



研究与开发

2D/3D 多模态医学图像配准算法研究

徐密¹, 诸葛斌¹, 袁非牛², 尹正虎¹, 董黎刚¹, 蒋献¹, 孙应绮³, 宋嘉琦¹,
史晓彤¹, 苏雷¹, 周屹博¹, 林诗凡¹

(1. 浙江工商大学信息与电子工程学院, 浙江 杭州 310020; 2. 上海师范大学信息与机电工程学院, 上海 201418; 3. 复旦大学附属儿科医院, 上海 201102)

摘要: 2D/3D 多模态配准在医学影像导航手术中起着重要作用, 主要用于提供术前三维图像和术中二维图像的实时信息, 帮助医生精准定位病灶, 规划手术路径, 从而提高手术的安全性和效率。提出了一种 2D/3D 多模态医学图像配准算法, 首先利用 Swin Transformer 优秀的特征提取能力, 构建了初始姿态估计模型, 实现姿态参数的快速预测; 接着, 为了提升整个配准方法的鲁棒性, 引入基于 Grangeat 关系的粗配准方法; 最后设计了基于梯度下降的精配准模块, 以提升整个配准过程的精确性, 且在该模块将 Sobel 微分算子与归一化互相关结合, 提升了参数优化过程中的灵敏度。实验结果表明, 所提配准方法在正位和侧位配准中的误差满足配准要求, 配准成功率有显著提升。

关键词: 2D/3D 图像配准; 多模态; 初始姿态估计; 粗配准; 精配准; 相似度测量

中图分类号: P237

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024070

Research on 2D/3D multimodal medical image registration algorithm

XU Mi¹, ZHUGE Bin¹, YUAN Feiniu², YIN Zhenghu¹, DONG Ligang¹, JIANG Xian¹,
SUN Yingqi³, SONG Jiaqi¹, SHI Xiaotong¹, SU Lei¹, ZHOU Yibo¹, LIN Shifan¹

1. School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310020, China

2. School of Information and Mechatronics Engineering, Shanghai Normal University, Shanghai 201418, China

3. Pediatrics Hospital of Fudan University, Shanghai 201102, China

收稿日期: 2023-12-12; **修回日期:** 2024-01-20

通信作者: 孙应绮, sunyingqiqq@163.com

基金项目: 国家自然科学基金资助项目 (No.62272308); 浙江省新型网络标准与应用技术重点实验室项目 (No.2013E10012); 浙江省“尖兵”“领雁”研发攻关计划项目 (No.2023C03202); 浙江省研究生一般项目 (No.1120XJ0622033); 国家级大学生创新创业训练计划项目 (No.202210353022, No.202310353040, No.202310353069); 浙江省大学生科技创新活动计划—新苗人才计划项目 (No.1120KZN0223068G)

Foundation Items: The National Natural Science Foundation of China (No.62272308), Zhejiang Key Laboratory of New Network Standards and Application Technology (No.2013E10012), Zhejiang Province “Spearhead” and “Leading Wild Goose” Research and Development Project (No.2023C03202), General Project for Postgraduates in Zhejiang Province (No.1120XJ0622033), National Innovation and Entrepreneurship Training Program for College Students (No.202210353022, No.202310353040, No.202310353069), Science and Technology Innovation Activity Program for College Students in Zhejiang Province Fresh Seedling Talent Program (No.1120KZN0223068G)



Abstract: 2D/3D multimodal alignment plays an important role in medical image navigation surgery, which is mainly used to provide real-time information of preoperative 3D images and intraoperative 2D images to help doctors accurately locate the lesions and plan the surgical paths, so as to improve the safety and efficiency of surgery. A 2D/3D multimodal medical image alignment algorithm was proposed, which firstly utilized the excellent feature extraction capability of Swin Transformer to construct an initial pose estimation model to realize the fast prediction of pose parameters. Then, in order to improve the robustness of the whole alignment method, a coarse alignment method based on the Grangeat relation was introduced. Finally, a fine alignment module based on gradient descent was designed. A fine alignment module based on gradient descent was designed to improve the accuracy of the whole alignment process, and the Sobel differential operator was combined with normalized correlation in this module to improve the sensitivity of the parameter optimization process. The experimental results show that the proposed alignment method meets the alignment requirements in the orthogonal and lateral alignment errors, and the alignment success rate is significantly improved.

Key words: 2D/3D image registration, multimodality, initial pose estimation, coarse registration, fine registration, similarity measurement

0 引言

多模态图像配准^[1]的目的是将不同成像方式产生的医学图像如计算机断层扫描 (computer tomography, CT) 影像、核磁共振图像 (magnetic resonance imaging, MRI)、X 射线 (X-ray) 图像等之间进行空间对齐, 以便融合或对比信息。目前多模态配准在医学图像处理与分析中有多种具有实际应用价值的, 例如临床诊断、疾病检测、影像引导骨科手术 (image guided orthopedic surgery, IGOS)^[2]等。本文的主要研究内容是 2D/3D 医学图像配准, 即将术前三维的 CT 影像数据和术中二维的 X 射线影像数据置于同一坐标系中, 进行图像配准, 以方便在术中提供图像引导, 从而提高手术的成功率^[3]。

传统 2D/3D 配准方法可分为基于灰度信息和基于特征信息的配准^[4]。基于灰度信息的方法通过优化图像灰度相似度来确定变换参数, 但对存在形变或灰度变化明显的图像效果较差^[5]。基于特征信息的方法通过提取图像中点、线等特征完成配准, 过程简单快速, 但容易丢失有价值的灰度、梯度等信息^[6]。在配准步骤中, 初始位姿估计是获取较好初始化参数的关键, 提高参数变换

范围进而提升配准效率^[7]。目前, 有手动选择术前和术中影像的解剖标志来获取三维位姿的方法, 但仅适用于部分骨部位, 并需医师具备丰富的经验^[8]。还有采用图形用户界面来实现图像配准参数初始化的方法, 通过可视反馈的软件工具移动相关医学图像, 为后续配准算法提供局部和全局参数估计结果, 更高效便捷^[9]。还有用来衡量配准效果好坏的数据, 即相似性测度, 常用的几种函数包括基于概率的互信息 (mutual information, MI)^[10]、基于过去信息的条件方差和 (sum of conditional variance, SCV)^[11]、在聚类中度量 2 个聚类结果的相近程度的归一化互信息 (normalized mutual information, NMI)^[12]、针对图像的局部区域进行计算的归一化互相关 (normalized cross-correlation, NCC)^[13]等。深度神经网络近年来逐渐应用于医学图像配准领域^[14]。例如, 卷积神经网络 (convolutional neural network, CNN) 广泛用于特征点检测^[15]、特征描述符^[16]及图像配准^[17]中。Eppenhof 等^[18]提出一种基于全监督思想的可变形配准方法, 通过 Bspline 生成图像的形变场, 利用 L1 损失缩小预测值和真实形变场的误差, 并反向传播优化网络。Miao 等^[19]提出一种具有自动注意力机制的多智能体方法, 用于实现

脊柱的三维CT和X射线图像配准。具体来说，通过观察局部区域，使用全卷积网络训练单个智能体，并在马尔可夫决策过程中进行配准。实验结果表明，该方法满足医学图像配准的精度需求，在目标配准误差上优于其他方法。

针对传统迭代配准方法中存在的配准速度慢和配准精度低问题，通过结合深度学习和其他前沿的配准技术，提出了一种2D/3D多模态医学图像配准算法（research on 2D/3D multimodal medical image registration algorithm, MMIR），该方法主要的贡献如下。

(1) 提出一种基于Swin Transformer的目标姿态预测（Swin Transformer-based target pose prediction, ST-TPP）模型，实现对初始搜索空间位姿参数的快速估计。通过将CT与X射线图像联合送入网络，增强了模型的泛化处理能力，网络主干利用3D卷积和Transformer结构，增强了对图像特征的提取能力。

(2) 研究基于梯度下降的精配准（accurate registration based on gradient descent, AR-GD）方法，针对相似性测度NCC精度较低、收敛域小的问题，引入Sobel算子，增加了算法收敛域，提高了配准精度。又基于可微分数字重建放射影像（digitally reconstructed radiograph, DRR）成像模块，使用训练过程进行迭代配准，在初始空间变换参数误差较小时，能够快速收敛到精确的配准结果。

1 配准方法

本文提出了一种2D/3D多模态医学图像配准算法，以满足临床需求，包括提供快速起始搜索参数的ST-TPP初始姿态估计模型、解决传统算法鲁棒性差的Grangeat关系的粗配准算法，以及解决传统方法配准精度低的梯度下降的精配准方法。

1.1 初始位姿估计

传统配准方法通过参数迭代搜索实现，当初始位姿与目标位姿有较大偏差时，会增加计算量、降低效率^[20]，同时也常会使配准结果限于局部极值，不能达到相似性测度的全局收敛^[21]。因此，大部分2D/3D配准研究中限制了初始偏差范围^[22]。传统的初始姿态计算复杂、耗时且鲁棒性低。本节改进了基于Swin-Transformer模型的方法，通过神经网络估计X射线图像的初始位姿，为后续配准提供更好的搜索参数。以下将详细介绍该模型。

1.1.1 空间投影转换基本原理

传统的DRR生成算法是不可微的，即不能通过反向传播DRR和目标图像的相似度损失来更新DRR的投影参数值，Gao等^[23]提出了一种空间投影转换器（projective spatial transformer, ProST）使得DRR过程可微，空间投影转换示意图如图1所示，基本原理如下。

给定一个CT图像 $V \in \mathbb{R}^{D \times W \times H}$ ，其体素大小为 $v_D \times v_W \times v_H$ ，定义一个原点在 V 中心的参考系 F^r 。全局使用深度(Dv_D)、宽度(Wv_W)和高度(Hv_H)的归一化坐标，这样 V 中的点就被包含在单位立方体 $(d, w, h) \in [-1, 1]^3$ 中。给定一个相机固有矩阵 $\mathcal{K} \in \mathbb{R}^{3 \times 3}$ ，然后将对应原点设定在坐标系 F^r 中的点 $(0, 0, \text{src})$ 处。如图1(a)所示是一个带控制点的空间网格 G 。因为，在没有CT体素的区域内的控制点不影响线积分，所以将网格点切割成一个覆盖精确CT空间的锥形结构。由于每条射线在 V 内有 K 个均匀间隔的控制点，因此可定义控制点矩阵为 $\mathbf{G} \in \mathbb{R}^{4 \times (M \cdot N \cdot K)}$ ，其中每列都是齐次坐标中的一个控制点。这些光线描述了一个锥形束的几何形状，它与检测平面相交，以 $(0, 0, \text{det})$ 为中心，垂直于 z 轴，由矩阵 $\mathcal{K}$ 确定。探测器平面的右上角位于点 $\left(\frac{p_M M}{v_W W}, \frac{p_N N}{v_H H}, \text{det}\right)$ 。

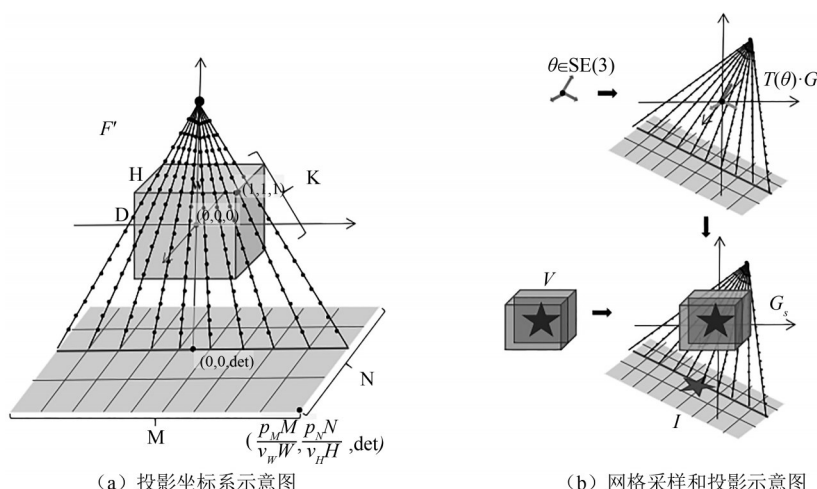


图1 空间投影转换示意图

ProST通过学习控制点 G 的变换来投影CT图像, 给定一个 $\theta \in \text{SE}(3)$, 可通过仿射变换矩阵得到了控制点的变换集 $T(\theta)$:

$$G_T = T(\theta) \cdot G \quad (1)$$

以及源点 $T(\theta) \cdot (0, 0, \text{src}, 1)$ 和检测平面的中心 $T(\theta) \cdot (0, 0, \text{det}, 1)$ 。由于这些控制点分布在体素之间, 所以需要 V 进行插值, 形成 G_s :

$$G_s = \text{interp}(V, G_T) \quad (2)$$

其中, $G_s \in \mathbb{R}^{M \times N \times K}$ 。最后, 沿每条射线进行积分得到一个二维图像 $I \in \mathbb{R}^{M \times N}$:

$$I^{(m,n)} = \sum_{k=1}^K G_s^{(m,n,k)} \quad (3)$$

1.1.2 ST-TPP模型搭建

Swin Transformer (ST) 是一种基于自注意力机制的神经网络模型, 是新型Transformer模型^[24]。它采用跨窗口的分组卷积代替全局自注意力, 降低了计算成本, 提高了效率和可扩展性。本文将其作为基线模型进行改造来预测X射线图像的初始姿态, 采用端到端构建思路, 模型输入为CT Volume和空间姿态参数rtvec, 输出为预测的姿态参数rtvec_p, 无须标注数据集, 降低了训练成本, 增强了泛化能力。以下介绍基于Swin Transformer的目标姿态预测 (Swin Transformer-based target pose prediction, ST-TPP) 模型的搭建过程。

ST-TPP网络结构如图2所示, 该模型主要由四层模块构成, 首先是DRR生成层, 该层根据输入的CT Volume和姿态向量rtvec, 经过空间投影转换后, 生成对应的DRR图像。接着是3D CNN层, 这一层主要完成数据的初步特征提取工作。接下来是核心模块层, 主要由ST的核心组件构成, 为了提升训练效率、降低模型的参数量, 在ST-TPP模型中, 分为3个阶段, Swin-Transformer块的数量设置为(2,2,6)。最后一层是全连接层, 起到回归的作用, 将上一层得到的高维特征映射成最终所需的六维向量, 其中最后一层采用两组回归, 分别实现对旋转和平移参数的回归。

接下来介绍各个模块的核心超参数设置。在ProST模块中, 令BatchSize=4, 表示一次生成4张DRR图像, 在DRR生成模块中, 需要对输入的CT进行预处理操作, 首先根据CT Volume的原始信息, 将体素尺寸修改为各向同性, 本文将体素大小修改成 $0.5 \times 0.5 \times 0.5$, 单位为mm; 为了增强单张DRR的图像特征, 这里将DRR图像复制成 N 通道, 保持通道数和CT通道一致; 在3D CNN模块中设置了4个3D卷积层, 通道数设置为(4,8,3,1), 卷积核尺寸为 $3 \times 3 \times 3$, 步长和边填充层数都设置为1, 该层的输入到输出采用了两个Skip Connection结构, 用来将3D CNN提取的特

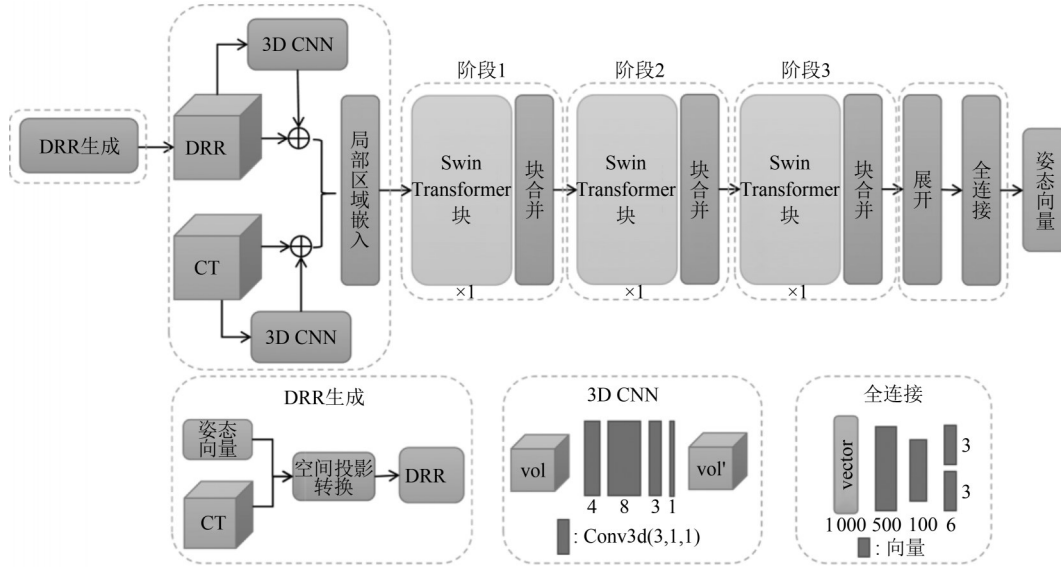


图2 ST-TPP网络结构

征和原始数据进行融合，最后将融合的数据按照通道所在的维度拼接在一起，作为一个整体送入ST模块中。

1.2 基于Grangeat关系的粗配准

传统的2D/3D配准方法主要通过迭代生成DRR图像，比较DRR与目标图像的相似度，优化找到最佳投影参数。但DRR生成过程缓慢，导致配准时间漫长。Frysch等^[25]提出了一种利用Grangeat关系进行配准的新方法，通过预计算简化了DRR生成，大幅提升配准速度，且精度和鲁棒性高。基于此，提出了基于Grangeat关系的粗配准（coarse registration based on Grangeat relation, CR-GR）方法。

1.2.1 Grangeat关系

配准算法的核心是利用了Grangeat关系，首先介绍其原理。在解析类的重建算法中，通过构造 $\delta'(x)$ 的中间函数，使得三维重建可以比较精确地进行，Grangeat关系见计算式：

$$-\int_{S^2} g_c(\hat{a}) \delta'(\hat{a} \cdot \hat{n}) dA = \frac{\partial}{\partial d} R_3[\mu](\hat{n}, d) \quad (4)$$

其中， $\delta'(x)$ 表示Dirac函数 $\delta(x)$ 的导数；向量 $\hat{a}$ 指向单位球面上的一个无穷小的面 dA 。

$R_3[\mu](\hat{n}, d)$ 表示对象的三维Radon变换，其中 $\hat{n}$ 是法向量， d 表示到平面原点的距离。函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 的 n 维Radon变换定义为所有超平面积分的集合，计算式如下：

$$R_n[f](\hat{n}, d) = \int_{\mathbb{R}^n} f(x) \delta(x \cdot \hat{n} - d) d^n x, \quad \hat{n} \in S^{n-1}, d \in \mathbb{R} \quad (5)$$

成像系统示意图如图3所示，展示了一种配备了单平板探测器的CT成像系统，若在笛卡尔探测器坐标 (u, v) 中描述式(4)的左侧，则投影数据变成 $g_c(u, v) = g_c(\hat{a}_{u,v})$ ：

从图3可以得到X射线投影和CT图像的中间函数 I_g 和 I_μ ，Grangeat关系可以写为：

$$I_g(\vartheta, s) = I_\mu(\vartheta, s) \quad (6)$$

$$I_g(\vartheta, s) = \frac{1}{\cos^2 \gamma_s} \frac{\partial}{\partial s} R_2[\tilde{g}_c](\hat{n}_{\vartheta, s}, s) \quad (7)$$

$$I_\mu(\vartheta, s) = \frac{\partial}{\partial d} R_3[\mu](\hat{n}_{\vartheta, s}, d) \quad (8)$$

其中， $R_2[\tilde{g}_c]$ 表示余弦加权投影图像的二维Radon变换， $\tilde{g}_c(u, v) = \cos \gamma_{u,v} g_c(u, v)$ ，其中 $\gamma_{u,v}$ 是指向 $\hat{a}_{u,v}$ 方向的锥角。 I_g 定义中的角度 γ 表示指向

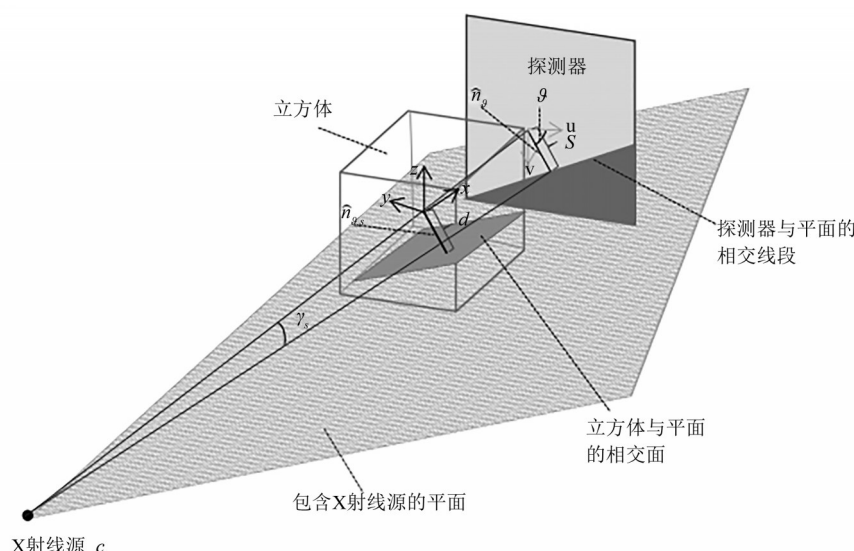


图3 成像系统示意图

探测器坐标 $s\hat{n}_g$ 的射线的锥角,是独立于 ϑ 的。在CT中间函数 I_μ 中,引入了平面法线 $\hat{n}_{g,s} \in S^2$,它与正弦图变量 (ϑ, s) 有固定的关系,即由 $\hat{n}_{g,s}$ 和 $d=c \cdot \hat{n}_{g,s}$ 所指定的平面被定义为包含射线源,所得到的交线由二维Radon参数 (ϑ, s) 表示。

1.2.2 目标函数与最优化问题

Grangeat关系表明,在理想条件下,X射线投影和CT的中间函数是相同的, I_g 和 I_μ 之间的任何差异都是由姿态不匹配造成的,这种不匹配用 E_{I_g, I_μ} 表示:

$$E_{I_g, I_\mu}(\mathbf{P}) = 1 - \text{corr}[I_g, I_\mu(\mathbf{P})] \quad (9)$$

在齐次坐标系中,刚性变换矩阵形式如下:

$$H_{R,t} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \in \mathbb{R}^{4 \times 4} \quad (10)$$

使用旋转矩阵 R 表示旋转,向量 t 表示平移,为了参数化 R 的3个自由度,选择一个向量 r ,即 $R=R(r)$,它的方向代表旋转轴,当更新投影位置时,新的投影矩阵为:

$$\tilde{P} = PH_{R(r),t} \quad (11)$$

最后,就得到一个需要优化的目标函数:

$$\arg \min_{r,t} E_{I_g, I_\mu}(PH_{R(r),t}) \quad (12)$$

上述计算式的优化问题类似于基于DRR的迭

代配准,不同的是当 $H_{R,t}$ 变化时,X射线变换必须随时重新计算,而 I_μ 可以从一个预先计算的中间函数空间进行插值,于是这里可以用一种新的方法,即预先计算出对象的所有可能的X射线变换,上文提到的中间空间具有Radon变换的维数,它保持了被变换对象的原始维数。

中间函数式(7)和式(8)需要计算二维和三维图像的Radon变换,它是在无限区间上定义的,但CT探测器的尺寸是有限的,所以要在以下参数范围内计算:

$$I_g(\vartheta, s) \text{ 的二维范围是: } s \in \left[-\frac{D_{\text{proj}}}{2}, \frac{D_{\text{proj}}}{2} \right],$$

$$\vartheta \in [0, \pi];$$

$$\frac{\partial}{\partial d} R_3[\mu](\theta, \varphi, d) \text{ 的三维范围是: } d \in$$

$$\left[-\frac{D_{\text{vol}}}{2}, \frac{D_{\text{vol}}}{2} \right], \theta \in [0, \pi], \varphi \in [-\pi, \pi];$$

D_{proj} 和 D_{vol} 分别表示CT和投影图像的对角线。基于Grangeat关系的配准框架如图4所示。可以将投影和CT图像中的信息编码到所谓的中间空间中,一旦得到图像数据,就可以预先计算这两个中间体,即图4中虚线部分,它们在整个几何优化过程中保持不变。

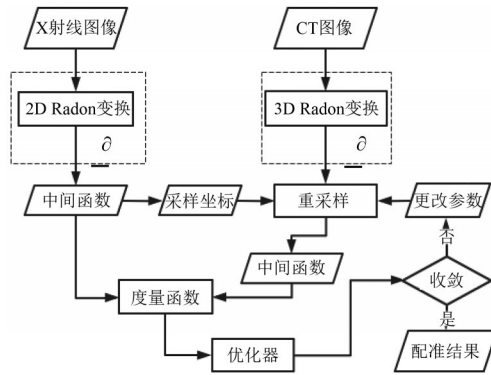


图4 基于Grangeat关系的配准框架

在图4中,通过对投影和CT图像预处理得到中间空间。在优化过程中,结合二维中间函数的采样坐标和几何变换,确定三维中间空间的重采样切片。使用度量函数比较CT中间函数和投影中间函数,并通过优化器改变几何修正以最小化度量值。

1.3 基于梯度下降的精配准

经过粗配准流程得到相对精确的配准结果,但在真实X射线图像配准中存在一定误差。为解决这一问题,介绍基于梯度下降的精准配准方法AR-GD,以提高配准的准确性。

1.3.1 相似度测量

配准的关键是衡量参考图像和浮动图像的相似度^[26]。归一化互相关是常用的配准方法,但容易陷入局部最优值。为了提高准确性,以考虑在NCC中加入图像的局部特征信息,如梯度和边缘信息。梯度表示组织之间的变化,边缘信息包含轮廓且不易受几何变换影响。利用高效的微分算子和边缘检测方法,将梯度和边缘信息应用于配准是一个有效的方法。

Sobel算子是一种常用的图像边缘检测算法,通过对图像进行卷积操作得到每个像素点的边缘梯度值^[27]。它通常由两个矩阵组成:用于检测水平和垂直方向的边缘信息。在实际应用中,Sobel算子通常与其他算法结合使用以提高精度和效率。

Sobel算子包含两组 3×3 的矩阵,分别为横

向及纵向,对于一张二维图像 I ,它在点 (x,y) 处的梯度计算式如下:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -2 & 0 & 1 \end{bmatrix} * I, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I \quad (13)$$

其中, I 代表原始图像, G_x 及 G_y 分别代表经横向及纵向边缘检测的图像梯度。

$$G(x,y) = \sqrt{G_x^2 + G_y^2} \quad (14)$$

其中, $G(x,y)$ 表示图像 I 在位置 (x,y) 的梯度幅值。为了增强对图像的特征提取性能,接下来将其与其他算法进行融合。

使用Sobel算子可以计算每个像素点的梯度向量,进而得到梯度方向图、梯度幅值图以及边缘图。但将归一化互相关直接应用于梯度幅值会存在图收敛范围窄的问题,因为忽略了大量图像信息。所以提出了一种基于Sobel算子的梯度归一化互相关(GNCC)方法,以更好地利用图像信息。

假定现在有两幅图像 I_a 和 I_b ,现在分别计算图像 I_a 在X方向上的梯度 $G_x(I_a)$ 、 $G_y(I_a)$ 、 $G_x(I_b)$ 、 $G_y(I_b)$,接着将这些计算结果结合归一化互相关中,形成GNCC,表达式如下:

$$\text{GNCC} = 1 - (\text{NCC}(G_x(I_a), G_x(I_b)) + \text{NCC}(G_y(I_a), G_y(I_b))) / 2 \quad (15)$$

其中,NCC代表计算归一化互相关的函数,首先将图像 I_a 和 I_b 在X方向的梯度图像进行相似度计算,然后将图像 I_a 和 I_b 在Y方向的梯度图像进行相似度计算,两幅图像越相似,NCC的计算结果越接近1,当图像完全匹配时,GNCC值为0,因此本文将GNCC作为AR-GD模型的相似度损失函数。

1.3.2 AR-GD模型

当粗配准过程完成后,空间变换参数的误差范围在 $\pm 5 \text{ mm}$ 和 $\pm 5^\circ$ 以内,由于误差较小,本文采用了一种基于梯度下降的方式来求取精确值。AR-GD模型结构如图5所示,模型结构比较简单,核心为两个模块,一个是DRR生成模块,另



一个是相似度计算模块，输入是粗配准参数、目标图像和CT图像。

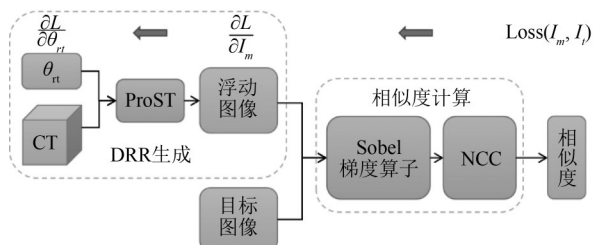


图5 AR-GD模型结构

在DRR生成模块中，ProST投影算法可以根据输入的姿态参数 θ_r 进行投影，生成对应的浮动图像 I_m ，然后相似度计算模块会计算浮动图像和目标图像的相似度损失，由于ProST模块是可微的，可以通过反向传播在不断迭代配准中来更新 θ_r 的值，使其朝着梯度下降的方向，快速逼近最优值。

式(16)展示了对 θ_r 进行基于梯度优化的反向传播梯度流。

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial I_m} \cdot \frac{\partial I_m}{\partial G_s} \cdot \frac{\partial G_s}{\partial G_T} \cdot \frac{\partial G_T}{\partial T(\theta)} \cdot \frac{\partial T(\theta)}{\partial \theta} \quad (16)$$

其中， L 表示 $\text{Loss}(I_m, I_t)$ ， G_s 表示线性插值后的点阵， G_T 表示经过仿射变换后的点阵， $T(\theta)$ 表示投影参数 θ 对应的仿射变换矩阵。

AR-GD精配准方法通过迭代寻优实现配准，损失函数由Sobel梯度算子和归一化互相关函数组合而成，将计算两幅图像的 X 、 Y 方向梯度图，再计算它们的相似程度，可加快收敛速度。

2 实验分析

首先介绍实验数据来源、实验所需的环境和相关细节，然后对整个模型的各个子模块分别进行实验分析，最后通过实验对比分析本文提出的模型的预测性能及表现，证明模型的有效性。

2.1 实验数据

实验数据共计20组，每组数据包括CT文件和两张X射线图像，以及每张X射线图像所对应

的拍摄姿态矩阵，第一组CT图像的渲染图如图6所示，第一组数据的X射线图像如图7所示。



图6 第一组CT图像的渲染图

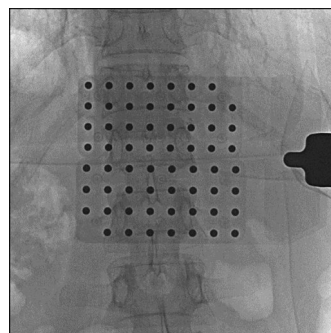


图7 第一组数据的X射线图像

2.2 实验模块和环境

配准算法的整体架构如图8所示，主要包括3个模块，分别为初始姿态估计模块、粗配准模块和精配准模块。初始姿态估计模块用来估计X射线图像的初始参数，解决了初始搜索参数与实际参数偏差过大导致配准效率低、只能达到局部最优解的问题^[20]。基于Grangeat关系的粗配准模块需要输入CT图像、X射线图像和基准姿态，输出粗配准结果。精配准模块使用基于梯度下降的优化方法，输入CT图像、X射线图像和粗配准参数，输出精确配准结果。相似度计算部分先用Sobel算子求取图像梯度特征，再用NCC计算梯度图像相似度，加快了求解收敛速度。

实验所涉及的环境：操作系统为Ubuntu 18.04和Windows10 64 bit，CPU为Intel Core i7-

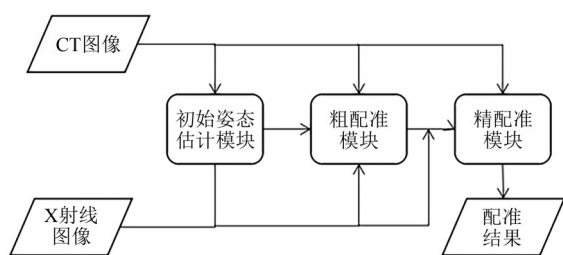


图8 配准算法的整体架构

10700 @ 2.90 GHz 八核，GPU 型号为 NVIDIA-A40 显存 24 GB，使用的语言有 Python 3.8 和 C++，使用的深度学习框架为 PyTorch 1.11.0，开发软件为 PyCharm 2020.1 和 Visual Studio 2017。

2.3 实验设计和结果

2.3.1 初始姿态估计

初始姿态估计的目标是预测 X 射线图像的投影姿态，提供精确的搜索范围，降低配准时间复杂度。训练所需数据集由 DRR 成像算法生成，经过风格化处理保持一致性，使模型对真实 X 射线图像预测准确。

DRR 图像由 6 个空间参数 (rx, ry, rz, tx, ty, tz) 决定，前 3 个参数代表沿着 x、y、z 轴旋转尺度，单位是 (°)，后 3 个参数代表按照 x、y、z 轴平移，单位是 mm，修改不同的参数，生成的图像也不同。在无约束情况下，平移参数无限制，旋转参数可以在 360° 内任取，但在骨科临床拍摄 X 射线图像时，C 臂机的拍摄姿态是在一定范围内运动的，所以当生成 DRR 图像时，应该与其保持一致。

实验前将 6 个参数分别在各自取值范围内以 10 为步长进行取值，并组合形成一个包含 6 个参数的列表，在所有参数取值并组合完毕后，就形成了一个包含约 20 万组投影参数的数据集。为减少数据量，此处按照等间隔抽取 1 万组参数，20 组 CT 共生成 20 万张 DRR 图像，最后按照 8:2 的比例将其分为训练集和测试集。在训练前还需要将 CT 的体素间距修正为各项同性即体素间距在 3 个方向上相同，另外将 CT 每张切片尺寸下采样为 224×224。以第一组数据为例，CT 的尺寸为 512×

512×176，体素间距为 0.351 5×0.351 5×1.25，单位是 mm，下采样后，CT 的尺寸变为 224×224×274，体素间距为 0.803 5×0.803 5×0.803 5，其他 CT 按照同样的方法进行预处理。在 DRR 成像模块中，X 光源到成像板的距离设置为 1 020，DRR 成像的尺寸设置为 224×224。

训练时超参数设置如下：Batch Size 设置为 4，Epoch 设置为 50，Optimizer 使用 Adam，lr 设置为 10^{-3} ，Weight Decay 设置为 10^{-4} ，Scheduler 选择 CosineAnnealingLR 调度器，其中 T_max=10，通过这些设置使得训练更快地收敛。模型的损失函数使用的是 MSE 均方误差函数，定义如下：

$$L = \frac{1}{n} \sum_{i=1}^n (rt_i - f(rt_i))^2 \quad (17)$$

其中， n 表示标签总数， rt_i 表示第 i 个标签， $f(rt_i)$ 表示第 i 个预测结果。图 9 展示了模型在训练过程中训练集的损失变化情况。

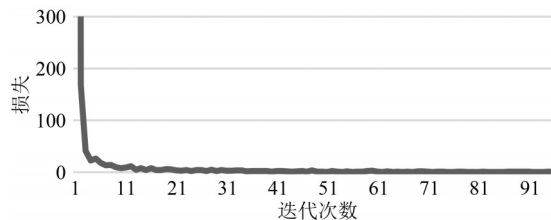


图9 训练集损失变化曲线

为验证模型性能，将其与另外两种方法进行了对比，分别是孙涛等^[20]提出的一种基于参数映射和图像分割的初始姿态估计方法 (PM-PE) 和金顺楠^[28]提出的基于卷积分类网络的粗配准方法 (GLNet)。在 5 个数据集上分别用上述 3 种方法重复实验 10 次，误差使用平均绝对误差 MAE 计算，10 次误差的均值作为最终误差 (平均值±标准差)，然后分别对旋转和平移参数进行误差分析，3 种方法的旋转误差对比见表 1。

可以看出 ST-TPP 模型在不同数据集上的表现都优于其他模型，在 5 个数据集上，3 个旋转参数误差均值为 12.15°，平移参数误差均值为



12.70 mm, 单个参数的估计误差在4附近, 远低于上文设定的10 mm或10°, 满足2D-3D配准的初始偏差波动范围。所以, 较传统卷积网络而言, Swin Transformer基于自注意力和移动窗口机制, 大幅提升了感受野, 降低了模型复杂度, 能够更加精准地提取图像特征, 结果更精确。

表1 3种方法的旋转误差对比

方法	数据集				
	Spinal01	Spinal02	Spinal03	Spinal04	Spinal05
PM-PE	16.54±1.79	17.83±1.83	15.61±1.66	16.32±1.71	15.57±1.62
GLNet	15.65±1.56	14.79±1.43	13.52±1.37	13.96±1.39	14.11±1.35
ST-TPP	12.52±1.23	12.91±1.34	11.38±0.95	12.46±1.27	11.52±1.22

2.3.2 粗配准实验

本节将对CR-GR粗配准方法进行效果和性能上的实验, 随机选取第11组CT图像进行实验, 实验时投影姿态以AP位投影参数为基准, 平移范围 $\max\text{Trans}=\pm 20$ mm, 旋转范围 $\max\text{Rot}=\pm 20^\circ$, 每组实验用100张DRR图像。

实验评价指标为成功率, 初步标准为配准误差不超过2 mm或2°。分为6个等级, 随机变化参数, 最难配准的等级六投影姿态距离基准姿态最远。初步实验结果分析得知当待配准参数在3个以内时, 配准成功率达到93%以上, 之后迅速降低。配准失败的数据分析表明, 投影参数与基准姿态偏差越大, 图像配准误差越大, 失败图像主要在 ± 20 mm或 $\pm 20^\circ$ 范围内。鉴于成功标准过高, 将达标要求改为配准结果与实际标签误差不超过5 mm或5°。然后重新实验, 不同难度等级的配准成功率如图10所示。

所以, CR-GR粗配准方法的主要误差范围在 ± 5 mm或 $\pm 5^\circ$ 范围, 满足2D-3D配准的粗配准偏差波动范围要求。为了验证CR-GR粗配准方法的性能指标, 对其在6参数变化下的配准时间进行了实验, 实验使用100张不同的DRR图像, 配准时间分布如图11所示。

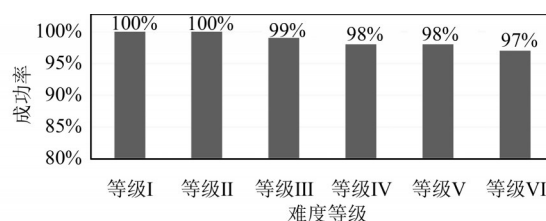


图10 不同难度等级的配准成功率

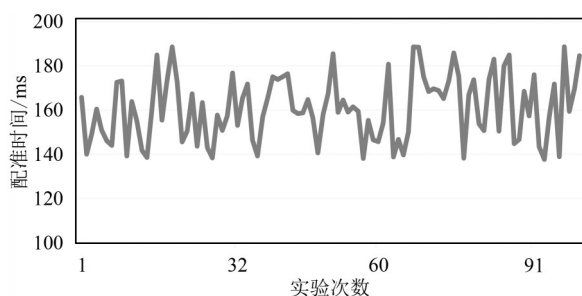


图11 配准时间分布

配准的平均时间在160 ms左右, 相较于其他非深度学习方法, CR-GR的配准速度是非常优秀的, 这得益于其配准之前对CT图像的预计算, 节省了配准时大量的DRR生成时间。

2.3.3 精配准实验

先对CT体数据沿着LR轴逆时针旋转90°, 保证投影姿态为(0,0,0,0,0,0)时, 和原先AP位的DRR成像结果相同。为验证该方法在粗配准后的最优性能, 以全零参数的DRR为目标图像, 用偏差最大的参数(5,5,5,5,5,5)生成待配准的浮动图像, 前3个参数代表沿着x轴、y轴、z轴的旋转, 单位是(°), 后3个参数代表沿着x轴、y轴、z轴的平移, 单位是mm。图12展示了两种姿态的DRR图像, 左图对应姿态参数(5,5,5,5,5,5), 右图对应姿态参数(0,0,0,0,0,0)。

基于上述图像开展实验, AR-GD方法使用SGD优化器, 将循环学习速率设置在 1×10^{-3} 和 3×10^{-3} 之间, 采用结合Sobel梯度算子的梯度NCC作为损失函数, 记作 L_N , 在配准过程中记录最近5次 L_N 的标准差, 记作 S_{LN} , 用来决定配准程序是否终止, 停止标准设置为 $S_{LN} \leq 1 \times 10^{-4}$, 实验目标是配准误差范围不超过1 (°) / (1 mm),

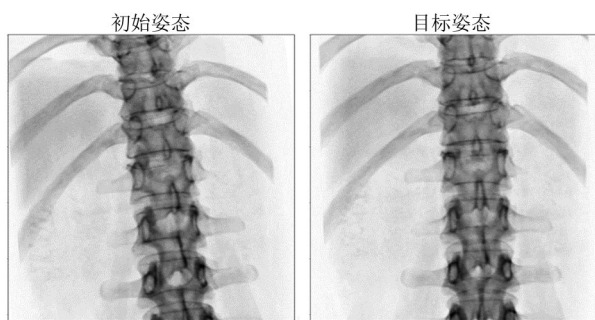


图12 待配准浮动图像和目标图像

配准过程损失如图13所示。

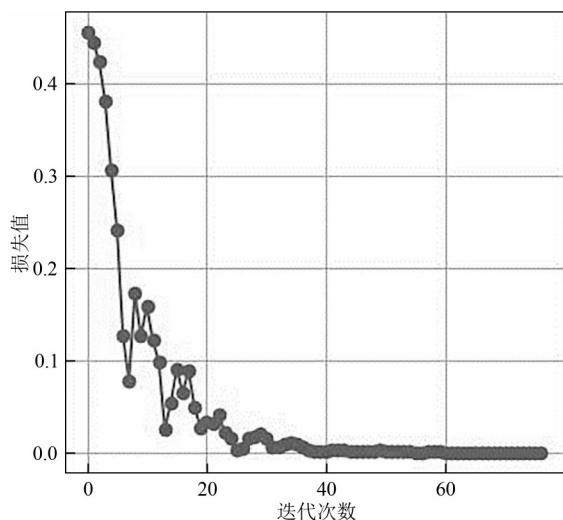


图13 配准过程损失

由图13可知,当迭代次数为35时,相似度损失值趋近于0,且在后续配准时仍然保持这种趋势,证明此时2D/3D图像已经配准。下面继续分析配准过程中平移参数和旋转参数的配准结果,旋转参数随配准过程变化如图14所示,平移参数随配准过程变化如图15所示。

图14显示了旋转参数配准结果和配准次数的变化,呈快速收敛的趋势,最初20次迭代中收敛速度较高,之后缓慢趋近于目标值,最终收敛。经过20次迭代后,误差达到 $\pm 1^\circ$,耗时16 s,相比于其他基于迭代的配准方法,AR-GD方法的配准速度是非常优秀的。图15显示了平移参数配准结果和配准次数的变化,与旋转参数保持一致,呈波浪式收敛趋势。在20次迭代后,误差范围达

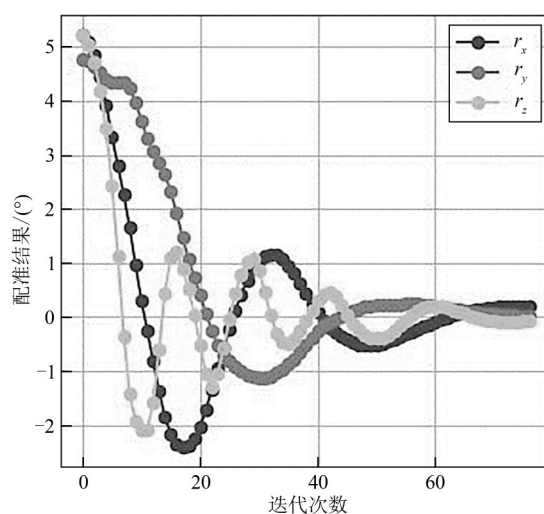


图14 旋转参数随配准过程变化

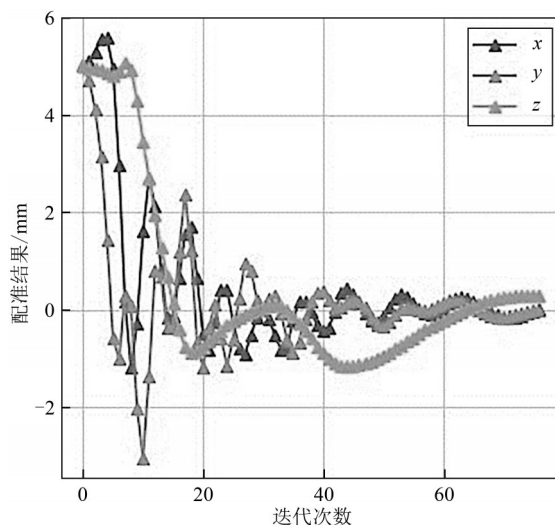


图15 平移参数随配准过程变化

到 ± 1 mm,满足精配准要求。

2.3.4 整体实验结果

(1) 配准评价指标

• 配准误差

为了体现配准误差值的直观性,这里使用MAE来度量,其计算式如下:

$$E = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (18)$$

其中, x_i 代表真实值的第*i*个参数, y_i 代表预测值的第*i*个参数。



• 配准时间

配准时间是评价配准算法好坏的重要原因之一,在保证配准精度的前提下,配准时间越低,算法的实用性越强。

• 配准成功率

配准成功率设定为配准成功的次数占配准总次数的比值,配准成功率高,证明算法的稳定性强,这里配准成功的标准设置为平移差值不超过 1 mm,旋转差值不超过 1°。

(2) 实验结果

图 16 所示为待配准的两幅 X 射线图像,分别是正位(AP 位) X 射线图像和侧位(LR 位) X 射线图像,其对应的真实成像参数分别为(-93,3,-1,-6,8,-7)、(-87,1,89,3,-1,5)。

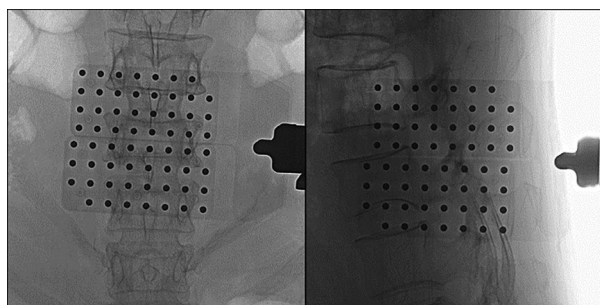


图 16 待配准正位(左)和侧位图像(右)

分别将这两张图送入 MMIR 模型,以正位图像配准结果为例,实验结果见表 2。

表 2 展示了正位的配准结果和误差值,可知平移和旋转参数的配准误差均小于 1 mm/ (1°),满足了配准要求。图 17 展示了两张图像配准后的效果,为了展示效果,将配准后的图像分割出脊柱部分,然后叠加在原始图像上,灰色阴影表示

配准后的图像。

为了验证本算法的性能,将其与传统的基于 Powell 优化的配准算法和基于深度学习的 GST-PG^[23]算法进行对比,主要比较它们在配准时间、配准误差和配准成功率上的差异。对比实验基于上文中的正位图像,随机生成 20 组变化范围介于 ± 10 mm/10° 的 DRR 图像,分别进行实验,结果取均值,3 种算法的性能对比见表 3。

可以看出,本文提出的 MIR-MF 配准方法在 3 个维度上均优于 Powell 和 GST-PG 两种配准算法,配准速度相较另两种方法有极大的提升,配准误差也最小。在 20 次实验中,本方法成功率为 100%,基于 Powell 的配准有 4 次失败,通过分析可知这是由于该算法在寻优过程中陷入了局部极值,配准误差较大。GST-PG 有一次失败,是由于其在 Z 轴方向的配准误差过大。

3 结束语

本文介绍了 2D/3D 多模态医学图像配准算法,在配准过程中,构建了预配准-粗配准-精配准 3 步配准策略,预配准模块凭借神经网络的优异性快速地预测出基准姿态,粗配准模块在基准姿态基础上再次配准,得到更加精确和鲁棒性较强的配准结果。为了使配准精度达到最优,精配准模块借鉴深度学习的训练过程来求得最优配准结果。最后,通过实验将提出的模型与其他常用模型进行对比。结果表明,本文所提方法更为优秀,在配准精度、相对速度和鲁棒性方面均有大幅提升。
致谢:感谢杭州三坛医疗科技有限公司为本研究提供临床试验数据。

表 2 正位(AP 位)图像配准结果

对比项	$r_x/(^\circ)$	$r_y/(^\circ)$	$r_z/(^\circ)$	t_x/mm	t_y/mm	t_z/mm
真实值	-93	3	-1	-6	8	-7
配准结果	-93.313 68	2.583 52	-0.812 74	-6.458 31	8.833 07	-6.377 61
误差值	0.313 68	0.416 48	0.187 26	0.458 31	0.833 07	0.622 39

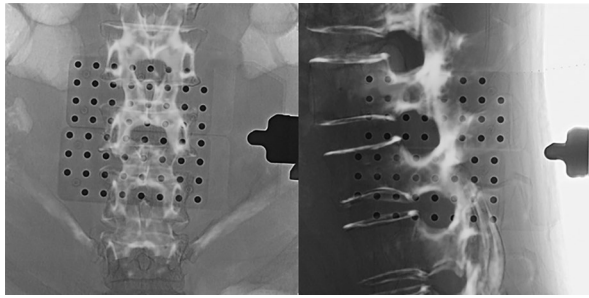


图17 配准效果

表3 3种算法的性能对比

算法	配准时间/s	配准误差	配准成功率
Powell	283.62	3.98	80%
GST-PG	167.06	0.87	95%
MIR-MF	20.71	0.41	100%

参考文献:

- [1] FAN X Q, ZHU Q Y, TU P X, et al. A review of advances in image-guided orthopedic surgery[J]. *Physics in Medicine and Biology*, 2022.
- [2] 张峰峰, 陈龙, 杨诗怡, 等. 机器人辅助脊柱手术三维定位导航[J]. *哈尔滨工程大学学报*, 2020, 41(12): 1735-1741.
ZHANG F F, CHEN L, YANG S Y, et al. Robot-assisted three-dimensional positioning and navigation in spinal surgery[J]. *Journal of Harbin Engineering University*, 2020, 41(12): 1735-1741.
- [3] 莫晓盈, 杨锋, 尹梦晓, 等. 医学图像配准的深度学习方法综述[J]. *小型微型计算机系统*, 2021, 42(8): 1706-1714.
MO X Y, YANG F, YIN M X, et al. Survey on deep learning in medical image registration[J]. *Journal of Chinese Computer Systems*, 2021, 42(8): 1706-1714.
- [4] VIERGEVER M A, MAINTZ J B A, KLEIN S, et al. A survey of medical image registration - under review[J]. *Medical Image Analysis*, 2016(33): 140-144.
- [5] MARTÍNEZ-RÍO J, CARMONA E J, CANCELAS D, et al. Deformable registration of multimodal retinal images using a weakly supervised deep learning approach[J]. *Neural Computing and Applications*, 2023, 35(20): 14779-14797.
- [6] HE H Q, CHEN M, CHEN T, et al. Matching of remote sensing images with complex background variations via Siamese convolutional neural network[J]. *Remote Sensing*, 2018, 10(2): 355.
- [7] AKTER M, LAMBERT A J, PICKERING M R, et al. Improved robustness of 3D CT to 2D fluoroscopy image registration using log polar transforms[C]//*Proceedings of the 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*. Piscataway: IEEE Press, 2016: 1-5.
- [8] HAMADEH A, LAVALLEE S, CINQUIN P. Automated 3-dimensional computed tomographic and fluoroscopic image registration[J]. *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery*, 1998, 3(1): 11-19.
- [9] GONG R H, STEWART J, ABOLMAESUMI P. Multiple-object 2-D-3-D registration for noninvasive pose identification of fracture fragments[J]. *IEEE Transactions on Bio-Medical Engineering*, 2011, 58(6): 1592-1601.
- [10] HIROTAKI K, MORIYA S, AKITA T, et al. Image preprocessing to improve the accuracy and robustness of mutual-information-based automatic image registration in proton therapy[J]. *Physica Medica: PM: an International Journal Devoted to the Applications of Physics to Medicine and Biology: Official Journal of the Italian Association of Biomedical Physics (AIFB)*, 2022(101): 95-103.
- [11] HEBDA-SOBKOWICZ J, ZIMROZ R, PITERA M, et al. Informative frequency band selection in the presence of non-Gaussian noise - a novel approach based on the conditional variance statistic with application to bearing fault diagnosis[J]. *Mechanical Systems and Signal Processing*, 2020(145): 106971.
- [12] SIDDIQI M H, ALI R, IDRIS M, et al. Human facial expression recognition using curvelet feature extraction and normalized mutual information feature selection[J]. *Multimedia Tools and Applications*, 2016, 75(2): 935-959.
- [13] VIEL C, VIOLLET S. Fast normalized cross-correlation for measuring distance to objects using optic flow, applied for helicopter obstacle detection[J]. *Measurement*, 2020(172): 108911.
- [14] SENTHIL PANDI S, SENTHILSELVI A, MARAGATHARAJAN M, et al. An optimal self adaptive deep neural network and spine-kernelled chirplet transform for image registration[J]. *Concurrency and Computation: Practice and Experience*, 2022, 34(27): e7297.1-e7297.16.
- [15] ZHANG Z Y, LIU Y, TIAN J W, et al. Study on reconstruction and feature tracking of silicone heart 3D surface[J]. *Sensors*, 2021, 21(22): 7570.
- [16] PYATOV V A, SOROKIN D V. Affine registration of histological images using transformer-based feature matching[J]. *Pattern Recognition and Image Analysis*, 2022, 32(3): 626-630.
- [17] EL-GAMAL F E Z A, ELMOGY M, ATWAN A. Current trends in medical image registration and fusion[J]. *Egyptian Informatics Journal*, 2016, 17(1): 99-124.
- [18] EPPENHOF K A J, PLUIM J P W. Pulmonary CT registration through supervised learning with convolutional neural networks[J]. *IEEE Transactions on Medical Imaging*, 2018, 38(5): 1097-1105.
- [19] MIAO S, PIAT S, FISCHER P, et al. Dilated FCN for multi-agent 2D/3D medical image registration[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2018, 32(1).
- [20] 孙涛, 郭可, 刘传耙, 等. 2D-3D配准中的初始位姿估计方法[J]. *天津大学学报(自然科学与工程技术版)*, 2022, 55(2): 143-150.



- SUN T, GUO K, LIU C B, et al. Initial pose estimation method in 2D-3D registration[J]. Journal of Tianjin University (Natural Science and Engineering Technology), 2022, 55(2): 143-150.
- [21] PICKERING M R, MUHIT A A, SCARVELL J M, et al. A new multi-modal similarity measure for fast gradient-based 2d-3d image registration[C]//Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway: IEEE Press, 2009: 5821-5824.
- [22] AKTER M, LAMBERT A J, PICKERING M R, et al. Robust initialisation for single-plane 3D CT to 2D fluoroscopy image registration[J]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2015, 3(3): 147-171.
- [23] GAO C, LIU X, GU W, et al. Generalizing spatial transformers to projective geometry with applications to 2D/3D registration[C]//Medical Image Computing and Computer Assisted Intervention, 2020: 329-339.
- [24] LIU Z, LIN Y, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 10012-10022.
- [25] FRYSCH R, PFEIFFER T, ROSE G. A novel approach to 2D/3D registration of X-ray images using Grangeat's relation[J]. Medical Image Analysis, 2021(67): 101815.
- [26] KANG M S, KIM K T. Automatic SAR image registration via tsallis entropy and iterative search process[J]. IEEE Sensors Journal, 2020, 20(14): 7711-7720.
- [27] VINCENT O R, FOLORUNSO O. A descriptive algorithm for sobel image edge detection[C]//Proceedings of informing science & IT education conference (InSITE). 2009, (40): 97-107.
- [28] 金顺楠. 面向术中CT导航的脊柱自动分割与配准技术研究[D]. 杭州: 杭州师范大学, 2021.
- JIN S N. Research on automatic spinal segmentation and registration technology for intraoperative CT navigation[D]. Hangzhou: Hangzhou Normal University, 2021.

[作者简介]



徐密 (1998-), 女, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为医学图像配准。



诸葛斌 (1976-), 男, 博士, 浙江工商大学信息与电子工程学院教授, 主要研究方向为医学图像配准与模式识别、互联网技术和云计算。



袁非牛 (1976-), 男, 博士, 上海师范大学信息与机电工程学院教授, 主要研究方向为图像处理、模式识别、人工智能、深度学习。

尹正虎 (1997-), 男, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为医学图像处理、人工智能。

董黎刚 (1972-), 男, 博士, 浙江工商大学信息与电子工程学院教授, IEEE和IEEE-CS成员, 中国电子学会高级会员, 浙江省计算机学会理事, 主要研究方向为智能网络、在线教育。

蒋献 (1988-), 男, 浙江工商大学信息与电子工程学院讲师、实验员, 主要研究方向为深度学习、在线教育。

孙应绮 (1988-), 女, 复旦大学附属儿科医院主治医师, 主要研究方向为影像学诊断。

宋嘉琦 (2002-), 男, 浙江工商大学萨塞克斯人工智能学院在读, 主要研究方向为医学图像处理。

史晓彤 (2004-), 女, 浙江工商大学计算机科学与技术学院在读, 主要研究方向为医学图像处理。

苏雷 (2003-), 男, 浙江工商大学信息与电子工程学院在读, 主要研究方向为医学图像处理。

周屹博 (2002-), 男, 浙江工商大学计算机科学与技术学院在读, 主要研究方向为数据科学。

林诗凡 (2000-), 女, 浙江工商大学信息与电子工程学院硕士生, 主要研究方向为医学图像配准。