

基于 AttentionNet 和 DenseUnet 的脊椎 CT 分割

田丰源, 周明全, 闫峰, 范力, 耿国华*

西北大学信息科学与技术学院, 陕西 西安 710127

摘要 在脊椎 CT 图像分割问题中, 由于脊椎与组织对比度过低和噪声的影响, 传统分割算法存在分割精度差和自动化程度低等问题。基于此, 提出一种通过 AttentionNet 定位脊椎, 然后使用改进的 DenseUnet 进行脊椎 CT 分割的方法。首先, 对所有脊椎 CT 样本数据进行裁剪、重采样、灰度值归一化等预处理操作; 再次, 对样本使用 AttentionNet 训练得到具有位置信息的 Attention 图; 然后, 对传统 DenseUnet 进行改进, 在每个 Dense block 加入 Shuffle 操作来增加网络的鲁棒性, 在每个 Dense block 后加入 1×1 卷积, 以降低通道数, 减少网络参数量; 接着使用改进后的 DenseUnet 对训练样本进行预训练, 得到具有先验信息的预测图; 最后, 将 Attention 图、预测图及原始图像融合为三通道的训练样本作为输入, 采用改进的 DenseUnet 训练分割模型, 并在测试集上进行验证, 最终实现脊椎 CT 自动分割。实验结果表明, 所提方法的分割精度优于传统 DenseUnet, 是一种有效的脊椎 CT 自动分割方法。

关键词 图像处理; 分割; Attention 图; 参数量; 预训练; 多通道融合

中图分类号 TP37

文献标志码 A

doi: 10.3788/LOP57.201008

Spinal CT Segmentation Based on AttentionNet and DenseUnet

Tian Fengyuan, Zhou Mingquan, Yan Feng, Fan Li, Geng Guohua*

School of Information Science & Technology, Northwest University, Xi'an, Shaanxi 710127, China

Abstract In the spinal computed tomography (CT) image segmentation problem, owing to the low contrast between the spine and tissues, and the influence of noise, the traditional segmentation algorithms have problems such as poor segmentation accuracy and low degree of automation. Aiming at solving the above-mentioned problems, a method of locating the spine through AttentionNet and then using improved DenseUnet to perform spinal CT segmentation is proposed herein. First, preprocessing operations such as cropping, resampling, and normalization of gray values are performed on all spinal CT sample data; the samples are trained using AttentionNet to obtain Attention maps with position information. Second, the traditional DenseUnet is improved, and each Dense block adds the Shuffle operation to increase the network robustness. After each Dense block, a 1×1 convolution is added to reduce the number of channels and network parameters. Third, the training samples are pretrained using the improved DenseUnet to obtain the prediction maps with prior information. Finally, the Attention map, prediction map, and the original images are fused into three-channel training samples as the input, and the improved DenseUnet is used to train segmentation model and is verified against the test set. Consequently, the spinal CT automatic segmentation is realized. The experimental results show that the segmentation accuracy of the proposed method is better than that of the traditional DenseUnet, and the proposed method is an effective automatic segmentation method for spinal CT.

Key words image processing; segmentation; Attention map; number of parameters; pre-training; multi-channel fusion

OCIS codes 100.2000; 100.3008; 100.4996

收稿日期: 2019-12-17; 修回日期: 2020-02-13; 录用日期: 2020-02-25

基金项目: 国家重点研发项目(2017YFB1402103)、国家自然科学基金(61672103, 61731015)、陕西省产业创新链项目(2016TZC-G-3-5, 2017ZDCXL-GY-03-01-01)、陕西省重点研发计划(2019ZDLSF07-02, 2019ZDLGY10-01)

* E-mail: ghgeng@nwu.edu.cn

1 引言

随着工作压力的增加,人们久坐不动,导致脊椎病越来越常见,所以预防和诊断脊椎病越来越重要。在传统医学诊断中,医生凭借自己的经验,结合 CT 影像进行诊断。但是单凭经验使用 CT 影像进行诊断时可能由于人为的主观臆断,对病灶位置产生误诊。为了避免传统诊断方法存在的问题,研究者们提出计算机辅助诊断的方法。计算机辅助诊断方法就是通过分割算法将脊椎 CT 影像中感兴趣区域(ROD)分割出来,以便医生观察病灶位置,制定治疗方案。脊椎的解剖结构复杂,在 CT 成像过程中易产生噪声,并且脊椎边界不明,所以脊椎分割问题具有挑战性。目前脊椎 CT 影像分割方法及其改进方法的主要目的在于提高分割准确率、分割效率及自动化程度。传统的图像分割算法主要包括基于阈值的分割方法^[1]、基于区域的分割方法^[2]、基于边缘信息的分割方法^[3]等,但这些方法对噪声敏感,鲁棒性差。研究者们对传统的方法进行改进,提出了新的分割方法。Sungkhun 等^[4]提出一种采用形态学 Otsu 阈值裁剪候选 ROI,再对 ROI 进行超像素聚集分割的方法,该方法提高了分割效率。李滚等^[5]通过隐马尔可夫随机场建立分类的先验分布,结合模拟退火算法和期望最大化算法,很大程度地提升了分割的精确性。以上方法虽然克服了传统分割方法的缺点,但仍存在需要人工干预、不能实现自动化分割的问题。

近些年来,深度学习在图像处理领域迅速发展,神经网络被应用于各种图像处理的场景中^[6-8]。2014年,随着全卷积神经网络(FCN)^[9]被提出,深度学习算法被广泛应用于图像分割中。为解决 FCN 对图像细节分割不好的问题,Ronneberger 等^[10]提出了 Unet,该网络通过跳跃连接层对上采样层和下采样层进行拼接来解决分割细节不好的问题。许多研究者们^[11-15]在这两种网络的基础上提出了新的分割网络。刘忠利等^[11]对 FCN 进行预训练,同时采用随机梯度下降法进行训练,提高了图像分割的准确率,但是该方法对边缘细节处理效果不佳。王亚刚等^[12]为避免层数过深引起网络退化,采用 Unet 与残差神经网络相结合的方法,从而增加了网络深度,提高了分割精确度。冯家文等^[13]提出了一种多源融合的 FCN,该网络对原始数据、数据边缘信息及灰度图像进行拼接,提高了分割精细度,并保证了算法的鲁棒性,但该网络对输入源有所限

制。Sekuboyina 等^[14]提出一种将 AttentionNet 与 Unet 结合的方法,该方法为分割提供了位置像素分布的先验信息,将分割注意力集中在脊椎骨组织上,从而提高了分割的精确度。Li 等^[15]将 Unet 与密集连接相结合,增加了特征的重用性,但是密集连接会导致训练所需的参数过多。

受到以上方法思路的启发,本文提出一种通过 AttentionNet 定位脊椎,使用改进的 DenseUnet 分割的方法。首先,对所有脊椎 CT 样本数据进行裁剪、重采样、灰度值归一化等预处理操作。然后,对样本使用 AttentionNet 得到具有位置信息的 Attention 图。通过在传统 DenseUnet 的每一个 Dense block 中加入打乱操作来增加网络的鲁棒性,在每一个 Dense block 后加入 1×1 卷积,以降低通道数,减少分割网络所需的参数量。使用改进的 DenseUnet 对样本进行预训练,得到具有先验信息的预测图,将 Attention 图、预测图及原始图像融合为三通道的训练样本。最后将融合后的训练样本作为输入,采用改进的 DenseUnet 训练分割模型,并在测试集上进行验证,最终实现脊椎 CT 自动分割。所提方法在经过预处理的 Ljubljanna 大学脊椎公开数据集^[16-17]中的分割结果优于传统 DenseUnet。

2 网络结构与算法原理

所提方法使用了两种网络,通过改进的 DenseUnet 进行预训练,并将结果作为训练样本的先验信息预测图;通过 AttentionNet 生成 Attention 图,并将该图作为训练样本的位置像素分布信息;对预训练结果、Attention 图及原始数据进行通道融合操作并送入 DenseUnet 中进行训练。多通道拼接结构如图 1 所示。

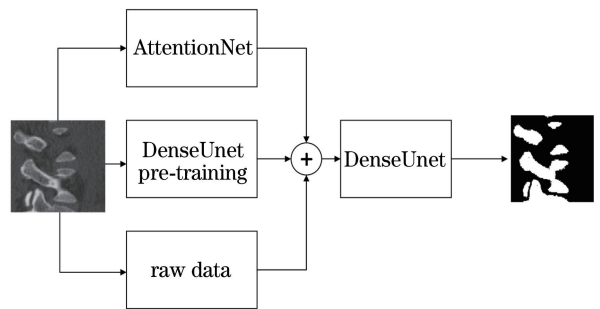


图 1 多通道拼接结构图

Fig. 1 Diagram of multi-channel splicing structure

图 1 中,首先将训练集的原始数据输入改进的 DenseUnet 并进行预训练,将得到的预测图作为训练样本的先验信息来预测图通道;同时将训练集的

原始数据输入 AttentionNet 并进行网络模型训练, 将得到的 Attention 图作为训练样本的位置像素分布信息通道;接着对原始数据通道、先验信息预测图通道及位置像素分布信息通道进行多通道融合, 得到一个多通道训练样本, 并将这一训练样本输入到改进的 DenseUnet 中进行网络分割模型训练。

2.1 AttentionNet

使用 AttentionNet^[14] 获得脊柱的大致定位, 为原始数据添加位置像素信息, 防止周围的噪声点对网络的训练产生负值影响, 从而提高分割精确度。在 AttentionNet 结构的基础上进行如下改进:

- 1) 在图像输入后加入上采样层, 防止池化作用导致的重要特征丢失;
- 2) 在前 4 层的卷积操作中使用 padding 方法^[18], 以保证卷积操作不改变图像大小;
- 3) 将膨胀率分别设置为 2, 3, 5, 避免出现栅格效应, 并保证 75% 的感知域被覆盖;
- 4) 为避免网络出现过拟合的情况, 在空洞卷积后采用 Dropout 技术^[19], 按照 0.5 的概率随机丢弃一部分网络节点。

网络结构如图 2 所示。首先通过最近邻插值对经过预处理的大小为 160×160 的数据进行上采样, 得到 640×640 大小的数据。接着进行 4 次卷积池化操作, 卷积核分别为 $5 \times 5, 7 \times 7, 5 \times 5, 3 \times 3$ 。为了避免卷积操作导致图像减小的情况, 在每一次卷积过程中均使用 padding 方法, 即对每个输入图像填充 2 个像素点。每一次池化均为最大值池化, 来保留图像的最强特征。经过 4 次卷积池化操作后得到的数据大小为 $40 \times 40 \times 64$ 。为了避免出现继续池化再次降低分辨率, 从而造成细节特征丢失的情况, 通过空洞卷积^[19] 来提高感知域。

空洞卷积的膨胀率设置不好会产生栅栏问题^[20], 会丢失局部信息。此次进行 3 次膨胀卷积, 叠加的膨胀卷积的膨胀率不能有大于 1 的公约数, 否则会发生栅格效应, 所以将膨胀率分别设置为 2, 3, 5, 这样可以保证参与计算的感受野能达 75% 以上。空洞卷积后进行池化操作, 得到的数据大小为 $20 \times 20 \times 64$ 。为了避免网络出现过拟合的情况, 在空洞卷积后采用 Dropout 技术^[21], 按照 0.5 的概率随机丢弃一部分网络节点。

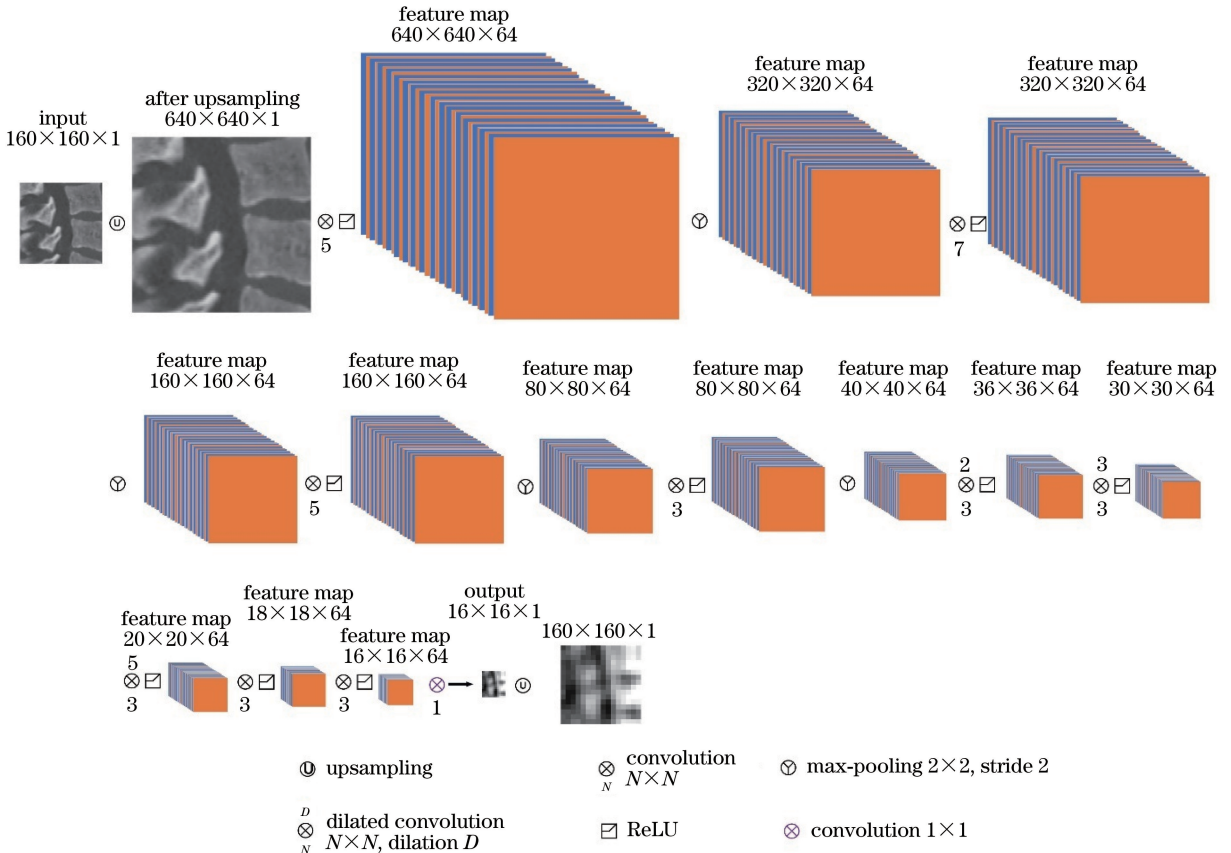


图 2 AttentionNet 结构
Fig. 2 Structure of AttentionNet

接着采用 2 次无 padding 方法的卷积操作,卷积核均为 3×3 ,得到的数据大小为 $16 \times 16 \times 32$ 。最后,使用卷积核为 1×1 的卷积降低通道数,最终得到大小为 $16 \times 16 \times 1$ 的输出数据。整个网络需要 98×10^4 个参数。在需要使用 Attention 图时,使用最近邻插值对其进行上采样即可得到 160×160 大小的 Attention 图。对标签进行 10×10 大小的分块,对每一块区域的前景像素,即脊椎像素进行统计,并计算分块中前景像素出现的概率。图 3 为 16 个像素点对应映射图中的 1 个像素点,得到一个 4×4 的映射图。按照图 3 对所得的概率进行映射,每 100 个像素点对应映射图中的 1 个点,相当于进行一个 10×10 的卷积运算,从而得到一个 16×16 的映射图。通过降低映射图像与输出图像间的方均误差(MSE)来训练 AttentionNet。

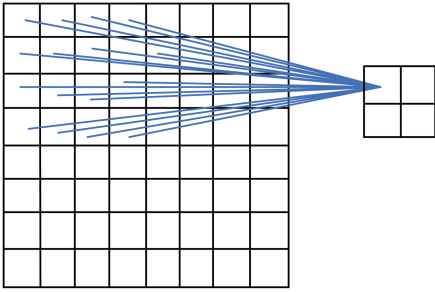


图 3 标签与映射图的映射关系

Fig. 3 Mapping relationship between label and mapping graph

2.2 改进的 DenseUnet

密集卷积网络(DenseNet)^[22]的提出,使研究者对在计算机视觉中的密集连接有较大关注。传统的 DenseUnet^[14]通过将 Unet 中的卷积层替换为 Dense 卷积块来完成与 DenseNet 的结合。Dense block 的主要作用是增加特征的重用性,卷积核的大小均为 3×3 ,具体的构造如图 4 所示。

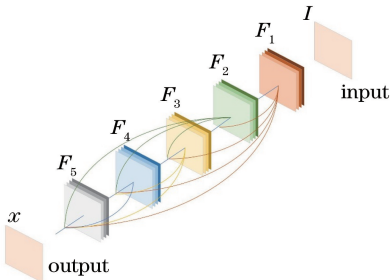


图 4 Dense block 结构

Fig. 4 Structure of Dense block

当 Dense block 中包含 5 个卷积层时,具体过程为:

1)使用第 1 个卷积对输入图像 I 进行操作,得到结果 F_1 ;

2)对 I 与 F_1 进行拼接,并使用第 2 个卷积进行操作,得到结果 F_2 ;

3)对 I, F_1, F_2 进行拼接,并使用第 3 个卷积进行操作,得到结果 F_3 ;

4)对 I, F_1, F_2, F_3 进行拼接,并使用第 4 个卷积进行操作,得到结果 F_4 ;

5)对 I, F_1, F_2, F_3, F_4 进行拼接,并使用第 5 个卷积进行操作,得到结果 x 。

计算过程使用的公式为

$$F_i = \text{ReLU}([I, F_1, \dots, F_{i-1}]), i = 1, 2, 3, \dots, n, \quad (1)$$

式中: $\text{ReLU}(\cdot)$ 为激活函数; i 为第 i 个卷积层; F_i 为 Dense block 中第 i 个卷积层的卷积操作结果; I 为 Dense block 的输入。

在传统 DenseUnet 中,每个 Dense block 包括 6 个阶段,每个阶段先进行 1 次 3×3 卷积操作,再进行 1 次 1×1 卷积操作。然后,后面的阶段 F_i 分别对前面的阶段 $[I, \dots, F_{i-1}]$ 进行 Concatenate 函数拼接操作。整个网络共需要 4697×10^4 个网络参数。

由于传统的 DenseUnet 训练时需要的参数过多,并且缺乏鲁棒性。所以对传统的 DenseUnet 进行如下改进,网络主体结构如图 5 所示。

1)每层 Dense block 包含的卷积层数各不相同。通过实验说明,将参数值设置为左右对称时可能会产生过拟合的问题。将 9 层 Dense block 包含卷积的层数分别设置为 5, 5, 6, 6, 7, 6, 6, 5, 5 时,分割的准确度在 0.8 上下浮动,损失值随着迭代次数增加而增加,出现过拟合的问题。所以分别将 9 层 Dense block 包含卷积的层数分别设置为 5, 5, 6, 6, 7, 7, 6, 6, 5, 防止网络出现过拟合的问题。

2)对 Dense block 的每个卷积进行拼接操作时,使用 Shuffle 方法进行拼接顺序打乱操作,增加网络的鲁棒性。改进后的公式为

$$F_i = \text{ReLU}(\text{Shuffle}[I, F_1, \dots, F_{i-1}])。 \quad (2)$$

3)在每个 Dense block 的最后,使用卷积核为 1×1 的卷积将通道数降为 24,防止出现经过多次卷积导致通道数过多影响训练效率、增加训练参数的现象。

经过 9 层的 Dense block 处理后,最后采用卷积核大小为 1×1 的卷积,并使用 Sigmoid 激活函数进行分类,就可以得到输出的数据。网络训练所需的参数量为 1900×10^4 个,远低于传统的 DenseUnet。

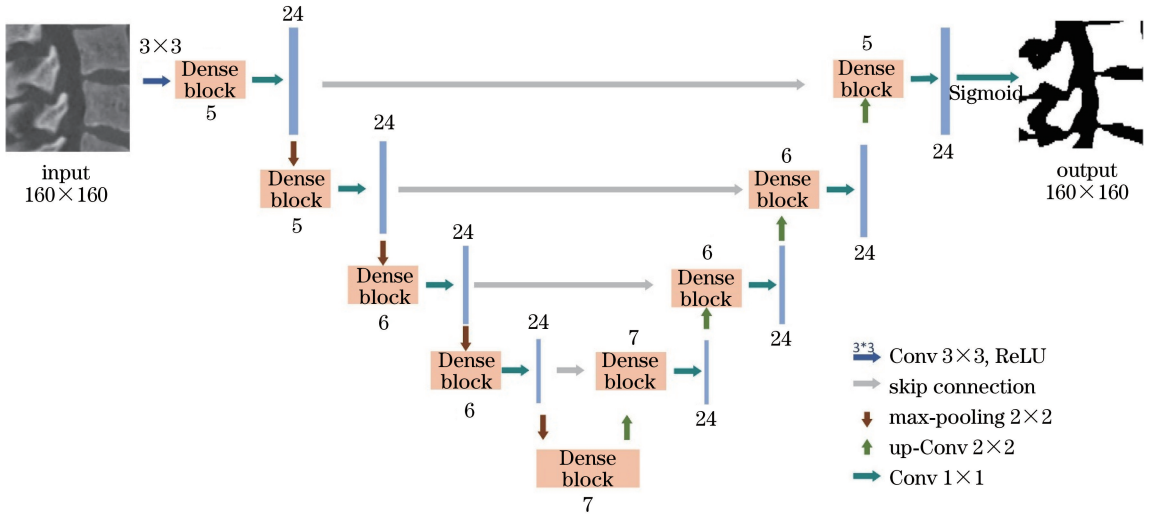


图5 DenseUnet 结构

Fig. 5 Structure of DenseUnet

2.3 网络训练

采用改进的 DenseUnet 进行 100 epochs 的预训练,将训练的 batch size 设置为 20,训练结束得到预测结果。接着训练 AttentionNet,将 batch size 设置为 20,进行 200 epochs 训练,训练结束得到 Attention 图。

对预训练得到的结果、Attention 图及原数据进行通道融合操作,得到一个三通道的训练样本。接着将拼接后的三通道训练样本送入改进的 DenseUnet 中进行训练,将 batch size 设置为 20,进行 200 epochs 训练。训练结束后得到最终的分割结果。为了防止在一定的迭代次数后出现损失停止下降并上升的情况,所有网络均采用自适应学习率。初始学习率设置为 0.001,每 40 epochs 进行一次学习率衰减,衰减率为 0.5,优化器为 Adam^[23]。

因为通过实验得出,使用 Dice_loss 函数对脊椎的分割效果最好,所以在 DenseUnet 中使用最简单的 Dice_loss 函数,表达式为

$$l_{\text{Dice}} = 1 - \frac{2 \times \| \mathbf{y}_{\text{pred}} \cap \mathbf{y}_{\text{true}} \|_1}{\| \mathbf{y}_{\text{pred}} \|_1 + \| \mathbf{y}_{\text{true}} \|_1}, \quad (3)$$

式中: \mathbf{y}_{pred} 为网络的预测输出; \mathbf{y}_{true} 为对应的标签; $\mathbf{y}_{\text{pred}} \cap \mathbf{y}_{\text{true}}$ 为预测的正确结果。

在 AttentionNet 中使用的损失函数为 MSE,公式为

$$l = \min_{\mathbf{y}_{\text{pred}}} \| \mathbf{y}_{\text{pred}} \|_1 - \| \mathbf{y}_{\text{true}} \|_1 \|_2^2, \quad (4)$$

式中: $\| \cdot \|_2$ 为二阶范数; $\min_{\mathbf{y}_{\text{pred}}}$ 为满足最小二阶范数的 \mathbf{y}_{pred} 。

3 实验与分析

实验环境基于 Windows 10 操作系统, CPU 型号为 Intel Core i7-6850K, 显卡为 Nvidia GeForce GTX 2080Ti, 显存大小为 11GB, GPU 环境为 CUDA-Toolkit8.0, 编程开发环境为 PyCharm, 深度学习框架为 Keras。

所使用的数据集为 Ljubljanna 大学脊椎公开数据集, 该数据集包含 10 组病人的脊椎 CT 数据。每组 CT 数据均有相应的 5 个骨组织的标签。使用第 2, 4, 5, 6, 7, 8, 9, 10 组 CT 数据作为训练集; 使用第 1 组 CT 数据作为验证集; 使用第 3 组 CT 数据作为测试集。

3.1 图像预处理

在预处理阶段对脊椎 CT 数据进行裁剪、重采样、灰度值归一化操作。首先通过对 CT 数据进行裁剪来去掉图像中的冗余部分; 接着对数据集进行重采样操作来统一 CT 数据的分辨率; 再对灰度值进行归一化来保证 CT 数据的灰度值为 0 ~ 255; 然后通过对 CT 数据进行裁剪来减少训练过程中的负参数; 最后对每张 CT 切片进行裁切, 得到 160 × 160 大小的 patch 块。

由于 CT 数据上存在没有标签的冗余部分, 这会对训练结果产生负影响。为了保证训练的准确性, 对 CT 数据进行裁剪。对 L1 和 L5 脊椎段进行裁剪, 来保证数据中的全部脊椎段均有与之相对应的标签。

由于每组 CT 数据的分辨率大小不同, 所以对

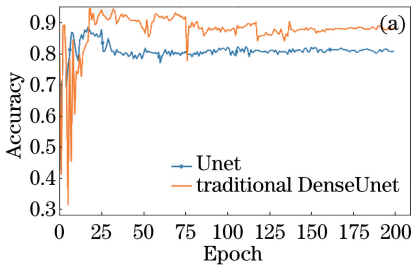
数据进行重采样,将分辨率统一为 $0.5 \text{ mm} \times 0.5 \text{ mm} \times 0.5 \text{ mm}$ 。CT 数据的灰度值范围比较广,所以将 CT 数据的最小灰度值设为 400,最大灰度值设为 3500。将灰度值小于 400 的值重置为 400,将灰度值大于 3500 的灰度值重置为 3500。然后再对像素值进行归一化,得到 $[0, 1]$ 之间的值,接着放缩到 $[0, 255]$ 。

接着对每张 CT 切片及对应标签进行裁切操作,按照 160×160 的大小和 160 的步长进行裁切。为了保证训练的准确性,通过设置阈值为 500 来忽略一些包含骨组织像素数量小于 500 的点。最终得到 3347 个训练集样本,509 个验证集样本及 350 个测试集样本。

3.2 实验结果与分析

3.2.1 Unet 和 DenseUnet 的性能对比

大部分论文都是以 Unet 为基线,对 Unet 进行改进,而本文使用 DenseUnet 为基线,所以对 Unet 和传统 DenseUnet 进行对比。分别采用准确率 (Dice)、IOU、及 VS (volumetric similarity) 作为评价标准。VS 原本是应用于三维分割的评价,这里将其推广到二维分割中,即计算分割结果与 GroundTruth 的像素数量比。三个评价指标的公式分别为



$$D_{\text{Dice}} = \frac{2 \times \| \mathbf{y}_{\text{pred}} \cap \mathbf{y}_{\text{true}} \|_1}{\| \mathbf{y}_{\text{pred}} \|_1 + \| \mathbf{y}_{\text{true}} \|_1}, \quad (5)$$

$$D_{\text{IOU}} = \frac{\| \mathbf{y}_{\text{pred}} \cap \mathbf{y}_{\text{true}} \|_1}{\| \mathbf{y}_{\text{pred}} \cup \mathbf{y}_{\text{true}} \|_1}, \quad (6)$$

$$D_{\text{VS}} = 1 - \frac{|V_{\text{true}} - V_{\text{pred}}|}{V_{\text{true}} + V_{\text{pred}}}, \quad (7)$$

式中: V_{true} 为标签中包含骨组织的像素点数量; V_{pred} 为预测结果中包含骨组织的像素点数量。

每个网络进行 200 epochs 训练,并在测试集上进行测试,分割结果如表 1 所示。可以看出,传统 DenseUnet 的三个评价指标均高于 Unet。

表 1 Unet 与传统 DenseUnet 分割结果

Table 1 Segmentation results of Unet and traditional

Network	DenseUnet		
	D_{Dice}	D_{IOU}	D_{VS}
Unet	94.19	89.68	96.49
Traditional DenseUnet	95.48	91.51	97.81

图 6 为 Unet、传统 DenseUnet 在验证集上的准确率、损失值与迭代次数的关系曲线。从图 6(a) 可以看出,传统 DenseUnet 在经过 200 epochs 迭代后,准确率稳定在 0.9 左右,远高于 Unet 的 0.8。从图 6(b) 可以看出,虽然传统 DenseUnet 的损失值为 0.1 ~ 0.2,但大部分小于 Unet 的 0.2。实验结果证明,传统 DenseUnet 的分割效果要好于 Unet。

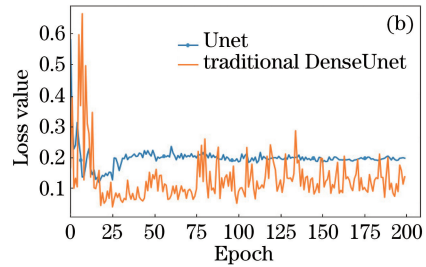


图 6 不同网络在验证集上的准确率、损失值与迭代次数的关系。(a) 迭代次数与准确率的关系;(b) 迭代次数与损失值的关系

Fig. 6 Relationship between accuracy, loss value, and iterations of different networks on validation set. (a) Relationship between iterations and accuracy; (b) relationship between iterations and loss value

使用误差分析图来表示两种网络得到的数据,如图 7 所示。图中第一列为训练集的数据,第二列为验证集的数据,第三列为测试集的数据。可以看出,训练集、验证集及测试集中传统 DenseUnet 的分割准确度浮动要远小于 Unet,并且在验证集和测试集中的中值均高于 Unet。所以采用传统 DenseUnet 作为基线。

3.2.2 所提方法与传统 DenseUnet 和 Unet 关于参数量的对比

Unet 和传统 DenseUnet 作为常见的分割网络,许多研究者都在这两个网络的基础上进行改进。在

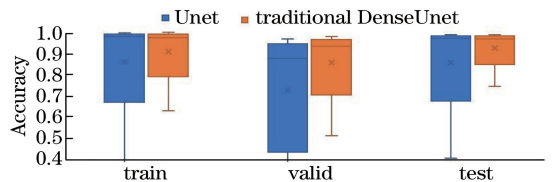


图 7 误差分析图

Fig. 7 Error analysis diagram

参数量上对所提方法和这两种网络进行对比。由于所提方法进行了预训练,并且训练了 AttentionNet,所以所提方法的参数量为 DenseUnet 预训练参数量、AttentionNet 参数量、DenseUnet 训练三通道样本参

数量之和,如表 2 所示。

可以看出,相比于传统 DenseUnet 的参数数量,所提方法减少了约 743×10^4 。虽然所提方法所需的参数数量多于 Unet,但在分割准确率上所提方法比 Unet 高 2.3%。

表 2 不同网络的参数量

Table 2 Number of parameters of different networks

Network	Number of parameters
Traditional DenseUnet	46978875
Unet	31030593
Proposed Method	39543451

3.2.3 不同通道的训练样本和基线的分割结果对比

所提方法共使用了三个通道的训练样本,分别为原始数据通道、预训练预测图通道及 Attention 图通道。本实验对三个通道进行两两组合,然后输入到改进的 DenseUnet 中训练分割模型,并对分割结果进行对比分析。由于每一个训练样本必须包含原始数据通道,所以通过两两组合后,能够得到两种双通道训练样本,即预训练预测结果与原数据通道融合的训练样本(记为 Dense_pred)和 Attention 图与原数据通道融合的训练样本(记为 Dense_atten)。将这两种训练样本与三通道训练样本(记为 Dense_end)分别送入改进的 DenseUnet 中进行训练,并对分割结果进行对比,同时与基线传统 DenseUnet 分割结果进行对比。四种样本分割结果的评价指标如表 3 所示。

表 3 不同训练样本的分割结果比较

Table 3 Comparison of segmentation results of different training samples unit: %

Training sample	D_{Dice}	D_{IOU}	D_{VS}
Traditional DenseUnet	95.48	91.50	97.81
Dense_atten	95.88	92.18	98.42
Dense_pred	94.03	89.00	97.77
Dense_end	96.42	93.19	98.00

从表 3 可以看出:三通道训练样本的 D_{IOU} 和 D_{Dice} 是最高的,相比于基线传统 DenseUnet, Dense_end 的 D_{Dice} 提高了 0.98%, D_{IOU} 提高了 1.85%;但是在 VS 评价指标中, Dense_end 略差于 Dense_atten; Dense_pred 的结果在三个评价指标中是最差的,均低于基线传统 DenseUnet。图 8 为四种训练样本的误差分析图。

从图 8 可以看出,三通道训练样本的最大值、最小值及中值均高于其他两种训练样本和基线,并且离散点数量也少于其他两种训练样本和基线,分割

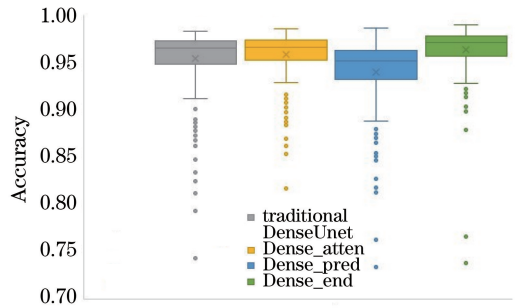


图 8 四种训练样本的误差分析图

Fig. 8 Error analysis diagram of four training samples

结果最好。Dense_atten 分割结果其次, Dense_pred 分割结果最差。对分割结果进行如下分析。

1) 本文的预训练结果是通过改进的 DenseUnet 进行 100 epochs 训练得出的,但是 100 epochs 的训练导致网络没有完全收敛,产生的预测图分割不完全,如图 9(c) 所示,而未分割完全的预测图给改进的 DenseUnet 分割模型训练提供了错误的先验信息,导致网络模型训练出现偏差。所以 Dense_pred 训练改进 DenseUnet 分割模型的结果要差于传统 DenseUnet。

2) Dense_end 比 Dense_pred 多了 Attention 图通道, Attention 图通道相当于提供了位置像素分布先验信息,将分割注意力集中在脊椎骨组织,从而忽略了会对分割模型训练产生负面影响的冗余信息,如图 9(d) 所示。所以 Dense_end 在训练分割模型时,受到预训练预测图分割不完全的影响小,分割模型时分割精度高。

3) 相比 Dense_atten, Dense_end 又加入了形状先验信息,形状先验信息的形状先验程度可以通过预训练中网络分割模型的 Dice 评价指标进行描述, Dice 值越高,形状先验程度就越高。所以 Dense_end 训练得到的分割模型的分割准确度高于 Dense_atten。

3.2.4 所提方法分析

所提分割方法包含三个主要步骤,即通过 AttentionNet 定位得到位置像素分布信息、通过预训练得到形状先验信息、使用三通道训练样本训练分割网络模型。

使用 AttentionNet 进行定位的本质就是对训练集生成位置像素分布信息图,不需要验证集和测试集对定位网络模型进行验证和测试,所以训练集的 Dice 值即为位置像素分布信息精度。通过增加训练样本数量来提高训练的精度,将训练集、测试集及验证集数据均输入到 AttentionNet,并且

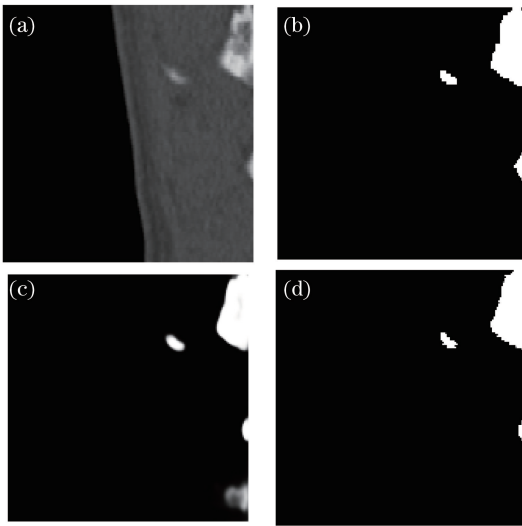


图9 03_365_2号测试样本。(a)原始数据;(b)标签;

(c)预训练预测图;(d) Dense_end 分割结果

Fig. 9 Test sample 03_365_2. (a) Raw data; (b) label;

(c) pre-trained prediction map; (d) segmentation result of Dense_end

以 batch size 为 20 进行 200 epochs 训练,训练的过程持续 2 h 30 min,网络的参数量为 98×10^4 个。在训练过程中得到的 Dice 值为 97.85%。在得到 AttentionNet 定位网络模型后,将训练集数据输入到定位网络模型中对脊椎骨组织进行定位,然而这些训练集数据已经为定位网络模型的训练提供了参数,所以输出的位置像素分布信息图具有较高的定位精度,能够准确地定位脊椎骨组织的位置。图 10 为图像数据与位置像素分布信息图,可以看出,像素的灰度值越高,则包含骨组织的像素点越多。

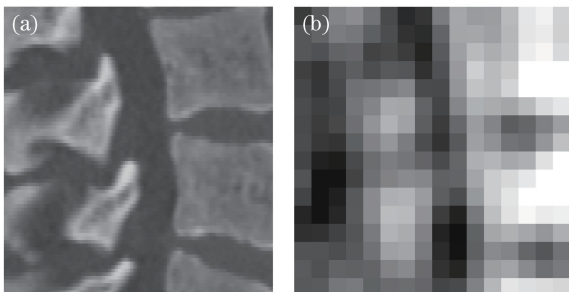


图 10 图像数据与位置像素分布信息图。(a)原图像;

(b)位置像素分布信息图

Fig. 10 Image data and location pixel distribution infographic. (a) Original image; (b) location pixel distribution infographic

在预训练过程中,将训练集输入改进的 DenseUnet,以 batch size 为 20 进行 100 epochs 的

预训练,整个预训练过程持续 2 h 55 min。预训练分割的精度可以用来度量形状先验程度,通过预训练得到的形状先验信息的形状先验程度为 89.22%。100 epochs 的预训练使得网络分割模型没有完全收敛,但可以得到脊椎骨组织的大致形状,这为训练样本提供了形状先验信息。

对位置像素分布信息通道、原始数据通道、形状先验信息通道进行多通道融合,再输入到改进 DenseUnet 中以 batch size 为 20 进行 200 epochs 的预训练,整个训练的过程持续 5 h 50 min。对训练得到的分割网络模型进行测试,得到分割精度的 Dice 值为 96.42%,高于传统 DenseUnet 的分割精度(95.48%),将分割精度提高了 0.98%。整个训练流程中需要的参数量为 3954×10^4 个,传统的 DenseUnet 中需要的参数量为 4697×10^4 个,参数量减少了 743×10^4 个。由于预训练和 AttentionNet 训练可以同时进行,所以所提方法的整体流程耗时为 8 h 45 min。

在训练传统 DenseUnet 时,由于计算机性能的影响,只能将 batch size 设置为 10。每一轮的训练时间在 2 min 30 s 左右,进行 200 epochs 的总耗时约为 8 h 20 min。相比传统的 DenseUnet,所提方法在训练时间上慢了 25 min。采用一种用时间换空间的策略,在增加了 25 min 训练时间的条件下,减少了 743×10^4 个参数量,并且分割精度提高了 0.98%。传统 DenseUnet 与所提方法的分割效果如图 11 所示。选取一张包含骨组织较少的切片和一张包含骨组织较多的切片进行展示,分别如图 11 第一行与第二行所示。从图 11 可以看出,所提方法的分割结果相比传统 DenseUnet 更接近于标签。

4 结 论

提出一种通过 AttentionNet 定位脊椎,使用改进 DenseUnet 进行分割的方法,以解决脊椎自动分割的问题。对样本使用 AttentionNet 得到具有位置信息的 Attention 图,在训练样本中加入 Attention 图通道,相当于在原始图像中添加 Attention 位置像素分布。通过在传统 DenseUnet 的 Dense block 中加入打乱操作来增加网络鲁棒性,在每一个 Dense block 后加入 1×1 卷积,以降低通道数,减少分割网络所需的参数量。对样本使用改进的 DenseUnet 进行预训练,得到具有先验信息的预测图。在训练样本中加入预训练,相当于在

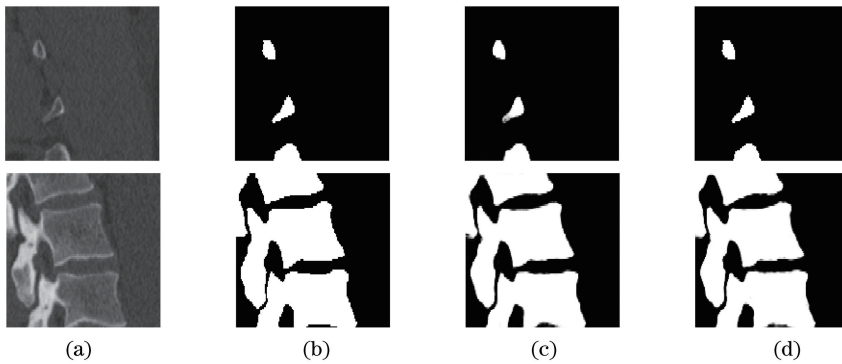


图 11 传统 DenseUnet 和所提方法的分割效果。(a)原图;(b)标签;(c)传统 DenseUnet;(d)所提方法
Fig. 11 Segmentation effect of traditional DenseUnet and proposed method. (a) Original image; (b) label;
(c) traditional DenseUnet; (d)proposed method

原始图像中添加脊椎形状先验信息。将 Attention 图、预测图及原始图像融合为三通道的训练样本,采用三通道样本训练改进的 DenseUnet 分割模型,得到分割结果的准确度相比于传统 DenseUnet 有所提升。虽然未收敛的预训练预测图会为网络训练产生负面影响,但 Attention 图的加入将分割注意力集中在脊椎骨组织,从而忽略了未完全分割的冗余部分。所提方法的误差分析图中存在两个离散点,所以网络的稳定性有待进一步提高。下一步的研究课题为如何提高网络分割的稳定性。

参 考 文 献

- [1] Jung C R, Scharcanski J. Robust watershed segmentation using wavelets[J]. *Image and Vision Computing*, 2005, 23(7): 661-669.
- [2] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. *Medical Image Analysis*, 2017, 42: 60-88.
- [3] Kaur G, Chhaterji J. A survey on medical image segmentation[J]. *International Journal of Science and Research*, 2017, 6(5): 1305-1311.
- [4] Sungkhun S, Rasmeequan S, Chinnasarn K, et al. Vertebral body segmentation using aggregate superpixels [C] // 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), July 13-15, 2016, Khon Kaen, Thailand. New York: IEEE, 2016.
- [5] Li G, Liu H, Ke S Q. Application of spine images segment and extract information [J]. *Electronic Design Engineering*, 2017, 25(16): 183-186.
李滚, 刘欢, 柯善群. 脊椎图像分割与椎体信息提取的算法及应用[J]. *电子设计工程*, 2017, 25(16): 183-186.

- [6] Liu J P, Yang Y Z, Chen M Y, et al. Image dehazing algorithm based on convolutional neural network and dynamic ambient light[J]. *Acta Optica Sinica*, 2019, 39(11): 1110002.
刘杰平, 杨业长, 陈敏园, 等. 结合卷积神经网络与动态环境光的图像去雾算法[J]. *光学学报*, 2019, 39(11): 1110002.
- [7] Feng Y, Yi B S, Wu C Y, et al. Pulmonary nodule recognition based on three-dimensional convolution neural network [J]. *Acta Optica Sinica*, 2019, 39(6): 0615006.
冯雨, 易本顺, 吴晨玥, 等. 基于三维卷积神经网络的肺结节识别研究[J]. *光学学报*, 2019, 39(6): 0615006.
- [8] Chen Y, Guo H G, Ai Y P. Single image dehazing method based on multi-scale convolution neural network [J]. *Acta Optica Sinica*, 2019, 39(10): 1010001.
陈永, 郭红光, 艾亚鹏. 基于多尺度卷积神经网络的单幅图像去雾方法[J]. *光学学报*, 2019, 39(10): 1010001.
- [9] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(4): 640-651.
- [10] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, M. Wells W, et al. *Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science*. Cham: Springer, 2015, 9351: 234-241.
- [11] Liu Z L, Chen G, Shan Z Y, et al. Segmentation of spine CT image based on deep learning[J]. *Computer*

- Applications and Software, 2018, 35(10): 200-204, 273.
- 刘忠利, 陈光, 单志勇, 等. 基于深度学习的脊柱CT图像分割[J]. 计算机应用与软件, 2018, 35(10): 200-204, 273.
- [12] Wang Y G, Wang M, Han J G, et al. R-U-Net neural network for automatic segmentation of femur area[J]. Mini-Micro Systems, 2019, 40(4): 839-844.
- 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[J]. 小型微型计算机系统, 2019, 40(4): 839-844.
- [13] Feng J W, Zhang L M, Deng X Y. Image segmentation based on multi-source fusion FCN[J]. Application Research of Computers, 2018, 35(9): 2877-2880.
- 冯家文, 张立民, 邓向阳. 基于多源融合FCN的图像分割[J]. 计算机应用研究, 2018, 35(9): 2877-2880.
- [14] Sekuboyina A, Kukačka J, Kirschke J S, et al. Attention-driven deep learning for pathological spine segmentation[M]//Glocker B, Yao J H, Vrtovec T, et al. Computational methods and clinical applications in musculoskeletal imaging. Lecture notes in computer science. Cham: Springer, 2018, 10734: 108-119.
- [15] Li X M, Chen H, Qi X J, et al. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes [J]. IEEE Transactions on Medical Imaging, 2018, 37(12): 2663-2674.
- [16] Ibragimov B, Likar B, Pernuš F, et al. Shape representation for efficient landmark-based segmentation in 3-D [J]. IEEE Transactions on Medical Imaging, 2014, 33(4): 861-874.
- [17] 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.
- [18] Liu G L, Shih K J, Wang T C, et al. Partial convolution based padding [EB/OL]. (2018-11-28) [2019-12-16]. <https://arxiv.org/abs/1811.11718>.
- [19] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [EB/OL]. (2016-04-30) [2019-12-16]. <https://arxiv.org/abs/1511.07122>.
- [20] Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation [EB/OL]. (2018-06-01) [2019-12-16]. <https://arxiv.org/abs/1702.08502>.
- [21] Koltchinskii V, Panchenko D. Complexities of convex combinations and bounding the generalization error in classification[J]. Annals of Statistics, 2005, 33(4): 1455-1496.
- [22] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [EB/OL]. (2018-01-28) [2019-12-16]. <https://arxiv.org/abs/1608.06993>.
- [23] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30) [2019-12-16]. <https://arxiv.org/abs/1412.6980>.