

中图法分类号: TP181; R687.1 文献标识码: A 文章编号: 1006-8961(2025)02-0601-14

论文引用格式: Shen A, Shen Y J, Jiang J F, Chen Z M, Huang R, He K J and Chen J. 2025. Semantic edge-guided 2D/3D registration for vertebrae in radiographs. Journal of Image and Graphics, 30(02):0601-0614(沈傲, 沈燕进, 蒋俊锋, 陈正鸣, 黄瑞, 何坤金, 陈杰. 2025. X线片椎骨语义边缘引导的2D/3D配准方法. 中国图象图形学报, 30(02):0601-0614)[DOI: 10. 11834/jig. 240001]

X线片椎骨语义边缘引导的2D/3D配准方法

沈傲¹, 沈燕进², 蒋俊锋^{1*}, 陈正鸣¹, 黄瑞¹, 何坤金¹, 陈杰²

1. 河海大学海上智能网信技术教育部重点实验室, 常州 213200; 2. 常州市第一人民医院, 常州 213003

摘要: **目的** 基于影像引导的脊柱手术机器人系统中, 2D/3D配准指的是将术前计算机断层扫描影像与术中X线片配准, 用于实现手术机器人对于人体组织的精准空间定位。常见做法是先用标志点进行粗配准, 再用灰度法修正位姿。标志点配准问题在于标志点识别精度不高且识别效率较低, 灰度法的捕获范围小且对初始位姿敏感。由于脊柱关节边缘重叠且术中X线片图像质量较低, 故利用物理边缘作为特征进行2D/3D配准精度不高。因此, 提出一种基于语义边缘提取的2D/3D配准方法。**方法** 首先, 提取X线片中成像清晰的椎弓根边缘和椎体两侧边缘作为语义特征进行2D/3D配准; 同时, 面向边缘提取任务, 研究一种间距约束的高效“U”形变形器网络, 该深度学习网络提高分割效率的同时, 保持了边缘分割的准确性, 并加入了椎骨间距约束损失的先验信息, 进一步提升了多椎骨语义边缘提取的精度。**结果** 模拟数据与真实数据上评估结果表明, 本文方法在配准精度与效率方面均优于现有方法; 位姿修正后, 本文方法平移误差小于1 mm, 旋转误差小于0.1°, 配准耗时在5 s左右, 能较好满足实际临床需求。**结论** 本文提出的基于椎骨语义边缘的2D/3D粗配准方法有效缩小了后续精配准过程的搜索空间, 从而提高了配准精度。在边缘提取方面, 将哈达玛乘积代替卷积操作的方式以及加入椎骨间距约束损失, 提高了语义边缘的提取效率和精度。因此, 本文方法能够较好满足2D/3D配准的精度与实时性需求。

关键词: 2D/3D配准; 脊柱手术机器人; 椎骨; 语义边缘提取; 视觉变形器网络(ViT)

Semantic edge-guided 2D/3D registration for vertebrae in radiographs

Shen Ao¹, Shen Yanjin², Jiang Junfeng^{1*}, Chen Zhengming¹, Huang Rui¹, He Kunjin¹, Chen Jie²

1. Key Laboratory of Maritime Intelligent Cyberspace Technology of Ministry of Education, Hohai University, Changzhou 213200, China;

2. The First People's Hospital of Changzhou, Changzhou 213003, China

Abstract: Objective In image-guided spine surgery-based robotic systems, 2D/3D registration refers to aligning preoperative 3D computed tomography (CT) images with intraoperative 2D X-rays to achieve precise spatial localization of the surgical robot for human tissues. The prevalent approach involves the use of landmark points for initial coarse registration and subsequently applying the intensity-based method to rectify the position. However, the landmark-based registration method typically uses a heatmap regression method, which can be GPU intensive. Simultaneously, the random field of view during intraoperative X-rays and the overlapping of human tissues on X-rays can result in the loss of tissue information. This loss can cause incorrect predictions or predictions with a considerable deviation of landmark points. The intensity-based method is considered the most accurate and efficient approach because it utilizes the entire image information. However, intensity-

收稿日期: 2024-01-30; 修回日期: 2024-06-14; 预印本日期: 2024-06-21

* 通信作者: 蒋俊锋 jigjf@hhu.edu.cn

基金项目: 河海大学海上智能网信技术教育部重点实验室中央高校业务费项目(B240203012); 江苏省重点研发计划资助(BE2022718)

Supported by: Key Laboratory of Maritime Intelligent Cyberspace Technology (Hohai University), Ministry of Education (B240203012); Jiangsu Province Key R&D Program, China (BE2022718)

based methods usually have problems such as a small capture range and sensitivity to the initial pose. The accuracy of 2D/3D registration when physical edges are used as features is low because of the overlap of spinal joint edges and the low quality of intraoperative radiographs. Here, a 2D/3D registration method based on semantic edge extraction is proposed.

Method The semantic edge-based 2D/3D registration method uses nonoverlapping pedicle edges and edges on both sides of the vertebra in the X-ray as semantic features for 2D/3D registration. The real-time detection Transformer model is first used to predict the bounding boxes of the vertebrae to be registered in intraoperative X-rays. The CT images of the vertebrae to be registered are then extracted from the known vertebrae masks. The semantic edges of the vertebrae from the intraoperative X-rays and the digitally reconstructed radiological images from the CT image projection are extracted by spacing constrained and efficient UNet Transformers (SCE-UNETR). Finally, the pose is updated iteratively by minimizing the reprojection error until convergence. The SCE-UNETR network uses a U-shaped structure for encoding and decoding; it comprises an encoder module, skip connection layers, and a decoder. The SCE-UNETR encoder consists of a vision Transformer (ViT) network, where the image features are connected to the decoder by skipping connection layers. The ViT network contains a total of 12 groups of Transformer blocks, each consisting of two normalization layers, a multihead attention layer, and a multilayer perceptron. In this study, the skip connection layer and the decoder of SCE-UNETR are illuminated by a multi-axis group Hadamard product attention module instead of a convolutional operation, which changes the quadratic complexity into linear complexity and reduces the quantity of network parameters. In addition, the network incorporates the loss of the vertebral spacing constraint as a priori information, which further improves the accuracy of multivertebra semantic edge extraction. **Result** The method proposed in this study was quantitatively and qualitatively evaluated from different perspectives, including semantic edge extraction, coarse registration, pose refinement, method comparison, registration accuracy, and registration time. For the semantic edge extraction task, the SCE-UNETR was compared with SwinUNETR, TransUNet, and UNETR. SCE-UNETR has approximately half the number of network parameters of the other networks. However, in terms of the metrics of segmentation overlap, average Dice, average intersection over union, and average 95% Hausdorff distance, the SCE-UNETR is better than those of the other three networks. In addition, all networks show some improvement in semantic edge extraction accuracy after the spacing constraint loss term is combined with Dice loss. For the 2D/3D registration task, the experiments demonstrate that the proposed method outperforms the state-of-the-art methods on both simulated and real data. Following pose refinement, the proposed method achieves a translation error of less than 1 mm and a rotation error of less than 0.1° . The registration process itself takes approximately 5 s, which is well suited to the clinical environment. **Conclusion** The proposed semantic edge-based 2D/3D registration method effectively reduces the search space of the subsequent pose refinement process, thus improving the registration accuracy. In terms of edge extraction, the network is illuminated, and the vertebrae spacing constraint loss is added, thereby improving the efficiency and accuracy of semantic edge extraction. Therefore, the proposed method can better meet the accuracy and real-time requirements of 2D/3D registration.

Key words: 2D/3D registration; spinal surgery robot; vertebra; semantic edge extraction; vision Transformer (ViT)

0 引言

图像引导的脊柱手术机器人系统借助高分辨率的医学影像精确定位患者的脊柱解剖结构,从而辅助医生完成手术(陈智强等,2018)。相比医生依靠传统经验的手术方式,这种方法能够显著提高手术的精确性、微创化,并有效减少愈合时间、术中辐射次数和手术风险,有助于手术标准化(徐敏铭,2018)。

脊柱手术的椎弓根定位精度需在 1 mm 以内,同

时对实时性要求高(Grupp等,2020a)。现有脊柱手术机器人存在定位精度不够高、定位时间过长等问题(Otake等,2012),其核心在于2D/3D配准算法的精度与效率尚不能满足临床需求。2D/3D配准指的是将病人术前的三维计算机断层扫描(computed tomography, CT)影像配准到术中二维X线片,实现术中不可见的人体组织的精准定位,如图1所示。

2D/3D配准方法通常先进行粗配准获得初始位姿,然后再利用精配准修正初始位姿。精配准方法通常分为基于灰度(Otake等,2012)、基于最小化点到面距离(Grupp等,2020b)和基于最小化点到投影

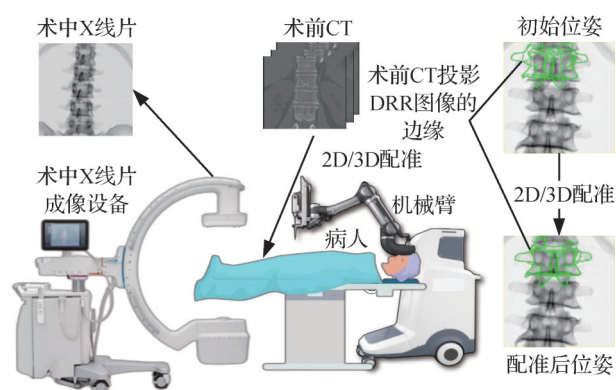


图1 脊柱机器人中2D/3D配准原理

Fig. 1 The principles of 2D/3D registration in spine robots

线距离(Jaganathan等, 2021) 3种类型。其中,在初始位姿较好时,基于灰度的方法是当前最精准高效的方法(Naik等, 2022),因其充分利用了整幅图像的信息。然而精配准方法通常存在捕捉范围(capture range)较小且对初始位姿敏感(Miao等, 2016)等问题。

粗配准方法通常利用标志点(黄永锋等, 2006)、边缘(Jaganathan等, 2023)等特征进行2D/3D粗配准,具体如下:1)基于标志点的2D/3D粗配准方法:标志点的提取通常利用热图回归(Yang等, 2020)的方法,该方法对于图形处理器(graphic processing unit, GPU)的开销较大;同时,术中X线片拍摄视角较为随机,X线片中的人体组织会重叠,丢失了人体组织的信息,因此会导致标志点预测错误或者偏差较大(Grupp等, 2020b)。2)基于边缘的2D/3D粗配准方法:对于骨骼边缘特征(Jang等, 2023),例如X线片中脊柱的中心线(Kadoury等, 2009)或股骨的轮廓(Gollmer等, 2007)等,可以利用熟知的Canny算子(Canny, 1983)、主动轮廓模型(Galibarov等, 2010)等传统算法进行提取。然而,X线片呈现的椎骨上下关节边缘通常会与其他区域重合,如图2所示,这种边缘特征提取不易。

本文从椎骨轮廓中选择了4种视觉上较为显著的椎骨语义边缘,即,左、右椎弓根轮廓与左、右边缘轮廓。X线片中这些边缘与上下椎骨不重合,且在成像质量通常不高的术中X线片中是较为清晰的特征。

语义边缘提取是计算机视觉领域的研究热点。随着UNet(Ronneberger等, 2015)和视觉变形器

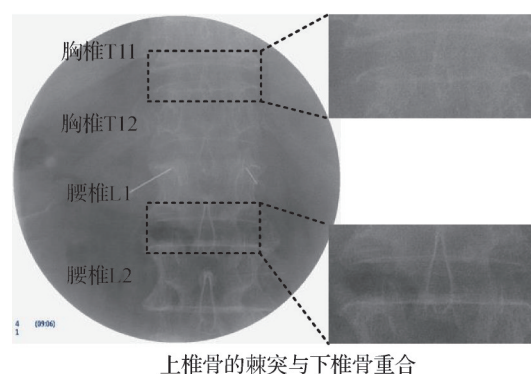


图2 真实X线片中椎骨上下边缘重叠区域

Fig. 2 The area of overlap between the upper and lower edges of the vertebrae in the X-rays

(vision Transformer, ViT)(Dosovitskiy等, 2020)的出现,利用深度学习分割模型提取待配准的特征区域已逐渐成为主流(Grupp等, 2020b)。目前,基于UNet改进的边缘提取算法有DEA-UNet(dense-edge-attention UNet)(Zeng等, 2022)、ET-Net(efficient Transformer enhanced UNet)(Zhang等, 2019)和DD-UNet(densely dilated UNet)(Wang等, 2023)等。此外,TransUNet(Transformer UNet)(王媛媛等, 2023)和CTN(contour Transformer network)(Lu等, 2021)则是基于ViT改进的网络。然而,深度学习分割模型,尤其是ViT模型,也存在GPU开销大、较为耗时等问题,不能较好满足术中实时性要求。综上,现有2D/3D配准算法的核心问题之一是如何快速精准地获取初始位姿。现有方法主要存在以下问题:1)现有基于标志点的方法精准度不高、GPU算力开销较大、效率较低,故难以满足临床需求;2)基于轮廓提取的方式在轮廓与其他区域重叠时,无法用传统的边缘提取算法计算;3)基于灰度法的捕获范围小、初始位姿不好的情况下,它容易陷入局部最优。目前,基于医学语义(以下简称为语义)边缘的脊柱粗配准方法尚未见公开发表。

本文提出一种基于语义边缘的2D/3D椎骨粗配准方法。目标是提取X线片中椎弓根边缘和椎体两侧边缘作为语义特征,并利用这种语义特征进行2D/3D配准;同时,基于现有计算机视觉中的边缘提取算法,提出了间距约束的高效“U”形变形器(spacing constrained and efficient UNet Transformers, SCE-UNETR)自动分割网络。本文方法能快速准确地获得CT影像的初始位姿,为后续精细配准减少搜索空间,避免陷入局部最优,从而更好地满足脊柱手术机

机器人的临床需求。

1 基于语义边缘深度提取的2D/3D粗配准方法

1.1 方法概述

基于图像灰度信息的2D/3D配准方法是通过最小化相似度误差(Unberath等, 2021), 计算CT影像的3D空间位姿。这里的相似度误差是指数字重建影像(digitally reconstructed radiology, DRR)(刘坤等, 2016)与X射线图像之间的相似度误差。DRR图像是利用光线跟踪(ray-tracing)算法(杜晓刚等, 2015)将三维体数据投影成二维透视图像得到的。基于最小化相似度误差的2D/3D配准数学模型为

$$T^* = \underset{T}{\operatorname{argmin}} D(L^\theta(DRR(V \times T)), L^\theta(I)) \quad (1)$$

式中, D 为相似度函数; L^θ 为参数为 θ 的椎骨边缘提取函数, DRR 将CT体数据 V 通过模拟光线跟踪算法生成的二维透视图像, V 为每节椎骨的CT体数据, I 为二维术中X线片, T 为CT影像的位姿, 由旋转和平移共6个自由度构成。本文利用归一化互相关(normalized cross-correlation, NCC)(麦永锋等, 2021)

作为相似度函数, 计算为

$$NCC(A, B) = \sum_{x,y} \frac{(A(x,y) - \mu_A)(B(x,y) - \mu_B)}{P\sigma_A\sigma_B} \quad (2)$$

式中, 输入的图像 A 和图像 B 都有 P 个像素, 平均值为 μ_A 和 μ_B , 标准差为 σ_A 和 σ_B , (x, y) 则表示像素的坐标。

本文方法总体流程如图3所示。首先, 利用实时检测变形器(Zhao等, 2024)模型预测出术中X线片的椎骨检测框, 根据检测框提取出待配准的腰椎, 每节具有医学语义标签, 例如腰椎分为L1、L2、L3、L4、L5(Netter, 2014); 然后, 根据现有的椎骨掩码提取出待配准椎骨的CT影像; 其次, 利用本文提出的SCE-UNETR网络分别提取出术中X线片与CT影像投影所得DRR图像的椎骨语义边缘; 最后, 利用式(1)计算出当前位姿变换 T 。其中, 优化式(1)的常用方法是协方差矩阵自适应演化算法(covariance matrix adaptation-evolution strategy, CMA-ES)(Otake等, 2012), 计算得到刚体变换 T 以更新椎骨的位姿, 随后生成新的DRR图像并提取语义边缘, 此过程不断迭代直至收敛, 最终输出CT椎骨的空间位姿变换 T 。以下对本文算法各模块进行详细介绍。

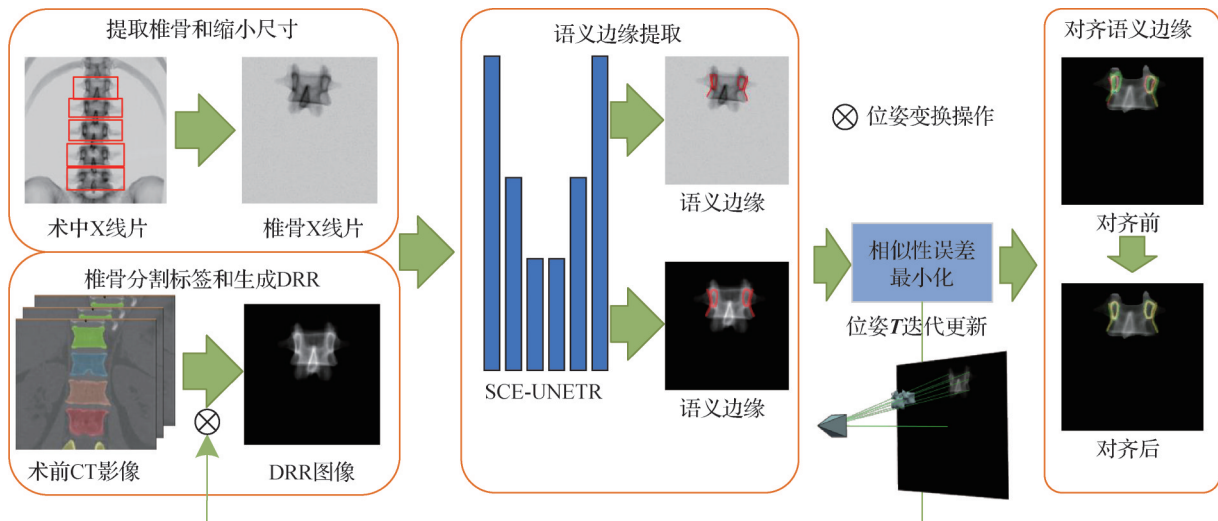


图3 基于语义边缘的2D/3D粗配准方法总体框架

Fig. 3 The framework for semantic edge-based 2D/3D coarse registration

1.2 SCE-UNETR结构

1.2.1 高效“U”形变形器分割框架

为提高分割效率并保持边缘分割准确性, 本文在UNETR网络(Hatamizadeh等, 2022)的基础上, 进一步提出了如图4所示的SCE-UNETR语义边缘分

割网络。该网络采用一种先编码后解码的“U”形结构, 总体结构由编码器模块、跳跃连接层和解码器构成, 具体介绍如下:

1) 编码器。SCE-UNETR的编码器由ViT网络(Dosovitskiy等, 2020)组成, 图像特征通过跳越连接

层连接到解码器。ViT网络共包含12组Transformer块,每组由两个归一化层(layer norm, LN)、一个多头注意力层和一个多层感知机组成。由于ViT网络采用一维线性编码,需要将图像经过嵌入层映射成一维序列。给定输入 $\mathbf{x} \in \mathbf{R}^{H \times W \times C}$ 重塑为若干二维图像块 $\mathbf{x}_v \in \mathbf{R}^{N \times (P^2 \cdot C)}$,其中 P^2 表示图像块的大小,本文设置为16, C 表示图像的通道数,设置为1,而 N 表示一维特征的长度,一般 $N = (H \times W)/P^2$,对 $\mathbf{x}_v$ 进行线性投影,计算式为

$$\mathbf{z}_0 = [\mathbf{x}_v^1 \mathbf{E}, \mathbf{x}_v^2 \mathbf{E}, \dots, \mathbf{x}_v^N \mathbf{E}] + \mathbf{E}_{\text{pos}} \quad (3)$$

式中, $\mathbf{E} \in \mathbf{R}^{D \times (P^2 \cdot C)}$ 为线性投影矩阵, $\mathbf{E}_{\text{pos}} \in \mathbf{R}^{N \times D}$ 则是包含位置信息的可微矩阵,其中 D 为投影的维度,最后将带有位置信息的图像块 $\mathbf{z}_0$ 输入ViT网络中,以提取特征。

2)跳跃连接层。SCE-UNETR的跳跃连接层共有4层。如图4所示,跳跃连接层从上至下分别由1层多轴哈达玛乘积注意力组(group multi-axis had-

amard product attention module, GHPA)(Ruan等, 2023)模块和3层反卷积组构成。第1层连接层无需上采样操作,因此,本文利用GHPA模块提取图像的低维特征以减少网络参数量。而后3层需对特征进行上采样,连接层由反卷积组构成。每组反卷积组由步长为 2×2 的反卷积层、步长为 3×3 的卷积层、LN层和ReLU(rectified linear unit)层构成。每层特征通过跳跃连接层传递给解码器,有助于解码器得到图像中低维且全局的信息。

3)解码器。对SCE-UNETR的解码器进行轻量化处理,其结构与UNet(Ronneberger等, 2015)类似,具体原理如下:首先,解码器与来自编码器中多种分辨率的特征相合并。解码器提取了Transformer模块中第3、6、9、12组序列作为连接的特征并重整为 $\frac{H}{P} \times \frac{W}{P} \times \frac{W}{P}$ (P 为下采样倍数)的特征图,在编码器的瓶颈处(即Transformer模块的最后一层的输出),经过

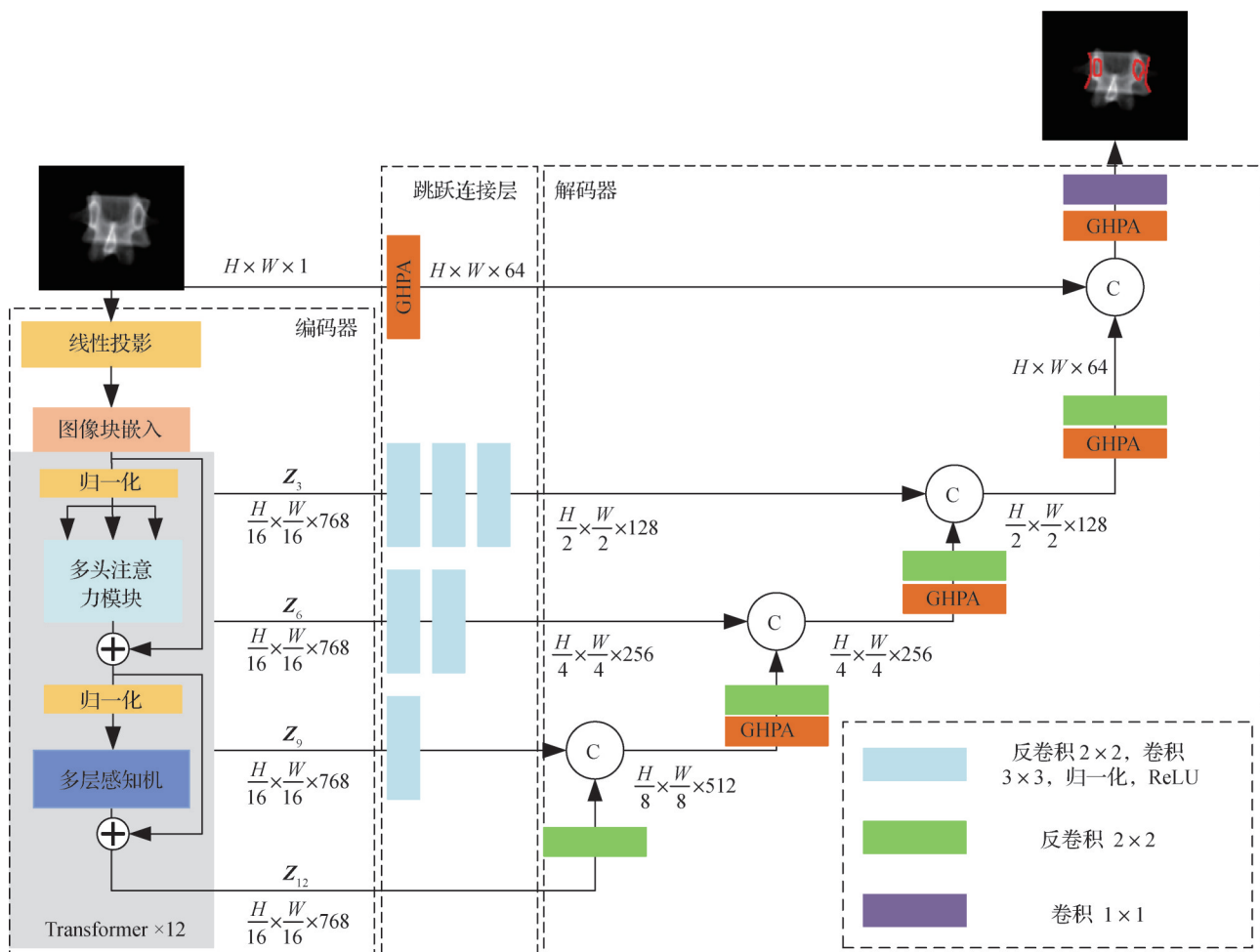


图4 SCE-UNETR网络结构

Fig. 4 The structure of SCE-UNETR

跳跃连接层将Transformer模块输出的特征图分辨率提高2倍;然后,将调整大小后的特征图与先前Transformer模块输出的特征图合并(例如, Z_6 与 Z_9 合并),并将它们输送到GHPA模块中提取不同表征的信息,再使用反卷积层进行上采样。重复此过程,直到恢复原始的分辨率。通过4层跳跃连接层,将特征图的分辨率从 $\frac{H}{P} \times \frac{W}{P}$ 恢复到 $H \times W$ 的全分辨率;最后,利用带有softmax激活函数的 1×1 卷积层将解码器输出的特征图转换为对每个像素点的分割预测。

1.2.2 高效哈达玛乘积注意力模块

为了减少参数量、保证2D/3D配准效率,本文将SCE-UNETR网络中跳跃连接层的第1层和解码器中的卷积操作替换为基于哈达玛乘积注意力机制(Hadamard product attention, HPA)(Zhou等,2021)的GHPA模块,如图5所示。

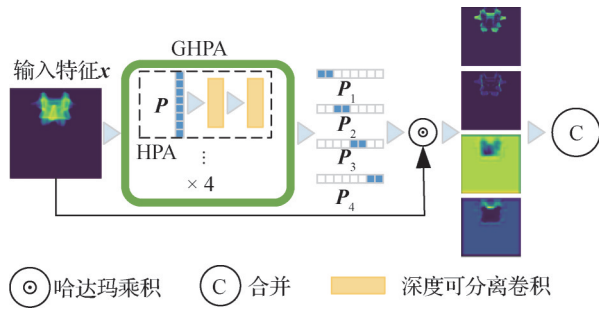


图5 GHPA模块结构

Fig. 5 The structure of the GHPA module

哈达玛乘积本质是逐元素乘积,相比于多头自注意力机制的二次方复杂度,GHPA只有线性复杂度。GHPA模块由4组HPA模块组成,将输入特征进行分组,对不同组进行HPA操作,从多个视角提取图像信息。GHPA的基本原理如下:首先,对于输入特征 x 和一个随机初始化的可学习权重 p ,使用双线性插值将 p 调整到与 x 相匹配的维度;然后, p 经过两层深度可分离卷积(Howard等,2017)操作进一步降低参数量并提取特征;最后将 x 和 p 进行哈达玛乘积操作,同时将4组特征合并以获得输出。

1.3 语义边缘提取的损失函数

本文设计了损失函数。该损失函数共包含Dice损失与椎体间距约束(spacing constraint, SC)损失项。具体为

$$L_{\text{total}} = L_{\text{Dice}} + L_{\text{SC}} \quad (4)$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i g_i}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^N g_i^2} \quad (5)$$

$$L_{\text{SC}} = \frac{1}{n} \sum_{i=1}^N (d_{\text{gt}} - d_{\text{pred}})^2 \quad (6)$$

式中, p_i 和 g_i 分别表示第 i 个像素的预测值和真实标签值, N 表示矩阵中像素值的数量。 d_{gt} 表示间距的真实值,而 d_{pred} 根据预测出的多椎体标签的间距得出。

本文设计Dice损失是为了提高预测语义边缘与真实语义边缘的相似性;设计SC损失是充分考虑相邻两节椎骨之间的语义边缘的包围盒空间距离要和目标距离保持一致。通过计算相邻椎骨语义边缘包围盒中心点之间的高度差来表达语义边缘的包围盒空间距离,具体如图6所示。

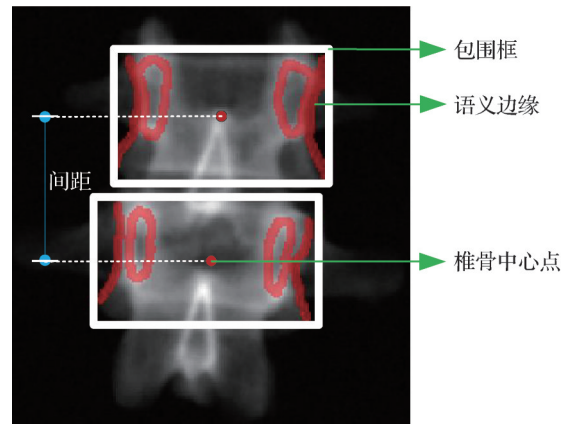


图6 计算椎骨中心的间距

Fig. 6 The spacing of the vertebral centers

2 实验及结果分析

2.1 实验流程与数据预处理

实验流程分为用于语义边缘提取的SCE-UNETR网络训练阶段和基于椎骨语义边缘的2D/3D配准阶段,具体如下:

1) 语义边缘提取网络(SCE-UNETR)训练阶段。本文利用11例CT影像通过随机变换CT的位姿生成210幅正位DRR图像;然后,将带语义边缘标签的DRR图像按4:1的比例,随机分为数量为168幅的训练集和数量为42幅的测试集;最后,将图像输入网络进行训练。SCE-UNETR网络训练超参数设置如下:网络训练的epoch数为100次,学习率的初始

值为0.1,每隔20个epoch衰减0.5。为了减少不同病人图像之间像素强度的变化,所有图像在输入网络前利用Z-Score进行归一化,使像素强度的尺度在某一范围。

2)基于椎骨语义边缘的2D/3D配准阶段。分别利用模拟数据与真实数据评估本文方法。模拟数据采用10例CT影像以及CT影像生成的正位DRR影像作为配套CT与模拟术中X光片;真实数据采用一组同一个病人的CT影像与一张术中正位X线片配套数据。其中,模拟数据生成具体如下:参考位姿(ground truth pose)是生成DRR图像所用的相机位姿;每例CT影像经过-10 mm与10 mm随机平移变换、-5°到5°分别绕 x 、 y 和 z 轴随机旋转变换。配准前需对数据进行预处理,如图7所示,预处理流程如下:首先,根据椎骨掩码提取出待配准椎骨的CT影像;然后,利用CT影像生成DRR图像;同时,为了模拟术中X线片图像的质量,本文对DRR图像添加了高斯噪声并调整CT图像的窗宽窗位;最后,将图像都缩小到 256×256 像素以提高配准效率。

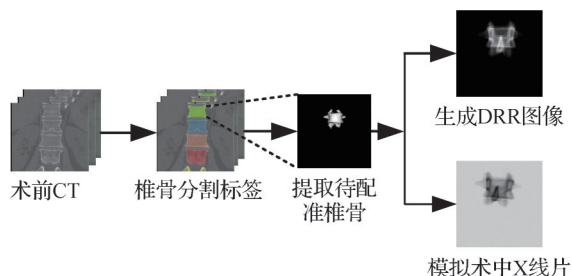


图7 2D/3D配准阶段预处理流程中椎骨提取与DRR图像生成

Fig. 7 The preprocess of 2D/3D registration involves extracting vertebral bones and generating DRR images

实验过程中所有临床影像数据来自常州市第一人民医院。CT影像的具体参数为:层厚0.5 mm,尺寸为 $512 \times 512 \times 601$ 像素。X线片尺寸为 512×512 像素,分辨率为0.3 mm,并提供了CT影像的椎骨掩码标签。此外,由放射科医生对生成的DRR图像进行语义边缘标签的标记,标记位置为椎弓根区域和椎骨的两侧。标注图像由另一名放射科专家进行逐一验证。

2.2 实验评估指标

本文分别对语义边缘提取和2D/3D配准两方面进行评估,具体如下:

1)语义边缘提取评估指标。利用Dice系数(Taha和Hanbury, 2015)、交并比(intersection over union, IoU)(Taha和Hanbury, 2015)、95%豪斯多夫距离(Hausdorff distance, HD)(Taha和Hanbury, 2015)和间距的平均绝对误差(mean absolute error, MAE)(Willmott和Matsuura, 2005)对模型进行评估。

2)2D/3D配准评估指标。本文从标志点误差、6自由度误差和配准耗时3个方面评估模拟数据。其中,6自由度误差利用MAE评估。标志点误差采用平均目标配准误差(mean target registration error, mTRE)(Van de Kraats等, 2005),具体为

$$mTRE(P_{CT}, T_{gt}, T_{reg}) = \frac{1}{K} \sum_{i=1}^K T_{gt} P_{CT_i} - T_{reg} P_{CT_i} \quad (7)$$

式中, T_{reg} 表示配准算法预测得到的变换矩阵, T_{gt} 表示真实的变换矩阵, P_{CT_i} 表示CT上第 i 个标志点的坐标, K 为标志点的数量,本文取10。

在真实数据上评估本文算法时,利用结构相似性(structure similarity index measure, SSIM)(Wang等, 2004)和归一化互相关(normalized cross-correlation, NCC)(麦永锋等, 2021)两种指标进行评估。

2.3 实验结果

本节分别从语义边缘提取、粗配准、方法对比、配准精度和配准时间等多个角度对本文方法进行定量与定性评估。

2.3.1 GHPA模块和SC损失项的有效性验证

为验证GHPA模块和SC损失项的有效性,利用基础网络UNETR,对结合不同优化方法的语义边缘提取结果进行评估。定量评估结果如表1所示,UNETR网络分别单独结合GHPA模块和SC损失项后,提取精度都有一定的提升,而结合GHPA模块网络参数量下降了约一半。在同时加入GHPA模块和SC损失项时(即SCE-UNETR),提取精度最高,且椎骨间距误差最小。

实验表明了GHPA模块和SC损失项能够降低网络参数量并提升网络提取语义边缘的精度。GHPA模块利用哈达玛乘积提取特征的同时,结合可分离深度卷积层使得输入特征通道数与输入通道数保持一致,该过程无参数量增加,并利用多头操作机制保证了特征信息的完整。而SC损失起到了正则项的作用,在网络提取多椎骨语义边缘时,对各个椎骨的语义边缘的间距进行了约束,提高了多椎体语义边缘提取的精度。

表1 不同优化方法的定量评估实验结果

Table 1 Experimental results of quantitative evaluation of different optimization methods

网络	模块	损失函数	平均 Dice/%	间距误差/mm	参数量/ 10^6
UNETR	无	Dice	89.18±5.54	2.88±3.55	672.49
	GHPA	Dice	90.52±1.29	2.13±2.02	348.46
	无	Dice+SC	90.08±1.68	1.57±0.92	672.49
SCE-UNETR(本文)	GHPA	Dice+SC	90.57±2.56	1.32±0.71	348.46

注:加粗字体表示各列最优结果。

2.3.2 基于不同网络架构的语义边缘提取的评估

1)定量评估。本文将 SCE-UNETR 网络与 SwinUNETR(Hatamizadeh 等,2021)、TransUNet 和 UNETR 进行对比,定量评估结果如表 2 所示,结论如下: SCE-UNETR 的模型参数量约是其他网络的一半,但

在分割重合度的指标平均 Dice、平均交并比(IoU)和平均 95% HD 方面,SCE-UNETR 网络语义边缘的提取精度均优于其他 3 种网络。此外,所有网络在 Dice 损失的基础上结合 SC 损失项后,语义边缘提取精度均有一定提升。

表2 语义边缘提取定量评估实验结果

Table 2 Experimental results of quantitative evaluation of semantic edge extraction

网络	损失函数	平均 Dice/%	平均 IoU/%	平均 95% HD	参数量/ 10^6
SwinUNETR	Dice	90.16±1.39	83.29±2.01	6.19±8.73	647.43
	Dice+SC	90.56±1.06	83.96±4.31	6.76±0.02	
TransUNet	Dice	91.09±1.04	84.65±1.55	6.04±0.02	730.39
	Dice+SC	91.36±1.08	85.15±4.36	4.15±5.66	
UNETR	Dice	90.62±1.66	84.00±2.43	6.05±7.89	672.49
	Dice+SC	91.99±1.28	86.11±3.60	5.91±0.02	
SCE-UNETR(本文)	Dice	91.13±1.26	84.73±1.88	3.75±5.95	348.46
	Dice+SC	92.37±1.28	86.67±2.83	2.21±0.02	

注:加粗字体表示各列最优结果。

2)定性评估。定性评估结果如图 8 所示,结论如下:在所有网络都利用 Dice 结合 SC 损失项的情况下,SwinUNETR 对于椎骨边缘不敏感,以及椎弓根和椎骨边缘连接处提取得不精细,TransUNet 提取的边缘都较粗,多处椎弓根和椎骨边缘区域存在重叠现象,本文提出的 SCE-UNETR 比基础网络 UNETR 更精准,并未出现语义边缘重叠和断裂的现象。

2.3.3 2D/3D粗配准定量评估

本节从配准误差和耗时两方面,与 SwinUNETR、TransUNet 和 UNETR 3 种边缘提取网络进行对比,配准定量评估结果如表 3 所示,在单个椎体配准和多个椎体同时配准的情况下,本文方法精度最高且配准耗时最短。

其他 3 种网络的本质是 ViT 网络的变体,该类型网络需要额外的位置编码来处理图像的空间结构,这大大增加了 ViT 网络的复杂性和参数量(Dosovitskiy 等,2020)。本文对 SCE-UNETR 网络进行了轻量化和提取性能的改进,提高了语义边缘的预测效率和精度,从而使得 2D/3D 粗配准更加高效和精准。

2.3.4 与其他配准方法对比

本节在模拟数据的单椎体和多椎体配准应用中,分别进行粗配准评估与精配准评估。在粗配准和精配准两个阶段,本文方法分别与基于标志点方法、基于物理边缘提取(Canny 算子提取)方法进行了综合对比。各方法平移误差与旋转误差的评估结果如图 9 与表 4 所示。利用本文方法进行粗配准结

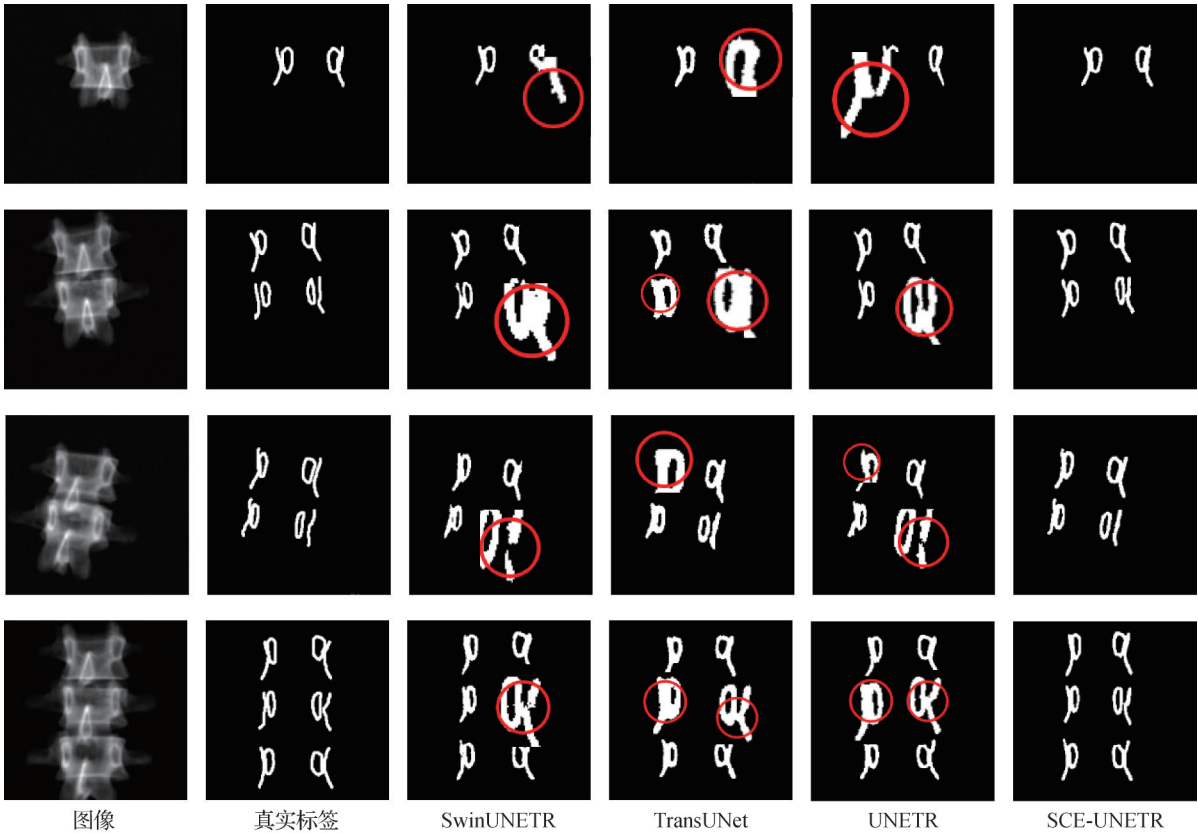


图8 椎骨语义边缘提取定性评估实验结果

Fig. 8 The experimental results of qualitative evaluation of vertebra semantic edge extraction

表3 不同提取网络对2D/3D粗配准影响的定量评估结果

Table 3 Results of quantitative evaluation of different extraction networks on 2D/3D coarse registration

网络	椎体	平均误差		mTRE	耗时/s
		平移/mm	旋转/(°)		
SwinUNETR	单	4.18±3.97	0.63±0.60	2.12±1.28	3.84±0.08
	多	4.11±4.23	0.93±0.80	4.34±1.53	9.34±0.73
TransUNet	单	3.71±4.12	0.49±0.55	2.82±0.89	3.39±0.32
	多	3.80±4.92	0.83±0.72	4.17±2.00	8.77±0.30
UNETR	单	3.13±3.54	0.51±0.40	1.79±1.62	4.46±0.91
	多	4.58±5.55	0.89±0.72	2.64±0.98	9.15±0.39
SCE-UNETR (本文)	单	1.70±0.96	0.16±0.02	1.12±0.88	2.88±0.09
	多	1.15±0.53	0.14±0.03	0.44±0.27	5.76±0.10

注:加粗字体表示各列最优结果。

合灰度法精配准的定性结果如图10所示。本文分别实验了单椎体和多椎体的应用场景,绿色轮廓表示椎体的术前CT体数据,背景为模拟术中X线片图像,图中展示了术前CT体数据经2D/3D配准后的最终位姿,可以观察到术前CT体数据基本都与术中X

线片图像对齐,椎骨的边缘和横突基本都已贴合,也表明配准完成。结论如下:在粗配准和精配准两个阶段中,本文方法在精度与效率方面均优于标志点方法与基于物理边缘提取的配准方法。

实验表明了基于语义边缘2D/3D粗配准方法相

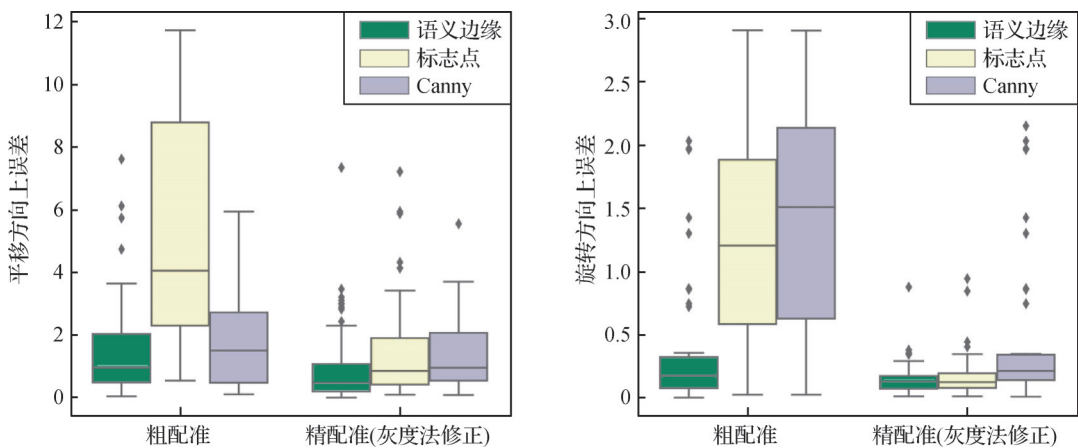


图9 不同配准方法的误差分布对比

Fig. 9 Comparison of the error distribution among different registration methods

表4 本文方法与基于标志点、基于物理边缘提取的2D/3D配准方法的对比

Table 4 Comparison of the proposed method with landmark points-based and physical edge extraction-based 2D/3D registration methods

方法	椎体	平均误差		mTRE	耗时/s
		平移/mm	旋转/(°)		
粗配准	标志点	单	4.33±3.49	1.31±0.87	2.62±1.86
		多	5.77±3.40	1.11±0.88	1.84±1.27
	Canny	单	4.62±0.93	2.17±0.50	1.79±1.62
		多	3.13±0.23	2.14±0.51	2.64±0.98
	语义边缘(本文)	单	1.70±0.96	0.16±0.02	1.47±0.70
		多	1.15±0.53	0.14±0.03	1.29±0.67
粗配准+精配准	标志点+灰度	单	1.75±1.02	0.16±0.02	1.44±0.66
		多	1.22±0.43	0.17±0.05	0.76±0.50
	Canny+灰度	单	1.38±0.24	0.20±0.03	1.34±0.63
		多	2.89±0.75	1.08±1.51	1.64±0.46
	语义边缘+灰度(本文)	单	0.96±0.52	0.08±0.07	1.12±0.88
		多	0.92±0.38	0.10±0.02	0.44±0.27

注:加粗字体表示各组各列最优结果。

比于其他方法在精度和效率上有一定的优势。基于标志点方法存在空间信息匮乏和计算量大等问题,导致粗配准精度低;基于物理边缘提取的方法会提取出冗余的图像特征,造成目标函数陷入局部最优,难以收敛。而基于语义边缘粗配准方法利用椎骨中显著的边缘信息进行配准,这种特征相比于标志点具有结构信息和显著的解剖特征,比Canny算子提取出的复杂边缘信息更简洁,从而提高了配准的精

度和效率。

在模拟数据的单椎体配准中,将本文方法与诸如梯度法(Grupp等,2020b)、Pro-ST(Gao等,2020)和自监督回归法(a patient-specific self-supervised, PSSS)(Zhang等,2023)3种当前性能较好的方法进行了对比,结果如表5所示。本文方法配准精度上优于其他3种配准方法,且耗时最少。

梯度法能充分利用图像中的灰度信息以得到精

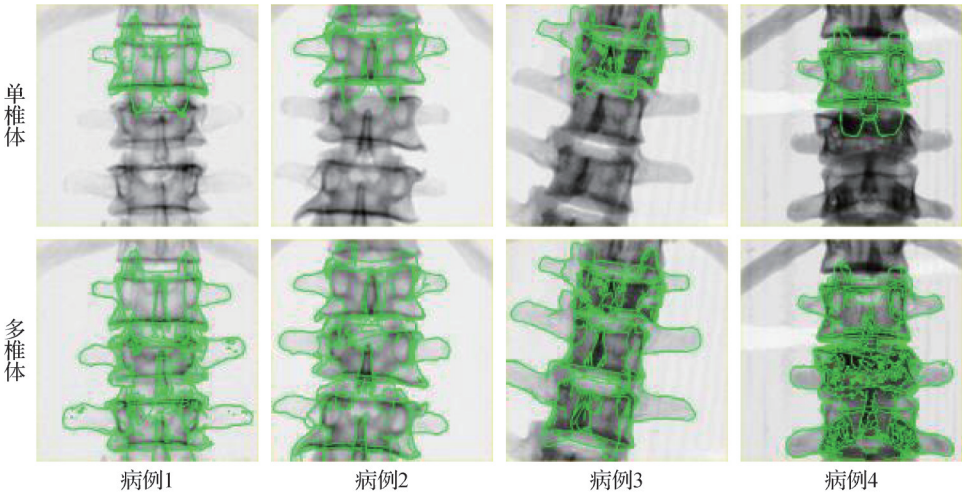


图10 本文方法位姿修正后的配准精度定性评估(绿色为DRR图像的椎骨轮廓)

Fig. 10 The qualitative assessment of registration accuracy after pose refinement of the proposed method
(vertebral contours of the DRR images in green)

表5 对比其他现有方法的评估结果

Table 5 Evaluation results against other existing methods

方法	平均误差		耗时/s
	平移/mm	旋转/(°)	
梯度法	1.75±1.02	0.16±0.02	5.46
Pro-ST	7.83±19.8	4.94±8.78	> 60.0
PSSS	5.23±2.01	0.61±0.44	11.36
本文方法	0.96±0.52	0.08±0.07	4.91

注:加粗字体表示各列最优结果。

确的变换位姿,但依赖于合适的初始位姿。Pro-ST利用卷积神经网络提取出图像中更高维的特征以进行粗配,但神经网络巨大的参数量增加了配准的耗时。PSSS利用直接回归法进行粗配,能够为精配提供较合适的初始位姿,但需要针对性地训练神经网络模型,泛化性较差。此外,这3种方法主要应用于盆骨影像中,对于腰椎这种解剖区域较小、灰度信息少且视野域较大的应用效果较差。本文方法针对椎骨区域的2D/3D配准进行了改进,利用自动提取的语义边缘先进行粗配,克服了视野域大难以获得合适初始位姿的问题,且语义边缘已包含足够的解剖特征和结构信息,提高了粗配准的精度,从而减少了后续精配准的搜索空间,提高了配准效率。

2.3.5 真实X线片上的评估

本节利用基于语义边缘粗配准结合灰度法在真实数据上进行了评估,分别评估了腰椎第4节(L4)和第5节(L5)的配准结果。真实数据上的定量评估

结果如表6所示,本文方法在单椎体配准和多椎体配准应用中表现稳定。

表6 真实X线片的配准评估结果

Table 6 Registration assessment results of the real X-rays

椎体	SSIM/%	NCC
L4(单椎体)	78.96	0.63
L5(单椎体)	79.16	0.49
L4与L5(多椎体)	84.23	0.51

真实数据上的定性评估结果如图11所示,用本文方法配准后,CT影像投影得到的DRR图像的椎骨轮廓(绿色)与真实X线片(固定不动)的椎骨轮廓基本重合。真实X线片存在噪声大、分辨率低和受植入物干扰等问题,本文方法利用语义边缘粗配准可降低图像质量对配准精度的影响。本文方法在粗配准过程中,只关注椎骨边缘并将椎骨的语义边缘对齐后即可得到合适的初始位姿。因此,在真实X线片中依旧能取得较准确的配准结果。

3 结 论

本文提出一种基于椎骨语义边缘的2D/3D粗配准方法,该方法充分结合了医学知识,利用椎骨的椎弓根和椎骨边缘作为配准特征,相比基于标志点和基于物理边缘提取的配准方法,能有效缩小后续精配准过程的搜索空间,从而提高配准精度。在边缘

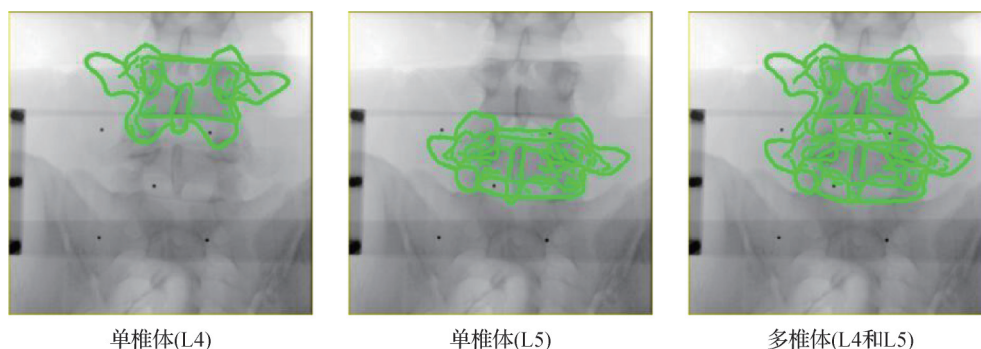


图 11 本文方法在真实 X 线片上的 2D/3D 配准结果(绿色为 DRR 图像的椎骨轮廓)

Fig. 11 The 2D/3D registration results of the proposed method on real X-rays (vertebral contours of the DRR images in green)

提取方面,本文提出了将哈达玛乘积代替卷积操作的方式,使得二次复杂性变成线性复杂性,有效降低了运算量,从而提高了 SCE-UNETR 网络的提取效率,因此,本文方法能够较好地满足 2D/3D 配准的精度与实时性需求。

本文方法存在以下两个问题:一方面,从粗配准到精配准修正位姿不是端到端网络;另一方面,由于术中 X 线片不易获取,采用 DRR 图像训练的椎骨语义边缘计算深度模型,但 DRR 图像与真实 X 线片之间存在域间隙(domain gap),这导致本文方法在真实 X 线片上的配准精度有一定程度下降。今后需要对以上问题作进一步改进。

参考文献(References)

- Canny J F. 1983. Finding edges and lines in images [EB/OL]. [2024-01-02]. <http://hdl.handle.net/1721.1/6939>
- Chen Z Q, Wang Z W, Fang L W, Jian F Z, Wu Y H, Li S and He H G. 2018. Real-time 2D/3D registration of vertebra via machine learning and geometric transformation. *Acta Automatica Sinica*, 44(7): 1183-1194 (陈智强, 王作伟, 方龙伟, 营凤增, 吴毅红, 李硕, 何晖光. 2018. 基于机器学习和几何变换的实时 2D/3D 脊椎配准. *自动化学报*, 44(7): 1183-1194) [DOI: 10.16383/j.aas.2017.c160711]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houselby N. 2020. An image is worth 16 × 16 words: transformers for image recognition at scale [EB/OL]. [2024-01-02]. <https://arxiv.org/pdf/2010.11929.pdf>
- Du X G, Dang J W, Wang Y P. 2015. Generation algorithm of digital reconstruction radiographs based on CUDA. *Computer Science*, 42(2): 301-305 (杜晓刚, 党建武, 王阳萍. 2015. 基于 CUDA 的数字重建影像生成算法. *计算机科学*, 42(2): 301-305) [DOI: 10.11896/j.issn.1002-137X.2015.2.064]
- Galibarov P E, Prendergast P J and Lennon A B. 2010. A method to reconstruct patient-specific proximal femur surface models from planar pre-operative radiographs. *Medical Engineering and Physics*, 32(10): 1180-1188 [DOI: 10.1016/j.medengphy.2010.08.009]
- Gao C, Liu X T, Gu W H, Killeen B, Armand M, Taylor R and Unberath M. 2020. Generalizing spatial transformers to projective geometry with applications to 2D/3D registration//*Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention*. Lima, Peru: Springer: 329-339 [DOI: 10.1007/978-3-030-59716-0_32]
- Gollmer S T, Lachner R and Buzug T M. 2007. Registration algorithm for statistical bone shape reconstruction from radiographs-an accuracy study//*Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Lyon, France: IEEE: 6375-6378 [DOI: 10.1109/IEMBS.2007.4353814]
- Grupp R B, Hegeman R A, Murphy R J, Alexander C P, Otake Y, McArthur B A, Armand M, and Taylor R H. 2020a. Pose estimation of periacetabular osteotomy fragments with intraoperative X-ray navigation. *IEEE Transactions on Biomedical Engineering*, 67(2): 441-452 [DOI: 10.1109/TBME.2019.2915165]
- Grupp R B, Unberath M, Gao C, Hegeman R A, Murphy R J, Alexander C P, Otake Y, McArthur B A, Armand M and Taylor R H. 2020b. Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration. *International Journal of Computer Assisted Radiology and Surgery*, 15(5): 759-769 [DOI: 10.1007/s11548-020-02162-7]
- Hatamizadeh A, Nath V, Tang Y C, Yang D, Roth H R and Xu D G. 2022. Swin UNETR: swin transformers for semantic segmentation of brain tumors in MRI images//*International MICCAI Brainlesion Workshop*. Springer: 272-284 [DOI: 10.1007/978-3-031-08999-2_22]
- Hatamizadeh A, Tang Y C, Nath V, Yang D, Myronenko A, Landman B, Roth H R and Xu D G. 2021. Unetr: Transformers for 3D medical image segmentation [EB/OL]. [2024-01-02]. <https://arxiv.org/pdf/2103.10504.pdf>

- Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2024-01-02]. <https://arxiv.org/pdf/1704.04861.pdf>
- Huang Y F, Li S and Zhao J. 2006. The registration of 2D/3D images based on external markers. *Journal of Shanghai Jiaotong University*, 40(7): 1108-1111. (黄永锋, 李杉, 赵俊. 2006. 基于外标记点的2D/3D图像配准方法. *上海交通大学学报*, 40(7): 1108-1111) [DOI: 10.16183/j.cnki.jsjtu.2006.07.009]
- Jaganathan S, Kukla M, Wang J, Shetty K and Maier A. 2023. Self-supervised 2D/3D registration for X-ray to CT image fusion//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 2787-2797 [DOI: 10.1109/WACV56688.2023.00281]
- Jaganathan S, Wang J, Borsdorf A, Shetty K and Maier A. 2021. Deep iterative 2D/3D registration//*Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention*. Strasbourg, France: Springer: 383-392 [DOI: 10.1007/978-3-030-87202-1_37]
- Jang J S, Kim J I, Ku B and Lee J H. 2023. Reliability analysis of vertebral landmark labelling on lumbar spine X-ray images. *Diagnostics*, 13(8): #1411 [DOI: 10.3390/diagnostics13081411]
- Kadoury S, Cheriet F and Labelle H. 2009. Personalized X-ray 3-D reconstruction of the scoliotic spine from hybrid statistical and image-based models. *IEEE Transactions on Medical Imaging*, 28(9): 1422-1435 [DOI: 10.1109/TMI.2009.2016756]
- Liu K, Lyu X Q, Gu Y, Yu H F, Ren G Y and Zhang M. 2016. The 2D/3D medical image registration algorithm based on rapid digital image reconstruction. *Journal of Image and Graphics*, 21(1): 69-77 (刘坤, 吕晓琪, 谷宇, 于荷峰, 任国印, 张明. 2016. 快速数字影像重建的2维/3维医学图像配准. *中国图象图形学报*, 21(1): 69-77) [DOI: 10.11834/jig.20160109]
- Lu Y H, Zheng K, Li W J, Wang Y R, Harrison A P, Lin C H, Wang S, Xiao J, Lu L, Kuo C F and Miao S. 2020. Contour transformer network for one-shot segmentation of anatomical structures. *IEEE Transactions on Medical Imaging*, 40(10): 2672-2684 [DOI: 10.1109/TMI.2020.3043375]
- Mai Y F, Sun Q C, Jia P F and Chen X J. 2021. 2D/3D medical image registration algorithm based on DRR and the similarity measure. *Beijing Biomedical Engineering*, 40(3): 263-272 (麦永锋, 孙启昌, 贾鹏飞, 陈晓军. 2021. 基于DRR及相似性测度的2D-3D医学图像配准算法. *北京生物医学工程*, 40(3): 263-272) [DOI: 10.3969/j.issn.1002-3208.2021.03.008]
- Miao S, Wang Z J and Liao R. 2016. A CNN regression approach for real-time 2D/3D registration. *IEEE Transactions on Medical Imaging*, 35(5): 1352-1363 [DOI: 10.1109/TMI.2016.2521800]
- Naik R R, Anitha H, Bhat S N, Ampar N and Kundangar R. 2022. Realistic C-arm to pCT registration for vertebral localization in spine surgery: a hybrid 3D-2D registration framework for intraoperative vertebral pose estimation. *Medical and Biological Engineering and Computing*, 60(8): 2271-2289 [DOI: 10.1007/s11517-022-02600-5]
- Netter F H. 2014. *Atlas of Human Anatomy*. 6th ed. Elsevier Health Sciences Press.
- Otake Y, Armand M, Armiger R S, Kutzer M D, Basafa E, Kazanides P and Taylor R H. 2011. Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: incorporation of fiducial-based C-arm tracking and GPU-acceleration. *IEEE Transactions on Medical Imaging*, 31(4): 948-962 [DOI: 10.1109/TMI.2011.2176555]
- Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Ruan J C, Xie M Y, Gao J S, Liu T and Fu Y Z. 2023. EGE-UNet: an efficient group enhanced UNet for skin lesion segmentation//*Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention*. Vancouver, Canada: Springer: 481-490 [DOI: 10.1007/978-3-031-43901-8_46]
- Taha A A and Hanbury A. 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(1): #29 [DOI: 10.1186/s12880-015-0068-x]
- Unberath M, Gao C, Hu Y C, Judish M, Taylor R H, Armand M and Grupp R. 2021. The impact of machine learning on 2D/3D registration for image-guided interventions: a systematic review and perspective. *Frontiers in Robotics and AI*, 8: #716007 [DOI: 10.3389/frobt.2021.716007]
- Van De Kraats E B, Penney G P, Tomazevic D, Van Walsum T and Niessen W J. 2005. Standardized evaluation methodology for 2-D-3-D registration. *IEEE Transactions on Medical Imaging*, 24(9): 1177-1189 [DOI: 10.1109/TMI.2005.853240]
- Wang Q Z, Huang W, Zhang Y, Li X Y, Ye X J and Hu K. 2023. Automatic segmentation of nasopharyngeal carcinoma in CT images using dual attention and edge detection//*Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece: IEEE: 1-5 [DOI: 10.1109/ICASSP49357.2023.10095233]
- Wang Y Y, Dong F, Shang L N, Zhang C and Xia S R. 2023. Research on the IVUS border detection method based on improved transunet. *Chinese Journal of Biomedical Engineering*, 42(1): 41-50 (王媛媛, 董芳, 尚丽娜, 张翠, 夏顺仁. 2023. 基于改进TransUnet网络的血管内超声图像边界提取方法研究. *中国生物医学工程学报*, 42(1): 41-50) [DOI: 10.3969/j.issn.0258-8021.2023.01.005]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image qual-

- ity assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 (4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Willmott C J and Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30 (1): 79-82 [DOI: 10.3354/cr030079]
- Xu M M. 2018. Effect of minimally invasive surgery and traditional open surgery on spinal trauma. *Medical Innovation of China*, 15 (34): 122-125 (徐敏铭. 2018. 脊柱创伤患者采用传统开放手术和微创手术治疗的对比研究. *中国医学创新*, 15 (34): 122-125) [DOI: 10.3969/j.issn.1674-4985.2018.34.033]
- Yang H, Fan B B and Guo L L. 2020. Anchor-free object detection with mask attention. *EURASIP Journal on Image and Video Processing*, 2020, (1): 1-17 [DOI: 10.1186/s13640-020-00517-3]
- Zeng Z H, Fan C D, Xiao L Y and Qu X L. 2022. DEA-UNet: a dense-edge-attention UNet architecture for medical image segmentation. *Journal of Electronic Imaging*, 31 (4): #043032 [DOI: 10.1117/1.Jei.31.4.043032]
- Zhang B C, Faghihroohi S, Azampour M F, Liu S T, Ghotbi R, Schunkert H and Navab N. 2023. A patient-specific self-supervised model for automatic X-ray/CT registration//*Proceedings of the 26th International Conference on Medical Image Computing and Computer Assisted Intervention*. Vancouver, Canada: Springer: 515-524 [DOI: 10.1007/978-3-031-43996-4_49]
- Zhang Z J, Fu H Z, Dai H, Shen J B, Pang Y W and Shao L. 2019. ET-net: a generic edge-attention guidance network for medical image segmentation//*Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention*. Shenzhen, China: Springer: 442-450 [DOI: 10.1007/978-3-030-32239-7_49]
- Zhao Y A, Lyu W Y, Xu S L, Wei J M, Wang G Z, Dang Q Q, Liu Y and Chen J. 2024. Detsr beat YOLOs on real-time object detection//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 16965-16974 [DOI: 10.1109/CVPR52733.2024.01605]
- Zhou J K, Wang P C, Wang F, Liu Q, Li H and Jin R. 2021. Elsa: Enhanced local self-attention for vision transformer [EB/OL]. [2024-01-02]. <https://arxiv.org/pdf/2112.12786.pdf>

作者简介

沈傲,男,博士研究生,主要研究方向为医学图形图像处理。

E-mail: 220220030001@hhu.edu.cn

蒋俊锋,通信作者,男,副教授,主要研究方向为计算机辅助设计与计算机图形学。E-mail: jiangjf@hhu.edu.cn

沈燕进,女,主管技师,主要研究方向为肌骨影像。

E-mail: syj1423@czfph.com

陈正鸣,男,教授,主要研究方向为计算机辅助设计与计算机图形学。E-mail: zmchen@hhu.edu.cn

黄瑞,男,副教授,主要研究方向为数字化设计与智能制造。

E-mail: huangrui@hhu.edu.cn

何坤金,男,教授,主要研究方向为计算机辅助设计和几何建模。E-mail: 19971456@hhu.edu.cn

陈杰,男,主任医师,主要研究方向为肌骨影像。

E-mail: chenjie2007@czfph.com