

## 基于多尺度边缘分割与混合注意力机制的脊柱CT图像分割

刘晶<sup>1</sup>, 徐皓<sup>1</sup>, 崔欣欣<sup>1</sup>, 田振宇<sup>1</sup>, 杨建兰<sup>2</sup>

1. 甘肃中医药大学信息工程学院, 甘肃 兰州 730000; 2. 泉州市正骨医院, 福建 泉州 362019

**【摘要】**脊柱疾病的前期主要通过计算机断层扫描技术进行筛查与初步判断。为解决脊柱CT图像目前存在的椎骨结构复杂、分割精度不足等问题,提出一种基于3D U-Net框架的脊柱CT图像改进分割网络,通过融合SE残差单元、椎骨边缘分割模型与改进混合通道-空间注意力机制,在VerSe 19、VerSe 20与CTSpine1K脊柱数据集上进行分割训练与测试。多次测试实验结果表明,本文模型在保证分割精度和分割效率有效提高的同时具有较好的泛化性与鲁棒性,在Dice相似系数、豪斯多夫距离与平均表面距离上相较于其他先进网络分割精度更高。本文模型在现有脊柱分割的网络中具有更强的分割性能,可为放射科医生提供有效临床信息。

**【关键词】**脊柱分割;3D U-Net;椎骨边缘分割;混合注意力机制

**【中图分类号】**R318;TP391.4

**【文献标志码】**A

**【文章编号】**1005-202X(2024)04-0463-09

## Spine CT image segmentation based on multi-scale boundary segmentation and hybrid attention mechanism

LIU Jing<sup>1</sup>, XU Hao<sup>1</sup>, CUI Xinxin<sup>1</sup>, TIAN Zhenyu<sup>1</sup>, YANG Jianlan<sup>2</sup>

1. School of Information Engineering, Gansu University of Chinese Medicine, Lanzhou 730000, China; 2. Quanzhou Orthopedic-Traumatological Hospital, Quanzhou 362019, China

**Abstract:** The early diagnosis of spinal diseases is mainly screened and initially diagnosed through computed tomography (CT). In view of the complex structure of vertebral bones and low segmentation accuracy in spinal CT images, a spinal CT image segmentation network based on 3D U-Net framework is proposed. The network which integrates squeeze-and-excitation residual module, vertebral boundary segmentation model, and improved hybrid channel-spatial attention mechanism is trained and tested on VerSe 19, VerSe 20, and CTSpine1K spinal datasets. Multiple experiments indicate that the model can effectively improve segmentation accuracy and efficiency while demonstrating good generalization and robustness. Compared with other advanced network models, the proposed network achieves higher segmentation accuracy in terms of Dice similarity coefficient, Hausdorff distance, and average symmetric surface distance. The proposed model exhibits superior segmentation performance among the existing spinal segmentation networks, providing radiologists with valuable clinical information.

**Keywords:** spine segmentation; 3D U-Net; vertebral boundary segmentation; hybrid attention mechanism

### 前言

脊柱是人体骨骼系统的重要组成部分,作为人体的支撑结构,承担着保护神经系统、支持身体重量、维持身体平衡等多种重要功能,是人体的中轴骨骼,由颈椎(C1~C7)、胸椎(T1~T12)、腰椎(L1~L6)、骶椎和尾椎组成。随着医学影像组学和计算机辅助

诊断系统(Computer Aided Diagnosis, CAD)的发展,脊柱分割在许多脊柱类疾病初步诊断中发挥着重要作用。为了辅助放射科诊断医师,实现计算机断层扫描(Computed Tomography, CT)脊柱图像的快速、稳定和准确的分割,为临床病理诊断、制定手术计划和术后评估提供有效信息,需要解决目前存在CT数据集小、分割模型训练速度慢等脊柱分割问题,针对目前的脊柱CT分割存在的难题需要进一步深入研究。

脊柱是一个复杂结构,包含多个椎体、椎间盘、椎弓根等部分,而CT成像技术可以清晰显示脊柱的骨骼结构。但脊柱周围肌肉神经等组织显示模糊,导致CT图像中脊柱骨骼结构与周围软组织对比度

**【收稿日期】**2023-12-08

**【作者简介】**刘晶,硕士研究生,研究方向:医学图像处理, E-mail: 2442204612@qq.com

**【通信作者】**杨建兰,硕士生导师,副教授,研究方向:医学影像识别与应用, E-mail: FJYJL@gszy.edu.cn

较低,在CT图像中会出现相似灰度特征,因此分割模型难以精准区分不同结构。针对CT图像中骨骼与周围软组织灰度值相似的问题,目前的解决方法有:改进U-Net模型的网络结构,通过U-Net编码器和解码器之间的跳跃连接实现图像的精细分割<sup>[1]</sup>;使用深度卷积神经网络逐层提取图像特征,在分割脊柱CT图像时通过捕获不同尺度的特征来提高分割效果<sup>[2]</sup>;使用金字塔结构或多分辨率输入等多尺度策略来提取不同层次的结构信息<sup>[3-4]</sup>。由于医学图像涉及患者隐私问题,限制了数据的共享和使用,导致很难获取大量标注的脊柱CT图像数据,针对数据量小的问题,目前的解决办法是迁移学习和旋转裁剪图像<sup>[5]</sup>。不同的CT设备扫描出的脊柱CT图像分辨率不同,不同分辨率的脊柱CT图像需要经过不同的归一化处理<sup>[6]</sup>,以确保脊柱分割模型在训练时能够稳定学习图像特征,直接将不同分辨率的脊柱CT图像输入到模型中,会导致模型训练时出现错位或输出失真,所以需要脊柱CT图像进行重采样预处理<sup>[7]</sup>。

当前主流的基于数据驱动的脊柱分割模型为有监督深度学习模型。例如,Wu等<sup>[8]</sup>提出一种基于二维(2D)视觉混合投影图像融合包络(LVLS-HVPFE)的三维(3D)腰椎定位与分割方法,用于腰椎CT扫描的椎体自动定位和分割。Nazir等<sup>[9]</sup>在2D U-Net的基础上,提出了嵌入式聚类切片U-Net(ECSU-Net)用于椎间盘的分割和分类。相对于部分脊柱的分割,全脊柱的分割更具有临床应用价值。Saeed等<sup>[10]</sup>提出一种名为3D MRU-Net的新移动残差U-Net椎骨分割网络,该分割网络基于编码器-解码器结构,使用带有残差模块的MobileNetv2作为编码器并进行特征提取,适用于全脊柱CT图像和部分脊柱CT图像的实例分割。Xu等<sup>[11]</sup>提出一种基于残差U-Net网络提取特征信息的网络和视觉Transformer结构相结合的脊柱分割网络,用于快速高效地自动分割多节椎骨。Meng等<sup>[12]</sup>提出解剖一致性周期循环的分割网络,运用解剖一致性的原则循环对任意视野CT图像的椎骨进行定位、分割和识别。该方法结合深度学习和脊椎的统计一致性先验知识,提高了分割模型在移行椎骨和病理病例情况下的稳健性。You等<sup>[13]</sup>提出基于Transformer和3D U-Net的单级网络-EG-Trans3DUNet分割网络来完成全自动3D脊椎分割,Transform编码器相比于纯U-Net结构能够更好地提取全局语义信息,有助于识别每个脊柱中的椎体数量,特别是第25个椎体,实现了准确分割每一块椎骨。Tao等<sup>[14]</sup>提出了一种基于深度学习的两阶段全自动标记和分割网络Spine-Transformers,第一阶段进行椎体标记,第二阶段进行精细的脊椎分割。

You等<sup>[15]</sup>提出了一种称为VerteFormer的单阶段模型,该模型包括视觉变换模块、边缘检测模块和全局信息提取模块。VerteFormer利用了视觉变换(Vision Transformer, VIT)在挖掘输入数据全局关系方面的优势,有效融合了椎骨的整体和局部特征,实现了每一块椎骨的准确分割。以上脊椎CT图像的分割方法都存在分割椎骨和软组织时边界不准确、脊柱CT图像分辨率不一致导致样本失真和算法参数过多的问题,本文提出一种基于2D U-Net框架的3D U-Net网络<sup>[16]</sup>,并且添加挤压(Squeeze-and-Excitation, SE)残差单元<sup>[17]</sup>、脊柱椎骨多尺度边界分割模块和混合通道-空间注意力机制来解决上述问题。

## 1 相关技术

### 1.1 基于3D U-Net框架的脊柱CT图像分割模型

本文分割模型采用的是基于U-Net框架的3D U-Net网络,由于脊柱CT数据集存在数据规模小、CT设备多样化、图像噪声差异化等问题,导致目前公开的脊柱数据集无法提供高质量数据样本进行大规模训练。而基于2D U-Net框架的3D U-Net设计解码器-编码器对称结构,利用重复卷积-反卷积操作对小规模的脊柱CT图像进行端到端的训练可以取得良好的训练结果。3D U-Net模型在脊柱CT图像的训练过程中,解码器部分通过卷积-堆叠池化从高分辨率的浅层图像提取出脊柱的深层局部特征,利用跳跃连接将下采样的低分辨率深层脊柱特征与上采样稀释的高分辨率图像进行特征融合,有效保留脊柱CT图像的局部浅层信息<sup>[18]</sup>。

虽然3D U-Net网络相较于RNN、Transformer等网络分割精度更高<sup>[19]</sup>,但3D U-Net网络在脊柱CT图像分割过程中存在以下问题:脊柱部分椎骨生物曲线复杂导致基础3D U-Net网络在提取脊柱边界深层特征时分割结果模糊;人体脊柱腰椎(L1~L5)中5块椎骨存在几何形状高度相似性,因此基础3D U-Net网络对于每一块椎骨无法进行一致性分割<sup>[11, 20]</sup>;部分人群存在第25节椎骨(L6)导致基础3D U-Net网络对于L6椎骨无法识别特征进行学习。针对以上问题,本文提出了如图1所示的改进措施,将基础3D U-Net网络融合SE残差单元与混合通道-空间注意力机制提高网络对脊柱CT图像深层特征的分割精度与脊柱定位识别能力<sup>[21-22]</sup>。同时设计脊柱椎骨边缘分割模块<sup>[23]</sup>,对编码器提取出的多尺度特征进行浅层信息融合与椎骨边界曲线分割,将脊柱椎骨边界特征与解码器稀释的脊柱深层局部特征融合实现脊柱各椎骨的边界一致性分割<sup>[24]</sup>。

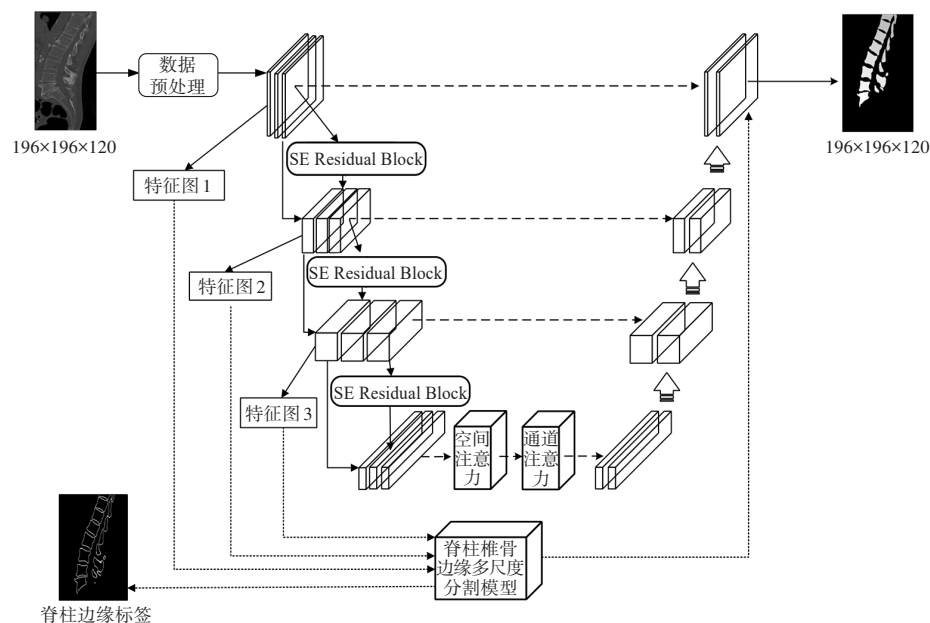


图1 基于3D U-Net的脊柱CT图像分割模型

Figure 1 Spine CT image segmentation model based on 3D U-Net

1.2 SE残差模块

SE残差单元由SE模块与残差结构组成,其结构如图2所示,在卷积层输出特征后分两条路径,一条路径利用残差结构将卷积层输入与输出相乘,脊柱CT图像是由24节椎骨连接构成的3D结构,椎骨前景区域相对于背景区域占比太小,因此通过残差结构提高网络对于脊柱数据集中不同细微局部特征的代表学习能力;另一条路径则通过SE模块动态调整脊柱深层特征的通道权重值,SE模块的原理如式(1)、(2)所示:

$\text{squeeze}(x^{H \times W \times C}) = \text{GlobalAvgPool}(x^{H \times W \times C})$  (1)

$\text{Output} = \text{Sig}(\text{FC}(\text{ReLU}(\text{FC}(f_{\text{squeeze}}(x^{H \times W \times C}))))$  (2)

其中,式(1)表示输入特征经过全局池化操作,式(2)表示经过挤压操作后的特征通过全连接层FC之后再经过ReLU激活函数最后经过FC全连接层与Sigmoid激活函数,对编码器输出的特征图 $x^{H \times W \times C}$ ( $H$ 、 $W$ 代表特征的高度与宽度, $C$ 代表特征通道维度)进行全局平均池化操作,计算每个通道的全局特征信息、压缩通道数量和特征信息。利用全局平均池化操作减少脊柱深层特征的通道信息后,再通过两层全连接结构FCN对深层特征进行通道权重计算与赋值,Sig表示Sigmoid函数激活通道权重并突出原始特征中需要关注的重点信息和抑制其它特征。将SE模块计算出来的各通道权重值分别和原特征图对应通道的2D矩阵相乘,增强后的特征图与原始的残差连接相加,形成最终的输出。

本文将SE残差单元嵌入编码器的卷积层之间,将卷积层提取出的粗糙深层特征进行关键特征粒度

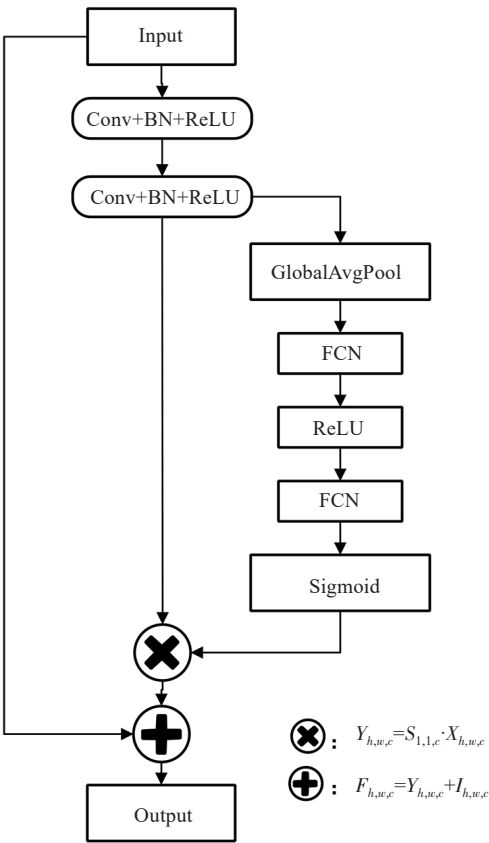


图2 SE残差模块结构

Figure 2 SE residual module

细化,实现脊柱深层局部特征的体素分割、精度细化并获得相对更优的深度特征输入到下一层结构,在分割脊柱椎骨小型突出的几何结构区域时拥有更好表现。



### 1.3 脊柱椎骨多尺度边缘分割模块

为了解决脊柱椎骨边界欠分割的问题,本文提出了基于卷积神经网络(Convolutional Neural Network, CNN)的脊柱椎骨多尺度分割模块。边缘分割模块的结构如图3所示,将编码器上层下采样提取出的包含丰富浅层信息的不同尺度特征图( $I_{L_i}, i=1,2,3$ )作为模块的输入,利用多重上采样反卷积操作对不同尺寸的特征进行尺度标准化,采取 $3 \times 3 \times 3$ 的卷

积核对包含不同尺度的浅层椎骨局部信息进行提取,通过 $1 \times 1 \times 1$ 特殊卷积核控制特征通道的数量,将提取的多维脊柱椎骨边界特征图进行特征通道拼接,拼接后的多维特征再次利用 $1 \times 1 \times 1$ 卷积核进行特征融合。将提取出的脊柱椎骨边界特征与解码器还原的脊柱椎骨局部特征相加,通过约束单个椎骨的边界特征来实现脊柱分割的局部特征与整体信息的深度融合,以确保脊柱中各椎骨的分割一致性。

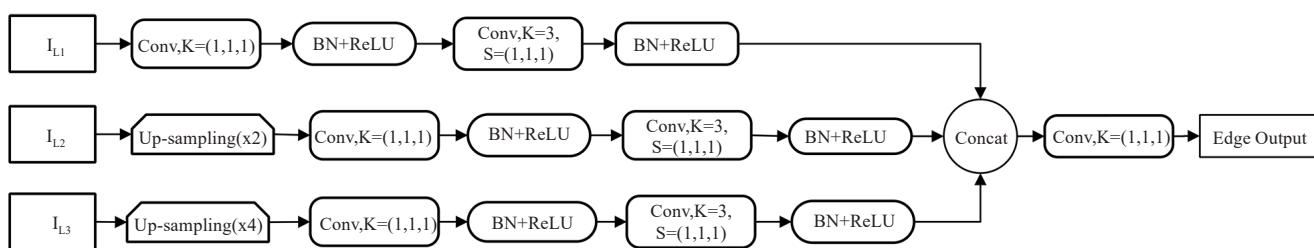


图3 脊柱椎骨边缘分割模块结构图

Figure 3 Spine vertebrae edge segmentation module

### 1.4 混合通道-空间注意力机制

脊柱CT图像横断面中的椎骨横截面积相较于其他背景区域比例太低,导致基于CNN框架的3D U-Net网络存在提取目标特征时学习缓慢、无法快速捕捉前景区域的问题。同时3D U-Net网络随着卷积层数的加深,对脊柱深度特征进行反复降维操作,会导致前景区域损失部分细节信息并且造成椎骨深层特征与周围软组织特征全局对比度低的欠分割问题。对于图像的高频区域特征提取会造成极大影响,为了缓解局部信息丢失与训练缓慢的问题。本文提出对编码器提取出的深层特征通过混合通道-空间注意力结构将低对比度的特征进行权重计算。将网络学习的注意力集中在椎骨特征的小范围区域内,减少对于非特征区域的学习,并进一步增强特征区域像素位置的表示,突出椎骨局部特征像素的位置信息与纹理细节信息。混合通道空间注意力结构由通道注意力模型、空间注意力模型与基于自注意力的Transformer结构组成,如图4、图5所示。

混合通道-空间注意力结构如图4所示,将编码器最深层的特征 $F^{H \times W \times C}$ 先输入到通道注意力模型中,同时将深层特征 $F^{H \times W \times C}$ 输入到最大池化与平均池化结构中,进行空间特征压缩得到 $F_{avg}^C$ 与 $F_{max}^C$ 两个空间背景信息参数,输入多层感知机(Multilayer Perceptron, MLP)计算通道特征注意力图,最后通过Sigmoid函数将得到的通道特征图进行激活得到 $M_c(F)$ 。通道注意力计算过程如公式(3)、(4)所示:

$$M(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (3)$$

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \quad (4)$$

其中, $W_0 \in \mathbb{R}^{C/r \times C}$ 且 $W_1 \in \mathbb{R}^{C \times C/r}$ 。将通道注意力特征输入到空间注意力模型中,空间注意力模型的结构如图4b所示,空间注意力机制集中学习特征图中的像素位置信息,通过最大池化与平均池化的串行操作,将通道注意力特征中的两个空间背景位置信息参数进行融合计算得到 $M_f(F)$ :

$$M_f(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (5)$$

再输入到卷积层进行低维特征映射,利用ReLU函数获得最终加权特征图 $M_s(F)$ :

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S])) \quad (6)$$

利用通道-空间注意力机制将原始深层局部特征去除噪声与无关组织信息等干扰,提高网络对于通道维度的特征信息关注。将加权特征图 $M_f(F)$ 作为 $\chi_i^{LT}$ 输入基于多头自注意力(Multi-headed Self-Attention, MSA)机制的局部Transformer结构,如图5所示,局部Transformer结构由两层MSA与MLP组成,通过局部Transformer结构可以同时考虑脊柱不同局部特征位置信息,并且利用MSA机制对深层特征进行多层次特征抽取解决脊柱椎骨局部特征高度相似的问题。局部Transformer结构的计算流程如公式(7)所示:

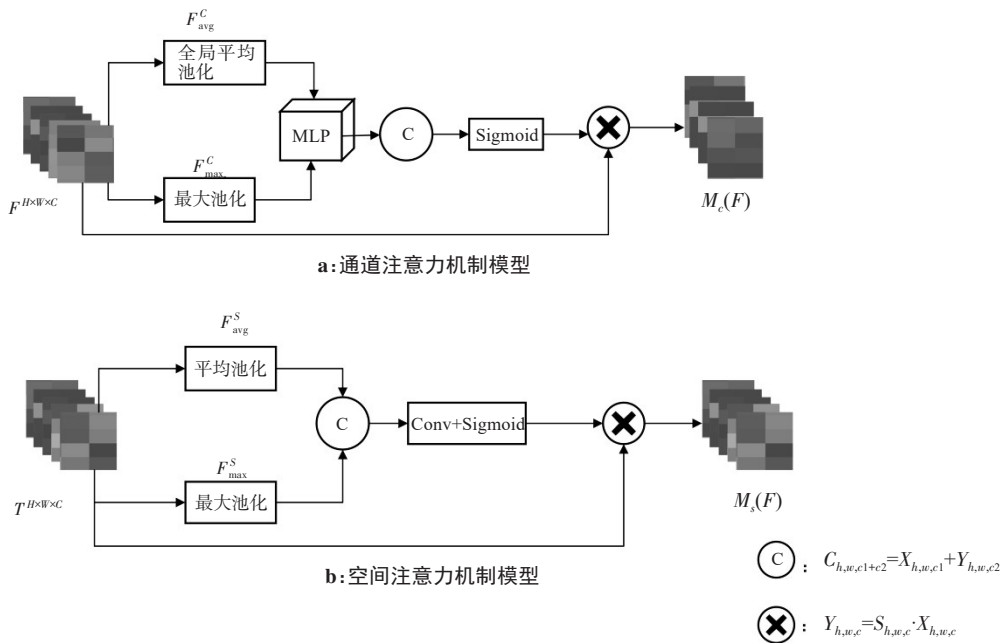


图4 混合通道-空间注意力机制

Figure 4 Hybrid channel-spatial attention mechanism

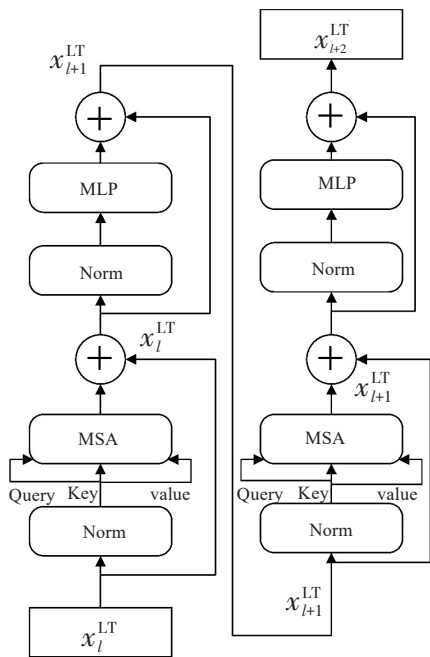


图5 基于自注意力机制的局部Transformer模块

Figure 5 Local Transformer module based on self-attention mechanism

$$\chi_{l+1}^{LT} = \text{MSA}\left(\text{Norm}\left(\chi_l^{LT}\right) + \chi_l^{LT}\right) \left(l = 0, 1, \cdots 11\right) \quad (7)$$

其中,  $l$  代表 Transformer 结构的层数, Norm 表示正归一化, 本文采用的是层归一以确保模型整体的稳定性与有效性。

1.5 评价指标

本文实验使用 Dice 相似系数(DSC)、豪斯多夫距离 (Hausdorff Distance, HD)、平均表面距离

(Average Symmetric Surface Distance, ASSD) 评测指标对脊柱 CT 图像分割结果进行分割精度评估。

DSC 指标是衡量模型预测集合与真实集合间相似度的正向指标, 取值范围为  $[0, 1]$ , 计算原理如式 (8) 所示:

$$\text{DSC}(P, T) = \frac{1}{N} \sum_{i=1}^N \frac{2|P_i \cap T_i|}{|P_i| + |T_i|} \quad (8)$$

其中,  $P$  表示模型预测结果,  $T$  代表专家分割结果,  $i$  表示脊柱椎骨的索引。

HD 评价指标通过计算脊柱分割预测结果与真实椎骨集合之间的空间距离以判断预测值与真实值的差异程度, 本文使用非对称 HD 衡量预测集合与真实集合的最大距离, 计算过程如式 (9) 所示:

$$\text{HD}(P, T) = \frac{1}{N} \sum_{i=1}^N \max \left\{ \sup_{p \in P_i} \inf_{t \in T_i} d(p, t), \sup_{t \in T_i} \inf_{p \in P_i} d(p, t) \right\} \quad (9)$$

其中,  $P_i$  表示索引为  $i$  的椎骨预测分割标签的表面距离集合,  $T_i$  代表真实标签索引为  $i$  的椎骨表面距离集合,  $d(p, t)$  表示  $P_i$  集合与  $T_i$  集合中  $p$  点与  $t$  点之间的欧几里德距离。HD 虽然在腰椎 L6 椎骨计算中产生较大离群值, 但是可以分析剩余椎骨的计算结果得到模型分割表现结果的平均表现。

ASSD 指标主要通过计算预测集合与真实集合之间的平均对称表面距离来评估模型的分割性能, 其计算流程如式 (10) 所示:

$$\text{ASSD} = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{P_i} \sum_{p=1}^{P_i} (p, B_i) + \frac{1}{G_i} \sum_{g=1}^{G_i} (g, P_i) \right) \quad (10)$$

其中,  $N$  代表样本的数量,  $P_i$  表示预测值边界  $i$  上体素的数量,  $G_i$  表示真实值边界  $i$  上体素的数量,  $d(p, B_i)$  代表预测边界  $i$  集合中体素值  $p$  到真实边界  $i$  的最短距离,  $d(g, P_i)$  表示真实边界  $i$  集中体素值  $g$  到预测边界  $i$  的最长距离。

## 2 脊柱图像分割实验

### 2.1 脊柱数据与实验环境

医学分割模型训练目前主要存在数据量规模小、数据样本质量参差不齐的难题, 目前主流的脊柱 CT 图像数据集中 CSI-Seg、Dataset-5、xVertSeg 数据集由于发表年份较早, 样本量相较于 VerSe 系列数据集与 CTSpine1K 数据集较少<sup>[25]</sup>。因此, 本文选用 VerSe 19、VerSe 20 与 CTSpine1K 数据集作为实验数据对提出的模型进行训练与测试。其中 VerSe 19 的样本总量为 160 例, VerSe 20 和 CTSpine1K 样本总量分别为 319 例与 1 005 例。本文实验采用 3:1:1 的比例将样本划分为训练集、验证集与测试集。

由于脊柱 CT 数据集 (VerSe 19、VerSe 20、CTSpine1K) 样本分辨率较高, 3 个数据集中 3D 图像平均尺寸到达 (512, 512, 594) 像素, 因此本文实验环境选用 NVIDIA A40 GPU, 显存为 48 GB, 内存选择 80 GB, 实验系统采用 Ubuntu 18.04, 环境配置为 Python 3.8+Pytorch 1.9.0 框架。根据多次预实验的结果, 本文实验训练 epoch 设置为 500 轮, batch-size 设置为 4, 学习率 lr 设置为 0.01%。

### 2.2 损失函数

针对脊柱 CT 数据集存在样本尺寸多样化、样本类别不均衡的问题。本文使用损失函数  $L$  如公式 (11) 所示,  $L$  主要由两部分损失函数组成, 整体标签损失函数  $L_s$  与局部脊柱分割损失函数  $L_e$ 。

$$L = \alpha L_s + \beta L_e \quad (11)$$

其中, 将局部脊柱损失函数作为整体分割损失函数的补充, 由于脊柱椎骨在训练分割的过程中会存在边界模糊和多节椎骨相似的问题, 并且椎骨的类别与数量严重失衡, 腰椎与胸椎末节 (T8~T12) 平均占比过高, 考虑引入平衡因子  $\alpha = 0.6, \beta = 0.4$  降低整体损失函数  $L_s$  对训练过程的影响, 利用局部脊柱损失函数  $L_e$  平衡分类。椎骨分割属于多标签任务, 其中单个训练样本之间包含的椎骨标签数量差距很大, 因此存在严重的标签类别不均衡现象, 为此笔者采用交叉熵与 Dice 损失函数结合的损失函数进行训练, 如公式 (12) 所示:

$$\begin{aligned} L_s &= \gamma L_{\text{CE}} + \delta L_{\text{dice}} \\ L_{\text{CE}} &= \frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(\hat{y}_{ic}) \\ L_{\text{dice}} &= 1 - \frac{1}{M} \sum_c \frac{2 \sum_i^{N^*} y_{ic} \hat{y}_{ic}}{\sum_i^{N^*} y_{ic} + \sum_i^{N^*} \hat{y}_{ic}} \end{aligned} \quad (12)$$

其中,  $L_{\text{CE}}$  作为二分类交叉熵损失函数,  $N$  为样本总数,  $M$  为标签类别总数,  $y_{ic}$  代表样本  $i$  的真实标签为  $c$  时取值为 1, 否则取值为 0,  $\hat{y}_{ic}$  代表样本  $i$  的预测标签值为  $c$  时的概率值。  $L_{\text{dice}}$  中  $N^*$  代表样本中体素的总数, 其余参数代表含义与  $L_{\text{CE}}$  相同。

对于局部脊柱损失函数, 笔者采用深监督的方式对椎骨边缘分割模块进行训练, 椎骨边缘分割模块将浅层椎骨的边缘轮廓语义信息分割为二值化标签, 为了降低椎骨边缘体素膨胀, 提高椎骨边缘轮廓浅层分割的精度, 采用二分类损失函数进行训练, 如式 (13) 所示:

$$L_e = -\lambda \sum_i y_i \log y_i - \mu \sum_i (1 - y_i) \log(1 - y_i) \quad (13)$$

由于椎骨边缘轮廓体素的占比远小于背景, 因此笔者将前景超参数  $\lambda$  设置为 0.8, 背景权重参数  $\mu$  设置为 0.2 以此降低  $L_e$  损失值。

### 2.3 实验结果

**2.3.1 数据预处理** 本文采用的数据集包含多节脊柱, 主要集中在胸椎与腰椎部位, 3 大数据集 (VerSe 19、VerSe 20、CTSpine1K) 共 793 例 CT 图像样本。CT 值范围为 100~3 000 HU, 而非前景区域 CT 值相对较小。因此, 我们对原始图像进行阈值分割来降低噪声与周边软组织对于目标区域的干扰, 将低通滤波的图像进行 ROI 区域裁剪, 尺寸为  $160 \times 160 \times 96$ 。

针对各样本存在的空间分辨率不同导致训练过程中分割效果不佳的问题, 本文将利用最邻近插值法将 CT 图像体素采样至  $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$  体素空间, 利用双边滤波对重采样后的图像进行去噪保证脊柱轮廓边缘的细节信息。

**2.3.2 VerSe 数据集分割结果** 本实验设计了 3 组对照训练, 针对 VerSe 19、VerSe 20、CTSpine1K 进行单独训练。针对 VerSe 系列数据集选择 3D U-Net 网络作为基准分割模型, 同时选取目前最先进的分割模型 (nnU-Net, Lessmann, Chen, Payer) 作为参照对象进行对比<sup>[26]</sup>。VerSe 系列数据集的测试实验结果如表 1 所示, 实验指标主要采用 1.5 节中提出的 DSC、HD 与 ASSD 3 类指标衡量实验模型的分割精度, 通过多次实验保证实验结果的准确性。

分析表 1 数据可以看到, 在 VerSe 19 的数据集上, 本文提出的脊柱分割模型 DSC 平均得分达到 88.92%, 相较于经典分割模型 nnU-Net 和目前最先进



表 1 VerSe 系列数据集实验结果对比

Table 1 Comparison of experimental results on the VerSe datasets

数据集	模型	DSC/%	HD/mm	ASSD/mm
VerSe 19	3D U-Net	71.88	24.59	9.7
	nnU-Net	83.79	9.91	6.9
	Lessmann	85.76	8.20	9.1
	Chen	82.56	9.98	8.3
	本文模型	88.92	7.73	6.2
VerSe 20	3D U-Net	80.75	14.72	11.3
	nnU-Net	89.59	10.20	6.2
	Payer	89.71	6.06	6.5
	Chen	91.23	7.15	7.9
	本文模型	92.41	6.22	5.8

的分割模型 Lessmann、Chen 模型分别提升 6.1%、3.6%、7.7%，平均提升 5.8%。表明本文提出模型在脊

柱的分割精度方面有了明显提升。在 HD 与 ASSD 指标上本文提出的模型取得最短距离 7.73 与 6.2 mm，相较于基准 3D U-Net 模型分别缩短了 68.5% 与 36.0%，验证本文模型在小规模脊柱 CT 数据集上具有鲁棒性。对比目前最先进的脊柱分割模型 nnU-Net、Lessmann 模型、Chen 模型，HD 与 ASSD 平均缩短了 16.7% 和 23.8%。但是在 VerSe 20 数据集中本文提出的模型在 HD 指标上相较于 Payer 模型未取得最短距离，本文提出的分割模型在边界分割上依然落后于 Payer 模型。但总体上，本文模型依旧取得了最优表现，证明在脊柱的平滑区域的分割与部分脊柱边缘区域的分割方面本文提出的模型具有良好的泛化能力。

在 VerSe 19 和 VerSe 20 两个数据集中各选择 2 例样本进行可视化实验，结果如图 6 所示，采用脊柱的矢状面作为实验结果展现，绿色矩形方框标记区域为不同分割模型与专家分割结果的不同之处。

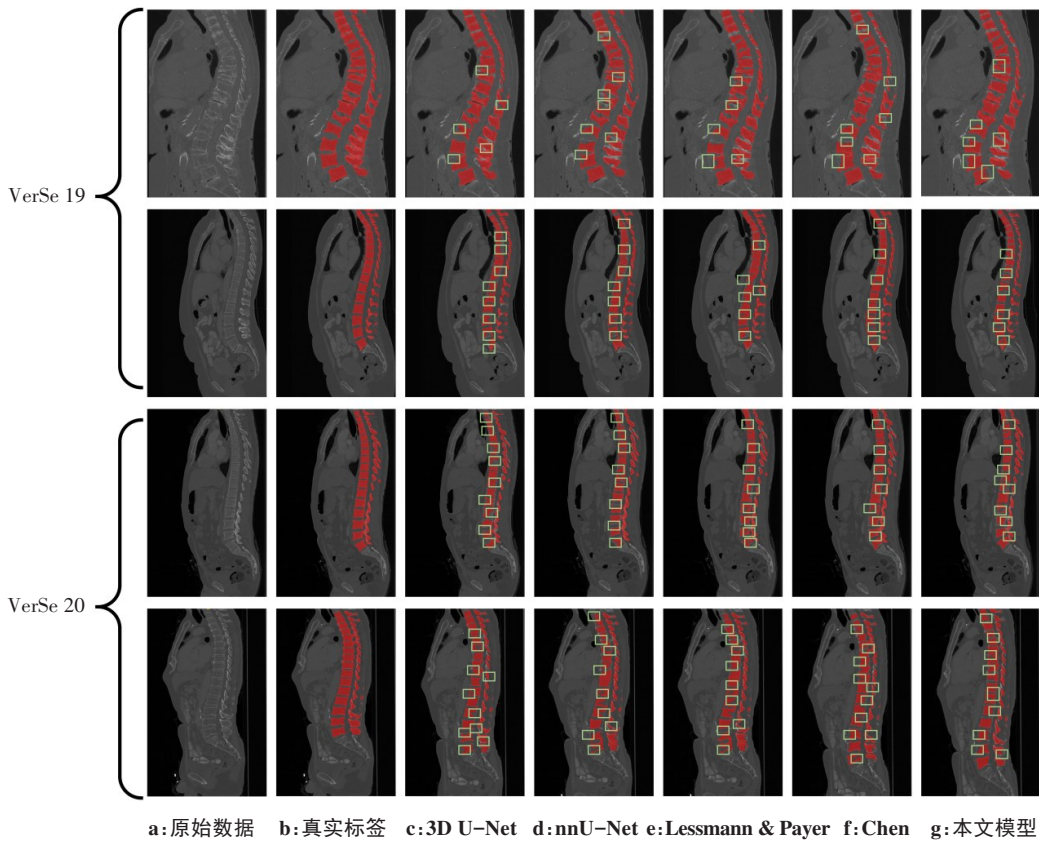


图 6 VerSe 19 与 VerSe 20 数据集实验结果模型

Figure 6 Experimental results on VerSe 19 and VerSe 20 datasets

从图 6 可以看出，基准网络 3D U-Net 的分割结果相较于其他模型假阳性区域较多，并且假阳性区域集中在脊柱椎骨高度相似区域，验证了基准模型 3D U-Net 在边界分割与脊柱椎骨定位方面缺乏有效学

习。相较于其他模型，本文提出的模型在边界区域的分割上有明显提高，脊柱椎骨各部位之间分割间隙明显，表明脊柱边缘分割模块在网络中对边界分割的有效性。

**2.3.3 CTSpine1K 数据集分割结果** 在 CTSpine1K 数据集上进行训练,测试集的分割结果如表 2 所示。分析表 2 可知,随着训练集与验证集中多种异常样本的增加,相对于 VerSe 系列数据集,本文提出的模型在 CTSpine1K 数据集上获得 83.89% 的平均 DSC 得分与 6.67 mm 的平均 HD,虽然没有达到 VerSe 系列的分割精度,但是相较于基准模型 3D U-Net 与目前的主流模型 nnU-Net 等依旧在 DSC 与 HD 上取得最优结果,验证了本文提出的模型在脊柱小规模数据集与大型数据集中有良好的泛化性能。

在 CTSpine1K 数据集上的测试实验结果如图 7 所示,选择 3 例患者样本进行分割结果可视化。由于 CTSpine1K 数据集由 4 个小型数据集组成,各数据集之间存在 CT 设备、成像质量、患者生理情况等多种差异,因此在 CTSpine1K 数据集上的测试分割精度相较于 VerSe 系列数据集存在一定差距。而本文提出的模型相较于

表 2 CTSpine1K 数据集实验结果对比  
Table 2 Comparison of experimental results on the CTSpine1K dataset

分割模型	DSC/%	HD/mm	ASSD
3D U-Net	74.73	16.93	6.4
nnU-Net	76.61	9.08	5.2
Lessmann	77.28	7.29	7.1
Chen	76.45	9.32	4.3
本文模型	83.89	6.67	4.4

nnU-Net、Lessmann、Chen 等目前主流的分割网络依旧取得了提升。从 Lessmann 模型(图 7e)与本文模型(图 7g)对比可知,利用混合通道-空间注意力不仅提高了网络对于脊柱各椎骨的定位,而且增强了脊柱的全局位置信息便于脊柱深层特征的识别,本文提出的模型分割结果中假阴性区域相对较少,DSC 得分取得最优。

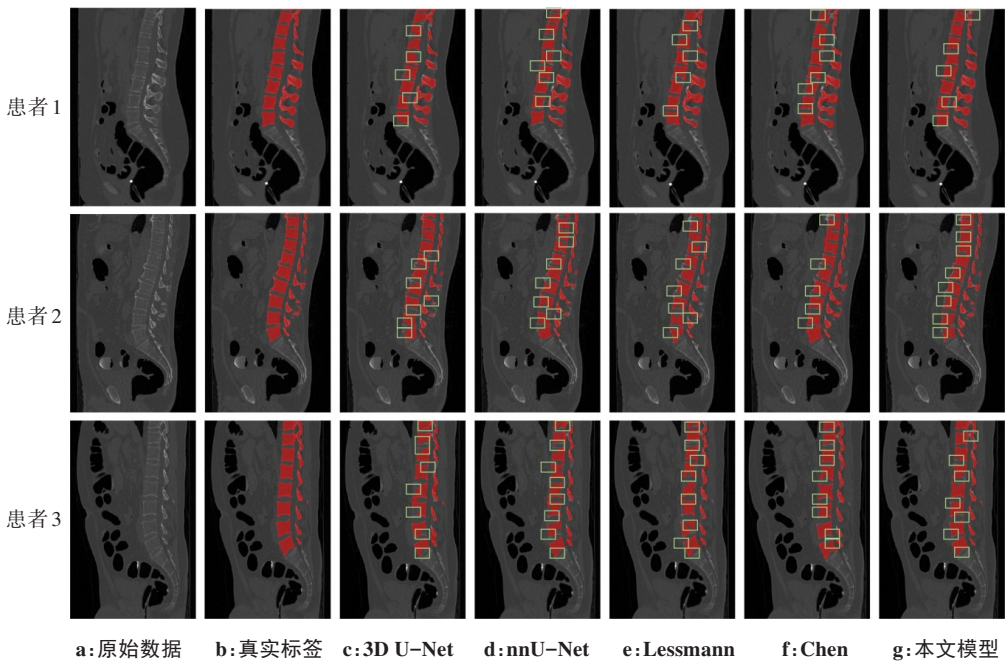


图 7 CTSpine1K 数据集实验结果  
Figure 7 Experimental results on the CTSpine1K dataset

2.4 消融实验

本文对所提出的模型进行消融实验以验证提出的 3 种改进方案的有效性。消融实验的结果对比如表 3 所示,由于 VerSe 19 与 VerSe 20 数据集样本组成近似,因此本文采用 VerSe 20 数据集作为代表进行消融实验。消融模块主要分为 3 部分:SE 残差模块、脊柱边缘分割 (SESM) 模块、ECA 注意力机制模块(通道注意力、空间注意力、自注意力)。选择原始 3D U-Net 作为空白对照组,在两组不同的数据集中进行基准分割。

首先分析 SE 残差模块在 3D U-Net 网络中的影响。

在 CTSpine1K 数据集和 VerSe 20 数据集中,添加 SE 残差模块的 3D U-Net+SE 的 DSC 得分相较于原始的 3D U-Net 网络分割精度分别提高了 2.07% 与 1.7%,HD 分别缩短了 44.5% 与 23.1%,验证了在编码器各层之间嵌入 SE 残差模块改变基础结构有助于提升脊柱的分割精度。在添加了 SESM 模块之后,在两组数据集中 DSC 得分提高至 78.76% 和 83.19%,HD 与 ASSD 也缩短至 9.54、7.84 mm 与 4.9、9.4 mm。SESM 模块将提取出的包含丰富边缘信息的浅层特征进行多尺度的边缘特征提取与融合,将脊柱边缘特征与提取出的深层特征相加,提高



表3 消融实验结果对比  
Table 3 Comparison of ablation experiment results

数据集	消融模块	DSC/%	HD/mm	ASSD/mm
CTSpine1K	3D U-Net	74.73	16.93	6.4
	3D U-Net+SE	76.28	9.39	6.3
	3D U-Net+SESM	78.76	9.54	4.9
	3D U-Net+ECA	74.92	7.06	6.9
	本文模型	83.89	6.67	4.4
VerSe 20	3D U-Net	80.75	14.72	11.3
	3D U-Net+SE	82.13	11.32	8.3
	3D U-Net+SESM	83.19	7.84	9.4
	3D U-Net+ECA	81.58	9.73	6.1
	本文模型	92.41	6.22	5.8

了脊柱边缘分割的精度,进一步细化了椎骨边缘的分离。实验证明在添加了SESM模块后,DSC、HD、ASSD指标提升,证明了脊柱边缘分割模块的有效性与鲁棒性。在基础3D U-Net模型中添加ECA模块,虽然DSC相较于原始的3D U-Net网络未取得显著提升,但在脊柱边缘分割评价指标上HD与ASSD明显缩短,相较于真实标签距离更短。ECA模块更侧重于脊柱整体位置信息的提取,在降低脊柱边缘分割模糊与脊柱椎骨定位模糊两方面具有明显的正向效果。本文模型在叠加了3个模块后取得了最优DSC得分与最短的HD,相较于原始的基准分割模型3D U-Net在叠加不同模块后进行多组实验,实验结果证明了本文所提改进方案的有效性与优越性。

3 结 论

本文模型在多种主流脊柱CT数据集上进行实验,将本文网络与目前主流分割网络进行对比实验。实验结果证明,本文模型相较于其他模型更优秀,在现有分割精度的基础上进一步提升。本文提出的基于多尺度边界分割与混合注意力机制的脊柱CT图像分割较好地解决了椎骨边界分割模糊的欠分割问题,为脊柱病理评估和计算机辅助手术治疗等各种临床实践提供大量有效的病理信息。

在实验过程中,笔者发现目前脊柱医学图像数据集中存在椎骨类别数量严重失衡的问题。如果训练样本的类别极少,模型在学习低频类别椎骨上会出现过拟合或欠拟合的情况,无法学习脊柱局部特征。需要找到一种有效的方法来减少椎骨类别严重失衡导致的欠分割问题。

【参考文献】

[1] Yang ZM, Wang Q, Zeng JC, et al. RAU-net: U-net network based on residual multi-scale fusion and attention skip layer for overall spine segmentation[J]. Mach Vis Appl, 2022, 34(1): 10.  
[2] Hassanzadeh T, Essam D, Sarker R. EEvoU-net: an ensemble of

evolutionary deep fully convolutional neural networks for medical image segmentation[J]. Appl Soft Comput, 2023, 143: 110405.  
[3] 宇文婉. 基于注意力驱动卷积神经网络的椎骨CT图像分割方法研究[D]. 哈尔滨: 哈尔滨理工大学, 2021.  
Yuwen W. Research on vertebral CT image segmentation based on attention-driven convolutional neural network[D]. Harbin: Harbin University of Science and Technology, 2021.  
[4] Schnider E, Wolleb J, Huck A, et al. Improved distinct bone segmentation in upper-body CT through multi-resolution networks[J]. Int J Comput Assist Radiol Surg, 2023, 18(11): 2091-2099.  
[5] 陈倩. 基于迁移学习的医学图像分割多任务算法研究[D]. 合肥: 中国科学技术大学, 2022.  
Chen Q. Multi-task algorithm based on transfer learning for medical image segmentation[D]. Hefei: University of Science and Technology of China, 2022.  
[6] Li XY, Luo GN, Wang W, et al. Curriculum label distribution learning for imbalanced medical image segmentation[J]. Med Image Anal, 2023, 89: 102911.  
[7] Golla AK, Lorenz C, Buerger C, et al. Cervical spine fracture detection in computed tomography using convolutional neural networks[J]. Phys Med Biol, 2023, 68(11): 115010.  
[8] Wu ZY, Xia GF, Zhang XH, et al. A novel 3D lumbar vertebrae location and segmentation method based on the fusion envelope of 2D hybrid visual projection images[J]. Comput Biol Med, 2022, 151(Part A): 106190.  
[9] Nazir A, Cheema MN, Sheng B, et al. ECSU-net: an embedded clustering sliced U-net coupled with fusing strategy for efficient intervertebral disc segmentation and classification[J]. IEEE Trans Image Process, 2022, 31: 880-893.  
[10] Saeed MU, Bin W, Sheng JF, et al. 3D MRU-net: a novel mobile residual U-net deep learning model for spine segmentation using computed tomography images[J]. Biomed Signal Process Control, 2023, 86, Part A: 105153.  
[11] Xu H, Cui XX, Li CF, et al. RU-nT: a network combining residual U-net and transformer for vertebral edge feature fusion constrained spine CT image segmentation[J]. IEEE Access, 2023, 11: 55692-55705.  
[12] Meng D, Boyer E, Pujades S. Vertebrae localization, segmentation and identification using a graph optimization and an anatomic consistency cycle[J]. Comput Med Imaging Graph, 2023, 107: 102235.  
[13] You X, Gu Y, Liu YY, et al. EG-Trans3DUNet: a single-staged transformer-based model for accurate vertebrae segmentation from spinal CT images[C]//2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). Piscataway, NJ, USA: IEEE, 2022: 1-5.  
[14] Tao R, Liu WY, Zheng GY. Spine-transformers: vertebra labeling and segmentation in arbitrary field-of-view spine CTs via 3D transformers[J]. Med Image Anal, 2022, 75: 102258.  
[15] You X, Gu Y, Liu YY, et al. VerteFormer: a single-staged Transformer network for vertebrae segmentation from CT images with arbitrary field of views[J]. Med Phys, 2023, 50(10): 6296-6318.  
[16] 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 International Publishing, 2015: 234-241.  
[17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.  
[18] Liu H, Sun YR, Cheng XY, et al. Prior-based 3D U-net: a model for knee-cartilage segmentation in MRI images[J]. Comput Graph, 2023, 115: 167-180.  
[19] Kim S, An S, Chikontwe P, et al. Bidirectional RNN-based few shot learning for 3D medical image segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2021: 1808-1816.  
[20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6000-6010.  
[21] Liu YC, Shao ZR, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[EB/OL]. (2021-12-10). <https://arxiv.org/abs/2112.05561>.  
[22] Li HF, Qiu KJ, Chen L, et al. SCAttNet: semantic segmentation network with spatial and channel attention mechanism for High-Resolution remote sensing images[J]. IEEE Geosci Remote Sens Lett, 2021, 18(5): 905-909.  
[23] Ma W, Gong CF, Xu SB, et al. Multi-scale spatial context-based semantic edge detection[J]. Inf Fusion, 2020, 64: 238-251.  
[24] Korez R, Ibragimov B, Likar B, et al. A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation[J]. IEEE Trans Med Imaging, 2015, 34(8): 1649-1662.  
[25] Deng Y, Wang C, Hui Y, et al. CTSpine1K: a large-scale dataset for spinal vertebrae segmentation in computed tomography[EB/OL]. (2021-07-05). <https://arxiv.org/abs/2105.14711>.  
[26] Sekuboyina A, Hussein ME, Bayat A, et al. VerSe: a vertebrae labelling and segmentation benchmark for multi-detector CT images[J]. Med Image Anal, 2021, 73: 102166.

(编辑:薛泽玲)