

基于 VGG-Net 的 X 射线全脊柱冠状面图像分割方法^{*}

申学泉¹ 张勇² 张润杰² 石琼芳³ 宋宇峰² 张权¹

(1. 中北大学生物医学成像与影像大数据山西省重点实验室 太原 030051; 2. 山西医科大学第二医院 太原 030001;
3. 太原市杏花岭区医疗集团中心医院 太原 030002)

摘要:在计算机辅助脊柱图像分析和疾病诊断应用中,从 X 射线脊柱图像中自动分割脊柱和椎骨是一个关键且具有挑战性的问题。为进一步提升脊柱图像分割精度,提出一种基于 VGG-Net 改进的模型。首先,将 VGG16 网络去掉了后面的全连接层,用作 U-Net 的特征提取网络;其次,为了增强图像的细节信息,在特征提取网络引入小波分解模块;最后,在上采样网络中设计了一种逐像素相减的自空间注意力模块(SUB-SSAM)机制,进一步提高网络模型识别关键特征的能力。实验结果表明,改进后的模型相较于原 VGG-Net 模型在平均交并比(mIoU)上提高了 2.39%、召回率(recall)提高了 0.96%、准确率(accuracy)提高了 1.31%,训练的该网络模型可以定位到每一块椎骨,准确分割椎体区域。

关键词:图像分割;U-Net;VGG-Net;小波分解;SUB-SSAM

中图分类号: TP391.4 文献标识码:A 国家标准学科分类代码: 510.40

Segmentation method of X-ray whole spine coronal image based on VGG-Net

Shen Xuequan¹ Zhang Yong² Zhang Runjie² Shi Qiongfang³ Song Yufeng² Zhang Quan¹

(1. Key Laboratory for Biomedical Imaging and Big Data of Shanxi Province, North University of China, Taiyuan 030051, China; 2. Second Hospital of Shanxi Medical University, Taiyuan 030001, China; 3. Central Hospital of Xinghualing District Medical Group, Taiyuan 030002, China)

Abstract: Automatic segmentation of the spine and vertebrae from X-ray spine images is a crucial and challenging task in computer-aided spine image analysis and disease diagnosis applications. To further improve the accuracy of spine image segmentation, this paper proposes an improved model based on VGG-Net. Firstly, the VGG16 network is modified by removing the fully connected layers and used as the feature extraction network for U-Net. Secondly, to enhance the detail information of the images, a wavelet decomposition module is introduced into the feature extraction network. Finally, a self-subtracted spatial self-attention module (SUB-SSAM) mechanism is designed in the upsampling network to enhance the network's ability to identify key features. Experimental results show that the improved model achieves a 2.39% improvement in mean intersection over union (mIoU), a 0.96% improvement in recall, and a 1.31% improvement in accuracy compared to the original VGG-Net model. The trained network model can accurately locate each vertebra and segment the vertebral area.

Keywords: image segmentation; U-Net; VGG-Net; wavelet decomposition; SUB-SSAM

0 引言

脊柱在人体中扮演着极其重要的角色^[1]。脊柱疾病在全球范围有很高的发病率,同时由于脊柱结构及其功能的复杂性,医生对脊柱疾病的诊断和评估是一个难点,通常根据医学指南或个人经验测量估计疾病参数。为了获

取准确的病灶信息,医生往往会在图像上做一些关键区域标注或者分割的工作,手动标注很耗时,而且测量的参数是可变的^[2]。

随着人工智能以及深度学习的发展,各种深度学习模型用于图像分割。Long 等^[3]首次提出全卷积神经网络(fully convolution network, FCN),将卷积神经网络中的

收稿日期:2023-07-21

* 基金项目:山西省应用基础研究计划项目(201901D111153)、生物医学成像与影像大数据山西省重点实验室开放研究基金项目资助

全连接层用卷积层代替,可以接受任意尺寸的输入图像,实现了端到端训练,像素到像素的预测,但得到的结果不够精细,没有充分考虑像素与像素的关系。Ronneberger 等^[4]提出了 U-Net 模型,模型的上采样和下采样结构能够实现信息的全局感知和局部细节的恢复,跳跃连接将相应层的特征图拼接,保留更多的语义信息,但模型容易过拟合,分割结果边界模糊。Badrinarayanan 等^[5]提出的 Seg-Net 模型在 VGG16 的基础上将对应编码器层的最大池化过程的池化索引进行对应解码器层的非线性上采样,产生的图是稀疏的,需要的内存和训练时间更少,但在上采样过程中有细节信息丢失,导致分割结果不够精准,并且对大尺寸图像的分割比较困难。Zhou 等^[6]提出的 U-Net++用短连接代替长连接,将不同尺度的特征进行特征叠加的方式整合,模型精度得到了很大的提高,结合深监督可以使参数量在精度可接受的范围内缩减。Peng 等^[7]提出的全局卷积网络(global convolutional network, GCN)模型整体采用上下采样的结构,在整体加入全局的可分离卷积模块和边界对齐模块,它同时考虑了分类与定位两个方面,但模型需要更多的计算成本,并且在小数据集上训练难度大。Chen 等^[8]提出的 DeepLabv3+模型在并行的带有空洞卷积的空间金字塔池化模块的基础上,加入解码器模块细化分割,但存在模型参数量庞大,不利于训练的问题。郭春麟等^[9]在 U-Net 模型的基础上用相邻两层跳跃连接代替单连接,并在上采样引入双通道注意力模块(CBAM),对图像深层信息进行注意力监督。上述模型中,改进后的 U-Net 网络在医学图像分割领域中有更加优异的分割性能,但在医学图像边缘信息的处理还有待提高,医学图像的细节特征信息利用还不够完善。

离散小波变换(DWT)有助于将输入分解为多个子带分量。事实上,DWT 已经被用于图像处理中,Sun 等^[10]将

DWT 和离散小波逆变换(IDWT)代替传统的下采样和上采样,在不损失信息的情况下有效地增加了网络感受野。Guo 等^[11]提出深度小波超分辨率(DWSR)作为预测小波与高分辨率图像的方法,利用一级 DWT 将超分辨率问题转化为小波系数的预测问题,由于小波系数带来的稀疏性,提高了模型的性能。Singh 等^[12]将 DWT 用于编码器的输入,来提取频域中的显著特征,可以将低频和高频分量及可以忽略的伪影分离开。

何晓云等^[13]在 U-Net 模型的解码器中嵌入注意机制,加大目标特征的权重,减缓无用信息的干扰;纪秋浪等^[14]提出一种多尺度注意力网络(MDA-Net)架构,引入多空洞残差双注意力模块,增强了图像细节特征的表征能力;邱云飞等^[15]在 DeepLabV3+模型中采用通道注意力融合高级特征,采用空间注意力提取低级特征,获取丰富的上下文信息,过滤背景信息,减少细节信息的丢失。

综上所述,为提高模型在医学图像分割任务中的性能,本文综合利用 DWT 和注意力等机制提出一种基于 VGG-Net 改进的模型。首先,在编码器引入小波分解模块,不仅进一步加强特征的提取,还有效丰富了特征;其次,在文献[16]提到的一对非对称卷积基础上,设计一种逐像素相减的自空间注意力模块(subtraction self-spatial attention module, SUB-SSAM),旨在提高模型识别关键特征的能力,在不改变网络深度的同时提升网络性能。

1 改进后的整体模型结构

本文在保持网络性能的前提下,将小波分解引入特征提取网络,增强网络对细节信息的提取,其次在上采样网络每一层后加入 SUB-SSAM 模块,使模型更加关注重要的特征,减弱无关特征的权重大小,模型整体结构如图 1 所示。

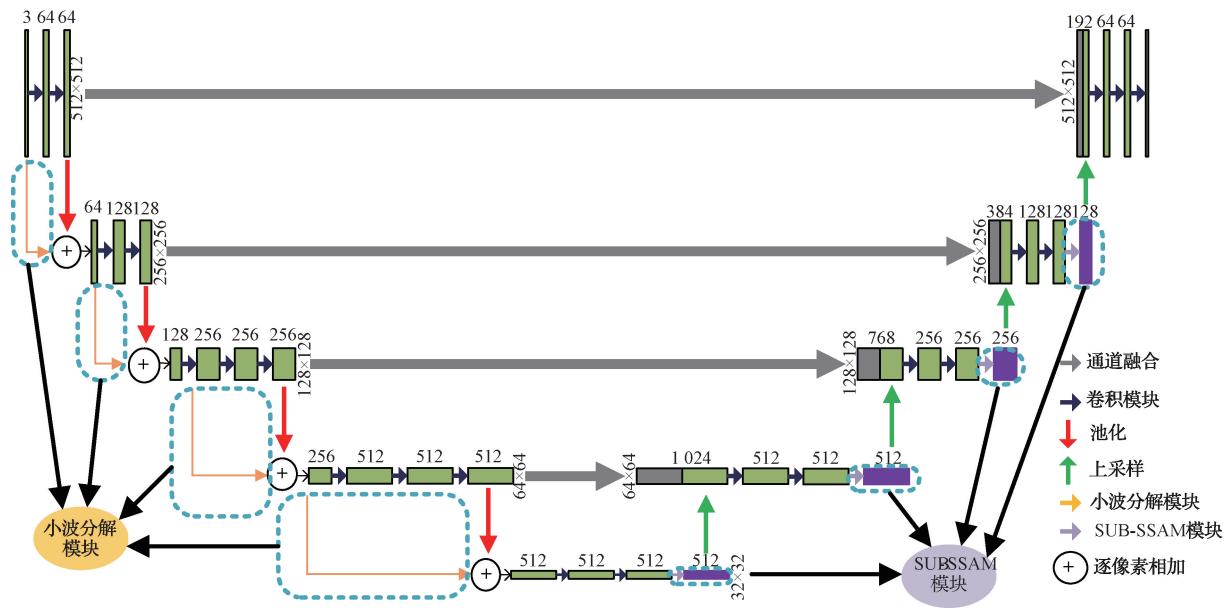


图 1 模型整体结构

2 算法设计

2.1 小波分解模块

特征提取主干采用 VGG16 替代原始 U-Net 下采样模块,每一层由池化层和卷积组成,具体 VGG16 提取网络模块参数如表 1 所示。在此基础上,在每一层的卷积之前(除第 1 层),池化层之后引入一个小波分解模块,得到的输出与池化层的尺寸通道数一致,并进行逐像素相加。小波分解模块具体结构如图 2 所示。

表 1 改进后的 VGG 网络模块参数

层	操作	输出
Layer1	(Conv2d+ReLU) × 2	512 × 512 × 64
	Max_pool	256 × 256 × 64
Layer2	DWT, (Conv2d+ReLU) × 2	256 × 256 × 128
	Max_pool	128 × 128 × 64
Layer3	DWT, (Conv2d+ReLU) × 3	128 × 128 × 256
	Max_pool	64 × 64 × 256
Layer4	DWT, (Conv2d+ReLU) × 3	64 × 64 × 512
	Max_pool	32 × 32 × 512
Layer5	DWT, (Conv2d+ReLU) × 3	32 × 32 × 512

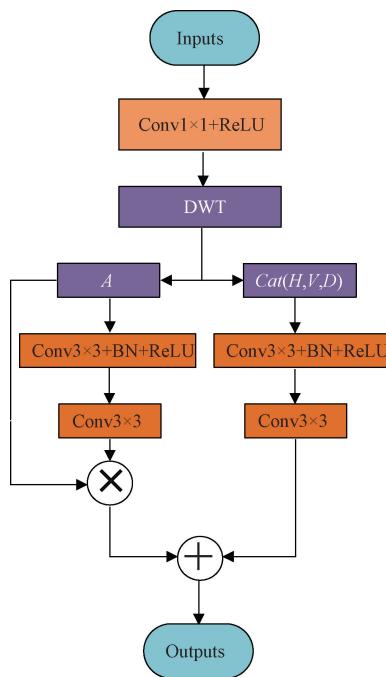


图 2 小波分解模块

DWT 可以将输入信息分解为多个分量,DWT 利用多个滤波器组以不同的分辨率分离特征向量的时间和频率分量。本文通过使用 4 个卷积 Haar 滤波器来实现二维的 DWT。这些滤波器或内核可以表示为 $k_{LL} = [1 \ 1; 1 \ 1]^T$, $k_{HH} = [-1 \ -1; 1 \ 1]^T$, $k_{HL} = [-1 \ 1; -1 \ 1]^T$ 和 $k_{LH} = [1 \ -1; 1 \ -1]^T$, 将输入图像 I 分解为 4 个子带,即 I_{LL} , I_{HL} , I_{LH} , 和 I_{HH} ,也称作平均分量 A 、水平分量

H 、垂直分量 V 和对角分量 D 。

$$A = I_{LL}(i,j) = I(2i-1,2j-1) + I(2i-1,2j) + I(2i,2j-1) + I(2i,2j) \quad (1)$$

$$H = I_{HL}(i,j) = -I(2i-1,2j-1) + I(2i-1,2j) - I(2i,2j-1) + I(2i,2j) \quad (2)$$

$$V = I_{LH}(i,j) = -I(2i-1,2j-1) - I(2i-1,2j) + I(2i,2j-1) + I(2i,2j) \quad (3)$$

$$D = I_{HH}(i,j) = I(2i-1,2j-1) - I(2i-1,2j) - I(2i,2j-1) + I(2i,2j) \quad (4)$$

本文首先将输入图像通过一个 1×1 的卷积升维,使其可以匹配下一层输入的通道数,其次将分离出的 H 、 V 、 D 子带图像进行通道拼接,与 A 子带图像分别进行卷积:

$$A, H, V, D = DWT(f_1(Inputs)) \quad (5)$$

$$Outputs = f_3(f_2(A)) \times A + f_3(f_2(Cat(H, V, D))) \quad (6)$$

式中: $Inputs$ 和 $Outputs$ 分别表示输入和输出; $f_1(\cdot)$ 表示 1×1 的卷积和 ReLU 激活函数; $DWT(\cdot)$ 表示离散小波变换; $f_2(\cdot)$ 表示 3×3 的卷积、BN 层和 ReLU 激活函数; $f_3(\cdot)$ 表示 3×3 的卷积;Cat(\cdot)表示进行通道拼接。

通过小波的二级分解增加了提取特征的多样性以及丰富性,从而可以更准确的提取图像特征。

2.2 SUB-SSAM 模块

Paszke 等^[16]设计的 Enet 模型将一对非称卷积加入到深度学习中,减弱了函数学习的过拟合的同时增加了感受野,并且卷积核的分解减少了大量的参数。在此基础上,本文尝试加入一种逐元素相减的方式削弱无关特征,加强重要特征,并结合设计的一种自空间注意力机制(SSA),SUB-SSAM 模块具体结构如图 3 所示。

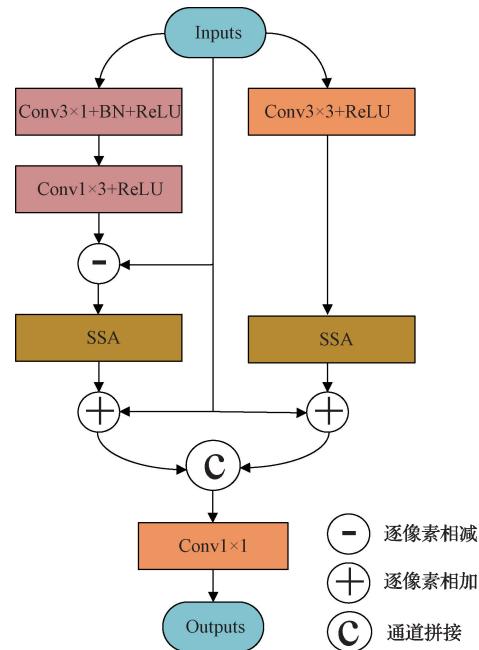


图 3 SUB-SSAM 模块结构

$$\text{Output1} = f_{\text{SSA}}((f_1(\text{Inputs}) - \text{Inputs}) + \text{Inputs}) \quad (7)$$

$$\text{Output2} = f_{\text{SSA}}(f_2(\text{Inputs}) + \text{Inputs}) \quad (8)$$

$$\text{Outputs} = f_3(\text{Cat}(\text{Output1}, \text{Output2})) \quad (9)$$

式中: Output1 和 Output2 分别表示左支路和右支路到通道拼接前的输出; f_{SSA} 表示自空间注意力模块, 结构如图 4 所示; f_1 表示一对 3×3 的非对称卷积; f_2 和 f_3 分别表示 3×3 的卷积和 1×1 的卷积; Cat 表示进行通道拼接。

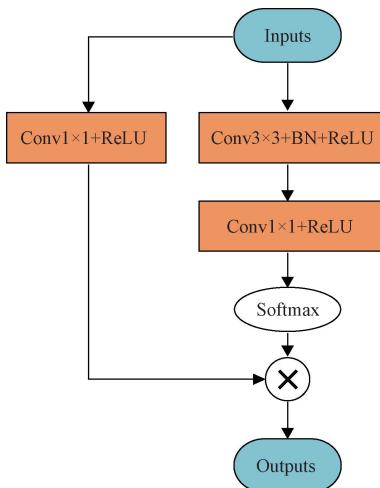


图 4 SSA 模块结构

3 实验结果分析

3.1 数据集及预处理

本文实验数据集为通过 labelme 软件进行手动标注的医学脊柱图像数据集, 标注后的标签转化成 VOC 格式。通过平移、镜像、旋转等方式对数据集进行扩充、增强, 该数据集共包含 1 638 张图像, 并按照 9 : 1 的比例划分训练集和测试集, 每一张图像的尺寸约为 3400×7300 , 在输入网络之前经过预处理, 采用在两侧添加灰度条不失真压缩的方式, 转化成尺寸大小为 512×512 的图像, 压缩后的图像如图 5 所示。



图 5 压缩后的图像

3.2 训练与评价指标

实验环境为 Intel(R) Core(TM) i7-9700K CPU@3.60 GHz

3.60 GHz, GPU 为 NVIDIA GeForce RTX 2070 SUPER, 内存 32 GB, 采用 Python3.7 和 Pytorch1.7.1。

本文模型输入为 $512 \times 512 \times 3$ 的 RGB 图像, 采用迁移学习的思想, 将预训练好的模型权重导入模型, 训练总次数 epoch 为 150 次, 前 40 次为冻结主干训练, 特征提取网络不发生变化, 仅对网络进行微调。Batchsize 和 Freeze_Batchsize 均设置为 2, 初始学习率为 10^{-6} , 每 10 次 epoch 对模型进行评估一次。如图 6 所示, 模型的训练损失和验证损失随迭代次数变化, 在 120 次迭代时趋于平稳, 模型收敛性良好。

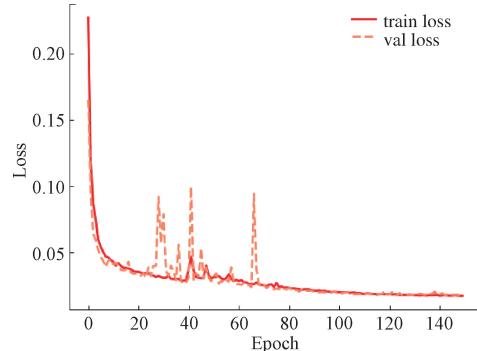


图 6 模型的损失曲线

实验采用平均交并比 (mean intersection over union, mIoU)、召回率 (recall)、准确率 (accuracy) 来客观评价算法效果。这 3 个评价指标是基于混淆矩阵进行计算, 混淆矩阵如表 2 所示。

表 2 混淆矩阵

真实情况	预测结果	
	正例	反例
正例	真正例 (true positive, TP)	假反例 (false negative, FN)
反例	假正例 (false positive, FP)	真反例 (true negative, TN)

各项指标在混淆矩阵的基础上计算如下:

$$IoU = \frac{TP}{TP + FP + FN} \quad (10)$$

$$mIoU = \frac{\sum (IoU)}{\text{类别数}} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

式中: TP 表示正例被正确预测的个数; FP 为反例被错误预测的个数; FN 为正例被错误预测的个数; TN 为反例被正确预测的个数。

3.3 消融实验

本文提出改进后的 VGG-Net 模型, 主要在编码器和

解码器方面进行改进。为了评估不同模块的改进以及改进模块直接的组合对模型性能的影响,设计一系列消融实验,使用 mIoU、Recall 和 Accuracy 作为评价指标,实验结果如表 3 所示。

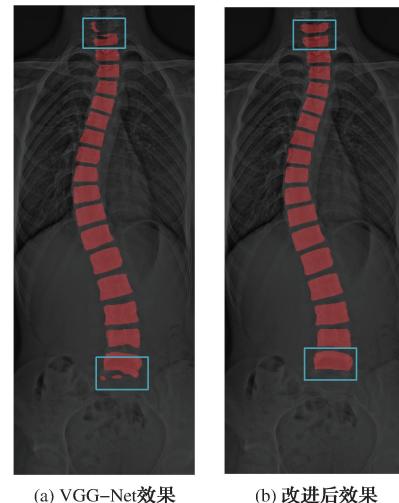
表 3 消融实验 (%)

模型	mIoU	Recall	Accuracy
VGG-Net	91.89	95.73	95.81
VGG-Net+SUB-SSAM	92.16	95.91	95.93
VGG-Net+小波	94.06	96.50	97.24
VGG-Net+小波+SUB-SSAM	94.28	96.87	97.38

由表 3 可知,本文提出的改进方法中,在编码器中引入小波分解模块,模型 mIoU、Recall 和 Accuracy 分别提升了 2.17%、0.77% 和 1.43%;解码器引入 SUB-SSAM 模块时,模型 mIoU、Recall 和 Accuracy 分别提升了 0.27%、0.18% 和 0.12%。将所有的改进的方法同时加入模型后,模型整体的 mIoU、Recall 和 Accuracy 分别提高了 2.39%、1.14% 和 1.57%。从实验结果分析可知,小波分解模块在分割精度及准确率上都有一定程度的提高,而 SUB-SSAM 在指标上虽然有所提高,但是提升不明显。图 7(a)为 VGG-Net 模型分割图,图 7(b)为加入小波和 SUB-SSAM 模块后的分割图。

3.4 实验结果与分析

不同模型的分割效果如图 8 所示,将分割图与原图 3 : 7 的比例融合,红色区域表示分割区域,所有模型对脊



(a) VGG-Net效果 (b) 改进后效果

图 7 模型前后的图像对比

柱区域均有较好的分割效果,在细节方面易出现欠分割、过分割、误分割和椎体连通等问题,但本文改进后的模型在分割细节上要优于其他模型。其中,Deeplabv3+、U-Net、GCN 3 个模型都存在胸椎欠分割的问题,Deeplabv3+模型腰椎过分割,U-Net 模型误分割、腰椎连通以及 GCN 模型腰椎连通。相对而言,本文改进后的模型未出现椎体连通区域和误分割区域,整体呈现良好的分割效果,区域完整视觉上更接近真实的标注结果。

由表 4 可知,对于 mIoU 指标,本文改进的模型的 93.9% 显著高于 Deeplabv3+、U-Net、GCN 模型的

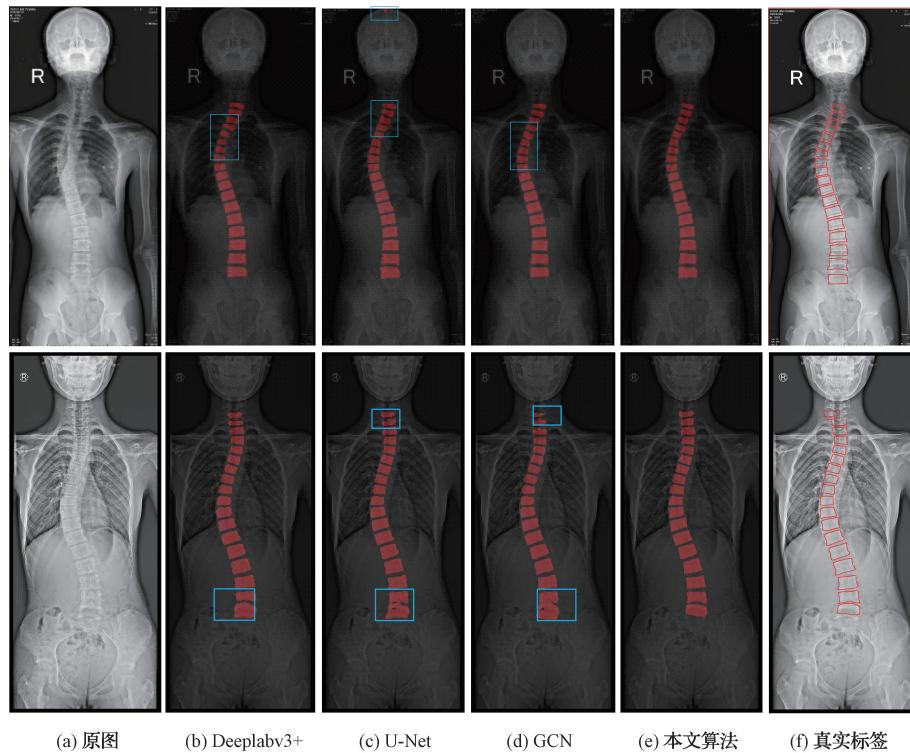


图 8 不同模型分割结果

88.81%、86.03%、86.34%；对于 Recall 指标，所改进的模型比 Deeplabv3+、U-Net、GCN 模型分别提高了 5.47%、8.25%、7.94%；对于 Accuracy 指标，所改进的模型比 Deeplabv3+、U-Net、GCN 模型分别提高了 2.57%、5.74%、4.75%。所提模型的有效性进一步得到验证。

表 4 不同模型的评价指标对比 (%)

模型	mIoU	Recall	Accuracy
Deeplabv3+	88.81	93.35	94.81
U-Net	86.03	93.36	91.64
GCN	86.34	92.71	92.63
本文算法	94.28	96.87	97.38

4 结 论

为了提高 X 射线全脊柱冠状面图像的分割精度的同时减少欠分割、误分割等问题，本文提出了一种基于 VGG-Net 改进的模型。该模型采用编解码结构，通过在编码器中引入小波分解模块有效地融合频域信息，获取更加丰富的特征，同时在解码器中引入 SUB-SSAM 模块，加强关键特征，削弱无关特征。实验在自制的数据集下进行，实验结果表明，本文算法可以精准分割椎体且解决了欠分割、误分割等问题。但在医学图像脊柱指标计算中，骶骨区域也需要精准分割出来，因此进一步优化边缘细节以及将骶骨区域分割出来是未来主要的工作。

参 考 文 献

- [1] HORNG M H, KUOK C P, FU M J, et al. Cobb angle measurement of spine from X-ray images using convolutional neural network[J]. Computational and Mathematical Methods in Medicine, 2019(9):1-18.
- [2] HAN Z, WEI B, MERCADO A, et al. Spine-GAN: Semantic segmentation of multiple spinal structures[J]. Medical image analysis, 2018, 50: 23-35.
- [3] 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.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference. Springer, 2015: 234-241.
- [5] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [6] ZHOU Z, SIDDIQUEE M M R, TAJBAKHSH N, et al. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 39(6): 1856-1867.
- [7] PENG C, ZHANG X, YU G, et al. Large kernel matters—improve semantic segmentation by global convolutional network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4353-4361.
- [8] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 801-818.
- [9] 郭春麟, 张勇, 刘祎, 等. 脊柱 X 射线图像分割方法[J]. 国外电子测量技术, 2022, 41(7):23-28.
- [10] SUN C, LAI H, WANG L, et al. Efficient attention fusion network in wavelet domain for denoising[J]. IEEE Access, 2021, 9: 53392-53400.
- [11] GUO T, SEYED MOUSAVI H, HUU V T, et al. Deep wavelet prediction for image super-resolution[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 104-113.
- [12] SINGH V K, KALAFI E Y, WANG S, et al. Prior wavelet knowledge for multi-modal medical image segmentation using a lightweight neural network with attention guided features[J]. Expert Systems with Applications, 2022, 209: 118166.
- [13] 何晓云, 许江淳, 陈文绪. 基于改进 U-Net 网络的眼底血管图像分割研究[J]. 电子测量与仪器学报, 2021, 35(10):202-208.
- [14] 纪秋浪, 王继红, 杨晨, 等. 多尺度双重注意力网络医学图像分割模型[J]. 国外电子测量技术, 2022, 41(6):65-71.
- [15] 邱云飞, 温金燕. 基于 DeepLabV3+ 与注意力机制相结合的图像语义分割[J]. 激光与光电子学进展, 2022, 59(4):130-139.
- [16] PASZKE A, CHAURASIA A, KIM S, et al. Enet: A deep neural network architecture for real-time semantic segmentation[J]. Computer Science, 2016, DOI:10.48550/arXiv.1606.02147.

作 者 简 介

申学泉,硕士研究生,主要研究方向为图像处理。

E-mail:shenxuequan2023@163.com

张权,博士,副教授,主要研究方向为图像处理、科学可视化。

E-mail:zhangibmet@nuc.edu.cn