

基于多尺度特征融合与注意力机制的脊柱 CT 图像分割^①



金顺楠¹, 周迪斌¹, 何斌², 顾静军³

¹杭州师范大学 杭州国际服务工程学院, 杭州 311121)

²(杭州三坛医疗有限公司, 杭州 310018)

³(浙江大学, 杭州 310058)

通讯作者: 金顺楠, E-mail: jsn794177840@163.com

摘要: 本文针对医学脊柱 CT 图像因骨密度不均匀、骨骼结构复杂或图像成像分辨率低等因素造成的分割精度较低的问题, 提出一种基于卷积-反卷积神经网络的 CT 图像脊柱分割方法. 通过引入多尺度残差模块及注意力机制改进 U-Net 网络, 训练特征模型并进行测试. 在真实数据集上的实验结果表明, 该方法能有效提高 CT 图像中脊柱的分割精度及分割效率, Dice 系数评估值为 0.97, IOU 系数评估值为 0.94.

关键词: 深度学习; U-Net; 残差网络; 注意力机制; 图像分割

引用格式: 金顺楠, 周迪斌, 何斌, 顾静军. 基于多尺度特征融合与注意力机制的脊柱 CT 图像分割. 计算机系统应用, 2021, 30(10): 280-286. <http://www.c-s-a.org.cn/1003-3254/8118.html>

Segmentation of Spine CT Images Based on Multi-Scale Feature Fusion and Attention Mechanism

JIN Shun-Nan¹, ZHOU Di-Bin¹, HE Bin², GU Jing-Jun³

¹(Hangzhou Institute of International Service Engineering, Hangzhou Normal University, Hangzhou 311121, China)

²(Hangzhou Santa Medical Technology Co. Ltd., Hangzhou 310018, China)

³(Zhejiang University, Hangzhou 310058, China)

Abstract: Medical spine CT images have low segmentation accuracy due to uneven vertebral bone density, complex vertebral structure and low imaging resolution. To tackle these problems, this study proposes a segmentation method for spine CT images with a convolutional-deconvolutional neural network. The multi-scale residual module and the attention mechanism are introduced to improve the U-Net network, and the feature model is trained and tested. Experimental results on real data sets show that this method can effectively improve the accuracy and the efficiency of spine CT image segmentation. The estimated results of Dice coefficient and Intersection Over Union (IOU) are 0.97 and 0.94, respectively.

Key words: deep learning; U-Net; residual network; attention mechanism; image segmentation

脊柱是人体不可或缺的重要组成部分, 上联颅骨, 下联髋骨, 具有支持躯干, 保护脊髓、内脏的重要功能. 随着人们生活节奏加快、工作压力繁重, 脊柱疾病患

者呈现出年轻化的趋势, 脊柱疾病的预防和治疗越来越重要. 脊柱解剖结构复杂、边界形状复杂, 计算机断层成像技术 (CT) 因成像快、分辨率高且能清晰呈现

① 基金项目: 浙江省教育厅一般科研项目 (Y202044994); 浙江省研究生教育学会科研项目; 杭州师范大学第二轮专业学位研究生课程教学案例库建设
Foundation item: General Scientific Research Program of the Department of Education of Zhejiang Province (Y202044994); Research Program of the Postgraduate Education Institute of Zhejiang Province; Construction Project of Teaching Case Databases for the Second Round of Professional Postgraduate Courses in Hangzhou Normal University

收稿时间: 2020-12-30; 修改时间: 2021-01-29; 采用时间: 2021-02-08

三维空间结构,而广泛地用于脊柱相关疾病的临床分析、诊断和计算机辅助治疗。

计算机图像辅助诊断技术的引入,可将脊柱CT成像中每帧图像的骨骼区域分割出来,结合三维重建将二维的数据三维化呈现出立体的影像,给医生提供病人脊柱姿态的丰富信息。医生能更加直观地对病人病变区进行观察与剖析,这对于提高脊柱疾病治疗的效率和成功率具有重要的推动作用。图像感兴趣区域的分割是计算机提供诊断辅助的基础,而医学CT图像数据量大,信噪比低,病人之间的差异性大等特点,使其图像分割问题与一般的图像相比,面临更多的课题和挑战。

早期的CT图像脊柱分割大多以基于先验知识和图像信息的传统分割算法为主,Yao等^[1]采用设置骨密度阈值获取初始脊柱分割,并分段拟合椎体模型对椎体进行精分割,但分割结果不能呈现更细化的骨结构。Huang等^[2]提出了一种基于边缘和区域水平集的方法用于椎骨分割,通过提取图像的梯度信息和局部特征,能够减小边界模糊对分割结果的影响。Li等^[3]提出一种基于区域相关性的自动初始化水平集方法用于腰椎CT图像分割,结合形态学滤波器和高斯混合模型,该方法能提取到相对精确的轮廓边界。传统算法一般需要人为干预,处理过程较为繁琐,不能做到完全的自动分割。同时为了保证临床手术实施的安全性,图像引导的脊柱手术通常需要极高的精度,但由于脊柱骨形状和结构的复杂变化,传统的图像分割方法在分割精度上并不满足要求。

近年来,计算机视觉技术飞速发展,卷积神经网络强大的非线性特征提取能力在一系列图像处理任务中取得了突出的表现。如何以最先进的技术为现代医学的进步提供动力,成为计算机科学和医学研究者的探究热点。郭树旭等^[4]将全卷积神经网络引入腹部CT肝脏分割,通过卷积神经网络进行特征提取和特征映射,使其分割结果具有较高的精度。刘忠利等^[5]将卷积-反卷积结构网络模型引入脊柱CT分割,实验证明该方法有较高的实时性和准确性。王亚刚等^[6]将深度残差网络引入U-Net网络实现股骨的自动分割,避免了欠分割与过分割问题,提升了分割效率与精度。李贤等^[7]将3D全卷积神经网络应用于脊柱CT分割,获得了较高的分割精度。采用数据驱动的深度学习方法已在图像算法领域取得了显著成就,但医学图像的分割追求着更高的分割准确率和分割效率。

本文基于深度学习理论与方法,结合对生物医学图像特征的分析,提出一种以U-Net上下采样层跳跃拼接结构为主框架,以多尺度残差块为基础模块,引入注意力机制自适应关注目标特征的分割网络模型。该网络能结合图像的多尺度特征提高分割精度,同时较低的参数计算量不会影响模型的训练速度。实验表明,该网络能实现脊柱CT的全自动分割,且相较于其他相似的经典深度模型,同等环境下该改进模型的准确率和准确效率均有较明显的提高。

1 相关技术

1.1 U-Net 卷积-反卷积框架

U-Net网络模型于2015年被提出^[8],是一种能基于少量图像数据进行训练的端到端的全卷积神经网络。它由编码器和解码器两部分组成,第1条路径用于特征提取,包含重复的卷积操作,通过堆叠池化和 3×3 的卷积来学习图像的特征;第2条路径为对称的扩展路径,通过反卷积操作一层层还原图像大小,利用跳跃连接与特征提取部分对应的特征层融合。与FCN相比,U-Net通过级联两条路径大小相等的特征图可以将高分辨率特征与解码器上采样输出特征连接,实现特征融合,有效地保留原图中的信息。

1.2 残差模块

2015年,He提出了残差网络^[9]这一概念来解决深层网络退化问题。残差模块是残差网络中重要的组成结构,通过使用恒等映射缩短了浅层特征与深层特征之间的路径,可以使得网络模型层数大大增加而不退化。常规的残差模块如图1,该模块使用两个 3×3 的卷积层对上层输入特征进行卷积操作得到残差映射,同时将原特征与卷积输出逐元素相加实现特征的融合。残差的计算公式如式(1)所示。

$$y = F(x, \{W_i\}) + W_s x \quad (1)$$

式中, x 和 y 分别为网络的输入和输出, W_i 为 x 的权重, $F(x, \{W_i\})$ 为网络进行学习的残差映射。网络学习拟合残差映射不仅可以简化学习目标和难度,当网络达到最优性能时,多余的网络层通过跨层连接可以实现上层特征的恒等映射,提高了深层网络的性能表现,解决网络加深梯度消失和爆炸的问题。

1.3 注意力机制

人类的视觉本质上是大脑对脑电信号的处理,人会根据自己的兴趣将注意力放到注意力焦点处进行仔

细观察, 大脑自动剔除不感兴趣的区域信息, 着重提取感兴趣区域的细节与特征. 神经网络的注意力机制借鉴了人类的这种视觉机制, 引入注意力机制能够捕获到更高级的语义信息, 增强模型对前景的学习能力, 提高模型的灵敏度和分割精度.

网络的特征提取通常是基于全局的, 目标物体和背景的特征权重系数没有很大差距, 这就导致了网络训练缓慢, 不能快速提取目标物体特征. 针对本文实验任务, 如图2所示的CT图像中, 目标骨骼区域只占到整幅图像的极小区域, 将注意力聚焦于该小部分目标区域, 不仅能加快训练的速度, 更能增强该位置上对象的表示, 突出特征细节. 因此可以在网络中增加注意

力模块, 通过给特征添加权重, 使网络减少对背景像素的关注, 加快对骨骼特征的学习.

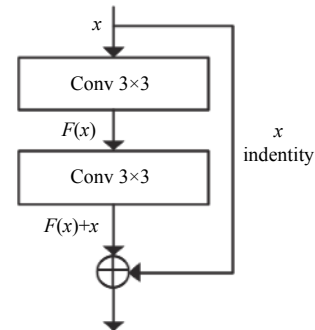


图1 常规残差模块

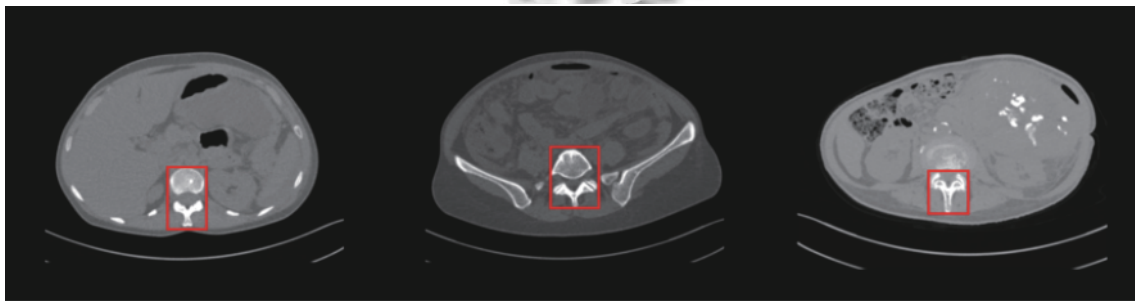


图2 脊柱CT分割目标区域占比

2 CT图像脊柱分割网络模型结构

针对脊柱CT图像分割任务的网络模型设计, 需解决如下问题:

(1) 医学图像分割是医学图像处理与分析领域的关键步骤, 其目的是将医学图像中具有某些特殊含义的部分分割出来, 为临床诊疗和病理学研究提供可靠的依据, 因此需要极高的精度.

(2) 脊柱CT图像中目标区域结构相对固定, 语义明确, 目标定位相对容易; 同时, 由于CT图像噪点多、灰度不均匀、伪影、不同软组织间灰度相近的特点, 精准分割难以实现.

(3) 医学图像的数据量相对较少, 若模型设计过于复杂, 参数量大, 网络极易过拟合, 影响模型精度.

U型网络能够同时结合低分辨率信息和高分辨率信息, 且相较于其他分割网络有更少的参数, 因此更适用于脊柱CT图像的分割任务.

经典U-Net网络中堆叠着一定数量的卷积操作, 作为低、中、高特征提取器. 越靠近底层的卷积神经网络拟合能力更强, 网络性能也更优异, 但随着网络的

加深, 图像的反复降维操作, 会一定程度上损失部分图像的细节信息, 难以获取到精细的边缘, 降低模型的最终性能.

为了从CT图像中提取到更多的细节信息, 缓解低对比度边缘的欠分割问题, 本文在U型网络中引入深度残差网络结构来优化卷积-池化结构的基础模块, 使其拥有多尺度特征表示能力, 改进后结构如图3所示.

该模块通过对瓶颈残差结构进行改进, 用一组更小的滤波器组替换全通道的 3×3 卷积核, 并将这些分组的滤波器以类残差的层次化方式连接, 来增加输出特征所能代表的尺度的数量. 对 1×1 卷积输出后的特征图按通道数均分为4块后分层进行特征提取, 特征计算结果如下:

$$y_i = \begin{cases} x_i, & i = 1 \\ K_i(x_i + y_{i-1}), & 4 \geq i > 1 \end{cases} \quad (2)$$

其中, x_i 为分块特征子集, K_i 为每层分块后进行的 3×3 卷积, y_i 为特征子集的输出. 经过分块卷积和残差

融合, 可得到不同感受野大小的特征输出: y_1 与输入特征相同, y_2 可表示为 3×3 卷积所提取的特征图, y_3 相当于 5×5 大小的感受野得到的特征图, y_4 相当于经过 7×7 大小的感受野得到的特征图. 模块中卷积输出与原特征图的跨层连接, 能缩短前后层之间的距离, 进一步提高特征的学习能力.

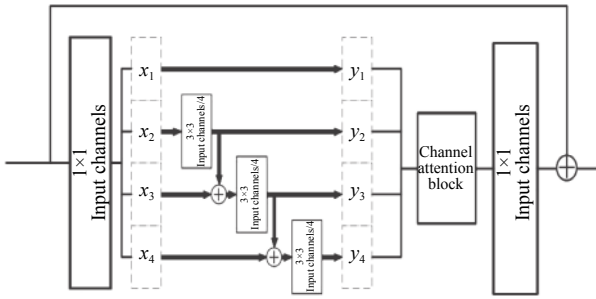


图3 多尺度残差结构

该残差结构采用先拆分后融合的策略, 能够在控制参数数量的同时增加每个网络层的感受野范围, 既不增加网络的计算负载, 还能从多个尺度上表示特征. 多尺度的融合有助于提取出更多细节信息, 降低脊柱目标和周围软组织对比度低对分割结果的干扰, 提高分割精度.

通道分层卷积的残差结构虽然可得到多尺度的特征表示, 但结构中各个特征图的通道并不存在依赖性. 不同的通道包含的特征信息存在差异且重要性不同, 因此, 图3模块中在多尺度残差融合特征后引入通道注意力机制对特征各通道的依赖性进行建模. 自适应地对特征进行逐通道调整, 网络就可以通过学习全局信息来有选择性地加强包含有用信息的特征, 使得下一层卷积能充分利用这些特征, 并对无用特征进行抑制, 从而提高网络的表示能力. 通道注意力模块如图4所示. 通道模块公式如下:

$$F_i = M_c(F) \otimes F \quad (3)$$

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (4)$$

(1) 上层传入特征, 在通道维度上使用全局最大池化和全局平均池化对特征进行降维, 得到2个不同的信道描述符;

(2) 按其通道维度将这两个不同空间背景的信道描述符合并送入一个两层的神经网络感知机;

(3) 将得到的两个特征图逐点相加, 并经过 Sigmoid

激活函数得到权重系数;

(4) 将原特征与权重系数对应元素逐个相乘得到特征图.

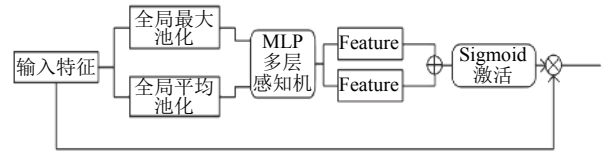


图4 通道注意力模块

医学 CT 图像本身易受噪声的干扰, 且脊柱 CT 图像因骨质不同以及组织混迭等因素导致影像图像灰度不均匀. 引入注意力机制, 可将上层特征中不相关的信息以及噪声的激活响应抑制, 只凸显相关信息的激活值, 一方面可以滤除背景噪声对梯度更新的影响, 另一方面可以减少上采样的边缘模糊问题. 该模块只给网络增加了极少的额外计算量, 却能带来显著的模型敏感度和准确率的提高. 综上所述, 本文结合残差单元与注意力机制提出一种针对 CT 脊柱分割的模型, 网络结构如图5所示.

3 实验设计

3.1 实验环境

本实验采用的硬件环境为 NVIDIA-RTX2060, 显存 6 GB. 软件环境为以 Python 3.7, TensorFlow-GPU 1.14.0 支撑的 Keras. 实验数据来自 California 大学的公开 CT 数据集^[10], 原图大小为 512×512 像素, 选用 1-6 组脊柱 CT 图像及对应标签共 3410 张作为训练与验证集, 选用第 7、8 组数据中共 400 张图像作为测试数据. 训练时超参数采用如下相同设置: batch_size 设定为 16; 初始学习率为 10^{-3} , 优化器选用 Adam, 在训练时自适应地调节学习率, 直至损失函数收敛.

3.2 评价指标

为了对本实验结果有一个真实可解释的评价, 本文选用医学图像分割常用的 Dice 相似性系数和 IOU 评测指标评估模型.

以 V_{gt} 表示真实标签中轮廓区域所包含的点集, $V_{prediction}$ 表示网络预测的轮廓区域点集. IOU 能计算出网络预测的结果与真实标签的交并比, 其公式如下:

$$IOU = \frac{V_{gt} \cap V_{prediction}}{V_{gt} \cup V_{prediction}} \quad (5)$$

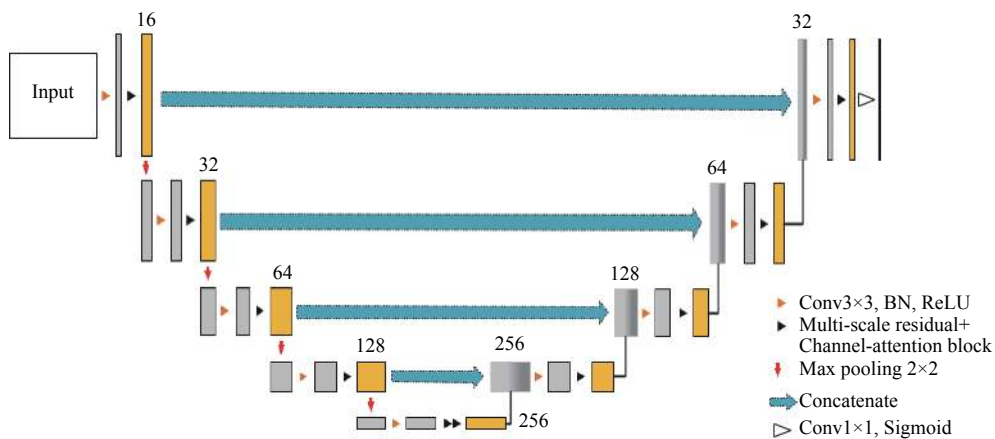


图5 改进的 U-Net 网络结构

Dice 相似性系数 (DSC) 是一种集合相似度度量函数, 用于计算真实标签与网络预测结果的相似度, 数值越接近于 1 图像的分割精确度越高. 公式如下所示:

$$DSC = 2 \frac{|V_{gt} \cap V_{prediction}|}{|V_{gt}| + |V_{prediction}|} \quad (6)$$

3.3 实验结果与分析

3.3.1 与其他分割网络的对比实验

实验主要分为训练和测试两个阶段, 首先在训练集上进行训练, 然后在测试集中进行模型测试, 得到实验结果. 为验证本实验模型相较于其他分割网络在脊柱 CT 图像分割任务的优越性, 将本文所提出的网络模型与原始 U-Net 网络、FCN、SegNet 网络进行对比.

同时, 分割任务中训练网络所使用的损失函数同样影响着网络的性能. 为测试损失函数对网络的影响, 将不同网络分别使用 Focal loss、二元交叉熵 (CE)、Dice loss 三种不同的损失函数进行训练, 分割精确度如表 1 所示.

分析表 1 可知, 相较于经典分割模型 FCN 与 SegNet, U-Net 级联卷积-反卷积结构对于医学图像的分割任务上有更好的表现, 这与预先设想一致. 与基础 U-Net 模型的分割结果对比, 本文引入残差与注意力机制对其进行改进后, 分割精度有了进一步的提高, 在 Dice loss 下的测试集的预测值与真实标签的相似性达到了 97.17%, 更满足医疗图像分割精确性的要求.

针对训练过程损失函数的实验分析可知, Focal loss 更适用于解决正负样本数量不均衡的图像分割任务, 本任务 CT 图像为一系列相关联的图像, 目标区域像素平均, 因此该损失函数在本任务中表现欠佳; 以

Dice 系数作为 Loss 进行直接优化, 训练中网络会更侧重对前景区域的挖掘, 因此在该损失函数的训练下分割结果精度较高, 但训练过程中 Loss 容易不稳定, 后续将考虑结合 CE 对其进行优化.

表 1 算法准确性对比 (%)

Network architectures	Evaluation	Loss functions		
		Focal loss	CE	Dice loss
SegNet	DSC	—	85.23	91.98
	IOU	—	76.90	87.35
FCN ^[4]	DSC	79.58	90.12	94.10
	IOU	66.92	82.23	89.00
U-Net ^[5]	DSC	85.41	94.07	96.15
	IOU	75.00	90.68	93.48
本文模型*	DSC	88.19	95.81	97.17
	IOU	79.07	92.00	94.52

从测试集中随机选取图像输入模型, 分别生成分割结果. 图 6 为不同网络模型分割的结果对比. 图 6(a) 为 CT 图像, 图 6(b) 为 CT 脊柱的真实标签, 图 6(c) 为 SegNet 分割结果, 图 6(d) 为 U-Net 分割结果, 图 6(e) 为本文网络结构分割结果. 将图中分割结果与图 6(b) 真实标签比较, 对于简单结构的脊柱切片, 不同的分割网络均能检测到 CT 图像中脊柱的目标区域, 成功实现目标定位. 其中, SegNet 只能实现目标的粗分割, 对细节处理较差. U-Net 对目标的分割效果较好, 但对于骨骼切片结构较为复杂的图像, 由于原图中脊柱与骨间软组织连接处的轮廓边界并不清晰, U-Net 对模糊边界的分割处理略显粗糙. 对比可见, 本文所用网络对骨间细小轮廓线的分割效果更好, 对模糊边界的处理更优. 本文中所用模型的多尺度残差结构与注意力机制的设置均是为了提高分割结果的细节表达, 实验效果图验

证, 该设置达到了预期设想.

将 CT 图像载入训练后的模型中得到脊柱分割序

列并进行三维重建, 便可得到清晰的 3D 脊柱模型, 如

图 7 所示.

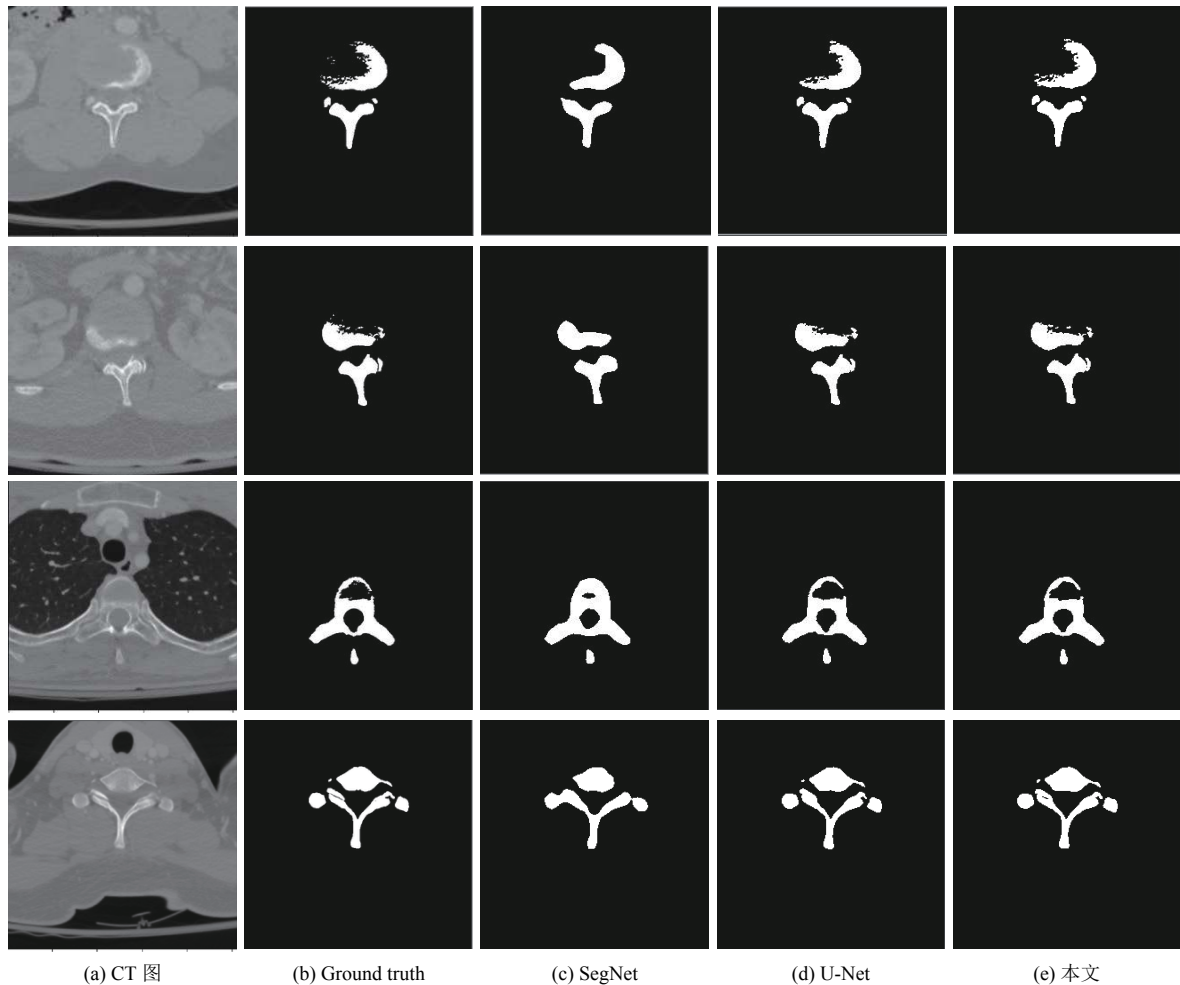


图 6 脊柱 CT 图像分割效果图

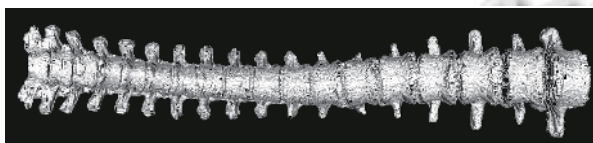


图 7 脊柱 CT 分割三维重建结果

3.3.2 网络各模块消融实验

为了验证各模块对整体模型的影响, 对 U-Net 原始模型、残差模块结构、注意力机制分别进行对比实验, 以证明网络中所有设置均为最优.

首先对上下采样基础模块结构对性能的影响进行比较. 为证明残差结构及多尺度特征提取的引入能提高分割精度, 改变基础模块的结构并测试分割精度, 实验结果如表 2 所示.

如表 2 所示, Case 1 不使用残差结构, 沿用经典 U-Net 的基础卷积模块, 分割结果略优于 Case 2 中使用 1 层残差结构的基础模块, Case 3 中增加了残差结构层数, 网络深度加深, 分割结果随之提高. Case 4 中上下采样均使用本文所提出的多尺度残差结构, 特征感受野更为丰富, 采用先降维后升维的方式能够降低计算量, 且注意力机制能自适应调整通道特征权值, 实验证明使用该模块具有更高的分割精度. Case 5 对网络中堆叠残差单元层数做对比实验, 实验结果所示, 采用 1 层残差单元作为基础模块有更高的分割精度, 而增加至 2 层其参数量增大过多, 不宜于网络的训练, 且分割精度明显下降, 故 3 层以上堆叠层数不做实验.

为了验证残差模块与注意力机制的组合能提高模

型精度,实验对各模块进行消融实验.实验结果如表3所示,为网络各模块的不同组合对性能的影响.

如表3所示,以原始U-Net卷积-反卷积结构作为基础架构并使用上文所提出的残差模块替换原始卷积模块可以提供更多地细节信息,对于低对比度边缘分割有所提高,其分割精度更高.模型添加注意力机制后,Dice系数升高,且IOU系数升高,证明其增强了模型的分割能力.本文将残差模块与注意力机制叠加引入U-Net,改进网络结构使得该模型在脊柱CT分割任务所得的Dice系数比原始U-Net的分割更高,分割效果更接近于标签的真实值.

表2 网络引入不同基础模块结构的结果对比

Case	Convolution unit			DSC
	Conv	Shortcut	Layers	
1	Conv3×3	×	1	0.9615
	Conv3×3			
2	Conv3×3	√	1	0.9560
	Conv3×3			
3	Conv3×3	√	2	0.9640
	Conv3×3			
4	Multi-scale ResUnit	√	1	0.9717
5	Multi-scale ResUnit	√	2	0.9628

表3 模型Dice系数表

Case	U型结构	Res-block	Attention block	DSC (%)	IOU (%)
1	√	√	×	95.60	92.33
2	√	×	√	96.80	94.07
3	√	√	√	97.17	94.52

4 结论与展望

本文提出了一种基于卷积-反卷积结构的神经网络模型,用于脊柱CT图像的全自动分割.该网络以U-Net结构为基本框架,引入分层残差连接的卷积模块及通道注意力机制,加强特征的传播,增强图像细节信息的分割效果.其特点在于采用了多尺度感受野融合的残差结构替代传统卷积模块逐层提取图像中目标区域特征,通过残差学习进一步细化了分割效果;并在特征融合后引入注意力机制加强了目标区域的权重,缓解了CT图像各组织结构对比度低、边界轮廓模糊等问

题对分割模型造成的干扰,提高了模型的鲁棒性.

本文所提网络结构对CT图像脊柱分割具有较高的分割精度和分割效率,但从实验结果中可发现网络设计仍有不足之处,例如对于脊柱与其下联肋骨存在误分割的情况,这也是下一步研究需要关注的重点.

参考文献

- 1 Yao JH, O'Connor SD, Summers RM. Automated spinal column extraction and partitioning. Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro. Arlington: IEEE, 2006. 390-393.
- 2 Huang JY, Jian FZ, Wu H, *et al.* An improved level set method for vertebra CT image segmentation. BioMedical Engineering OnLine, 2013, 12(1): 48. [doi: 10.1186/1475-925X-12-48]
- 3 Li Y, Liang W, Tan JD, *et al.* A novel automatically initialized level set approach based on region correlation for lumbar vertebrae CT image segmentation. Proceedings of 2015 IEEE International Symposium on Medical Measurements and Applications. Turin: IEEE, 2015. 291-296.
- 4 郭树旭, 马树志, 李晶, 等. 基于全卷积神经网络的肝脏CT影像分割研究. 计算机工程与应用, 2017, 53(18): 126-131. [doi: 10.3778/j.issn.1002-8331.1611-0523]
- 5 刘忠利, 陈光, 单志勇, 等. 基于深度学习的脊柱CT图像分割. 计算机应用与软件, 2018, 35(10): 200-204, 273.
- 6 王亚刚, 王萌, 韩俊刚, 等. 一种自动分割股骨区域的R-U-Net神经网络. 小型微型计算机系统, 2019, 40(4): 839-844. [doi: 10.3969/j.issn.1000-1220.2019.04.029]
- 7 李贤, 何洁. 3D全卷积网络在脊柱分割中的应用. 电子科技, 2018, 31(11): 75-79.
- 8 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015. 234-241.
- 9 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770-778.
- 10 Yao JH, Burns JE, Muñoz H, *et al.* Detection of vertebral body fractures based on cortical shell unwrapping. Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin Heidelberg: Springer, 2012. 509-516.