs[J]. Journal of Biomedical Engineering, 2020, 37(2): 311-316.
- [15] Bria A, Marrocco C, Tortorella F. Addressing class imbalance in deep learning for small lesion detection on medical images[J]. Comput Biol Med, 2020, 120: 103735.
- [16] Devlin J, Chang MW, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv: 1810.04805, 2018.
- [17] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[J]. arXiv preprint arXiv: 2010.11929v1, 2020.
- [18] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[J]. arXiv preprint arXiv: 2111.06377, 2021.
- [19] Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [20] Chen J, Lu Y, Yu Q, et al. TransUNet: transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv: 2102.04306, 2021.
- [21] Torralba A, Russell BC, Yuen J. LabelMe: online image annotation and applications[J]. Proc IEEE, 2010, 98(8): 1467-1484.
- [22] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv preprint arXiv: 1706.03762v3, 2017.
- [23] Djork-Arné C, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs)[C]//International Conference on Learning Representations, 2016.
- [24] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks[J]. Adv Neural Inf Process Syst, 2012, 25(2): 1097-1105.

(编辑:薛泽玲)