

DOI:10.19651/j.cnki.emt.2209623

基于深度学习的脊椎 CT 图像分割^{*}

黄 昆 张俊华 普 钟
(云南大学信息学院 昆明 650500)

摘要: 脊椎 CT 图像分割是脊椎三维重建可视化的关键。针对脊椎 CT 图像中脊椎边缘模糊,结构复杂,形状多变等问题,基于深度学习方法提出一种双解码器网络。该网络在编码解码网络 U-Net 结构基础上增加了一条结构相同的并行解码分支,两个解码分支可以互补地提取图像特征。并且,在编码与解码之间加入双重特征融合模块,解决网络在下采样和上采样过程中造成的语义信息丢失问题。同时用密连混合卷积模块代替原始卷积模块,提高网络对多尺度特征的提取能力。此外加入高效注意力模块,使网络在空间上注重学习感兴趣区域,在通道上抑制无关信息。在 CSI2014 公开数据集上进行测试,Dice 系数达到 0.970,Jaccard 系数达到 0.945,召回率达到 0.962。实验结果表明,该网络能够提高脊椎分割精度,具有较好的泛化能力,可以满足临床脊椎 CT 图像分割需求。

关键词: 脊椎分割;深度学习;双解码器网络;双重特征融合模块;密连混合卷积模块;高效注意力模块

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Vertebra CT image segmentation based on deep learning

Huang Kun Zhang Junhua Pu Zhong
(School of Information, Yunnan University, Kunming 650500, China)

Abstract: Vertebra CT image segmentation is the key to the visualization of vertebra 3D reconstruction. Aiming at the problems of blurred vertebra edge, complex structure and changeable shape in vertebra CT images, a dual-decoder network is proposed based on deep learning method. The network adds a parallel decoding branch with the same structure on the basis of the U-Net structure of the encoding and decoding network, and the two decoding branches can extract image features complementary. Moreover, a dual feature fusion module is added between encoding and decoding to solve the problem of semantic information loss caused by the network downsampling and upsampling. At the same time, the original convolution module is replaced by the densely connected hybrid convolution module to improve the network's ability to extract multi-scale features. In addition, an efficient attention module is added to make the network focus on learning regions of interest in space and suppress irrelevant information in channels. Tested on the CSI2014 public dataset, the Dice coefficient reaches 0.970, the Jaccard coefficient reaches 0.945, and the Recall rate reaches 0.962. The experimental results show that the network can improve the accuracy of vertebra segmentation, has better generalization ability, and can meet the needs of clinical vertebra CT image segmentation.

Keywords: vertebra segmentation; deep learning; dual decoder network; dual feature fusion module; densely connected hybrid convolution module; efficient attention module

0 引 言

由脊椎构成的脊柱是人体骨骼的重要部分。脊椎疾病患病率高且后果严重,2016年中国导致伤残寿命年损失排第一位的疾病是脊椎疾病^[1]。许多严重的脊椎疾病需要通过手术进行治疗,脊椎 CT 图像的分割是脊椎三维重建可视化的关键。由于临床脊椎 CT 图像中脊椎的边缘比较模

糊,并且脊椎存在结构复杂,形状多变等特点,而胸椎与肋骨的解剖位置非常相近,这些情况都增加了脊椎分割的难度。

目前,脊椎 CT 图像分割方法主要有基于传统方法和基于深度学习两类。传统脊椎 CT 图像分割方法,如 Aslan 等^[2]使用统计水平集方法分割脊椎,但该方法需要利用期望最大化算法进行初始化和参数估计。Rasoulia 等^[3]提

收稿日期:2022-04-15

^{*} 基金项目:国家自然科学基金(62063034)、云南大学研究生实践创新项目(2021Z50)资助

出了一种基于配准的脊椎 CT 图像分割技术,并建立了相应的统计的多脊椎形状+姿势模型。该模型可以捕获脊椎形状和姿态的变化,同时减少配准参数的数量。Lim 等^[4]提出了一种改进水平集的脊椎分割框架,该框架通过 Willmore 流提取局部几何形状特征,并利用先验形状核密度估计器获取所需的先验知识,从而能够处理因噪声污染导致信息缺失的图像。Pereanez 等^[5]提出了一种基于统计形状分解和条件模型的脊椎分割方法。该方法首先利用统计形状分解获取脊椎各部分的统计模型,然后使用条件模型来模拟各子部分之间的统计相互关系。Korez 等^[6]提出了一种自动检测和分割脊椎的框架。它采用一种基于插值理论的技术,在脊椎图像中检测出单个脊椎的位置作为脊椎分割的输入,然后利用形状约束变形模型方法进行分割。但是这些算法过于复杂,计算处理时间长,分割精度不够高等,导致这些方法难以应用于临床的脊椎分割。

随着人工智能的发展,基于深度学习的神经网络因优异的性能在医学图像处理中取得了不错的成果^[7-9]。许多国内外学者提出了基于深度学习的脊椎 CT 图像分割方法,Zhang 等^[10]提出基于 PointNet++ 的颈椎分割。该方法需对 CT 图像进行预处理,以及对 PointNet++ 分割结果做后处理,这虽然提高了分割的准确度,但是仅使用了点云边缘信息,强度信息利用率较低,导致对一些精细的结构分割不全。Kuok 等^[11]提出基于全卷积密集网络的胸腰椎分割。该方法将脊椎分割结果和椎间平面检测结果相结合重建出了脊椎的三维模型。但是由于数据集较少,导致网络泛化性能不够。Vania 等^[12]提出全自动脊椎分割方法,该方法将卷积神经网络和全卷积网络相结合,并利用类冗余作为软约束,降低了模型出错概率。然而由于冗余类的存在,提高了网络学习难度,导致分割精度不高。Chuang 等^[13]提出了一种迭代的脊椎实例分割模型,该模型虽然减少了内存的消耗,但是网络分割精度不高。刘侠等^[14]提出将 U-Net 与图割算法相结合的方法。即利用改进的 U-Net 模型进行粗分割,然后对粗分割结果利用图割算法做后处理。该方法分割的脊椎轮廓精度高,但存在脊椎分割不完整的情况。

2015 年 Ronneberger 等^[15]提出了 U-Net 网络,它由编码器,解码器以及跳跃连接三部分组成。编码器进行多次下采样和卷积操作,降低图像分辨率的同时获取图像上下文信息。解码器先是通过多次上采样操作来恢复图像的分辨率,然后通过跳跃连接将编码器获得的底层信息与恢复的高层信息相融合,最后通过多次卷积操作提取融合后的特征。U-Net 在医学图像分割任务中取得了很好的效果,使其得到了广泛应用。为了满足不同任务的实际需求,不同学者对 U-Net 进行了改进,提出了一系列改进型网络。如 Oktay 等^[16]提出 Attention U-Net,该网络设计了注意力门控机制(attention gate, AG),然后将 AG 模块替换 U-Net 的跳跃连接,有效地抑制了网络对无关区域的学习,同时使

网络关注学习与任务有关区域。Zhou 等^[17]提出 U-Net++, 该网络在 U-Net 基础上设计了嵌套结构和密集跳跃连接,解决了不同场景应用对网络深度的要求,同时使解码器更好地融合多尺度信息。Huang 等^[18]提出 U-Net3+, 它采用全尺度跳跃连接将不同尺度特征图的语义信息相结合,同时通过深度监督从全尺度聚合的特征图中学习层次表示。

针对上述脊椎 CT 图像分割中存在的问题,本文设计了一种双解码器网络。该网络首先通过双解码器结构互补地提取图像细微特征,同时在编码与解码之间用双重特征融合模块代替简单的跳跃连接以解决语义丢失问题。其次利用密连混合卷积模块替换普通卷积,以增强网络对脊椎多尺度特征的提取能力。最后通过高效注意力模块使网络注重学习兴趣区域同时抑制背景信息,从而提高网络对脊椎的分割精度。

1 相关工作

1.1 脊椎分割网络结构

本文针对脊椎分割问题而设计出双解码器网络(dual decoder network, DD-Net),其结构如图 1 所示。受细粒度识别双线性卷积神经网络^[19]的启发,将两个结构相同的特征提取器提取到的特征进行融合。本文网络的主体是在 U-Net 编码解码结构基础上加以改进的双解码器结构(dual decoder structure, DDS)。即网络解码部分是由两个结构相同的解码器组成,一个解码分支寻找目标位置,另一个分支对目标进行识别。这种设计使得网络结构变宽,可以使网络互补地学习到丰富的图像特征。虽然网络结构变得复杂,但网络总体架构为一个输入一个输出,同时网络可以通过端到端地进行训练。

DD-Net 网络的输入是大小为 256 像素×256 像素的脊椎 CT 灰度图像,输出是脊椎分割得到的二值图像。网络的编码分支一共进行四次下采样,网络的两个并行解码分支分别进行 4 次相同的上采样。每次下采样都是大小为 2×2 的最大池化操作,而每次上采样都是卷积核大小为 2×2,步长为 2×2 的转置卷积操作。在进行下采样和上采样操作之前,网络需要通过一个密连混合卷积模块(densely connected hybrid convolution module, DCHCM)以提取图像多尺度特征。为了让网络学习更多图像底层的语义信息,每经过一次下采样,DCHCM 的通道数增加一倍,而每经过一次上采样,DCHCM 的通道数减少一倍,保持了网络的对称性。同时,为了减少图像细节特征的丢失,保留更多的语义信息,在同一级的下采样前与上采样后之间,通过双重特征融合模块(dual feature fusion module, DFFM)将编码器提取的低层语义特征和解码器提取的高层语义特征进行融合。最后将两个并行解码分支的输出进行融合,再经过高效注意力模块(efficient attention module, EAM)去除由特征融合造成的冗余特征,然后通过

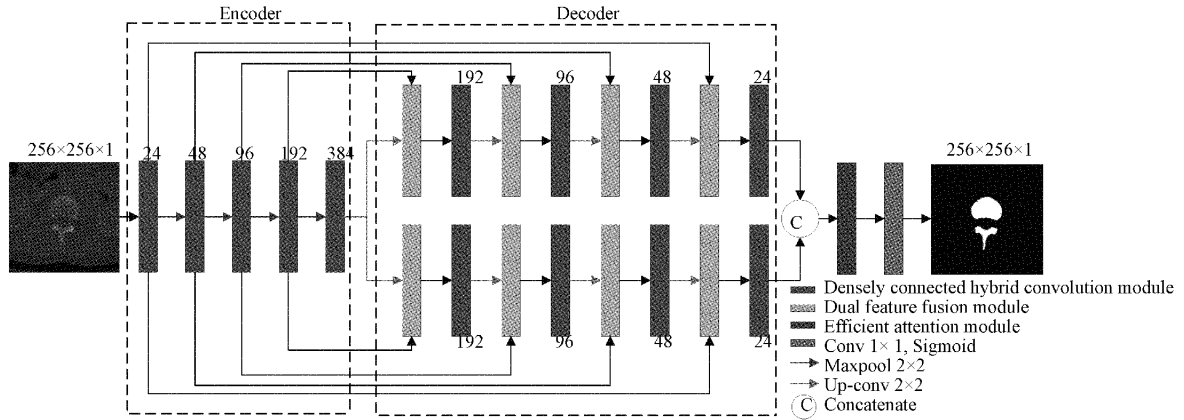


图 1 DD-Net 结构图

一个 1×1 的卷积操作,以及 Sigmoid 函数激活输出即可得到脊椎分割结果。

1.2 双重特征融合模块

U-Net 为了改善下采样和上采样过程造成的语义信息丢失问题,在同级编码器与解码器之间进行跳跃连接。跳跃连接只是将两个特征图进行简单的通道拼接,导致特征融合程度不高。为了尽可能减少语义信息的丢失,本文在跳跃连接上进行改进,设计了 DFFM。特征融合主要有特征图通道拼接(Concatenate)与特征图逐像素相加(Add)两种方式,Concatenate 方式增加了特征图通道数,而每个特征图信息没有增加,Add 方式增加了每个特征图的信息,而特征图通道数不变。本文设计的 DFFM 整合了这两种特征融合方式,其结构如图 2 所示。

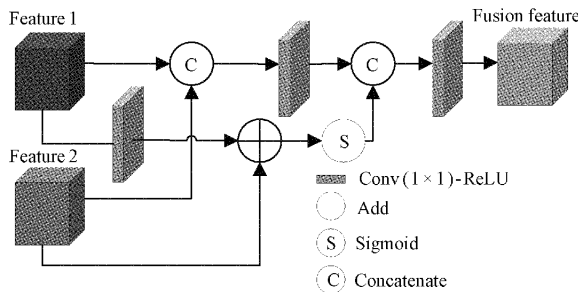


图 2 DFFM 结构图

特征 1 和特征 2 分别进行 Concatenate 和 Add 两种融合方式,得到第一级融合特征 F_C 和 F_A 然后将 F_C 和 F_A 通过 Concatenate 方式得到第二级融合特征 F_F 。特征融合过程原理如式(1)~(3)所示。

$$F_C = f^{1 \times 1}([F_1; F_2]) \quad (1)$$

$$F_A = \sigma(f^{1 \times 1}(F_1) \oplus F_2) \quad (2)$$

$$F_F = f^{1 \times 1}([F_C; F_A]) \quad (3)$$

式中: F_1, F_2 分别为输入特征图 1 和 2, F_C 为 Concatenate 方式得到的特征图, F_A 为 Add 方式得到的特征图, F_F 为最终得到的融合特征图。 $f^{1 \times 1}$ 表示卷积核大小为 1×1 的卷积操作, \oplus 表示 Add 操作, $[\]$ 表示 Concatenate 操作, σ

为 Sigmoid 函数。本文设计的 DFFM 通过对输入特征在空间和通道上的双重融合,以减少无用背景噪声造成的干扰,同时增强了上下文信息。

1.3 密连混合卷积模块

Dense-Net^[20] 为了加强特征在网络中的传播和利用率,设计了密集连接模块,该模块将每一层输入与输出特征进行通道拼接作为下一层的输入,从而能够改善网络训练过程中梯度消失的问题。但是密集连接模块存在多尺度特征提取能力不足的问题,而空洞卷积^[21]能够很好地解决这一问题。空洞卷积的主要思想是在卷积核之间插入像素值为 0 的像素以增大卷积核,从而能够在不增加参数的条件下扩大感受野,不同空洞率下空洞卷积的感受野如图 3 所示。空洞卷积的卷积核大小计算方式如式(4)所示。

$$K = (r - 1)(k - 1) + k \quad (4)$$

式中: K 为空洞卷积核大小, r 为空洞率, k 为原始卷积核大小。

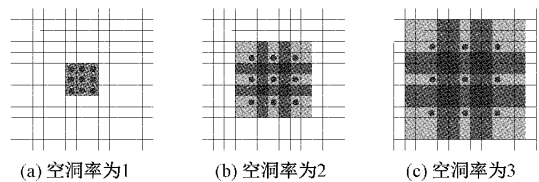


图 3 卷积核大小为 3×3 的空洞卷积感受野示意图

本文将密集连接思想和空洞卷积相结合,设计了具有三个特征提取层结构的 DCHCM,结构如图 4 所示。第一个和第三个特征提取层是由卷积核大小为 3×3 ,空洞率为 2 的空洞卷积(dilated convolution, Dilated Conv)和批归一化(batch normalization, BN)以及校正线性单元(rectified linear unit, ReLU)组成,而第二个特征提取层是由卷积核大小为 3×3 的普通卷积(Convolution, Conv),BN 以及 ReLU 组成。这种在两个空洞卷积之间插入一个普通卷积的混合卷积组成方式,有效地避免了连续进行空洞卷积导致的棋盘效应。假设输入每次经过每层特征提取层后产生的特征通道个数为 k ,则设计的 DCHCM 的输出特征通

道数为 $3k + k_0$ ，其中 k_0 为输入特征通道的个数。本文设计的 DCHCM 能够减少网络的参数量，扩大了感受野，能够提高网络提取多尺度特征的能力。

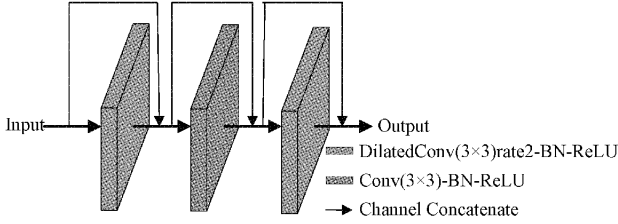


图 4 DCHCM 结构

1.4 高效注意力模块

卷积注意力模块^[22] (convolutional block attention module, CBAM) 因其模块化结构, 轻量级参数, 广泛应用于语义分割网络。但 CBAM 在本文网络中效果不明显, 因此本文在 CBAM 基础上设计出适用于脊椎分割的 EAM。如图 5 所示, EAM 结构分为通道注意力模块和空间注意力模块, 两个模块采用串联方式连接。输入特征图 $F \in R^{C \times H \times W}$ 经过全局平均池化和全局最大池化分别得到特征因子 F_{avg}^c 和 F_{max}^c , 与 CBAM 不同的是, EAM 去掉了多层

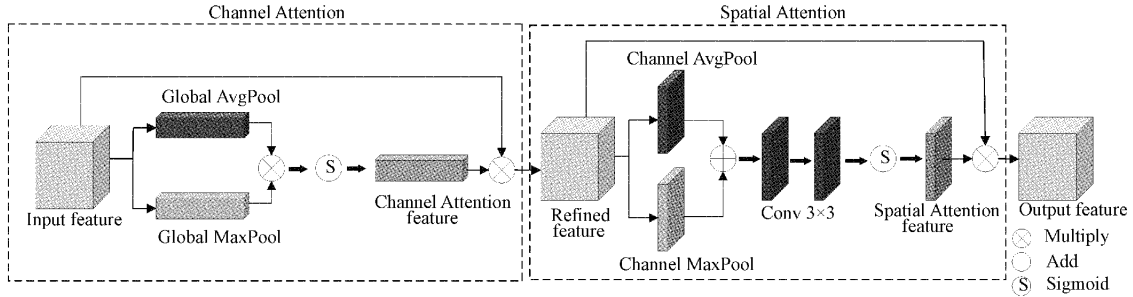


图 5 EAM 结构

1.5 损失函数

本文主要是将脊椎与非脊椎组织分割开, 本质上可以看作二分类问题, 即将像素分为前景和背景两类。所以本文采用二元交叉熵作为损失函数来监督网络的训练, 其表达式如式(8)所示。

$$L = - \sum_{i=1}^N y^{(i)} \log \hat{y}(i) + (1 - y^{(i)}) \log (1 - \hat{y}(i)) \quad (8)$$

式中: N 为样本总数, i 为样本, $y^{(i)}$ 为第 i 个样本的标签值, $\hat{y}(i)$ 为第 i 个样本的预测值。

2 实验设计

2.1 数据集与预处理

本文实验数据来自 CSI2014 脊椎分割挑战^[23] 提供的公开数据集。该数据集一共有十个病例的脊椎 CT 图像数据, 并含有专家标注的胸椎与腰椎标签, 病例数据具体情况如表 1 所示。为了使数据集符合网络的输入要求, 对原始数据进行切片并将格式转换为图片, 同时将图片位深调

感知机, 进一步降低了模型的复杂度。然后直接对 F_{avg}^c 和 F_{max}^c 进行相乘操作, 再经过 Sigmoid 激活得到通道注意力特征图 $M_c \in R^{C \times 1 \times 1}$, 最后 M_c 与 F 进行相乘操作, 生成通道特征权重系数重新标定后的特征图 F' 。 F' 经过通道上的平均池化和最大池化分别得到特征因子 F_{avg}^s 和 F_{max}^s , 为了使网络更加关注空间位置信息, 将 CBAM 中对 F_{avg}^s 和 F_{max}^s 的通道拼接改为相加操作。再经过两个 3×3 的卷积操作以及 Sigmoid 激活得到空间注意力特征图 $M_s \in R^{1 \times H \times W}$, 最后将 M_s 与 F' 进行相乘操作, 生成最后的输出特征图 F'' 。通道注意力特征图和空间注意力特征图以及输出特征图的计算过程分别如式(5)~(7)所示。

$$M_c(F) = \sigma(F_{avg}^c \otimes F_{max}^c) \quad (5)$$

$$M_s(F) = \sigma(f^{3 \times 3}(f^{3 \times 3}(F_{avg}^s \oplus F_{max}^s))) \quad (6)$$

$$F' = F \otimes M_c(F) \quad (7)$$

$$F'' = F' \otimes M_s(F)$$

式中: σ 表示为 Sigmoid 函数, $f^{3 \times 3}$ 表示卷积核大小为 3×3 的卷积操作, \otimes 表示逐元素相乘操作, \oplus 表示逐元素相加操作。注意力模块通过将输入特征图在通道与空间上的权重系数重新标定, 使网络在空间上注重学习感兴趣区域, 通道上抑制无关信息, 从而提高网络的分割精度。

整为 8。为了减少网络训练过程计算量, 将所有图片裁剪成 256 像素 \times 256 像素大小, 并将灰度进行归一化。

表 1 数据集分布

数据组	年龄	像素间距/ mm	切片厚度/ mm	切片数/ 张
Case1	23	0.312 5	1.0	559
Case 2	22	0.312 5	1.0	507
Case 3	27	0.312 5	1.0	560
Case 4	28	0.353 5	1.0	625
Case 5	19	0.312 5	1.0	601
Case 6	26	0.343 7	1.0	562
Case 7	21	0.312 5	1.0	509
Case 8	16	0.312 5	1.0	548
Case 9	23	0.361 3	1.0	572
Case 10	25	0.312 5	1.0	552

2.2 实验环境

实验环境主要配置:CPU 为 Intel(R) Xeon(R) W-2102CPU@2.90 GHz;GPU 为 NVIDIA GeForce RTX 2080;RAM 为 16GB;操作系统为 Windows10 专业工作站版。实验过程以 Python3.7 为开发语言,Keras2.3.1 为框架,Tensorflow2.0 为后端,Mimics Research21.0 为脊椎三维重建软件。

2.3 网络训练

实验数据集根据病例数划分为训练集、验证集、测试集为 8:1:1,即选取 8 个病例数据为训练集,1 个病例数据为验证集,1 个病例数据为测试集。从而得到训练集图片数量 4 471 张,验证集 572 张,测试集 552 张。为了防止过拟合,提高网络的泛化能力,在训练过程中对训练集进行了数据增强,数据增强的方式有随机旋转,水平和垂直平移,随机错切,随机缩放等。

网络训练时,batch size 设置为 2,初始学习率设为 0.000 5。当验证集的损失连续 5 轮不再下降时,学习率衰减为原来的 20%,最小学习率设置为 0.000 001,采用 Adam 优化器进行训练。

2.4 评价指标

本文采用的评价指标主要有 Dice 系数(Dice),Jaccard 系数(Jaccard)^[24],召回率(Recall)。Dice 系数又叫做重叠指数,是验证医学图像分割中最常见的指标。Dice 系数越大,表示预测结果越接近标签,网络性能越好。它的定义如式(9)所示。

$$Dice = \frac{2TP}{FP + 2TP + FN} \quad (9)$$

Jaccard 系数用于评价样本集之间的相似程度。Jaccard 系数越大,分割结果越好。它的定义如式(10)所示。

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

召回率是二分类问题的常用评价指标,同样适用于本研究。定义如式(11)所示。

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

式中:TP 表示正确预测出脊椎的像素数,FN 表示将脊椎预测为背景的像素数,FP 表示将背景预测为脊椎的像素数。

3 实验结果与分析

3.1 空洞率大小对网络性能的影响

本文为了加强网络多尺度特征提取能力,设计了密连混合卷积模块。该模块中两个空洞卷积是网络提取脊椎多尺度特征的关键,本文对比了两个空洞卷积在不同空洞率下对网络性能的影响,结果如表 2 所示。由表 2 可知,空洞率大小对网络性能的影响比较明显,空洞率越大,空洞卷积的感受野越大。随着空洞率增大,网络分割精度先提

高后下降。这表明对于脊椎这种相对较小的目标来说,适当地扩大感受野能提高网络分割精度,但当感受野太大会使网络忽略掉部分脊椎信息,从而导致网络性能下降。为使网络性能达到最优,本文选定两个空洞卷积的空洞率大小都为 2。

表 2 不同空洞率下的分割结果

空洞率	Dice	Jaccard	Recall
1,1	0.966	0.939	0.951
1,2	0.969	0.944	0.961
2,2	0.970	0.945	0.962
2,3	0.948	0.918	0.935
3,3	0.947	0.913	0.927

3.2 消融实验

为了验证本文设计的 DDS,DFFM,DCHCM,EAM 对网络性能的影响,设计了相应的消融实验,结果如表 3 所示。其中 Model_1 在原始 U-Net 的基础上加入 DDS,Model_2 在 Model_1 的基础上加入 DFFM,Model_3 在 Model_2 的基础上加入 DCHCM,Model_4 在 Model_3 的基础上加入 EAM。从表 3 中可看出,原始 U-Net 的 Dice 系数为 0.937,Jaccard 系数为 0.888,召回率为 0.895。Model_1 与 U-Net 相比,性能得到全面提升,表明本文设计的 DDS 能够互补地学习图像特征,从而提高分割精度。Model_2 的 Dice 系数提升最大,在 Model_1 的基础上提升了 1.4%,表明将编码与解码获得的特征通过高效特征融合模块在空间与通道维度上进行融合,明显改善了语义信息丢失问题。Model_3 性能在 Model_2 的基础上进一步得到提高,表明通过特征的重用以及空洞卷积使网络学习到多尺度特征,从而提高分割精度。Model_4 相对于 Model_3 性能更进一步,表明 EAM 通过在空间上使网络注重学习感兴趣区域,同时在通道上抑制无关信息,进而提高了分割精度。

表 3 不同模块对网络性能的影响

模型	Dice	Jaccard	Recall
U-Net	0.937	0.888	0.895
Model_1	0.945	0.902	0.907
Model_2	0.959	0.927	0.934
Model_3	0.965	0.936	0.950
Model_4	0.970	0.945	0.962

图 6 展示了不同改进模型对四张示例测试图像的分割结果,可以直观地展示各模块对网络性能的提升程度。从图 6(c)中可看出原始 U-Net 分割存在明显欠分割现象,即椎体分割不完整。图 6(d)在加入双解码器结构后,欠分割现象有所改善。图 6(e)在加入双重特征融合模块后,可

以看到网络由于语义信息丢失问题的改善,欠分割现象明显减少,只是欠缺细节部位的分割。图 6(f)在加入密连混合卷积模块后,分割结果进一步得到改善。图 6(g)在加入

高效注意力模块后,分割结果最为完整且边缘平滑,相比其他模型分割结果更接近标签。在细微的结构,如脊椎横突部分,本文方法(Model_4)也能很好地分割出来。

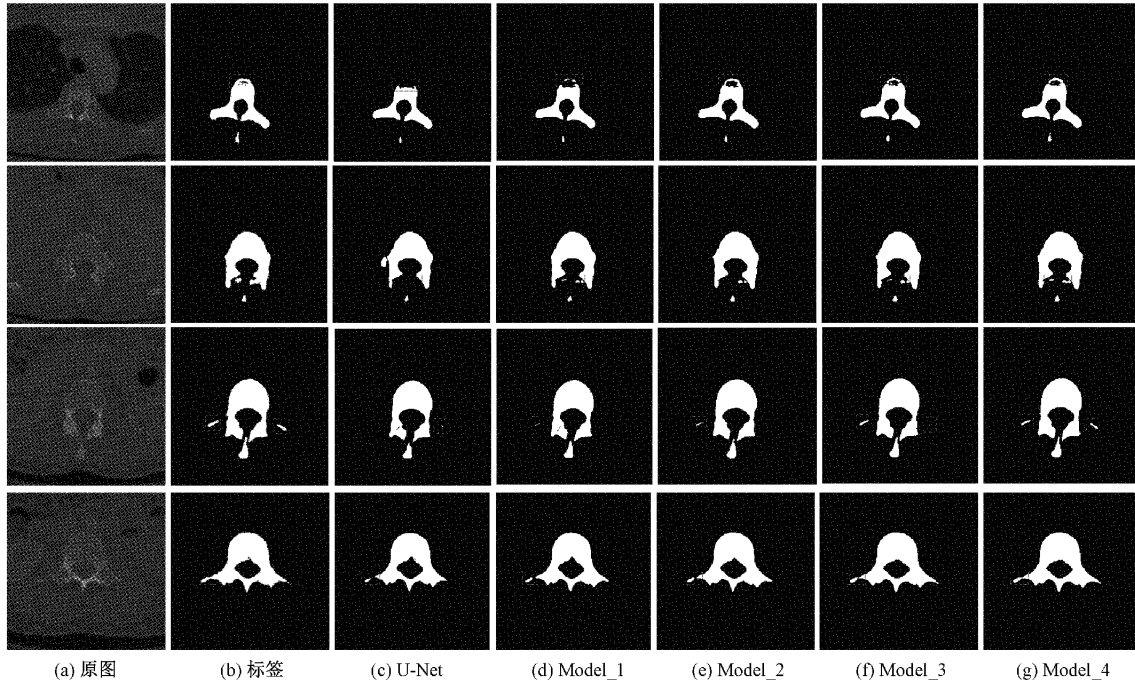


图 6 不同模型分割结果

脊椎 CT 图像分割的临床目的是更好、更快地实现脊椎的三维重建可视化,从而用于脊椎手术过程的指导。将不同模型的分割结果进行三维重建,结果如图 7 所示。可以看到,

本文方法(Model_4)分割结果重建出来的胸腰椎最完整,最接近标签的重建结果。进一步验证了本文方法分割性能较高,同时表明本文方法分割结果能够满足脊椎三维重建需求。

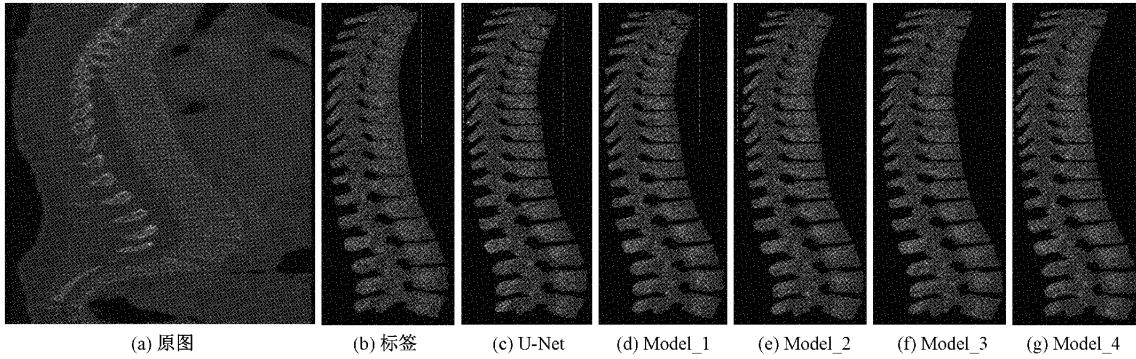


图 7 不同模型重建结果

3.3 EAM 和 CBAM 对比实验

本文在 CBAM 基础上进行改进,设计了适用于脊椎分割的注意力模块 EAM,在相同条件下与 CBAM 做对比实验,以评估本文设计的 EAM 性能。实验结果如表 4 所示,本文设计的 EAM 在三项指标上都要优于 CBAM,表明本文设计的 EAM 性能在脊椎分割网络中要优于 CBAM。

为了更加直观地展示 EAM 与 CBAM 的性能差异,本文将它们对两张测试图像进行热力图可视化,其结果如图 8 所示。由图 8 可知,EAM 相比于 CBAM 更加关注脊

表 4 EAM 与 CBAM 性能比较

模型	Dice	Jaccard	Recall
CBAM	0.965	0.937	0.948
EAM	0.970	0.945	0.962

椎区域,且边缘更加清晰,进一步证明了 EAM 性能要高于 CBAM。

3.4 与其他脊椎分割方法的对比实验

为了客观地评估本文方法的性能,在同一数据集下,

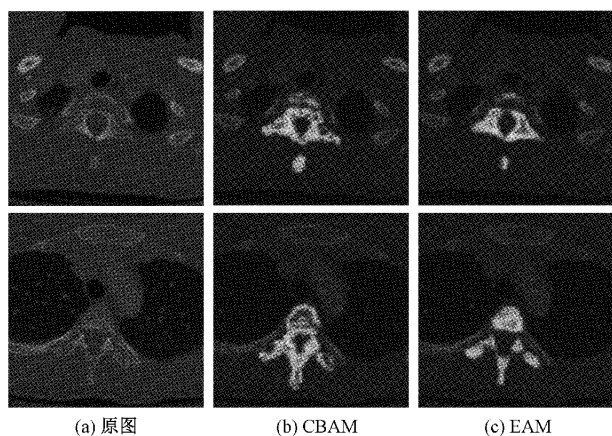


图 8 热力图可视化结果对比

将本文方法与同样采用深度学习方法的文献[25-28]分割结果进行比较,结果如表 5 所示。Li 等^[25]利用空洞残差模块替换 U-Net 跳跃连接方式,同时在解码器中加入 CBAM 模块。缓解了网络中语义丢失问题,但是存在特征提取能力不足的问题,导致分割精度不高。Novikov 等^[26]将双向卷积长短期记忆引入 U-Net 结构中,使网络能够捕获脊椎切片之间的时空关系,但并未改善语义信息丢失的问题。Lessmann 等^[27]提出一种迭代实例分割方法,将已经分割出的脊椎作为未分割脊椎的先验知识,但是存在级联问题,在分割过程中单个椎体分割不准确会导致后续所有椎体的分割出现问题。Wang 等^[28]在训练过程中引入自适应去噪学习策略,以避免噪声标签对训练的影响,但存在脊椎多尺度特征提取能力弱的问题。本文方法 Dice 系数要优于其他方法,证明了本文提出的网络结构对脊椎分割性能更好。

表 5 本文方法和深度学习脊椎分割方法分割结果对比

方法	Dice	Recall
Li ^[25]	0.893	0.960
Novikov ^[26]	0.949	—
Lessmann ^[27]	0.963	—
Wang ^[28]	0.958	0.955
DD-Net(本文方法)	0.970	0.962

3.5 与其他先进分割网络的对比实验

本文不仅与专门设计的脊椎分割网络进行对比,而且在相同实验条件下,与 U-Net, Attention U-Net, U-Net++, U-Net3+ 先进分割网络做对比实验,实验结果如表 6 所示。

由表 6 可知 U-Net 的 Dice 系数为 0.937, Jaccard 系数为 0.888, 召回率为 0.895, 可以看出 U-Net 在脊椎分割中表现一般。这是因为 U-Net 编解码器采用了普通卷积且跳跃连接为简单的通道拼接方式, 导致 U-Net 存在特征提

表 6 不同网络分割结果对比

网络	Dice	Jaccard	Recall	参数量/ 10^7
U-Net	0.937	0.888	0.895	31.05
Attention U-Net	0.962	0.931	0.943	0.85
U-Net++	0.964	0.941	0.959	4.87
U-Net3+	0.965	0.937	0.951	6.75
DD-Net(本文方法)	0.970	0.945	0.962	4.44

取能力不足和特征融合不足的问题。Attention U-Net 通过将注意力门控机制 AG 模块代替 U-Net 的跳跃连接, 使网络更加关注目标区域。U-Net++ 在 U-Net 的基础上设计了密集跳跃连接结构, 在解码器上聚合了不同尺度的语义特征。U-Net3+ 延续了 U-Net++ 的改进思想, 设计了全尺度的跳跃连接。进一步聚合了不同尺度特征图中的高级语义信息与低级语义信息。这三种改进网络都只是针对 U-Net 特征融合不足的问题进行改进, 网络的特征提取能力并未得到改善。相对于这些改进网络, 本文设计的 DD-Net 网络通过双解码器结构, 密连混合卷积模块, 增强了网络对特征的提取能力。而双重特征融合模块和高效注意力模块, 增强了网络对特征的融合能力。

由表 6 可知本文提出的网络的 Dice 系数为 0.970, Jaccard 系数为 0.945, 召回率为 0.962。本文方法在三个评价指标都要优于其他网络, 表明本文方法分割性能更好。在模型参数指标上看, 本文方法参数量要多于 Attention U-Net, 但与 U-Net++, U-Net3+ 在同一个量级, 且要明显少于 U-Net, 表明本文方法在提高分割精度的同时减轻了计算负担。

3.6 网络泛化性能评估

为了验证本文方法的泛化能力, 对 4 个临床脊椎 CT 数据进行测试, Dice 系数结果如表 7 所示。其中 Case1 和 Case2 是健康的脊椎数据, Case3 和 Case4 是患病的脊椎数据。由表可知, 各网络对健康脊椎数据分割精度要高于患病脊椎数据。这是因为提供数据的患者患有骨质疏松以及脊椎压缩性骨折, 导致脊椎变形严重, 网络难以将其分割出来。本文方法仅在 Case3 上的 Dice 系数略低于 U-Net3+, 而在 Case1, Case2 和 Case4 上 Dice 系数要高于其他网络。综合来看, 本文方法泛化性能要优于其他网络。

表 7 不同网络泛化性能对比

网络	Case1	Case2	Case3	Case4
U-Net	0.910	0.922	0.741	0.902
Attention U-Net	0.903	0.927	0.825	0.838
U-Net++	0.909	0.921	0.825	0.888
U-Net3+	0.910	0.934	0.830	0.910
DD-Net(本文方法)	0.917	0.938	0.827	0.914

4 结 论

本文针对脊椎 CT 图像分割问题,提出了 DD-Net。该网络以编码解码网络为基本架构,构建了双解码器结构,以互补地学习图像特征。同时在编码与解码之间加入双重特征融合模块,将编码器提取的低级语义信息和解码器提取的高级语义信息在空间与通道两个维度上进行融合,解决语义信息丢失问题。并且用密连混合卷积模块代替普通卷积操作,增强网络对多尺度特征提取能力。最后加入高效注意力模块,使网络在空间上关注感兴趣区域,通道上消除无关信息。实验结果表明,本文方法脊椎分割精度明显优于其他分割方法,具有较好的泛化性能,能够达到临床脊椎分割要求。后续工作中,本文可以通过扩充数据集或者对网络结构进行改进以进一步提高网络的泛化性能,使网络能够对多种脊椎进行分割。

参考文献

- [1] 曾新颖,齐金蕾,殷鹏,等. 1990~2016 年中国及省级行政区疾病负担报告[J]. 中国循环杂志, 2018, 33(12): 1147-1158.
- [2] ASLAN M S, FARAG A A, ARNOLD B, et al. Segmentation of vertebrae using level sets with expectation maximization algorithm[C]. Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2011, March 30-April 2, 2011, Chicago, Illinois, USA. Chicago: IEEE, 2011: 2010-2013.
- [3] RASOULIAN A, ROHLING R, ABOLMAESUMI P. Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape + pose model[J]. IEEE Transactions on Medical Imaging, 2013, 32(10): 1890-1900.
- [4] LIM P H, BAGCI U, BAI L. Introducing willmore flow into level set segmentation of spinal vertebrae[J]. IEEE Transactions on Biomedical Engineering, 2012, 60(1): 115-122.
- [5] PEREANEZ M, LEKADIR K, CASTRO-MATEOS I, et al. Accurate segmentation of vertebral bodies and processes using statistical shape decomposition and conditional models[J]. IEEE Transactions on Medical Imaging, 2015, 34(8): 1627-1639.
- [6] KOREZ R, IBRAGIMOV B, LIKAR B, et al. A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation[J]. IEEE Transactions on Medical Imaging, 2015, 34(8): 1649-1662.
- [7] 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.
- [8] 赵其杰,周安稳,朱俊豪,等. 基于 PE-Vnet 网络的三维骨骼图像分割方法[J]. 仪器仪表学报, 2020, 41(7): 243-251.
- [9] 杨志秀,韩建宁,于本知,等. 一种改进的 DeepLab V3+ 的医学图像分割方法[J]. 国外电子测量技术, 2021, 40(9): 18-23.
- [10] ZHANG L, WANG H. A novel segmentation method for cervical vertebrae based on PointNet ++ and converge segmentation[J]. Computer Methods and Programs in Biomedicine, 2021, 200(2): 105798.
- [11] KUOK C P, HSUE J Y, SHEN T L, et al. An effective CNN approach for vertebrae segmentation from 3D CT images[C]. 2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC). San Francisco: IEEE, 2018: 1-6.
- [12] VANIA M, MUREJA D, LEE D. Automatic spine segmentation from CT images using convolutional neural network via redundant generation of class labels[J]. Journal of Computational Design and Engineering, 2019, 6(2): 224-232.
- [13] CHUANG C H, LIN C Y, TSAI Y Y, et al. Efficient triple output network for vertebral segmentation and identification [J]. IEEE Access, 2019, 7: 117978-117985.
- [14] 刘侠,余鸿波,李冰,等. 基于改进 U-Net 模型下的椎骨 CT 图像分割方法[J]. 哈尔滨理工大学学报, 2021, 26(3): 58-64.
- [15] 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. Munich: Springer, Cham, 2015: 234-241.
- [16] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas [EB/OL]. (2018-04-11) [2022-03-20]. <https://arxiv.org/abs/1804.03999>.
- [17] ZHOU Z W, SIDIQUEE M M R, TAJBAKNSH N, et al. UNet ++: A nested U-Net architecture for medical image segmentation [C]. Proceedings of International Workshop on Deep Learning in Medical Image Analysis and International Workshop on Multimodal Learning for Clinical Decision Support, Heidelberg: Springer, 2018: 3-11.
- [18] HUANG H, LIN L, TONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation [C]. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and

- Signal Processing (ICASSP), Barcelona, Spain: IEEE, 2020: 1055-1059.
- [19] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition[C]. Proceedings of the IEEE International Conference on Computer Vision Santiago: IEEE, 2015: 1449-1457.
- [20] HUANG G, LIU Z, LAURENS V D M, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu: IEEE, 2017: 4700-4708.
- [21] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2022-03-20]. <https://arxiv.org/abs/1511.07122>.
- [22] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), Berlin Germany: Springer, Cham,2018: 3-19.
- [23] YAO J, BURNS J E, MUNOZ H, et al. Detection of vertebral body fractures based on cortical shell unwrapping[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, Berlin: Springer-Verlag, 2012:509-516.
- [24] CHANG H H, ZHUANG A H, VALENTINO D J, et al. Performance measure characterization for evaluating neuroimage segmentation algorithms[J]. Neuroimage, 2009, 47(1): 122-135.
- [25] LI W, TANG Y M, WANG Z, et al. Atrous residual interconnected encoder to attention decoder framework for vertebrae segmentation via 3D volumetric CT images [EB/OL]. (2021-04-08) [2022-03-20]. <https://arxiv.org/abs/2104.03715>.
- [26] NOVIKOV A A, MAJOR D, WIMMER M, et al. Deep sequential segmentation of organs in volumetric medical scans [J]. IEEE Transactions on Medical Imaging, 2018, 38(5): 1207-1215.
- [27] LESSMANN N, GINNEKEN B V, JONG P D, et al. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification[J]. Medical Image Analysis, 2019, 53: 142-155.
- [28] WANG Z, ZHANG Z, VOICULESCU I. RAR-U-Net: A residual encoder to attention decoder by residual connections framework for spine segmentation under noisy labels [C]. 2021 IEEE International Conference on Image Processing (ICIP), Anchorage: IEEE, 2021: 21-25.

作者简介

黄昆, 硕士研究生, 主要研究方向为深度学习, 医学图像处理。

E-mail: kunhuang@mail. ynu. edu. cn

张俊华(通信作者), 博士, 教授, 主要研究方向计算机视觉, 医学图像处理。

E-mail: jhzhang@ynu. edu. cn

普钟, 硕士研究生, 主要研究方向为深度学习, 医学图像处理。

E-mail: puzhong@mail. ynu. edu. cn