Automated and Fully Differentiable Image Registration boosted by Deep Learning Segmentation Mask

Jinhao Wang[1], Florian Vogl[1], Saša Ćuković[1], William Taylor[1]
[1]Laboratory of Movement Biomechanics, Institute for Biomechanics, ETH Zurich, Switzerland
Email: jinhao.wang@hest.ethz.ch

Summary

Traditional 2D/3D image registration suffers from a limited capture range and high sensitivity to noise, particularly in natural bone registration [1]. This study presents a fully differentiable 2D/3D registration framework, integrating volumetric rendering and deep learning-based segmentation masks. By focusing the similarity metric on anatomically relevant regions, segmentation masks significantly enhance registration accuracy and convergence stability. Experimental results show that masks reduce translation and rotation errors by more than 50%, increase the capture range by 30%, and improve success rates by up to 25%. The framework achieves GPU-accelerated convergence within 20–30 seconds per frame, making it a robust and efficient solution for clinical applications.

Introduction

2D/3D image registration is essential for accurate kinematic analysis, but traditional methods struggle with limited capture range, slow convergence, and sensitivity to image artifacts. Previous work using edge-based similarity metrics has demonstrated poor performance in natural bone registration due to the absence of distinct boundaries. This study proposes a fully differentiable registration pipeline, integrating PyTorch3D-based volumetric rendering with deep learning-enhanced segmentation masks, allowing gradient-based optimization and a focused similarity metric.

Methods

We incorporated a fully differentiable volumetric renderer into our 2D/3D image registration framework using PyTorch3D to simulate X-ray attenuation [2]. A gradient-based optimizer (Adam) iteratively refines the pose by minimizing the Gradient Cross-Correlation (GCC) similarity metric, enhanced with deep learning-based segmentation masks. The Deeplabv3 segmentation network extracts anatomical structures, improving robustness by focusing on relevant regions [3].

The registration process consists of coarse initialization (simulated annealing) followed by fine-tuning with Adam. Experiments were conducted with and without segmentation masks using single-plane fluoroscopic images of knee joints, evaluating accuracy, capture range, and convergence efficiency for both natural bone and implants.

Results and Discussion

Segmentation masks significantly enhanced registration accuracy, reducing femur translation errors from 1.43 mm to 0.15 mm and rotation from 2.57° to 0.81°. For tibia, errors decreased from 2.04 mm to 0.92 mm and 5.07° to 2.10°, respectively. Success rates improved from 76.7% to 94.6% (femur) and 65.8% to 83.1% (tibia) by eliminating irrelevant structures.

The capture range expanded from 8 mm/5° to 12 mm/8° (femur) and 5 mm/5° to 10 mm/5° (tibia). Loss landscape analysis showed masks smoothed optimization, reducing local minima and improving convergence. PyTorch3D-based GPU acceleration allowed registration to converge in 20–30 seconds per frame, significantly outperforming traditional gradient-free methods.

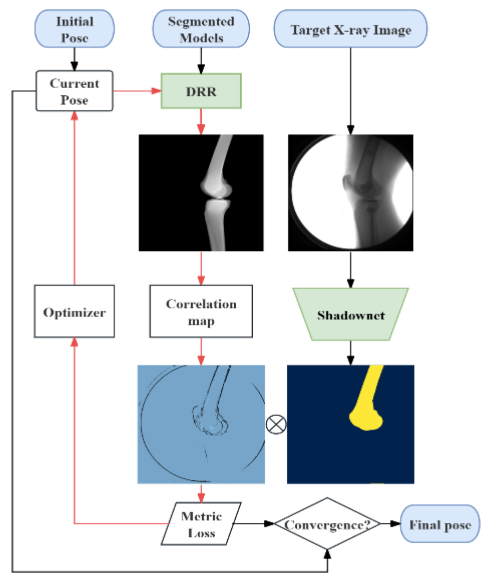

Figure 1: Schematic plot of a fully differentiable image registration using volumetric renderer and segmentation mask.

Conclusions

This study demonstrates that integrating segmentation masks into a fully differentiable 2D/3D image registration pipeline improves accuracy, robustness, and computational efficiency. Compared to traditional approaches, segmentation masks enhance similarity metric stability, reduce registration errors, and expand capture range. Future work should explore multi-plane imaging and advanced initialization strategies to further optimize tibia registration and extend the method to broader clinical applications [4].

References

[1] Postolka B et al., Med Eng Phys, 77:107–113, 2020.
[2] Liu S et al., *NeurIPS*, **32**:1–10, 2019.
[3] Vogl F et al., *PLoS ONE*, **17**:e0266094, 2022.
[4] Guan S et al., *IEEE TMI*, **35**:326-336, 2016.