

基于自注意力的双通路全脊柱 X 光图像分割模型

师文博¹, 杨环¹, 西永明², 段文玉¹, 徐同帅², 杜钰堃²

1. 青岛大学计算机科学技术学院, 山东 青岛 266071; 2. 青岛大学附属医院崂山院区脊柱外科, 山东 青岛 266000

【摘要】全脊柱X光图像(包含脊柱、骶骨及髌骨)分割是目前脊椎疾病智能诊断中首要关键的环节。针对U-Net语义分割算法在全脊柱X光图像多区域分割精度较差的问题,提出一种双通道语义分割算法DAU-Net,通过空间通道与语义通道分别学习空间信息特征与图像语义特征,并在解码器端对两类特征进行融合,获取脊柱X光图像中更精准的分割边界。在空间通道中,使用空洞卷积及残差模块扩大视野域并保留更多远端特征信息。此外,将自注意力机制引入语义通道,并设计不同的自注意力编码与自注意力解码模块构建全局关联信息,实现对多个目标骨骼区域语义分割。实验结果表明,DAU-Net能够有效提高脊柱X光图像上的分割精度,相比U-Net、ResU-Net、Attention U-Net、U-Net++,Dice系数分别提高4.00%、1.90%、4.60%、1.19%。

【关键词】脊柱图像分割;U-Net;语义分割;双通道网络;自注意力机制

【中图分类号】R318;R816.8

【文献标志码】A

【文章编号】1005-202X(2022)11-1385-08

Self-attention based dual pathway network for spine segmentation in X-ray image

SHI Wenbo¹, YANG Huan¹, XI Yongming², DUAN Wenyu¹, XU Tongshuai², DU Yukun²

1. College of Computer Science and Technology, Qingdao University, Qingdao 266071, China; 2. Department of Spinal Surgery, Laoshan Branch, the Affiliated Hospital of Qingdao University, Qingdao 266000, China

Abstract: The segmentation of X-ray images of the entire spine including the spine, sacrum and iliac bone is the essential step for the intelligent diagnosis of spine diseases. A semantic segmentation network named dual pathway with self-attention for refined U-Net (DAU-Net) is proposed to solve the problem of poor accuracy of U-Net semantic segmentation algorithm in multi-region segmentation in full-spine X-ray image. DAU-Net adopted spatial pathway and semantic pathway to learn spatial information and semantic information separately, and then combines these two types of features at the decoder, thereby obtaining more accurate segmentation boundaries in full-spine X-ray images. In the spatial pathway, dilated convolutions and residual blocks are used to expand the receptive field and capture the long-range dependency feature information. Furthermore, the self-attention mechanism is applied in the semantic pathway, and different self-attention encoders and self-attention decoders are designed to construct global association to achieve the semantic segmentation of multiple target bone regions. The experimental results show that DAU-Net can effectively improve the segmentation accuracy in full-spine X-ray images, and its Dice coefficient is 4.00%, 1.90%, 4.60% and 1.19% higher than those of U-Net, ResU-Net, Attention U-Net and U-Net++, respectively.

Keywords: spine image segmentation; U-Net; semantic segmentation; dual pathway network; self-attention mechanism

前言

脊柱是人体最重要的骨骼之一,拥有支撑身体结构、保护身体器官的能力,同时在机械活动、负荷

转移过程中也发挥重要的作用。如果人体脊柱发生畸形,则可能会影响身体各种器官的正常发育^[1-2]。脊柱侧弯是最常见的一种脊柱畸形,也是目前危害我国青少年儿童的常见病之一。若不及时治疗,会给青少年带来各种健康问题。在脊柱侧弯临床诊断中,全脊柱X光片(通常包含脊柱、骶骨及髌骨区域)是最重要的诊断依据。传统疾病诊断方案中,医生需手动测量多个医学参数,如Cobb角、顶椎偏距、椎体旋转度、肩高度、骶骨中垂线等。然而手动标记及参数计算是一件十分耗时的任务。此外,由于不同拍摄条件下的X光图像差异较大,医生可能会得出不

【收稿日期】2022-07-11

【基金项目】山东省重点研发计划(2019GGX101021);山东省泰山学者项目(ts20190985)

【作者简介】师文博,硕士研究生,研究方向:计算机视觉、医疗图像处理,E-mail: shi_wenbo@outlook.com

【通信作者】杨环,博士,副教授,研究方向:计算机视觉、深度学习,E-mail: cathy_huanyang@hotmail.com

同的诊断结果,从而影响侧弯疾病类型的诊断。因此,基于X光片的脊柱侧弯智能诊断成为近年来研究的热点问题,其中全脊柱X光片的自动分割是首要关键的一个环节,旨在实现脊柱、骶骨及髂骨的多区域分割,为后续的参数测量及侧弯分型奠定基础。

近年来,以卷积神经网络(Convolutional Neural Network, CNN)为基础的深度学习语义分割网络逐步进入医疗图像分割领域^[2-5]。在脊柱图像处理领域里, Wu等^[6]使用BoostNet模型对脊柱侧弯每节椎骨的4个坐标点标记,用于对脊柱侧弯Cobb角度的计算。Grigorieva等^[7]基于U-Net模型和先验知识实现了对脊柱X光图像的粗略分割。赵阳等^[8]实现了基于U-Net模型与多尺度特征融合的脊柱分割算法。上述方法仅实现了X光脊柱图像的椎体定位或者精度较低的椎体分割,且无法对诊断分型需要的其他骨骼特征进行语义分割。

Han等^[9]基于循环神经网络与对抗生成网络模型理论,提出一个用于脊柱矢状位核磁共振(MRI)图像的分割框架,实现脊柱与椎间盘的多分类语义分割。文献[10-12]基于U-Net、残差网络以及通道注意力机制实现脊柱CT或MRI图像语义分割。毛孝鑫等^[13]基于Transformer机制实现CT图像上的脊柱分割。以上方法虽可以完成脊柱分割,但依然存在3个问题:(1)仅实现了脊柱与椎体的部分分割,对于其它判断脊柱疾病的特征区域(如骶骨和髂骨)没有进行语义分割;(2)常见的分割算法主要应用于CT、MRI图像,但X光图像容易受到骨骼、内脏器官等影响,图像更容易有叠加的阴影,因此上述分割算法在X光图像上的分割精度较差;(3)由于卷积核大小的影响,基于CNN的语义分割模型更多关注提取局部视野域,无法建立整张图像特征信息的整体关联模型,使得语义分割精度提升受到阻碍。

近年来,为了在神经网络中建立整个图像的整体特征关联,自注意力机制逐渐走入医疗图像分割领域,与CNN相比,自注意力机制能获取远端的特征信息,允许模型根据输入的内容动态捕捉远端特征^[14]。Gao等^[15]、Valanarasu等^[16]和Ji等^[17]分别在医疗图像会议上提出基于融合自注意力机制与卷积神经网络的语义分割模型,并在皮肤病、脑部、腺体分割等领域取得良好表现。鉴于自注意力机制在全局信息提取方面的优越性,本研究提出一种基于自注意力与卷积神经网络的双通道语义分割模型DAU-Net,实现了对全脊柱X光图像脊柱、骶骨、髂骨部分的语义分割。该模型的主要创新如下:(1)构造双通道全脊柱X光图像分割框架,分别学习图像中的语义信息与空间信息,在保证分类精准的同时提高分割精度,进一步改善由于重复编码与解码操

作带来的空间信息丢失问题;(2)在空间通道中,使用残差模块替代卷积层,提取更多更深层的特征信息,并使用空洞卷积代替池化操作,扩大视野域并保留更多远端特征信息;(3)设计不同的自注意力编码(Self-Attention Encoder, SAE)模块与自注意力解码(Self-Attention Decoder, SAD)模块,并分别加入语义通道的编码器解码器结构中,捕获特征图的全局关联信息;(4)将空间通道解码器的特征作为SAD的空间位置补充,弥补自注意力模块空间信息的丢失,实现对多个目标骨骼区域语义分割,获取更加精准的分割边界。

1 双通道语义分割DAU-Net

1.1 整体网络架构

为了实现全脊柱X光图像多个骨骼区域的同时分类以及提高其分割精度,本研究提出一种基于自注意力机制与CNN的双通道语义分割算法DAU-Net,该算法使用两条通道对全脊柱X光图像进行语义分割,模型整体结构如图1所示。首先使用语义分割网络(Semantic Pathway)对图像的不同目标区域进行特征提取并实现对脊柱、髂骨的分类;其次使用空间信息网络(Spatial Pathway),针对所有目标区域进行空间特征提取,使用较浅的卷积分割模型实现背景与整体目标区域的分割。语义信息网络中,包含一个基于编码器解码器结构,使用跳跃连接、卷积残差模块和自注意力机制的语义分割模型,实现对不同目标区域语义特征的分,改善不同目标骨骼之间粘连的问题;空间信息网络包含一个基于卷积残差模块与空洞卷积神经网络的分割模型,用于获取所有目标检测区域的特征信息,弥补网络层数增加、自网络层数增加以及注意力模型结构在解码过程中丢失的空间位置信息,改善分割结果边缘精度差的问题。

1.2 空间信息网络

空间信息网络用于提取全脊柱X光图像粗略空间位置信息,因此不需要深层次卷积去提取细致语义特征。本研究使用一个深度为2的卷积网络提取特征信息,在空间编码器(Spatial Encoding, SE)部分中,每层使用一个卷积残差模块^[18](图2)和一个5×5、步长为2的空洞卷积提取粗略语义信息。在卷积残差模块中,首先使用批标准化对输入特征归一化,其次采用两个连续的卷积核为3×3、步长为1的卷积层提取特征信息,每一个卷积层后使用LeakyReLU作为非线性激活函数。最后将输入特征图与残差模块输出特征图进行逐元素相加,构建恒等映射条件,提升特征学习能力。使用空洞卷积层替代池化下采样操作,在保留更大图像视野域的同时将图像尺寸比原图像缩小一倍。SE1、SE2的卷积通道数为分别为32、64。

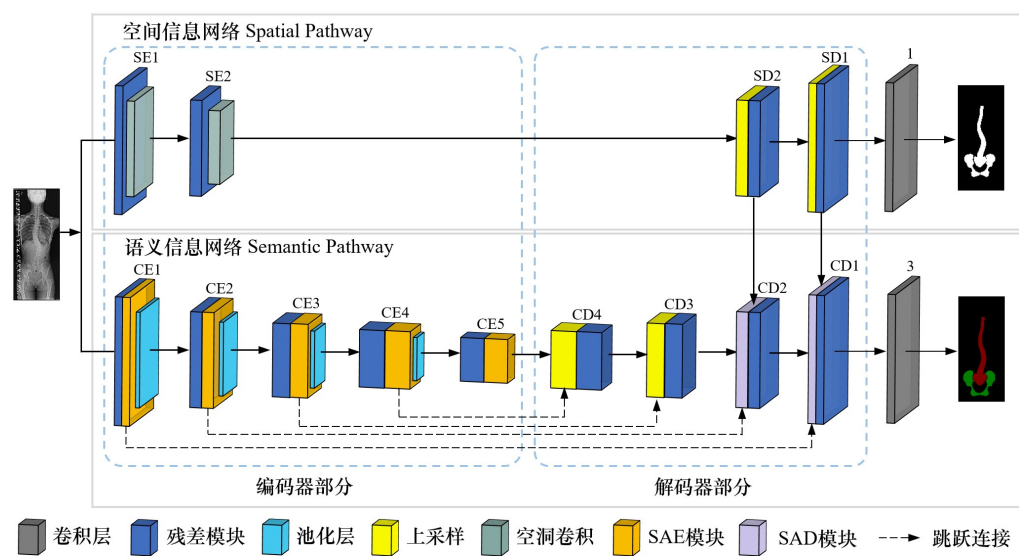


图 1 双通道语义分割模型

Figure 1 Dual pathway semantic segmentation network

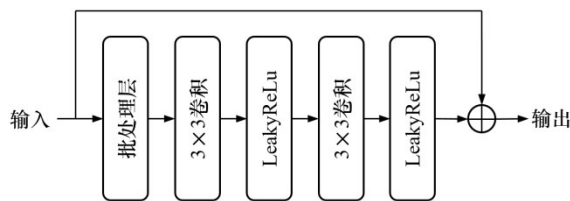


图 2 卷积残差模块结构

Figure 2 Convolution residual block structure

在解码器部分,本研究使用 2 层空间解码层 (Spatial Decoding, SD)将图像逐步上采样还原为原图像尺寸,在上采样操作中,使用一个卷积核为 4×4、步长为 2 的转置卷积将图像尺寸放大一倍,再使用残差模块对上采样后的图像特征进行提取。SD2、SD1 的卷积通道数分别为 64、32。最后使用一个通道数为 1 的 1×1 的卷积核,融合不同通道的特征信息,用于计算损失。

1.3 语义信息网络

语义信息网络通过学习脊柱 X 光图像中的语义信息,将脊柱、髌骨等骨骼与背景分离。本模型使用一个深度为 5 的编码器、解码器结构,并使用跳跃连接的方式将对应层的编码特征信息与解码的特征信息级联,用以丰富图像特征信息。在语义编码阶段 (Context Encoding, CE),每层编码器由卷积残差模块、SAE 模块和 2×2 最大池化层构成。首先使用卷积残差模块提取图像中的局部特征信息,其次使用 SAE 模块提取本层中的全局特征信息,最后使用 2×2 最大池化层缩小图像尺寸,减少特征维度。CE1 至 CE5 的卷积通道数分别为 32、64、128、256、256。

在语义解码器 (Context Decoding, CD)中,CD4、

CD3 使用一个卷积核为 4×4、步长为 2 的转置卷积将图像上采样,并与跳跃连接将下采样的特征图级联,再使用残差卷积模块获取更多的特征信息,卷积通道数分别为 256、128。在 CD2、CD1 中,使用 SAD 模块替代原有的单独转置卷积操作,将深层解码器的特征信息、跳跃连接层的特征信息以及空间信息网络的特征信息有效融合,改善由于 SAD 模块自身导致的空间信息丢失的问题,卷积通道数分别为 64、32。最后使用一个 1×1 的卷积模块对特征图进行降维,输出为 n 分类特征通道,本研究中 n 设置为 3,分别为脊柱含髌骨、髌骨与背景。

1.4 自注意力模块

在语义信息网络中,编码器、解码器中分别使用不同的自注意力模块。本研究的自注意力模块主要分为 SAE 模块与 SAD 模块,均由残差结构与多头自注意力模块 (Multi-Head Self-Attention, MHSA) 构成,自注意力模块的整体结构如图 3 所示。在自注意力模块中,首先对输入特征使用批标准化对图像特征进行归一化,使用 MHSA 提取全局特征信息并使用 LeakyReLU 进行激活操作,然后与原图像进行逐像素相加操作得到中间融合特征,其次使用 1×1 的卷积操作实现跨通道的信息融合,再使用 LeakyReLU 对图像特征进行非线性激活,最后将激活后的特征图与输入图像进行逐像素相加,构建恒等映射条件,从而获取含有更丰富的特征信息的自注意力模块输出特征图。

然而,在语义信息网络中编码器与解码器部分使用的 MHSA 不同 (图 4)。传统 MHSA 在自然语言处理中,凭借能够捕获远端关联特征信息,取得了优

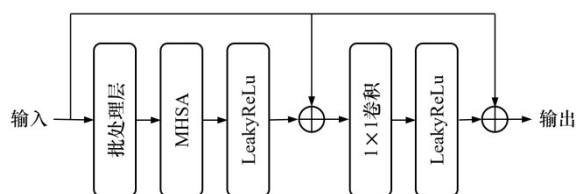
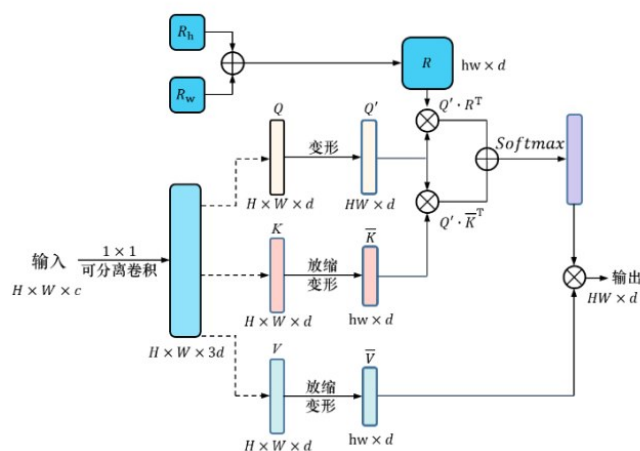


图3 自注意力模块整体结构

Figure 3 Self-attention block architecture

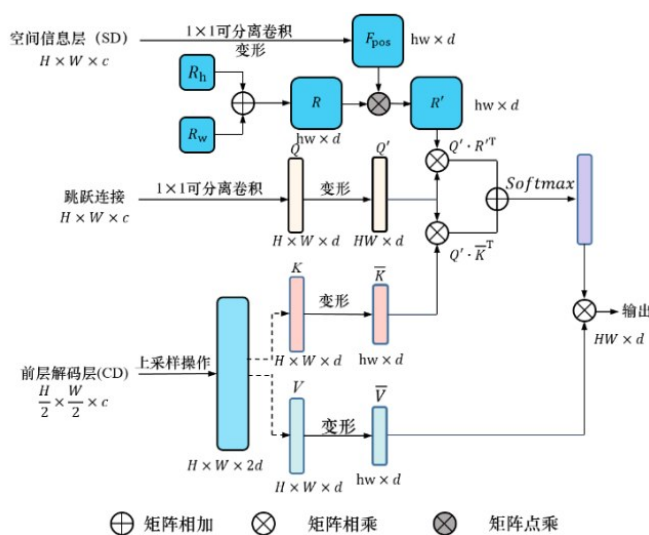
异的性能。但在图像领域中,除了分割目标边缘区域外,图像大部分的像素都可以划分为目标区域与非目标区域,因此使用传统的自注意力机制对于逐个像素计算成对注意力是低效且冗余的。受 Gao 等^[15]工作的启发,本研究 SAE 模块使用 MHSA 下采样模块(图 4a)。其输入特征图 $X \in R^{H \times W \times c}$,



a: MHSA 下采样模块

其中 H, W 分别代表特征图的长、宽, c 代表特征图的通道数,使用 1×1 的深度分离卷积分别生成特征矩阵 Q, K, V , 并且满足以下公式: $Q, K, V \in R^{H \times W \times d}$, 其中 d 是每个头对应的通道数,满足函数 $c = d \times n$, n 为超参数 MHSA 的头数。将特征矩阵 K, V 通过双线性插值法下采样,将其映射为 $h \times w$ 的特征矩阵 \bar{K}, \bar{V} , 将 Q, \bar{K}, \bar{V} 变形为序列,其中 Q' 的尺寸为 $r \times d, r = H \times W; \bar{K}, \bar{V}$ 的尺寸为 $r' \times d, r' = h \times w$, 本研究中 $h = w = 8, n = 4$ 。通过图像下采样操作,大幅度降低自注意力机制在图像领域中所需要的计算量,基础自注意力公式可设计为:

$$\text{Attention}(Q', \bar{K}, \bar{V}) = \text{softmax} \left(\frac{Q' \bar{K}^T}{\sqrt{d}} \right) \bar{V} \quad (1)$$



b: MHSA 上采样模块

图4 MHSA 编码模块与解码模块

Figure 4 Encoding and decoding blocks of multi-head self-attention (MHSA)

由于自注意力机制是将特征信息序列化后构建特征的全局关联信息,因此存在特征信息的空间位置丢失的问题^[19]。本研究通过引入相对位置信息编码的方式改善问题,则自注意力模块输出计算如式(2)所示:

$$\text{Attention}(Q', \bar{K}, \bar{V}) = \text{softmax} \left(\frac{Q' \bar{K}^T + Q' R^T}{\sqrt{d}} \right) \bar{V} \quad (2)$$

其中, R 是满足尺寸 $hw \times d$ 的空间特征信息矩阵,是在 MHSA 模块添加额外的相对编码位置信息矩阵^[19]。满足公式 $R = R_h + R_w$ 。针对任意两个像素点 $i = (i_x, i_y)$ 与 $j = (j_x, j_y)$, R_h 与 R_w 满足公式 $R_h[i, j] = r_{i_y - j_y}^h$, $R_w[i, j] = r_{i_x - j_x}^w$ 。 $r_{i_y - j_y}^h$ 与 $r_{i_x - j_x}^w$ 分别代表图像像素 i 与 j 的水平方向与竖直方向的距离差。此外,将 $Q' \bar{K}^T$ 与

$Q' R^T$ 相对位置空间信息编码逐像素相加的方式融合,再使用 softmax 函数对其进行激活,最后再与矩阵 \bar{V} 相乘,获得 MHSA 下采样的输出。

SAD 模块使用 MHSA 上采样模块(图 4b),其输入信息为 3 个部分,包括编码器跳跃连接的特征信息、前层解码器的特征信息与空间通道的特征信息,如图 1 中 CD2 的输入为 CE2 跳跃连接的特征输出、CD3 输出及 SD2 输出。在 MHSA 上采样模块中,采用多种融合方式,将此 3 部分特征信息进行深度融合。

对跳跃连接输入使用一个 1×1 的可分离卷积对图像通道调整生成特征图 Q ; 将前层解码器的特征信息使用卷积核为 4×4 、步长为 2 的转置卷积,将图像尺

寸扩大一倍,获得 \mathbf{K} 、 \mathbf{V} ,同样满足公式 \mathbf{Q} 、 \mathbf{K} 、 $\mathbf{V} \in R^{H \times W \times d}$ 。此外,MHSA上采样模块中矩阵 \mathbf{Q}' 、 $\bar{\mathbf{K}}$ 、 $\bar{\mathbf{V}}$ 处理方式与MHSA下采样中的操作及尺寸一致。

此外,为了添加更多的位置空间信息改善分割精度,本研究将空间通道的特征SD与二维位置编码信息 \mathbf{R} 进行融合。首先,对SD特征图使用 1×1 可分离卷积进行维度变换,再使用双线性下采样与变形操作,得到 F_{pos} ,满足 $F_{\text{pos}} \in R^{h \times w \times d}$ 。将 F_{pos} 与空间信息编码 \mathbf{R} 进行点乘融合,得到 $\mathbf{R}' = \mathbf{R} \cdot F_{\text{pos}}$ 。

将跳跃连接获得的语义特征,经过分别与解码器特征及空间信息特征进行矩阵相乘,得到 $\mathbf{Q}\bar{\mathbf{K}}^T$ 及 $\mathbf{Q}\mathbf{R}'^T$,通过逐像素相加的方式进行融合,然后将两者融合经softmax激活函数映射后,与 $\bar{\mathbf{V}}$ 矩阵相乘,得到MHSA上采样模型的输出,满足公式:

$$\text{Attention}(\mathbf{Q}', \bar{\mathbf{K}}, \bar{\mathbf{V}}) = \text{softmax}\left(\frac{\mathbf{Q}'\bar{\mathbf{K}}^T + \mathbf{Q}\mathbf{R}'^T}{\sqrt{d}}\right)\bar{\mathbf{V}} \quad (3)$$

1.5 损失函数

本研究使用二元交叉熵损失函数,分别对空间信息与语义信息网络预测图计算损失。在空间信息网络中,将预测结果通过阈值为0.5的二值化处理输出预测图,并与二值化处理的真值图进行计算。在语义信息网络中,分别对3个不同的通道进行二值化处理,计算损失函数。损失函数总表达式为:

$$\text{loss} = \alpha \times \text{BCE}(\text{pred}_{\text{loc}}, \text{GT}_{\text{loc}}) + \sum_{i=1}^3 \text{BCE}(\text{pred}_{\text{sem}}^i, \text{GT}_{\text{sem}}^i) \quad (4)$$

式中, pred_{loc} 与 GT_{loc} 分别对应空间网络预测图与真值图, $\text{pred}_{\text{sem}}^i$ 、 GT_{sem}^i 分别对应语义信息网络不同通道中的预测图与真值图, α 是超参数,本研究设置为0.3。

2 结果与分析

2.1 脊柱X光图像数据集

本研究所用到的脊柱X光图像均来自于青岛大学附属医院影像科,由专业的临床医师使用开源标注工具Labelme^[20]对于需要分割的脊柱(颈椎第7节到腰椎第5节)、骶骨以及髌骨部分进行手工标注,为方便标注,把脊柱和骶骨作为一个连通部分进行标注,两个髌骨作为一类标记。此数据集共包含320张X光脊柱图像(图5),以及对应的320份标签文件。由于医院机器以及主治医师保存数据习惯不同,数据集中原始图像分辨率分布差异较大,训练集中图像最小为 $2\,300 \times 1\,440$ 像素,最大为 $7\,940 \times 3\,000$ 像素,因此在训练模型之前需要对图像进行尺寸归一化,通过双线性插值法对图像按照长边进行等比例缩放

为 512×256 ,将短边缺失的部分填充0,保证图像尺寸的统一性。将数据集随机选取80%做训练集(256张)以及20%为测试集(64张),并通过10折交叉验证的方式对模型进行验证。由于训练集图像数量较少,因此使用水平翻转、水平位移、高度位移以及图像旋转等操作进行数据增强至2 600张。

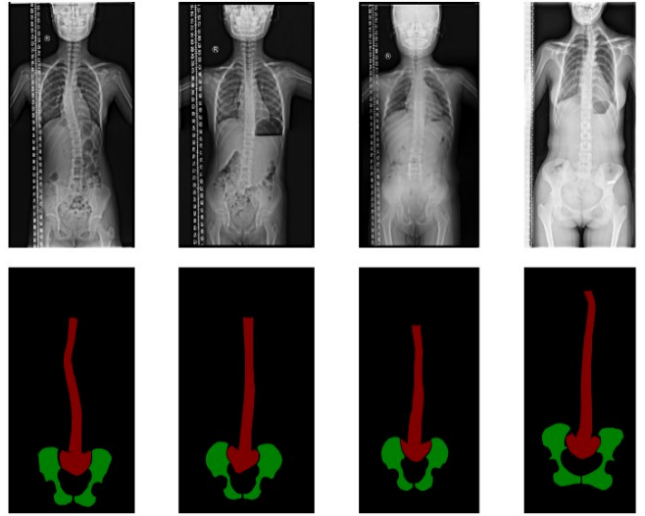


图5 脊柱X光图像数据集

Figure 5 Examples from the spine X-ray image dataset

2.2 模型训练

本研究实验使用TensorFlow2.0^[21]深度学习框架,系统处理器为Intel W2265,内存为32 GB;同时使用英伟达GPU显卡加速网络运算速度,型号为NVIDIA GTX 2080,显存为8 GB。使用Adam^[22]优化器对网络参数进行优化,初始学习率为0.001,超参数 β_1 、 β_2 为默认值0.9与0.999 9。训练时批处理大小设置为3,迭代次数为150。本研究使用基于模型修改的混合二元交叉熵作为损失函数。对于评价指标,本研究主要使用Dice系数与敏感度(Sensitivity, SE)、特异性(Specificity, SP)、精准率(Precision, PR)进行评价^[5,23],其表达式如下:

$$\text{Dice} = \frac{2|\mathbf{R}_1 \cap \mathbf{R}_2|}{|\mathbf{R}_1| + |\mathbf{R}_2|} \quad (5)$$

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{SP} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (7)$$

$$\text{PR} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

式中, \mathbf{R}_1 代表模型预测矩阵, \mathbf{R}_2 代表标签标注矩阵。TP、FP、FN、TN分别代表预测图中像素为真阳性、假阳性、假阴性、真阴性的概率。

Dice系数是医学图像分割的标准算法,能够显

示模型分割性能;敏感度与漏诊率相关,当敏感度越高时,漏诊率越低,即图像中欠分割的区域越小;特异性与诊断中误诊率相关,特异性越高,在疾病诊断时误诊越低,即图像中过分割的区域越小;准确度指的是预测被分割物体正确分类的比例。以上评估函数分割部分均为脊柱含骶骨与髌骨的正确分类概率的平均值。

2.3 实验结果

首先设置 3 组消融实验分别对提出的各个模块进行实验测试,其中表 1 展示了各模块对 DAU-Net 模型性能的影响。使用 U-Net 模型作为消融实验的基准模型且对以下模型进行对比:(1)将 U-Net 模型中的卷积层替换为图 2 所示的卷积残差模块,实验中称为 ResU-Net+;(2)在 ResU-Net+ 的基础上,添加 SAE、SAD 模块,在 SAD 中仅使用相对空间位置编码的方式,而不使用空间通路的信息对 SAD 模块进行信息弥补,实验中称为 AResU-Net+;(3)在 ResU-Net+ 的基础上添加空间信息网络部分,使用逐元素相加的方式对空间信息网络与语义信息网络的特征进行融合,称为 DU-Net;(4)在 DU-Net 的基础上添加 SAE、SAD 模块,即本研究提出的 DAU-Net。

从表 1 中的各项指标上得出,ResU-Net+模型在添加卷积残差模块后,分割精度随着网络深度的加深有一定的提升。AResU-Net+的Dice、敏感度、精准度有所下降,特异性有所提升,在不引入空间信息网络的条件下,效果不如 ResU-Net+,但 Dice 系数依然比原始 U-Net 提高 2.1%。而在 ResU-Net+网络中添加空间信息网络后,DU-Net 模型能够提高分割精度 Dice、敏感度与准确率的结果,但是在特异性上反而有所下降。在添加了自注意力机制后,DAU-Net 所有评估指标均有进一步的提升,证明自注意力机制与空间信息网络可以有效提升模型分割精度。

表 1 采用不同策略消融实验结果(%)

模型	Dice	敏感度	特异性	精准率
U-Net	87.89±3.29	96.25±2.63	98.91±2.05	94.18±1.95
ResU-Net+	90.10±4.67	96.73±2.35	98.97±1.40	95.33±1.35
AResU-Net+	89.99±2.68	96.56±1.28	99.06±1.44	93.78±2.06
DU-Net	90.35±3.25	96.91±1.79	98.36±1.47	95.74±1.85
DAU-Net	91.89±2.42	97.47±1.40	99.27±1.12	96.11±1.76

本研究同时测试了自注意力模块中 MHSA 超参数 n 对结果的影响,分别测试了 $n=2,4,6$ 时模型的表现,实验结果如表 2 所示,实验表明参数 $n=4$ 的时

候,Dice、敏感度、特异性以及精准率均达到最好效果。

表 2 多头自注意力参数对精度影响(%)
Table 2 Effect of MHSA parameters on accuracy (%)

模型	Dice	敏感度	特异性	精准率
$n=2$	90.67±2.57	97.23±1.77	99.18±1.67	95.65±2.05
$n=4$	91.89±2.42	97.47±1.40	99.27±1.12	96.11±1.76
$n=6$	91.15±2.68	97.13±1.28	99.20±1.44	95.61±2.06

此外,同样验证了本研究提出的损失函数里超参数 α 对结果的影响。测量 α 从 0 开始,以 0.1 间隔到 $\alpha=1.0$ 为止模型的性能,实验结果如图 6 所示,实验结果受到 α 影响较小,但当 $\alpha=0.3$ 时,实验结果取得最优。同时,本研究对比了 DAU-Net 与 U-Net^[2]、ResU-Net^[24]、Attention U-Net^[25]、U-Net++^[26] 算法在测试集上的分割结果,数据对比如表 3 所示。表 3 中的数据显示本研究算法在不同的评估函数下均取得领先成绩,充分证明了该算法在全脊柱 X 光图像多区域分割的可靠性。

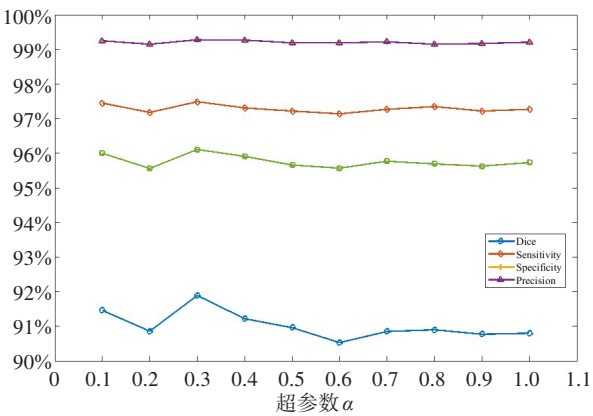


图 6 超参数 α 对性能的影响
Figure 6 Effects of hyperparameter α on performance

表 3 不同算法在测试集上的结果(%)

模型	Dice	敏感度	特异性	精准率
U-Net	87.89±3.29	96.25±2.63	98.91±2.05	94.18±1.95
ResU-Net	89.99±3.17	96.63±2.01	98.54±1.50	95.10±1.89
Attention U-Net	87.29±4.18	95.54±2.20	98.08±1.62	93.98±2.27
U-Net++	90.70±2.62	97.15±1.79	98.81±1.04	95.48±1.35
DAU-Net	91.89±2.42	97.47±1.40	99.27±1.12	96.11±1.76

实际分割效果图如图 7 所示,可以看出 DAU-Net 在脊柱与骶骨部分分割效果更加精准。如图 7 中黄

色框体所示,DAU-Net存在较少的过分割现象,能够在髌骨部分下侧对于两个髌骨之间的区域表现更加明显,能够准确区分目标骨骼与背景。同时,图7白

色框体显示,相比其他模型,DAU-Net有较少欠分割的情况,在骶骨区域分割的粘连程度较小,在脊柱区域分割也比其他算法更加完整。

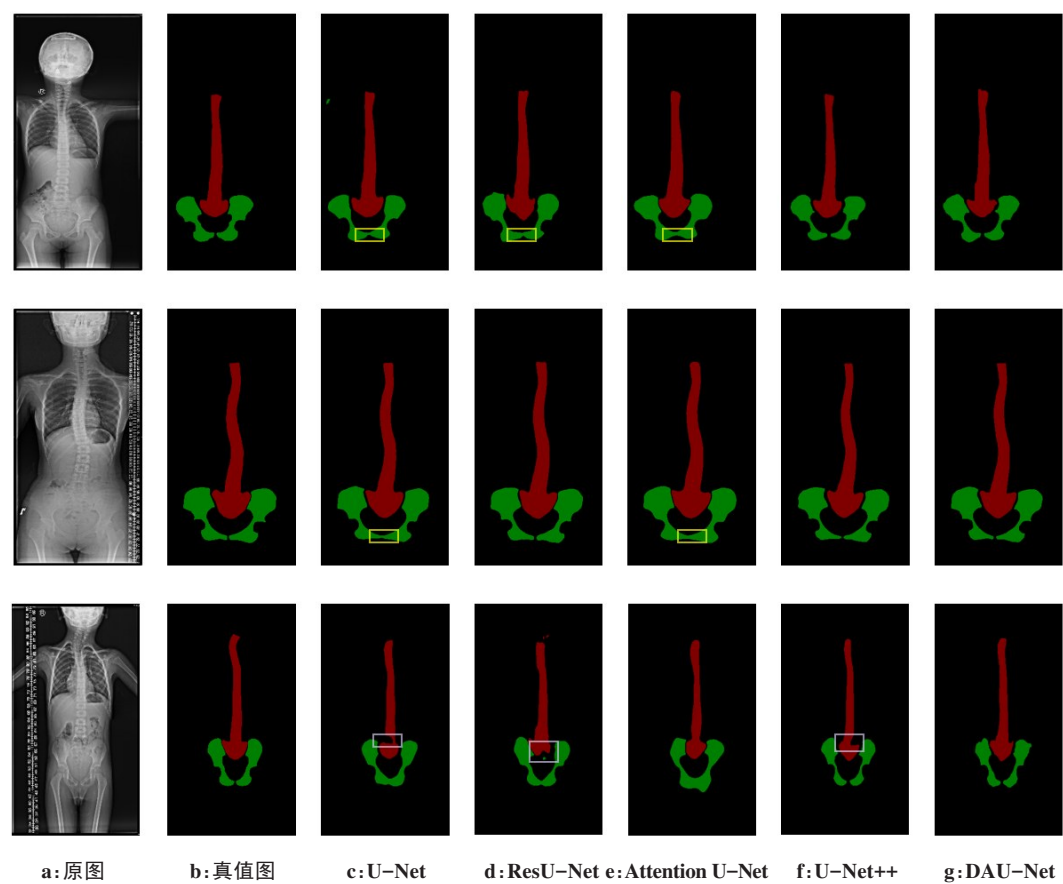


图7 不同算法分割结果
Figure 7 Segmentation results of different methods

3 结 语

本研究提出一个基于双通道网络、卷积残差模块与自注意力机制的深度学习模型,用于全脊柱X光图像语义分割。该模型通过双通道网络分别学习X光图像的空间特征信息与语义特征信息,并在语义信息网络中使用自注意力机制提取图像中远距离特征信息之间的关联,提升特征提取能力;在空间信息网络中使用大步伐的空洞卷积替代池化层,在扩大视野域的同时保留更多的空间位置信息,进一步提升分割精度。实验结果证明DAU-Net在复杂的全脊柱X光图像中实现更加精准的语义分割,在脊柱边缘与两个髌骨粘连部分分割效果有明显改善。下一步工作,研究方向将会在数据集中添加左右倾斜(Bending)状态数据集,继续提升模型的精准程度,为医生进行脊柱侧弯决策诊断分型提供更加全面的语义分割信息内容的支撑。

【参考文献】

[1] 邱贵兴, 庄乾宇. 青少年特发性脊柱侧弯的流行病学研究进展[J]. 中华医学杂志, 2006, 86(11): 790-792.
Qiu GX, Zhuang QY. The epidemiological research progress of adolescent idiopathic scoliosis[J]. Chinese Medical Journal, 2006, 86 (11): 790-792.

[2] 郭春麟, 张勇, 刘伟, 等. 脊柱X射线图像分割方法[J]. 国外电子测量技术, 2020, 41(7): 23-28.
Guo CL, Zhang Y, Liu W, et al. Spinal X-ray image segmentation method[J]. Foreign Electronic Measurement Technology, 2020, 41(7): 23-28.

[3] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-assisted Intervention. Springer: Cham, 2015: 234-241.

[4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.

[5] 张宇. 基于深度学习技术的视网膜血管分割[D]. 广州: 华南理工大学, 2020.
Zhang Y. Segmentation of retinal vessel based on deep learning technology[D]. Guangzhou: South China University of Technology, 2020.

[6] Wu H, Bailey C, Rasoulinejad P, et al. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer: Cham, 2017: 127-135.

- [7] Grigorieva I, Vyunnik N, Kolpinsky G. The construction of an individualized spinal 3D model based on the X-ray recognition[C]//2018 23rd Conference of Open Innovations Association (FRUCT). IEEE, 2018: 143-149.
- [8] 赵阳, 张俊华. 多尺度特征融合的脊柱X线图像分割方法[J]. 计算机工程与应用, 2021, 57(8): 214-219.
Zhao Y, Zhang JH. Spine X-ray image segmentation method based on multi-scale feature fusion[J]. Computer Engineering and Applications, 2021, 57(8): 214-219.
- [9] Han Z, Wei B, Mercado A, et al. Spine-GAN: semantic segmentation of multiple spinal structures[J]. Med Image Anal, 2018, 50: 23-35.
- [10] 周瑞文. 基于U-net的卷积神经网络的脊柱MRI图像分割[J]. 绿色科技, 2021, 23(8): 3.
Zhou RW. Spinal MRI image segmentation based on U-net convolutional neural network [J]. Green Technology, 2021, 23(8): 3.
- [11] 金顺楠, 周迪斌, 何斌, 等. 基于多尺度特征融合与注意力机制的脊柱CT图像分割[J]. 计算机系统应用, 2021, 30(10): 280-286.
Jin SN, Zhou DB, He B, et al. Spine CT image segmentation based on multi-scale feature fusion and attention mechanism [J]. Computer Systems & Applications, 2021, 30(10): 280-286.
- [12] 张辰翰. 基于深度学习的脊柱CT图像分割[D]. 成都: 电子科技大学, 2021.
Zhang CH. Spine CT image segmentation based on deep learning[D]. Chengdu: University of Electronic Science and Technology of China, 2021.
- [13] 毛孝鑫, 宋烨, 郝泳涛. 基于Transformer的脊柱CT图像分割[J]. 电脑知识与技术, 2021, 17(20): 124-126.
Mao XX, Song Y, Hao YT. CT image segmentation of the spine based on Transformer[J]. Computer Knowledge and Technology, 2021, 17(20): 124-126.
- [14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.
- [15] Gao Y, Zhou M, Metaxas DN. Utnet: a hybrid transformer architecture for medical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 61-71.
- [16] Valanarasu JM, Oza P, Hacıhaliloglu I, et al. Medical transformer: gated axial-attention for medical image segmentation [J]. arXiv preprint arXiv: 2102.10662, 2021.
- [17] Ji Y, Zhang R, Wang H, et al. Multi-compound transformer for accurate biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2021: 326-336.
- [18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [19] Bello I, Zoph B, Vaswani A, et al. Attention augmented convolutional networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3286-3295.
- [20] Torralba A, Russell BC, Yuen J. Labelme: online image annotation and applications[J]. P IEEE, 2010, 98(8): 1467-1484.
- [21] Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems [J]. arXiv preprint arXiv: 1603.04467, 2016.
- [22] Kingma DP, Ba J. Adam: a method for stochastic optimization [J]. Neural Networks, 2014, 15(4): 95-103.
- [23] 陈颖颢, 李晗, 周雪婷, 等. 融合空洞卷积与注意力的胃癌组织切片分割[J]. 中国图象图形学报, 2021, 26(9): 2281-2292.
Chen YS, Li H, Zhou XT, et al. Segmentation of gastric cancer tissue slices combining hollow convolution and attention [J]. Journal of Image and Graphics, 2021, 26(9): 2281-2292.
- [24] Diakogiannis FI, Waldner F, Caccetta P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data [J]. ISPRS, 2020, 162: 94-114.
- [25] Oktay O, Schlemper J, Folgoc LL, et al. Attention U-net: learning where to look for the pancreas[J]. arXiv preprint arXiv: 1804.03999, 2018.
- [26] Zhou Z, Siddiquee MM, Tajbakhsh N, et al. Unet++: redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Trans Med Imaging, 2019, 39(6): 1856-1867.

(编辑:陈丽霞)