

多尺度并行全卷积神经网络的肺计算机断层扫描图像非刚性配准算法

林立昊, 易见兵*, 曹锋, 方旺盛

江西理工大学信息工程学院, 江西 赣州 341000

摘要 图像配准广泛应用于图像引导的肺肿瘤放射治疗,但现有算法对形变较大的图像配准效果不佳。针对该问题提出一种算法利用多尺度并行下采样模块缩减图像大小,得到多尺度低分辨率特征图;并采用金字塔空洞卷积模块提取图像特征,以提高模型的感受野。该算法通过自适应通道注意力模块调整神经网络对不同形变特征的偏重,以解决模型偏重于较大形变而对小形变配准效果不佳的问题;同时在损失函数中加入平滑性约束来提高形变场的平滑性;并通过训练样本数据扩增的方法提高算法模型的稳定性和泛化性。在 DIR-lab、Creatis 数据集测试中,提出算法的目标配准误差 (TRE) 分别为 1.71、1.50 mm,而全卷积神经网络 (FCN) 算法的一次迭代 TRE 分别为 2.83、2.01 mm,实验结果表明提出算法的 TRE 明显小于 FCN 算法,且其泛化性和稳定性也较好。

关键词 医用光学与生物技术; 计算机断层扫描图像配准; 通道注意力; 非刚性配准; 多尺度; 空洞卷积

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP202259.1617004

Non-Rigid Registration Algorithm of Lung Computed Tomography Image Based on Multi-Scale Parallel Fully Convolutional Neural Network

Lin Lihao, Yi Jianbing*, Cao Feng, Fang Wangsheng

School of Information Engineering, Jiangxi University of Science and Technology,
Ganzhou 341000, Jiangxi, China

Abstract Image registration is widely used in image-guided lung tumor radiotherapy, but the existing algorithms are not effective against large deformation images. Therefore, this paper proposes an algorithm using a multiscale parallel down sampling module to reduce the image size and obtain a multiscale low-resolution feature map, and the pyramid dilated convolution module is used to extract image features to improve the model's receptive field. The algorithm adjusts the bias of the neural networks on different deformation features through the adaptive channel attention module to solve the problem in which the model is biased and poor for large and small deformation registration, respectively. Simultaneously, a smoothness constraint is added to the loss function to improve the deformed field smoothness. The training set data-augmentation method is also used to improve the model's stability and generalization. The proposed algorithm has target registration errors of 1.71 mm and 1.50 mm in the DIR-lab and Creatis datasets, respectively, whereas the one-iteration fully convolutional neural network (FCN) has target registration errors (TREs) of 2.83 mm and 2.01 mm in the above datasets, respectively. The experimental results show that the TRE of the algorithm is significantly smaller than that of the FCN algorithm, and the generalization and stability performances of the algorithm are also improved.

Key words medical optics and biotechnology; computed tomography image registration; channel attention; non-rigid registration; multi-scale; dilated convolution

收稿日期: 2021-08-06; 修回日期: 2021-09-08; 录用日期: 2021-09-24

基金项目: 国家自然科学基金(61862031)、江西省自然科学基金(20181BAB202004)、江西省教育厅科技项目(GJJ190458, GJJ200818)、赣州市科技计划项目(GZKJ20206030)

通信作者: *yjianbing8@163.com

1 引言

随着生物医学与计算机等技术的不断发展,通过计算机断层扫描(CT)等医学成像技术能够得到肺部清晰的解剖结构和组织特征,为临床中确定病人肿瘤位置提供了有效信息。由于呼吸运动导致不同相位CT中的肺部肿瘤位置偏移较大,在临床中医疗工作者为了获得更加准确的肿瘤位置信息,通常需要通过图像配准技术将两幅或多幅图像对齐,辅助医生进行诊断、分析。

传统医学图像配准多采用基于局部区域图像强度相似性寻优方法或基于图像特征配准方法,前者通过不断迭代寻找空间变换的最优解,在处理单模态和形变较小图像配准问题时有较好效果;后者大多采用特征点进行匹配,如定向快速特征检测(ORB)、尺度不变特征变换(SIFT)、加速鲁棒性特征(SURF)等算法,该类算法在多模态和图像形变剧烈的情况下都有较好的表现^[1-4]。以上方法经过多年的研究已经取得不错的效果,但这些算法执行时耗费的时间较长,通常配准一组3D图像需要数十分钟,实时性较差。

近年来,利用深度学习技术进行图像配准逐渐成为医学图像配准领域的研究热点,这些方法通过搭建卷积神经网络实现图像配准。丛明等^[5]使用基于监督学习的方法搭建了核磁共振(MR)图像配准网络,该网络能自动从标注数据中学习图像间映射关系,并取得了不错的效果。Hu等^[6]提出以标签驱动的弱监督学习实现图像配准,该方法一定程度上绕过了基于图像强度的相似性度量方法。但弱监督学习不善于处理多模态和时间序列上的形变图像,且这类方法需要医学专家对海量数据进行手动标注,对标注人员的专业性和工作量是一项艰巨的挑战。在Jaderberg等^[7]提出空间变换网络(STN)后,无监督学习成为学者们的主要研究方向。Balakrishnan等^[8]提出基于U形卷积神经网络(U-Net)来实现单模态配准的体素变形(voxelmorph)方法,并在脑部数据集上取得了较好的效果。de Vos等^[9]采用B样条插值结合卷积神经网络搭建了深度学习图像配准框架(DLIR),并在3D图像配准上取得了满意的效果。薛湛琦等^[10]提出基于低分辨率辅助特征和卷积神经网络的无监督脑部图像配准方法。卢佳毅^[11]将注意力机制应用于3D肺部CT配准。无监督学习只需要采集足够的待配准训练集图像,通过网络训练即可获得配准模型,这省去了大量标注数据的工作。但目前基于无监督学习的肺部CT图像配准算法仍存在局限,如在3D图像配准时对形变较大的图像配准效果较差。

针对传统的医学图像配准算法耗时较长、有监督深度学习的图像配准需要大量的标注数据,而当前无监督深度学习的图像配准算法对形变较大的图像配准效果一般的问题,本文在全卷积神经网络(FCN)算

法^[12]的基础上提出了一种基于多尺度并行全卷积神经网络(D-FCN)的肺CT图像非刚性配准算法。相较于以往的算法,本文主要贡献为:

1) 针对传统算法对形变较大的图像配准效果不佳的问题,提出利用多尺度并行下采样和金字塔空洞卷积扩大感受野,并通过提取多层不同分辨率的特征来提高算法对较大形变图像的配准效果,且算法的计算量较小;

2) 针对算法网络提取的大形变特征(肺部轮廓等)比较丰富,而小形变特征(毛细血管等)比较匮乏的问题,提出将3次下采样后的特征图像通过自适应通道注意力模块,实现对微小形变的特征自适应加权,以提高图像配准的精度;

3) 为提高配准精度和形变图像的平滑性,在图像相似性度量的基础上加入平滑性约束来刻画损失函数,并通过高斯噪声和椒盐噪声攻击等数据扩增方法来提高网络模型的稳定性和泛化性。

2 基于 D-FCN 的肺 CT 图像非刚性配准

2.1 算法综述

本文提出算法的流程如图1所示,具体为:输入一组3D配准图像(浮动图像和固定图像),经过D-FCN学习后输出一个密集位移矢量场(DVF),该矢量场表示浮动图像中的每个像素点与固定图像中对应位置的位移;接着浮动图像根据位移矢量场表示的每个像素点的位移信息,利用空间变换网络(STN)生成一副新的形变图像;最后根据损失函数(包括相似性度量和平滑性约束)计算固定图像和形变图像之间的差异性并更新模型的参数,经过多次迭代后得到最终的形变模型。

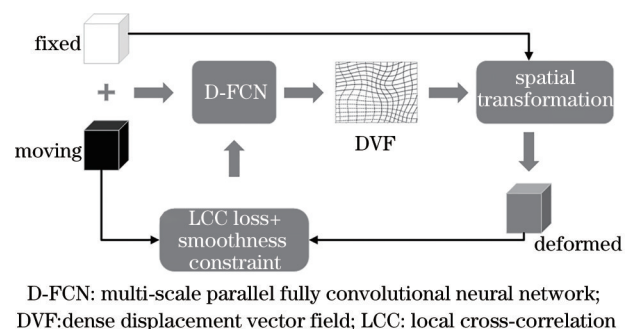


图1 D-FCN肺图像配准算法流程图

Fig. 1 Flow chart of D-FCN lung image registration algorithm

2.2 网络结构

2.2.1 多尺度并行下采样模块

3D肺CT图像信息含量丰富、内容复杂,且不同病人、不同设备间采集的图像差异较大,这要求网络具有良好的泛化性和自适应性。大部分深度学习算法均采用小窗口进行特征提取,如采用3*3*3大小的卷积

提取特征,需要多次计算才能获得较大形变特征,配准效率低。基于以上原因,本文充分考虑了肺 CT 图像配准任务在大形变特征的需求,在金字塔池化模块 (PPM)^[13]上进行了改进,形成多尺度并行下采样模块以实现模型多尺度特征的提取。

如图 2 所示,该模块首先对输入图像分别进行步长为 1、2、4 的平均池化,得到 3 种不同尺度的特征图,再通过批量归一化 (BN) 层和 ReLU 激活函数,将 3 种特征图和输入图像同时通过三线性插值,得到不同尺

度下分辨率为原图像 1/2 大小的特征图,最后将 3 种不同尺度特征图以及原特征图进行特征连接,输出通道数为 $4 \times C$ 的级联特征图, C 表示特征图的通道数。图 2 中, $H \times W \times D$ 表示特征图的分辨率,池化层和卷积层的卷积核均为 $3 \times 3 \times 3$ 。图 3 为同分辨率下的不同尺度特征图,图 3(a) 为输入模型的原始图像,图 3(b) 包含较多纹理、形状特征,信息熵较大,图 3(c) 和 3(d) 多为强度特征,信息熵较小,让这些特征图进入下一层,将有利于模型对多尺度图像的理解。

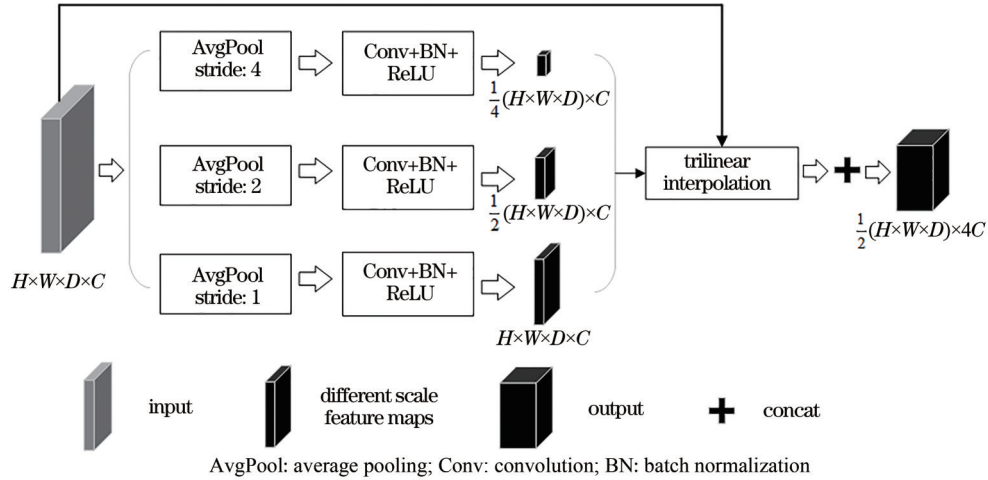


图 2 多尺度并行下采样模块

Fig. 2 Multi-scale parallel down-sampling module

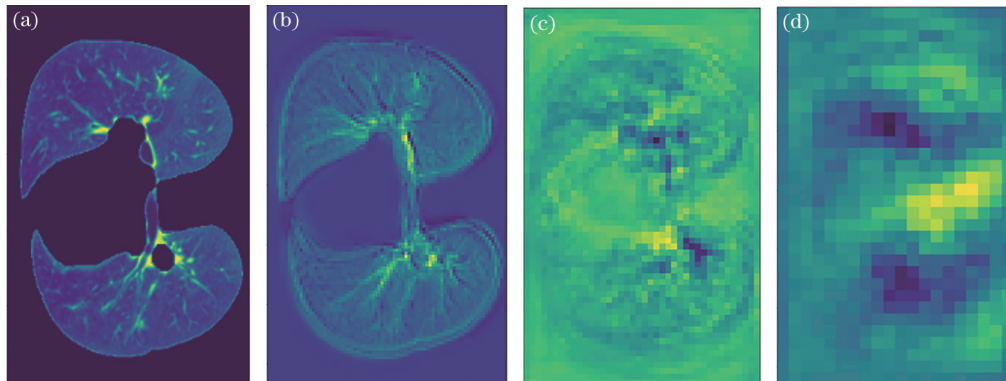


图 3 同分辨率不同尺度下的肺部特征图。(a)原图像;(b)池化核为 1;(c)池化核为 2;(d)池化核为 4

Fig. 3 Lung feature maps at the same resolution and different scales. (a) Original image; (b) pooling core is 1;

(c) pooling core is 2; (d) pooling core is 4

2.2.2 金字塔空洞卷积模块

经多尺度并行下采样后,普通卷积难以同时处理包含众多不同尺度信息特征图,且下采样采用池化层浓缩特征,容易造成信息丢失影响配准效果,因此本文采用金字塔空洞卷积模块进行特征提取来解决上述问题。空洞卷积目前已广泛地应用在目标检测和语义分割等任务中^[14-15],相较于普通卷积,它需要设定一个扩张因子控制卷积核中的空洞。当扩张因子为 1 时,空洞卷积和普通卷积操作一样;当扩张因子大于 1 时,卷积核中的每个算子会略过相应数量的点,形成一个范

围更大且稀疏的卷积核。早期在语义分割等分类任务中,空洞卷积(扩张因子为 1)只是用来提取图像特征,在下采样时使用池化层压缩数据,这样做的优点是它能对特征降维并保持性质不变,缺点是池化层本身是卷积运算,该方法增加了计算量且模型较大。图 4 表示空洞卷积效果图,图 4(a)~4(c)表示不同扩张因子的卷积核,图 4(d)~4(f)表示不同扩张因子下卷积计算时的感受野,颜色越深的部分其感受野越强烈。感受野的计算公式为

$$r_n = r_{n-1} + (k - 1) * \prod_{i=1}^{n-1} S_i, \quad (1)$$

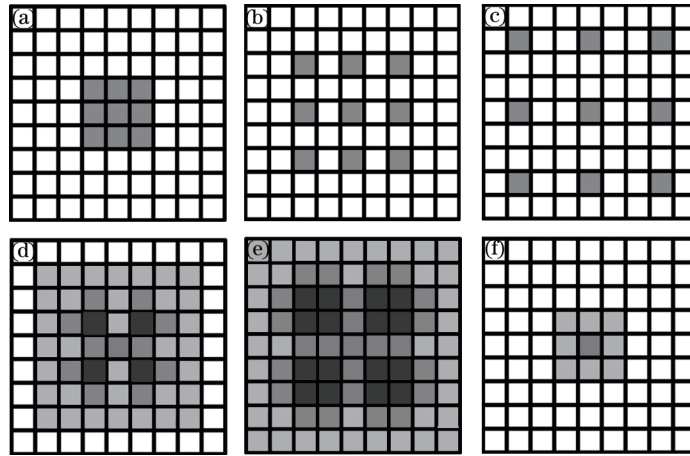


图 4 空洞卷积示意图。(a)扩张因子为 1;(b)扩张因子为 2;(c)扩张因子为 3;(d)扩张因子为 1 时的感受野;(e)扩张因子为 2 时的感受野;(f)扩张因子为 3 时的感受野

Fig. 4 Schematic diagram of dilated convolution. (a) Dilated factor is 1; (b) dilated factor is 2; (c) dilated factor is 3; (d) receptive fields with dilated factor of 1; (e) receptive fields with dilated factor of 2; (f) receptive fields with dilated factor of 3

式中： r_n 表示第 n 层感受野； r_{n-1} 表示第 $n-1$ 层感受野； S_i 表示第 i 层卷积或池化的步长； k 表示卷积核大小。

如图 5 所示,通过将不同扩张因子的空洞卷积堆叠,金字塔空洞卷积可以同时获取特征图中不同尺度的信息,该模块已在 Deeplab 等模型^[14]中取得了不错的效果。模块中所有空洞卷积均采用 $3 \times 3 \times 3$ 卷积核,步长为 1,为保证图像特征不损失信息,空洞卷积的 padding 参数需设置为和扩张因子相同的大小,每个空洞卷积层都包含 BN 层和 ReLU 激活层。根据扩张残差网络(DRN)的结论^[15],一般将扩张因子设为连续等距扩大的数值可以有效避免棋盘格效应,所以为平衡运算效率和配准效果,且经实验验证,本文将扩张因子分别设为 1、2、4、8。

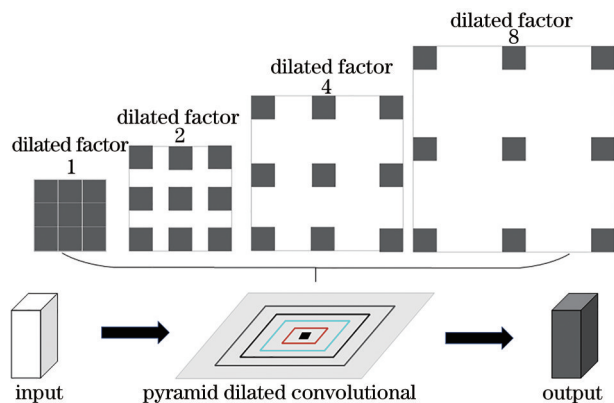


图 5 金字塔空洞卷积模块

Fig. 5 Pyramid dilated convolution module

2.2.3 自适应通道注意力模块

注意力机制可以引导模型偏重于信息量大的特征图,提升模型性能的同时避免了增加网络深度。由于本文模型通过多尺度并行下采样模块和金字塔空洞卷积层提取特征后,表达形变较大特征图(如肺部轮廓等全局特征)数量明显多于微小特征(如血管形变等局部

特征),但本文模型需要更多表达微小形变的特征来提高配准精度。为解决此问题,本文提出将 3 次下采样后的特征图像通过自适应通道注意力模块,实现对微小形变特征的自适应加权,舍弃部分较大形变特征。图 6(a)和 6(b)分别表示 $1/8$ 分辨率特征图中包含较大形变和微小形变信息的图像。

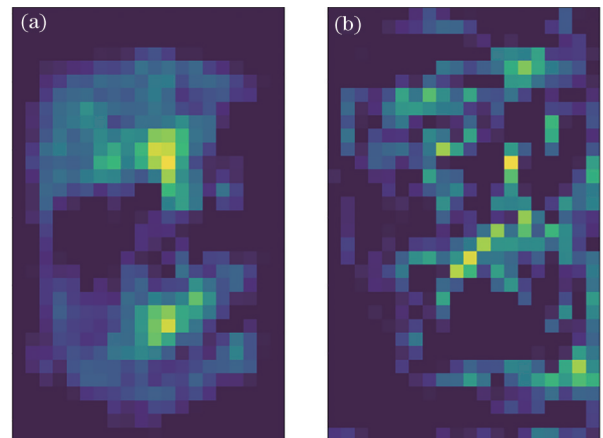


图 6 经过 3 次下采样后 $1/8$ 分辨率下的不同特征图。(a)较大形变特征图;(b)微小形变特征图

Fig. 6 Different feature maps at $1/8$ resolution after three down sampling. (a) Large deformation feature map; (b) small deformation feature map

常见的通道注意力机制往往作为特征提取的辅助模块参与训练,计算每个特征图的权重,最终将效果最好的特征图保留并传入下一层,该机制的有效性已在压缩激励网络(SE-net)等许多工作^[16-17]中得到了验证。但通道注意力机制在 SE-net 等算法中有较深的网络层数作为支撑,本文采用的 FCN 深度不足,较难完全发挥通道注意力机制强劲的通道筛选功能。基于对信号处理中软阈值函数消除冗余信息的理解,以及深度残差收缩网络的启发^[17],本文通过对通道注意力

机制设置一个以图像差异性为基准的软阈值 τ , 将其改进为自适应通道注意力模块, 使模型可以主动舍弃不感兴趣的特征图。如图 7 所示, 将特征图输入自适应通道注意力模块后, 依次通过全局平均池化 (AvgPool)、全连接层 (FC)、ReLU 层、FC 和 Sigmoid 层, 得到各通道的权重值, 再将权重和特征图同时输入软阈值层, 根据软阈值 τ 调整各通道特征的权重。将重新分配完权重的特征图输入卷积核为 $3 \times 3 \times 3$ 的卷积层, 舍弃权重较低的大形变特征图, 得到数量为原通道数 $1/2$ 的特征图, 平衡模型对大形变的侧重, 从而提高模型性能。

在自适应通道注意力模块中, 输入注意力模块前的特征图为 4 种不同尺度图像级联的矩阵, 其矩阵中

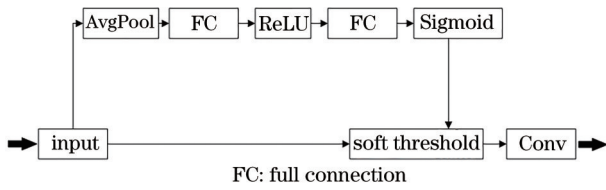


图 7 自适应通道注意力模块
Fig. 7 Adaptive channel attention module

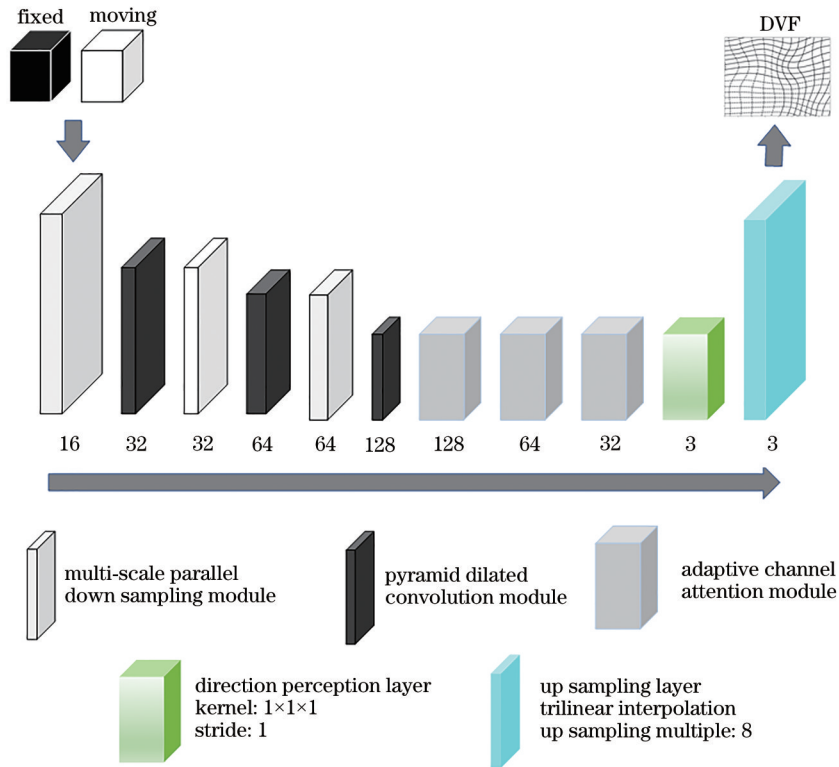


图 8 空洞 FCN 结构图
Fig. 8 Structure diagram of dilated FCN

该网络模型能接受任意尺寸的输入图像, 可以对图像进行像素级分类, 利用空洞卷积扩大感受野并提取多尺度图像特征。网络中的下采样层采用 3 个多尺度并行下采样模块, 每次下采样后跟随一个金字塔空

前 3 个子块主要为较大形变特征图, 第 4 个子块为较小形变特征图。在软阈值层中, 通过随机选取大形变子块和小形变子块中的特征图各 10 对, 计算这 10 对强度差异 D_{diff} 均值得到 τ , 其中 D_{diff} 的计算公式为

$$D_{diff}(A, B) = \sqrt{\sum (A_i - B_i)^2}, (i = 1, 2, \dots, n), (2)$$

式中, A_i 和 B_i 分别表示大形变特征图 A 和微小形变特征图 B 的像素强度值。接着计算当前大形变特征图 A 和微小形变特征图 B 的差异 D_{diff} , 若其差异大于软阈值 τ , 则判断该特征图为较大形变特征图, 对其加权自适应参数 k , 在本文实验中 k 取 0.9 效果较好, 反之, 其权重值不变, 则有

$$W_{weights}^* = \begin{cases} k * W_{weights}, & D_{diff} \geq \tau \\ W_{weights}, & 0 < D_{diff} < \tau \end{cases} (3)$$

2.3 D-FCN

本文提出了一种 D-FCN, 如图 8 所示, 该网络在 FCN^[12] 的结构基础上, 采用 3 层多尺度并行下采样模块获得多尺度低分辨率图像, 通过金字塔空洞卷积模块进行多尺度特征提取, 再通过 3 层自适应通道注意力用于特征处理, 最后经过方向感知层和上采样层得到输入图像对的密集位移矢量场。

洞卷积模块进行特征提取。经过 3 个下采样层和特征提取层后配准图像对的大小是原始图像的 $1/8$, 然后进入 3 层自适应通道注意力模块, 在此模块中, 算法会从不同通道中多次挑选合适的图像特征。由于 3D 图

像配准需要得到每个像素点的位移信息,本文采用一个 $1 \times 1 \times 1$ 卷积层代替 CNN 的 FC 层,线性预测像素在 X、Y、Z 这 3 个方向上的偏移。最后图像对通过上采样层,由缩放因子为 8 的三线性插值函数生成与输入图像相同分辨率的 DVF。

2.4 损失函数

图像配准的主要目标是获得空间变换 φ , 找到浮动图像 M 和固定图像 F 间逐像素的对应关系。传统方法通常将图像配准转化为优化问题,重复迭代以识别出使图像对齐最大化的变换,这一过程常使用图像的差异性测度描述。对于形变图像配准,空间变换通常以具有空间平滑度约束的密集位移矢量场进行表述。与 voxelmorph^[8]类似,本文模型的损失函数 $L(\cdot)$ 由图像差异性度量 $D(\cdot)$ 和正则项 $R(\cdot)$ 组成, λ 表示正则项系数,如下式所示:

$$L(F, M, \varphi) = D[F, \varphi(M)] + \lambda R(\varphi). \quad (4)$$

在医学图像配准中,有多种可用于判别配准精度的差异性度量方法,如基于图像强度均方差、交叉互相关和互信息^[18]等。前两种通常用于单模态图像配准,且基于图像强度均方差的方法受图像强度分布和对比度影响较大,鲁棒性欠缺;基于互信息的配准方法在传统多模态图像配准中较常使用,但目前尚未有研究表明其在深度学习中的性能。本文采用局部交叉互相关函数作为图像差异性损失函数。要说明的是在 voxelmorph 算法^[8]中给出了局部和全局交叉互相关的详细公式,在本文中采用 $5 \times 5 \times 5$ 窗口的局部互相关函数效果较好。通过最小化浮动图像和固定图像之间的差异性度量 $D(\cdot)$ 来训练模型,可以得到其形变场 DVF。差异性度量公式如下式所示:

$$D[F, \varphi(M)] = 1 - CC[F, \varphi(M)], \quad (5)$$

式中, CC 表示交叉互相关函数。

在训练时使差异性度量 D 最小化,即将浮动图像最大化地向固定图像配准,但这样得到的 DVF 空间平滑性效果较差。因此需要在损失函数中加入正则化函数来约束这种过拟合,由于在图像配准中各点误差较小,不容易产生梯度爆炸,并且 3D 图像预测中各点需

要考虑矢量的方向性,基于以上因素本文选择了 L2 函数作为空间正则项 $R(\varphi)$, 如下式所示:

$$R(\varphi) = \frac{1}{n} \sum \|M - \varphi(M)\|_2^2. \quad (6)$$

3 配准实验设置及结果分析

本文实验的硬件环境:CPU 为 Intel Xeon(R) E5-2630 v4, 显卡为 4 块 Nvidia Tesla P100; 软件环境:操作系统为 Ubuntu16.04, 深度学习框架为 Pytorch。在算法的网络结构中使用 Adam 优化器进行训练,学习率设为 0.001, 损失函数正则项系数 λ 取 0.2, batch-size 为 4, epoch 大小设为 100, 在数据集扩增后,为提高训练效率采用 DataParallel 工具包进行多显卡(GPU)并行训练。

3.1 数据集和数据增强

为验证算法的有效性,本文采用美国德克萨斯大学安德森分校的癌症中心公开的 DIR-lab 肺数据集^[19]和 Léon Bérard 癌症中心公开的 Creatis 数据集^[20]对算法进行测试。DIR-lab 数据集由 10 个肺癌病人的 4D-CT 图像组成,每组 CT 包括从最大呼入(T00)到最大呼出(T50)的 10 个相位点的 3D CT 图像; Creatis 数据集包括 6 个非小细胞肺癌病人的 4D-CT 图像,其中 case1 的最大呼入相位为 T10, 最大呼出相位为 T60, 其余 case 的最大呼入相位为 T00, 最大呼出相位为 T50。为评估算法的性能,在最大呼入和最大呼出的两个相位点, DIR-lab 数据集和 Creatis 数据集分别给出了 300 个和 100 个左右的专家标志点来描述两个相位的对应位置。

如图 9 所示,为提高算法的抗过拟合性,本文对原始数据分别通过增加高斯噪声和椒盐噪声来扩增样本数据,图 9(a)为原始图像,图 9(b)为加入均值为 0、标准差为 10 的高斯噪声图像,图 9(c)为加入 $0.05 \times$ 图像总像素点个数的椒盐噪声图像。将原来 DIR-lab 数据集 100 个 3D 肺 CT 数据扩增到 300 个, Creatis 数据集 60 个 3D 肺 CT 数据扩增到 180 个。将 DIR-lab 所有同一个 case 的图像两两配对(除验证集数据)得到 8690

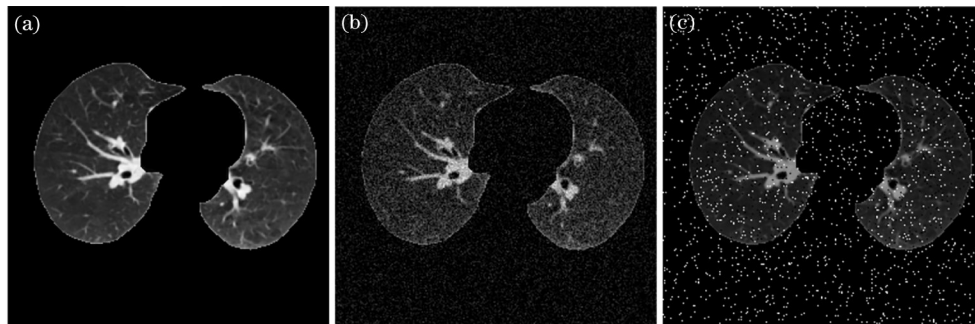


图 9 原始肺 CT 图像及在噪声攻击下的图像(轴状面)。(a)原始图像;(b)加入高斯噪声图像;(c)加入椒盐噪声图像
Fig. 9 Original lung CT image and images under noise attack (axial surface). (a) Original image; (b) image with Gaussian noise; (c) image with salt and pepper noise

组图像对作为训练集,将 10 组原始 CT 图像的最大呼入(浮动图像)和最大呼出(固定图像)图像对作为验证集;在 Creatis 数据集中,将最大呼入设为浮动图像,最大呼出设为固定图像,训练集共 5214 个图像对,6 组原始肺 CT 图像对为验证集。

本文的模型可以配准任意大小的图像对,但为了简化实验,对所有数据都以相同的方式进行预处理。利用区域生长分割算法去除肺 CT 图像中的骨骼,仅保留肺实质,再根据肺 CT 图像强度的特点设定相应阈值去除噪声。将 DIR-lab 数据集图像大小按给定的原始像素尺寸重采样为 $248 \times 160 \times 96$, Creatis 数据集图像大小重采样为 $280 \times 200 \times 288$ 。最后将上述两个数据集进行归一化操作,使下采样时能充分利用每一次卷积计算,减少计算误差。

3.2 实验结果分析

本文从强度差异性、图像配准误差、计算复杂度来评价提出算法在 DIR-lab 数据集和 Creatis 数据集中的性能。

3.2.1 算法在 DIR-lab 数据集 中的 评价

1) 强度差异性

为评价浮动图像到固定图像的配准效果,采用浮动图像根据形变场得到的形变图像和固定图像之间的强度差异来度量,两者之间差异越小,说明本文算法的配准效果越好。选取 DIR-lab 数据集的 case1 来进行该项性能测试,如图 10 所示,图 10(a)~10(c) 分别为固定图像、待配准的浮动图像和配准后的形变图像,图 10(d)、10(e) 分别表示浮动图像和形变图像相对于固定图像的像素强度差值,黑色部分表示差异值小,配准基本到位,白色部分表示差异值较大,尚未完全对齐。由图可见,图 10(e) 中肺下部轮廓相对于图 10(d) 的大面积差异已基本消除,这表示肺轮廓基本配准到位,且肺实质内的血管等白色部分也有明显改善,其强度差异较小,肺血管部分基本实现对齐。由上可知,经过本文算法得到的形变图像与固定图像强度差异有明显减小,血管和肺的边界部分基本实现对齐,说明算法在该评价方式下的效果较好。

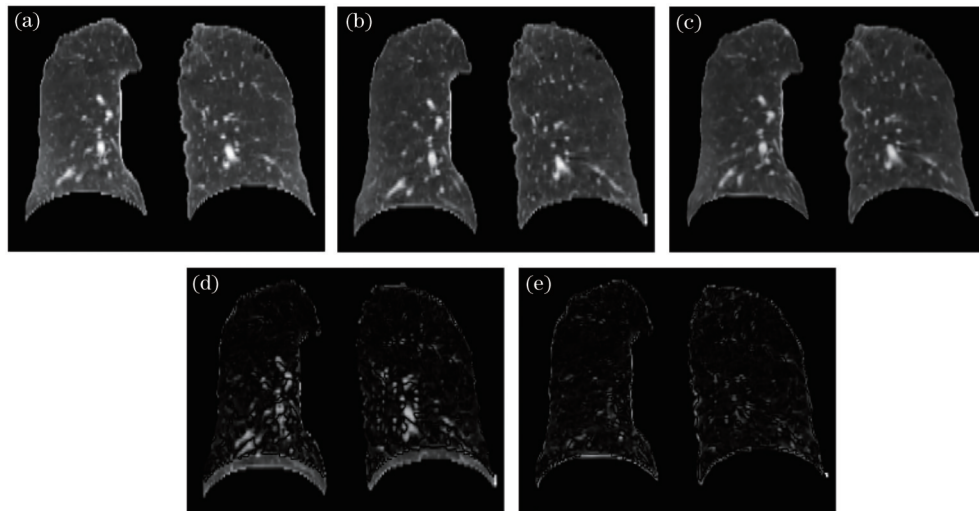


图 10 DIR-lab 数据集 case1 配准结果(冠状面)。(a)固定图像;(b)浮动图形;(c)形变图像;(d)配准前差异;(e)配准后差异
Fig. 10 Registration results of case1 in DIR-lab dataset (coronal plane). (a) Fixed image; (b) moving image; (c) deformed image; (d) difference before registration; (e) difference after registration

2) 目标配准误差

目标配准误差(TRE)^[21]指的是浮动图像的专家标志点形变后的位置和固定图像专家标志点的位置之间的欧氏距离,距离越小,表明该算法的配准精度越高。为验证本文各改进步骤的有效性,分别对多尺度并行下采样模块、金字塔空洞卷积模块和自适应通道注意力模块进行消融实验。实验结果如表 1 所示,给出了算法各改进步骤在 DIR-lab 数据集的 300 个专家点的 TRE 的平均值(标准偏差),表中未采用模块均以同等深度 $3 \times 3 \times 3$ 卷积层代替。从表 1 可以看出,提出的各改进步骤在图像配准中都对算法性能有一定的提升,特别是自适应通道注意力模块的作用比较明显。

由于多尺度并行下采样中得到了 4 个不同尺度的

特征图,采用 4 个不同扩张因子的空洞卷积叠加,形成金字塔空洞卷积模块进行特征提取,其中扩张因子为 1 的空洞卷积层用于提取原分辨率特征图,扩张因子为 2、4、8 的空洞卷积分别用于提取 3 种不同尺度的特征。本文通过实验验证了不同扩张因子在模型中对配准精度的影响,见表 2。由表 2 可见,扩张因子不宜设得过大,一般超过 16 时,会使采样到的数据越来越稀疏,且较大的扩张因子容易出现局部信息丢失、信息之间依赖性损失较大等问题,上述因素将导致配准精度下降;而扩张因子过小则无法发挥空洞卷积在多尺度特征提取中对较大形变的优势,如 case8、case10 的配准效果较差。经实验综合考虑,选择扩张因子为 1、2、4、8 最适合本文任务。

表 1 提出算法各改进步骤在 DIR-lab 数据集上的 300 个专家点 TRE(标准偏差)

Table 1 Target registration errors (standard deviations) of 300 expert points in DIR-lab dataset for each improvement step of proposed algorithm

Multi-scale parallel down sampling	Pyramid dilated convolution	Adaptive channel attention	Spatial regularization	Data augmentation	Target registration error (standard deviation) /mm
					3.45 (2.19)
✓					2.77 (1.77)
✓	✓				2.60 (1.97)
✓	✓	✓			1.96 (1.29)
✓	✓	✓	✓		1.90 (1.34)
✓	✓	✓	✓	✓	1.71 (1.20)

表 2 模型中不同扩张因子在 DIR-lab 数据集上的 300 个专家点 TRE(标准偏差)

Table 2 Target registration errors (standard deviations) of 300 expert points in DIR-lab dataset with different dilated factors in module

unit: mm

Case	Dilated factors with 1,2,3,4	Dilated factors with 1,2,4,8	Dilated factors with 1,4,8,16	Dilated factors with 1,8,16,24
1	1.35 (0.62)	1.13 (0.63)	1.33 (0.64)	1.92 (0.86)
2	1.46 (0.55)	1.04 (0.50)	1.50 (0.58)	1.88 (0.87)
3	1.63 (0.79)	1.54 (0.77)	2.48 (1.65)	3.44 (1.91)
4	1.81 (0.97)	1.66 (0.96)	2.77 (1.79)	3.24 (1.78)
5	1.90 (1.28)	1.76 (1.24)	2.54 (1.94)	2.91 (2.13)
6	1.96 (1.10)	1.90 (1.19)	2.94 (2.46)	3.83 (2.11)
7	1.95 (1.29)	1.78 (1.06)	3.43 (2.92)	3.96 (2.57)
8	4.68 (3.59)	2.94 (3.15)	7.52 (7.16)	9.04 (7.08)
9	1.86 (0.81)	1.74 (0.88)	2.38 (1.21)	2.96 (2.56)
10	2.09 (1.96)	1.70 (1.70)	2.58 (2.74)	3.31 (3.01)
Mean	2.06 (1.29)	1.71 (1.20)	2.95 (2.31)	3.64 (2.48)

为评价本文提出算法的整体性能,将提出算法与其他经典算法在 DIR-Lab 数据集上进行了测试,经典算法包括 Eppenhof^[22]、R-Net^[11]、DLIR^[9]、FCN^[12]等算法。为保证实验的公平性,评价的各算法都采用无监督的深度学习方法,且 FCN 算法采用其一次迭代的结果。实验结果如表 3 所示,表示各配准算法在 DIR-lab 数据集的 300 个专家点的平均 TRE(标准偏差),提出

算法在 DIR-lab 数据集上的最大呼入和最大呼出相位的 300 个专家标志点平均 TRE 为 1.71 mm,标准偏差为 1.20 mm,说明该算法的配准精度较高、总体稳定性能较好。从表 3 中可见,提出算法相较于其他算法在 case8 的配准精度提升最大,case1~case3 由于形变较小配准难度较低,精度提升不大,其余 case 的配准精度相较其他算法均有 1 mm 左右的提升。这主要是由于

表 3 各配准算法在 DIR-lab 数据集上的 300 个专家点 TRE(标准偏差)

Table 3 Target registration errors (standard deviations) of 300 expert points in DIR-lab dataset of each registration algorithm unit: mm

Case	Initial	Eppenhof	R-Net	DLIR	FCN	D-FCN
1	3.89 (2.78)	1.65 (0.89)	1.50 (1.05)	1.27 (1.16)	1.15 (0.61)	1.13 (0.63)
2	4.34 (3.90)	2.26 (1.16)	1.74 (1.24)	1.20 (1.12)	1.13 (0.64)	1.04 (0.50)
3	6.94 (4.05)	3.15 (1.63)	2.36 (1.04)	1.48 (1.26)	1.60 (0.94)	1.54 (0.77)
4	9.83 (4.86)	4.24 (2.69)	3.13 (1.60)	2.09 (1.93)	2.10 (1.38)	1.66 (0.96)
5	7.48 (5.51)	3.52 (2.23)	2.92 (1.70)	1.95 (2.10)	2.26 (1.79)	1.76 (1.24)
6	10.89 (6.97)	3.19 (1.50)	2.95 (1.83)	5.16 (7.09)	2.93 (2.69)	1.90 (1.19)
7	11.03 (7.43)	4.25 (2.08)	3.52 (2.00)	3.05 (3.01)	3.45 (3.17)	1.78 (1.06)
8	14.99 (9.01)	9.03 (5.08)	5.52 (3.40)	6.48 (5.37)	8.60 (7.52)	2.94 (3.15)
9	7.92 (3.98)	3.85 (1.86)	3.22 (1.57)	2.10 (1.66)	2.53 (1.82)	1.74 (0.88)
10	7.30 (6.35)	5.07 (2.31)	3.07 (2.15)	2.09 (2.24)	2.56 (2.07)	1.70 (1.70)
Mean	8.46 (6.58)	4.02 (3.08)	2.94 (1.80)	2.64 (4.32)	2.83 (3.67)	1.71 (1.20)

本文采用多尺度并行全卷积网络进行配准,对形变较大的配准图像效果更佳,且小形变配准效果未受影响,表 3 中的配准结果正好验证了该结论。

图 11 给出了提出算法在 DIR-lab 数据集 10 个 case

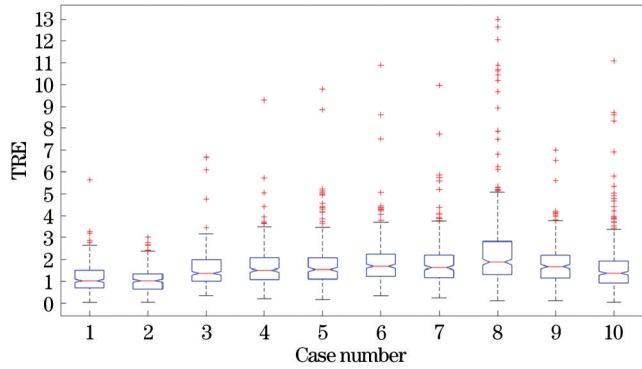


图 11 提出算法在 DIR-lab 数据集 10 个 case 的最大呼入到最大呼出相位上得到的 TRE 箱线图

Fig. 11 Box plot of target registration errors obtained by proposed algorithm from maximum inhalation to maximum exhalation in 10 cases in DIR-lab dataset

上的最大呼入到最大呼出相位上 300 个专家标志点的 TRE 箱线图,横轴表示 10 个 case,纵轴表示每个 case 对应的 300 个专家标志点的 TRE 值。由图可见,大部分 case 的专家标志点 TRE 的异常值较少,case8 由于形变特别剧烈,且因噪声原因部分位置配准效果一般,故有少数专家点的 TRE 较大。

3.2.2 算法在 Creatis 数据集 中的 评价

1) 强度 差异 性

图 12 为该实验在 Creatis 数据集 case1 中的配准效果图。图 12(a)~12(c)分别为固定图像、待配准的浮动图像和配准后的形变图像,图 12(d)、12(e)为配准前后的强度差异图。由图对比可见,配准后的图 12(e)肺下部轮廓差异相较于配准前的图 12(d)差异有大幅减小,且图 12(d)中表示肺实质各血管差异的白色部分在图 12(e)中也已减淡。由于 Creatis 数据集中肺部形变较小,其差异对比不如 DIR-lab 数据集明显,但仍可见形变图像基本配准到位,血管实现对齐,除轮廓边缘和少数组织有部分差异外,配准后整体差异较小。

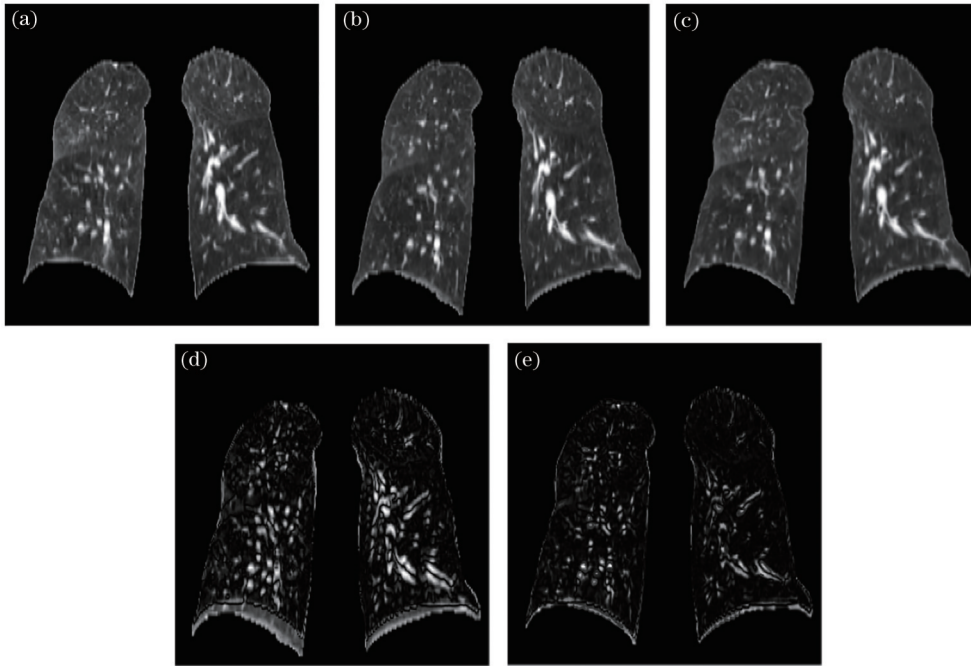


图 12 Creatis 数据集 case1 配准结果(冠状面)。(a)固定图像;(b)浮动图像;(c)形变图像;(d)配准前差异;(e)配准后差异
Fig. 12 Registration results of case1 in Creatis dataset (coronal plane). (a) Fixed image; (b) moving image; (c) deformed image; (d) difference before registration; (e) difference after registration

2) TRE

为验证模型的泛化性,本文在 Creatis 数据集上进行了与 DIR-lab 数据相同的实验,由于尚未发现有其他基于深度学习的配准算法采用该数据集,本文仅在 FCN 算法^[12]上复现了该实验,RPM^[23]为传统的点集匹配算法,实验结果见表 4。由表 4 可见,本文提出算法在 TRE 上相较于 RPM^[23]和 FCN^[12]算法均有较大提高,且在形变相对最剧烈的 case2 中有明显提升,这也

验证了该算法对大形变图像配准的效果。图 13 为 100 个左右专家点的 TRE 的箱线图,从图中可见各 case 配准点的 TRE 的异常值较少,体现了模型较好的鲁棒性。

3.2.3 算法在 DIR-lab 和 Creatis 数据集 中的 计算 复杂度

为评价本文提出算法的计算复杂度,表 5 引用了 RPM^[23]、R-Net^[11]、DLIR^[9]、FCN^[12]配准单幅 3D 图像

表 4 各配准算法在 Creatis 数据集上的 100 个专家点 TRE(标准偏差)

Table 4 Target registration errors (standard deviations) of 100 expert points in Creatis dataset of each registration algorithm unit: mm

Case	Initial	RPM	FCN	D-FCN
1	6.34 (2.95)	1.84 (1.56)	1.40 (0.57)	1.34 (0.47)
2	14.04 (7.20)	3.88 (2.91)	3.81 (3.06)	1.74 (1.10)
3	7.67 (5.05)	2.69 (2.66)	1.85 (1.38)	1.57 (0.87)
4	7.33 (4.89)	1.89 (1.85)	1.68 (1.48)	1.64 (0.98)
5	7.09 (5.10)	2.54 (2.24)	1.73 (1.22)	1.26 (0.84)
6	6.68 (3.68)	2.01 (1.49)	1.56 (1.09)	1.45 (0.81)
Mean	8.15 (5.60)	2.47 (2.27)	2.01 (1.46)	1.50 (0.85)

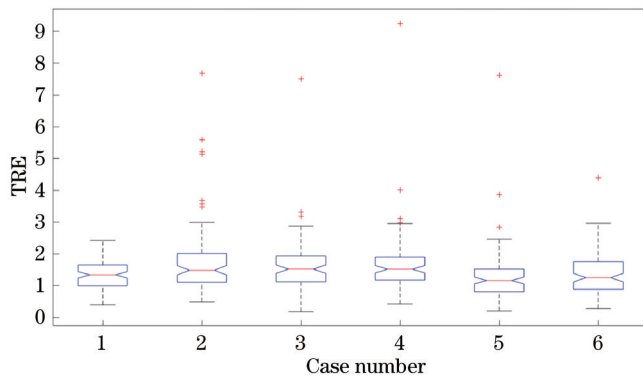


图 13 提出算法在 Creatis 数据集 6 个 case 的最大呼入到最大呼出相位上得到的 TRE 箱线图

Fig. 13 Box plot of target registration errors obtained by proposed algorithm from maximum inhalation to maximum exhalation in 6 cases of Creatis dataset

表 5 各算法在一组肺 CT 图像配准的耗时

Table 5 Running time of each algorithm to register a set of lung CT images

Algorithm	Time /s
RPM	900.00
R-Net	≈1.00
DLIR	0.30±0.05
FCN	0.63
D-FCN	0.10±0.05

对的运行时间,并与提出算法进行比较。从表 5 可得,提出算法相对于传统点集匹配算法的运行时间明显减少;和当前深度学习算法比较,其运行时间也更短,R-Net^[11]和 DLIR^[9]是 U 形卷积结构,其网络深度和卷积层数量均比提出算法大,配准单幅图像在 1 s 和 0.3 s 左右,FCN^[12]层数与提出算法相当,且其需要多次迭代才能配准完全,约 0.63 s 左右。而提出算法在一组图像的配准时间为 0.1 s 左右,由此可见提出算法模型的计算复杂度较低,基本可以满足当前临床的实时性要求,因此该算法在临床中具有一定的应用价值。

4 结 论

提出了一种基于 D-FCN 的肺 CT 图像配准算法,该算法在网络中加入多尺度并行下采样模块、金字塔空洞卷积模块和自适应通道注意力模块以提高模型性能;通过改进损失函数、数据扩增等方法,增强模型的平滑性和泛化性。实验结果表明该算法在数据集 DIR-lab 和 Creatis 中的实时性好、配准精度较高、泛化能力较强,因此具有一定的临床应用价值。由于提出算法用于训练的数据量较少,算法的配准精度较传统配准算法还有一定的差距,在今后的研究工作中,将从两个方面展开:1) 采集规模较大的数据集;2) 改进算法的网络结构和损失函数,测试更多不同数据集,以提高算法的泛化性及配准精度。

参 考 文 献

- [1] 苏博,陶芬,李可,等.同步辐射纳米 CT 图像配准方法研究[J].物理学报,2021,70(16):160704.
Su B, Tao F, Li K, et al. Image alignment for synchrotron radiation based X-ray nano-CT[J]. Acta Physica Sinica, 2021, 70(16): 160704.
- [2] Tareen S A K, Saleem Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK [C]//2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), March 3-4, 2018, Sukkur, Pakistan. New York: IEEE Press, 2018: 17732676.
- [3] 刘鹏南,徐冬冬,白春梦.基于尺度不变特征变换的异源图像配准方法[J].激光与光电子学进展,2021,58(24):2410003.
Liu P N, Xu D D, Bai C M. Scale-invariant feature transform-based heterogeneous image registration method [J]. Laser & Optoelectronics Progress, 2021, 58(24): 2410003.
- [4] 易见兵.基于点集匹配的 4D CT 肺运动估计算法研究[D].深圳:深圳大学,2017.
Yi J B. 4D CT lung motion estimation by robust point matching[D]. Shenzhen: Shenzhen University, 2017.
- [5] 丛明,吴童,刘冬,等.基于监督学习的前列腺 MR/TRUS 图像分割和配准方法[J].工程科学学报,2020,42(10):1362-1371.
Cong M, Wu T, Liu D, et al. Prostate MR/TRUS image segmentation and registration methods based on supervised learning[J]. Chinese Journal of Engineering, 2020, 42(10): 1362-1371.
- [6] Hu Y P, Modat M, Gibson E, et al. Label-driven weakly-supervised learning for multimodal deformable image registration[C]//2018 IEEE 15th International Symposium on Biomedical Imaging, April 4-7, 2018, Washington, DC, USA. New York: IEEE Press, 2018: 1070-1074.
- [7] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems 28: Annual Conference on Neural

- Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada. [S.l.: s.n.], 2015: 2017-2025.
- [8] Balakrishnan G, Zhao A, Sabuncu M R, et al. VoxelMorph: a learning framework for deformable medical image registration[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(8): 1788-1800.
- [9] de Vos B D, Berendsen F F, Viergever M A, et al. A deep learning framework for unsupervised affine and deformable image registration[J]. *Medical Image Analysis*, 2019, 52: 128-143.
- [10] 薛湛琦, 王远军. 基于低分辨率辅助特征和卷积神经网络的脑部图像配准方法[J]. *光学技术*, 2021, 47(1): 80-86.
Xue Z Q, Wang Y J. Brain image registration method based on low-resolution auxiliary features and convolutional neural network[J]. *Optical Technique*, 2021, 47(1): 80-86.
- [11] 卢佳毅. 基于卷积神经网络的肺部 CT 图像配准方法[D]. 武汉: 华中科技大学, 2019.
Lu J Y. Lung CT image registration based on convolutional neural network[D]. Wuhan: Huazhong University of Science and Technology, 2019.
- [12] Fang Q M, Gu X M, Yan J C, et al. A FCN-based unsupervised learning model for deformable chest CT image registration[C]//2019 IEEE Nuclear Science Symposium and Medical Imaging Conference, October 26-November 2, 2019, Manchester, UK. New York: IEEE Press, 2019: 19565241.
- [13] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [14] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 833-851.
- [15] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 636-644.
- [16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [17] Zhao M H, Zhong S S, Fu X Y, et al. Deep residual shrinkage networks for fault diagnosis[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(7): 4681-4690.
- [18] 曾安, 王烈基, 潘丹, 等. 基于 FCN 和互信息的医学图像配准技术研究[J]. *计算机工程与应用*, 2020, 56(18): 202-208.
Zeng A, Wang L J, Pan D, et al. Research on medical image registration technology based on FCN and mutual information algorithm[J]. *Computer Engineering and Applications*, 2020, 56(18): 202-208.
- [19] Castillo R, Castillo E, Guerra R, et al. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets[J]. *Physics in Medicine and Biology*, 2009, 54(7): 1849-1870.
- [20] Vandemeulebroucke J, Rit S, Kybic J, et al. Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs[J]. *Medical Physics*, 2011, 38(1): 166-178.
- [21] Fitzpatrick J M, West J B. The distribution of target registration error in rigid-body point-based registration[J]. *IEEE Transactions on Medical Imaging*, 2001, 20(9): 917-927.
- [22] Eppenhof K A J, Lafarge M W, Moeskops P, et al. Deformable image registration using convolutional neural networks[J]. *Proceedings of SPIE*, 2018, 10574: 105740S.
- [23] Yi J B, Yang X, Chen G L, et al. Lung motion estimation using dynamic point shifting: an innovative model based on a robust point matching algorithm[J]. *Medical Physics*, 2015, 42(10): 5616-5632.