

## 基于神经网络的双X射线影像2D-3D配准算法

沈延延<sup>1</sup>, 冯汉升<sup>2</sup>

1. 安徽大学电气工程与自动化学院, 安徽 合肥 230601; 2. 中国科学院等离子体物理研究所, 安徽 合肥 230031

**【摘要】**针对基于迭代优化的传统2D-3D医学图像配准算法运行速度慢,难以达到实时配准的要求,本研究提出一种实时2D-3D配准方法。通过将空间刚体变换参数分解到两个平面上,将2D-3D配准简化为两个步骤,包含2D-2D近似刚体配准与单参数2D-3D刚体配准。同时利用深度卷积神经网络拟合患者X射线影像残差与其对应姿态差异间的非线性映射关系,从X-DRR图像对的残差回归出空间刚体变换参数。经由头颅CT数据训练后的网络,在0.04 s内完成了高精度的双X射线配准。本研究提出的配准方法满足了放疗过程中进行实时2D-3D配准工作的要求。

**【关键词】**双X射线影像;2D-3D配准;卷积神经网络;几何分解

**【中图分类号】**R318

**【文献标志码】**A

**【文章编号】**1005-202X(2020)03-0293-06

### 2D-3D double X-ray image registration method based on neural network

SHEN Yanyan<sup>1</sup>, FENG Hansheng<sup>2</sup>

1. School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China; 2. Institute of Plasma Physics, Chinese Academy of Sciences, Hefei 230031, China

**Abstract:** Based on the situation that the traditional 2D-3D medical image registration algorithm based on iterative optimization can not realize real-time registration due to slow running speed, a real-time 2D-3D registration method is proposed. By decomposing spatial rigid transformation parameters into two planes, 2D-3D registration is simplified into two steps, including 2D-2D approximate rigid registration and single-parameter 2D-3D grid registration. Meanwhile, deep convolutional neural network is used to fit the nonlinear mapping between X-ray images residual and its corresponding postural difference, and the space rigid transformation parameters are regressed from the residual of the X-DRR image pair. The network trained by head CT data can complete the high-precision double X-ray registration within 0.04 s. The proposed registration method can satisfy the requirements of real-time 2D-3D registration during radiotherapy.

**Keywords:** double X-ray image; 2D-3D registration; convolutional neural network; geometric decomposition

### 前言

现代医学中,放疗是恶性肿瘤的重要治疗手段。但放射线在轰击病灶的同时也会损害肿瘤组织周围的正常细胞,因此需要在治疗过程中使用图像引导放疗(Image-Guided Radiotherapy, IGRT)技术保证放疗精确程度,从而提高放疗对肿瘤细胞的杀灭效果,并降低副作用。2D-3D医学图像配准技术可以进行患者摆位误差的计算,是极其重要的一环。

传统基于灰度的2D-3D配准算法通常采用迭代优化的方式,通过搜索CT的空间姿态并进行投影生成数字重建放射影像(Digital Reconstruction Radiograph, DRR),对比DRR图像与治疗中获取的X射线影像的相似度来判断当前CT姿态与患者姿态间的差异<sup>[1]</sup>。由于DRR图像的生成非常耗时,导致基于灰度的2D-3D配准算法较慢。针对这一问题, Mu<sup>[2]</sup>采取块投影法,利用已有的DRR图像帮助生成新的DRR图像,减少DRR图像生成的耗时。Tornai等<sup>[3]</sup>则将GPU并行计算应用到2D-3D配准中加快计算过程。而Pan等<sup>[4]</sup>采取阈值分割来排除不感兴趣区域,减小计算数据的大小。Lei等<sup>[5]</sup>则利用一对正交X射线影像上两个单面板的粗配准结合作为2D-3D配准初始化减小迭代次数。Ghafurian等<sup>[6]</sup>提出一种新的图像梯度概率密度直方图作为图像的特征,

**【收稿日期】**2019-10-21

**【基金项目】**中国科学院合肥物质科学研究院“十三五”规划重点支持项目(kp-2017-24)

**【作者简介】**沈延延,硕士研究生,从事医学图像处理研究,E-mail: yanyan\_sh@foxmail.com

**【通信作者】**冯汉升,主要从事质子治疗系统研究,E-mail: hsfeng@ipp.ac.cn

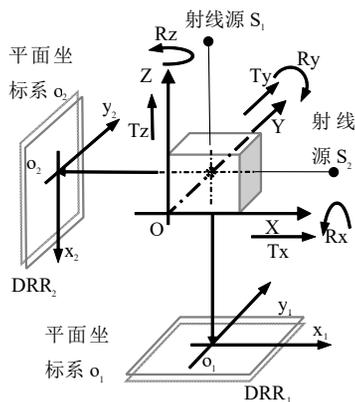
避免搜索优化6个自由度的变换向量。晋青鹏等<sup>[7]</sup>分析解剖特征与DRR图像变化的关系进行非刚性配准。同时随着深度学习的发展,Miao等<sup>[8-10]</sup>最先提出并改进使用卷积神经网络(Convolutional Neural Network, CNN)从图像残差中计算CT空间刚体变换参数。Pinheiro等<sup>[11]</sup>则是将CNN回归与传统方法结合到一起,迭代地对新生成的DRR图像使用CNN回归。Pei等<sup>[12]</sup>同样使用CNN并采取多尺度特征融合的方法综合提取局部和总体特征进行配准。

本研究提出一种基于CNN的双X射线影像配准方法,利用两个具有一定夹角的X射线影像在平面上的配准,将空间刚体变换参数的一部分转化为平面上的近似刚体变换,同时利用3个神经网络分组进行回归预测的方法简化2D-3D配准,从而达到快速进行2D-3D配准的目的。

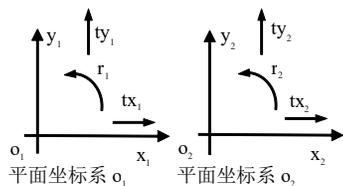
## 1 方法

### 1.1 几何分解

模拟拍摄双X射线影像时的机械结构,建立如下空间坐标系O及两个X射线影像所在的平面坐标系 $O_1$ 、 $O_2$ 。其中射线源 $S_1O_1$ 与射线 $S_2O_2$ 间绕Y轴的夹角为 $\theta_{re}$ ,且该两个射线都平行于平面X-O-Z。几何示意如图1所示。



a: 空间坐标系与一对正交平面坐标系的几何关系



b: 2个正交平面坐标系及其平面刚体变换参数

图1 空间坐标系与平面坐标系

Fig.1 Spatial and plane coordinate systems

根据该几何关系可将空间刚体变换的6个参数中的5个(包括沿X、Y、Z轴的平移 $T_x$ 、 $T_y$ 、 $T_z$ ,以及沿

X、Y轴的旋转 $R_x$ 、 $R_y$ )分解到两个X射线影像所在平面坐标系中,反之也可由两个平面坐标系计算空间坐标系的 $T_x$ 、 $T_y$ 、 $T_z$ 、 $R_x$ 、 $R_y$ 。

首先,从空间参数分解到平面参数:

$$\begin{aligned} tx_1 &= scale_1 \times Tx, ty_1 = scale_1 \times Ty, r_1 = Rz, \\ ty_2 &= -scale_2 \times Ty, r_2 = Rx \times \sin(\theta_{re}) + Rz \times \cos(\theta_{re}), \\ tx_2 &= scale_2 \times (Tx \times \cos(\theta_{re}) - Tz \times \sin(\theta_{re})) \end{aligned} \quad (1)$$

然后,从平面参数反演出除 $R_z$ 以外的5个空间参数 $T_x$ 、 $T_y$ 、 $T_z$ 、 $R_x$ 、 $R_y$ :

$$\begin{aligned} Tx &= 0.5 \times \left( \frac{tx_1}{scale_1} + \frac{tx_2 \times \cos(\theta_{re})}{scale_2} \right), \\ Ty &= 0.5 \times \left( \frac{ty_1}{scale_1} + \frac{ty_2}{scale_2} \right), Tz = -\frac{tx_2 \times \sin(\theta_{re})}{scale_2}, \\ Rx &= r_2 \times \sin(\theta_{re}), Ry = r_1 \end{aligned} \quad (2)$$

射线源到DRR平面的距离即源像距(SID),射线源到等中心点的距离即源轴距(SAD),上述公式中放缩系数 $scale_1$ 与 $scale_2$ 的计算如下:

$$\begin{aligned} scale_1 &= \frac{SID}{SAD + Ty}, \\ scale_2 &= \frac{SID}{SAD + (Tx \times \sin(\theta_{re}) + Tz \times \cos(\theta_{re}))} \end{aligned} \quad (3)$$

在该坐标系下,模拟拍摄一对夹角为 $\theta_{re}$ 的X射线影像,生成两幅DRR图像,由坐标系分解可以得到在两个平面上的X-DRR图像对配准结果,计算出空间参数的 $T_x$ 、 $T_y$ 、 $T_z$ 、 $R_x$ 、 $R_y$ 。

### 1.2 分步配准

对应射线下的X-DRR图像对之间的差异主要由CT姿态的空间变换造成,因此从X-DRR图像对的残差到空间的姿态变换是存在着一个高度非线性的映射。而神经网络在理论上是可以无限逼近任意非线性函数的,因此本研究采用神经网络对空间刚体变换参数进行回归预测。为简化配准问题的复杂程度,将配准分解成两个步骤进行。由前文分析可知刚体变换参数中的5个组成元素反映在平面变换中,根据参数类型的不同将其分为两组分别进行回归,包含平面内参数 $P_{in} = \{tx_1, tx_2, ty_1, ty_2, r_1, r_2\}$ ,平面外参数 $P_{out} = \{R_2\}$ 。

由于参数 $T_x$ 、 $T_y$ 、 $T_z$ 、 $R_x$ 、 $R_y$ 并非独立作用在两个平面上,其中平移参数在一个平面上表现为平移,在另一个平面上为放缩。例如 $T_x$ 在平面坐标系 $o_1$ 上表现为平移 $tx_1$ ,在平面坐标系 $o_2$ 上表现为放缩系数 $scale_2$ 。旋转参数在一个DRR图像上表现为旋转,同时使得射线源到另一个DRR图像像素点的光路所经过的CT体素点发生改变,进而造成对应DRR像素值改变以及图像轮廓的形变(如 $R_x$ 与 $r_2$ 以及 $DRR_1$ )。

因此 X-DRR 图像对之间是一种非线性变换。当初始位置到目标位置的差值较小, 即算法收敛范围较小时可近似成刚体变换。 $P_{in}$  是两个平面上 X-DRR 图像对之间的近似刚体变换参数。

按难易顺序先后回归  $P_{in}$  和  $P_{out}$ , 并且根据  $P_{in}$  回归结果对 DRR 图像进行对应的刚体变换, 进而尽可能地消除已知参数对图像的影响, 使得在后续网络中仅包含未知参数造成的图像差异。其算法流程如图 2 所示。

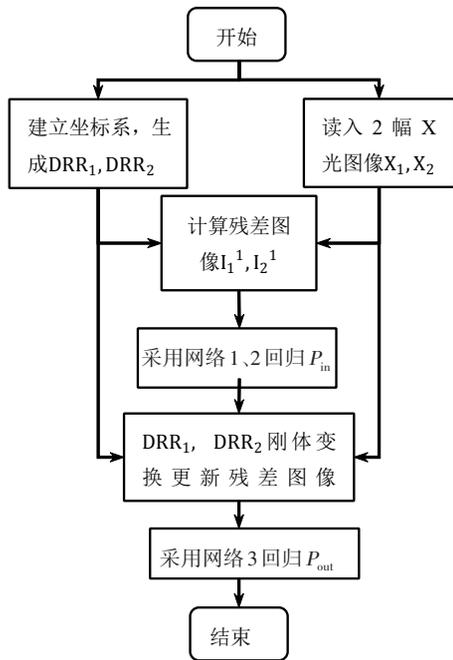


图 2 分步回归

Fig.2 Step-by-step regression

### 1.3 网络结构

在第 1.2 节中使用了 3 个独立训练的 CNN 回归网络。其中网络 1、2 的目的分别是回归两个平面上的参数, 因此它们的网络框架完全一致。输入为 X-DRR 图像对的残差图像, 为了减小输入图像大小, 将  $512 \times 512$  的残差图像分割为互有重叠部分的 4 个  $258 \times 258$  的区域, 重叠在一起后输入一个仅少量卷积层组成的扩展网络 CNN-Pre。将扩增的特征图输入由 Dense Block<sup>[13]</sup> 组成的特征提取网络, 并且每一个 Dense Block 的输入都是由之前所有的 Dense Block 的输出经过不同的池化层将特征尺寸统一后连接再经过  $1 \times 1$  卷积层的特征融合操作得到。最后将提取到的所有特征进行全局平均池化, 再把池化结果输入全连接层, 输出患者摆位误差。网络结构如图 3 所示, 网络 1 的 CNN-Pre Block 内部结构见表 1。

网络结构经过前两个网络的回归, 对  $DRR_1$ 、 $DRR_2$  图像独立进行了刚体变换来消除已知参数的影响。由于前一步的回归精度问题, 必然存在参数误差, 而且空间刚体变换参数并非独立作用于两个平面上, 故而在回归  $P_{out}$  前对两个平面上图像进行独立的刚体变换不足以完全消除已知参数的影响, 即输入  $P_{out}$  回归网络的两个平面上的残差图像  $I_1$ 、 $I_2$  中还包含  $P_{in}$  回归误差以及平面刚体变换带来的干扰。因此需要一个拟合能力更强的回归网络, 在网络 1、2 的基础上进行改进出网络 3。

通过对网络 1、2 在其 CNN-Pre 部分进行结构修改来提高该网络的特征提取能力, 修改后的 CNN-Pre

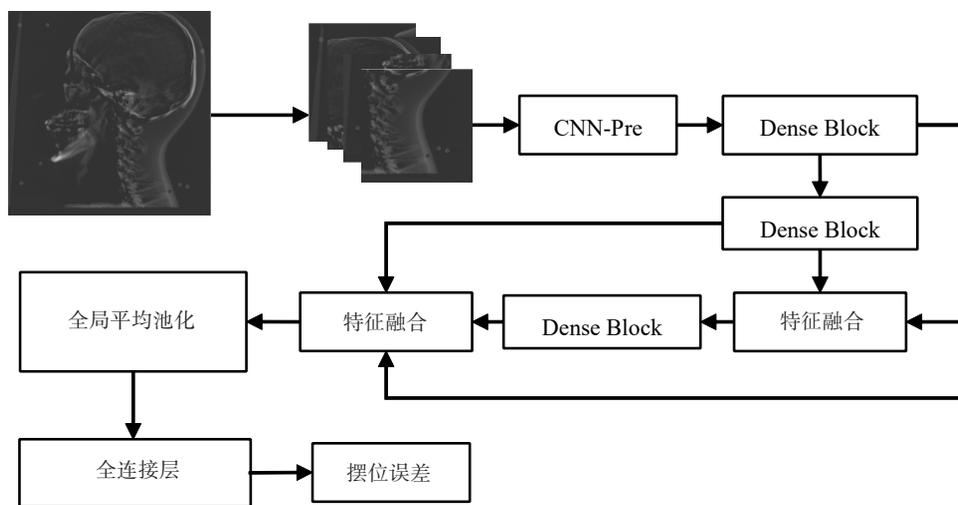


图 3 网络主要框架

Fig.3 Main architecture of network

如图 4 所示。其中, Conv2D-1 使用深度可分离卷积<sup>[14]</sup>; Conv2D-2 使用  $1 \times 1$  的卷积核减少特征数量;

Conv2D-4 使用  $5 \times 5$  空洞卷积增大感受野<sup>[15]</sup>; Conv2D-5 使用  $1 \times 1$  卷积核进行特征融合; 其它卷积核为  $3 \times 3$ 。

表1 网络1的CNN-Pre Block内部结构

Tab.1 Internal structure of CNN-Pre Block of network 1

层	输入	输出
Batch Normalization	4@258×258	4@258×258
2D Convolution	4@258×258	64@256×256
ReLU	64@256×256	64@256×256
Average Pooling	64@256×256	64@128×128
Batch Normalization	64@128×128	64@128×128
2D Convolution	64@128×128	128@128×128
ReLU	128@128×128	128@128×128
Average Pooling	128@64×64	128@64×64

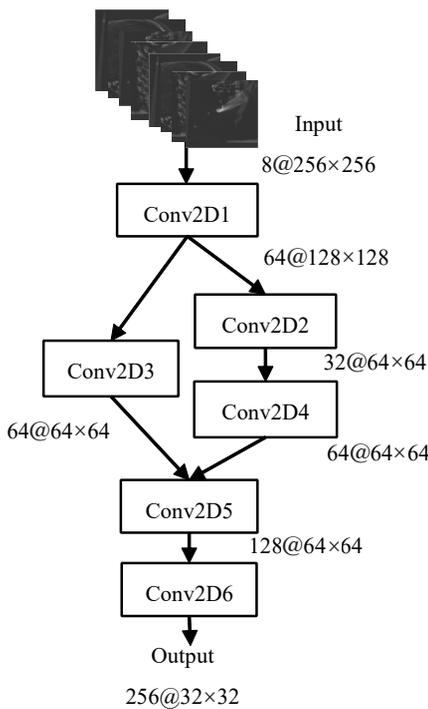


图4 网络3中CNN-Pre结构图

Fig.4 Structure of CNN-Pre of network 3

所有卷积前都添加 Batch Normalization<sup>[16]</sup>层, 激活函数为 ReLu<sup>[17]</sup>, 同时 Conv2D-1、Conv2D-2、Conv2D-3、Conv2D-6使用2×2的平均池化。

对于网络3的输入图像需要进行预处理。首先去除两个残差图像四周边缘的32个像素宽度丢弃平面变换时在图像边缘引入的误差, 然后再按照相互重叠32个像素宽度的方式将剩余图像分割为4个256×256的区块, 最后将两个平面上残差图像分割结果重叠为8@256×256, 作为网络的输入图像。

## 2 实验

### 2.1 数据获取

在本实验中, 采用DRR图像模拟X射线影像以获取准确参数用于训练。针对CT尺寸为512×512×283, 物理尺寸为(400×400×283) mm<sup>3</sup>的头颅CT生成约26万组训练数据<sup>[18]</sup>。每一组数据包含在初始参数下生成的2幅DRR图像, 在目标参数下生成的2幅模拟X射线影像, 所有图像大小都是512×512。设置两次拍摄时, 绕Z轴旋转的夹角为90°。取95%的图像数据作为训练集, 剩余部分作为测试集。

### 2.2 实验结果

在上述训练数据下, 在显卡为Tesla P100、CPU为Intel E5-2680的服务器上使用深度学习框架Pytorch 0.4编写训练。训练轮数Epoch为32, 每一轮中不重复地随机抽取的mini Batch大小为64, 直到所有组数据都被使用过。训练中采取神经网络常用的Adam优化方法<sup>[19]</sup>。损失函数采用L2范数即刚体变换参数的均方误差。网络1训练结果如图5所示, 其中学习率LR按式(4)进行调整。

$$LR = \begin{cases} 1 \times 10^{-4}, & 0 < \text{Epoch} \leq 8 \\ 5 \times 10^{-5}, & 8 < \text{Epoch} \leq 16 \\ 1 \times 10^{-5}, & 16 < \text{Epoch} \leq 24 \\ 1 \times 10^{-6}, & 24 < \text{Epoch} \leq 32 \end{cases} \quad (4)$$

为对比该配准方法的配准效果, 所采用的对比算法为CUDA加速迭代式的双X射线影像配准(CUDA)。在第2.1节所述头颅数据上1万多幅测试图像中随机挑选50组, 配准结果如表2所示。

其中采用的配准结果衡量标准为mTRE<sup>[20]</sup>与配准参数平均误差ME。

mTRE计算方法如下:

$$mTRE = \frac{1}{N} \sum_{i=1}^N \|T_{reg} Q_i - T_{gd} Q_i\| \quad (5)$$

取CT中包含N个点的点集Q, 计算其在配准结果的变换 $T_{reg}$ 与目标真实变换 $T_{gd}$ 下的对应点距离误差均值, 同时取CT数据的物理尺寸长度1%即2.83 mm作为配准成功的标准。ME作为配准精度衡量标准, ME(P)计算方式如下:

$$ME(P) = \frac{1}{N} \sum_{i=1}^N \|P'_i - P_i\| \quad (6)$$

P与P'是包含N个平移配准参数的向量, 分别为配准参数与真实参数, 其中平移 $P_T = [Tx, Ty, Tz]$ , 旋转 $P_R = [Rx, Ry, Rz]$ , 分别计算平移与旋转的ME。

### 2.3 分析

实验结果表明, 本文方法在ME度量下, 其平移参数回归误差为0.047 mm, 相对迭代算法下降了81.20%; 旋转参数回归误差为0.057, 相对CUDA算法下降了68.51%, 有显著提升。在mTRE标准下, 本

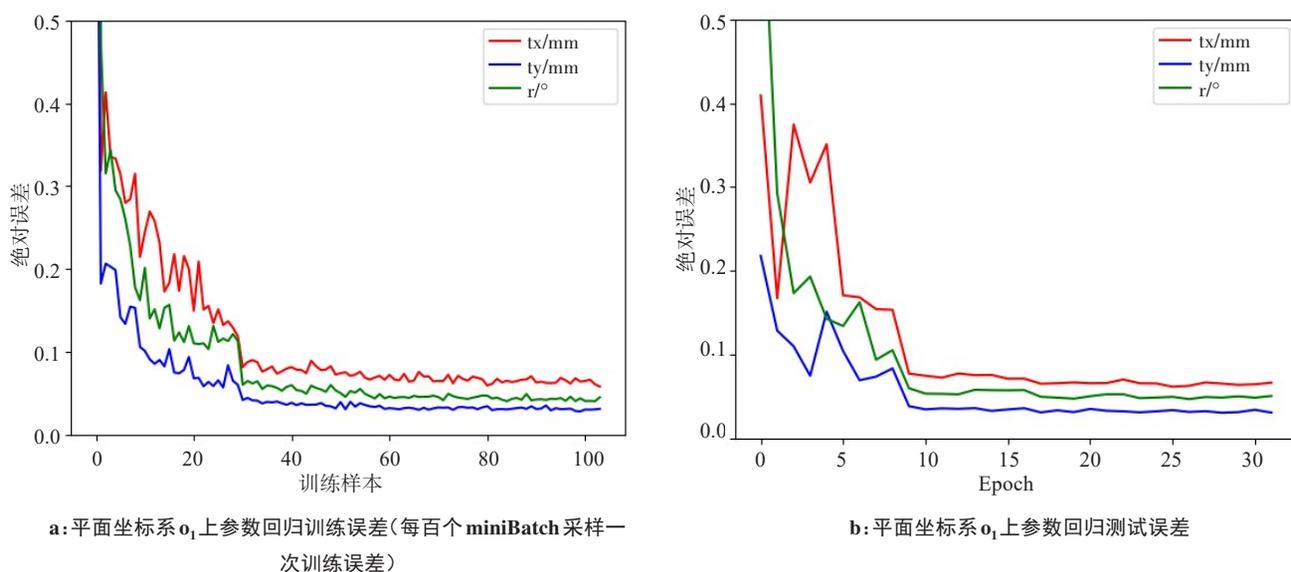


图5 网络1训练结果

Fig.5 Training result of network 1

表2 实验结果

Tab.2 Experience results

方法	mTRE/mm	成功率/%	耗时/s	ME( $P_T$ )/mm	ME( $P_R$ )/°
CUDA	0.951 1±1.715 5	98	13.386	0.250	0.181
本文	0.307 2±0.144 9	100	0.040	0.047	0.057

文方法配准误差为 0.462 mm, 仅为 CUDA 算法的 32.32%。因此本文方法的配准精度在多个指标下相对传统算法都有较大提升。

X-DRR 图像间的相似性度量是关于空间变换参数的高度非线性函数。因此在传统迭代优化过程中非常容易陷入局部最优值, 造成配准误差较大。而由于采用了 CNN 回归的方式, 本文方法在配准过程中没有传统迭代算法的优化过程, 依靠神经网络强大的拟合能力, 从训练数据中寻找 X-DRR 残差图像到空间参数的映射关系; 又由于空间参数分步回归, 进一步简化了这种映射, 使得神经网络能回归出更加精确的最优解。

在 50 组随机数据中平均耗时仅为 0.04 s, 即达到 25FPS 的处理速度, 远快于 CUDA 方法, 达到实时配准的标准。对于训练好的神经网络, 在配准中仅需要对于残差图像做一次单向的运算, 无需迭代生成 DRR 图像和计算图像相似度, 从而规避了传统迭代优化算法中最耗时的部分, 具有非常快的配准能力。

### 3 结论

在 IGRT 系统的工作中, 2D-3D 配准速度的快慢直接影响患者的治疗体验。本研究提出的利用双 X

射线进行空间坐标分解, 简化空间坐标系变化与 X-DRR 残差图像映射关系; 利用含有 Dense Block 结构的 CNN 的强大非线性拟合能力, 分步回归出空间刚体变换参数。相对基于灰度的迭代优化配准算法, 在配准精度和速度都有明显提升, 尤其是其配准速度已经达到实时配准的标准, 在临床治疗中具有一定实用潜力。

### 【参考文献】

- [1] MARKELJ P, TOMAŽEVIČ D, LIKAR B, et al. A review of 3D/2D registration methods for image-guided interventions[J]. Med Image Anal, 2012, 16(3): 642-661.
- [2] MU Z. A fast DRR generation scheme for 3D-2D image registration based on the block projection method[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016: 169-177.
- [3] TORNAI G J, CSEREY G, PAPPAS I. Fast DRR generation for 2D to 3D registration on GPUs[J]. Med Phys, 2012, 39(8): 4795-4799.
- [4] PAN H, ZHOU C, ZHU Q, et al. A fast registration from 3D CT images to 2D X-ray images [C]. 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA). IEEE, 2018: 351-355.
- [5] LEI Y, YAN Z. An improved 2D-3D medical image registration algorithm based on modified mutual information and expanded Powell method[C]. 2013 IEEE International Conference on Medical Imaging Physics and Engineering. 2013: 24-29.
- [6] GHAFURIAN S, HACHALILOGLU I, METAXAS D N, et al. A

- computationally efficient 3D/2D registration method based on image gradient direction probability density function[J]. *Neurocomputing*, 2017, 229: 100-108.
- [7] 晋青鹏, 夏威, 高欣. 基于形变量和相似性测度相关联的肺部2D/3D图像弹性配准算法[J]. *航天医学与医学工程*, 2018, 31(6): 630-636. JIN Q P, XIA W, GAO X. 2D/3D non-rigid registration algorithm of lung images based on deformation correlated with similarity measure [J]. *Space Medicine & Medical Engineering*, 2018, 31(6): 630-636.
- [8] MIAO S, WANG Z J, LIAO R. A CNN regression approach for real-time 2D/3D registration[J]. *IEEE Trans Med Imaging*, 2016, 35(5): 1352-1363.
- [9] MIAO S, WANG Z J, ZHENG Y, et al. Real-time 2D/3D registration via CNN regression[C]. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE, 2016: 1430-1434.
- [10] MIAO S, PIAT S, FISCHER P, et al. Dilated FCN for multi-agent 2D/3D medical image registration [C]. Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [11] PINHEIRO H N, REN T I, SCHEIB S, et al. 2D/3D megavoltage image registration using convolutional neural networks[J]. *Comput Sci*, arXiv: 1811.11816.
- [12] PEI Y, ZHANG Y, QIN H, et al. Non-rigid craniofacial 2D-3D registration using CNN-based regression [M]. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, Cham, 2017: 117-125.
- [13] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 4700-4708.
- [14] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 1251-1258.
- [15] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. *Comput Sci*, arXiv: 1511.07122.
- [16] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. *Comput Sci*, arXiv: 1502.03167.
- [17] HARA K, SAITO D, SHOUNO H. Analysis of function of rectified linear unit used in deep learning [C]. 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, 2015: 1-8.
- [18] WU J. ITK-based implementation of two-projection 2D/3D registration method with an application in patient setup for external beam radiotherapy[J]. *Insight J*, 2010: 784.
- [19] KINGMA D P, BA J. Adam: a method for stochastic optimization[J]. *Comput Sci*, 2014 arXiv: 1412.6980.
- [20] VAN DE KRAATS E B, PENNEY G P, TOMAZEVIC D, et al. Standardized evaluation methodology for 2D-3D registration[J]. *IEEE Trans Med Imaging*, 2005, 24(9): 1177-1189.

(编辑:谭斯允)