Intraoperative 2D/3D Registration via Spherical Similarity Learning and Differentiable Levenberg-Marquardt Optimization

Minheng Chen[1,2]; Youyong Kong, PhD[1]

[1]Southeast University, [2]University of Texas at Arlington

WACV 2026 TUCSON, AZ 3/6 - 3/10

Abstract

- Intraoperative 2D/3D registration is a process aimed at aligning intraoperative 2D images, such as X-ray images, with corresponding preoperative CT scans. It is a crucial step in providing surgical planning guidance and navigation positioning for minimally invasive procedures. This process allows for the precise placement of implants or surgical instruments during spinal surgeries, such as percutaneous vertebroplasty and pedicle screw internal fixation.
- In literature, with the development of machine learning technology, several regression-based and landmark-based methods have been proposed, demonstrating promising performance. However, they suffer from either *the vulnerability to minor image changes*, leading to prediction instability, or *demand an extensive amount of clinically unsatisfactory and nearly impractical well-annotated data* for training. These limitations significantly hamper the clinical applicability of these methods. Recently, inspired by the traditional optimization-based methods, Gao et al. [1,2] proposed a fully differentiable framework that approximates the geodesic between two points in a special Euclidean group in three dimensions (SE(3)) by using neural networks. This framework increases the capture range of registration while being more resilient to changes in local features.
- Existing similarity learning-based frameworks' shortcomings: 1) current methods opt to approximate the manifold in Riemannian geometry within Euclidean space. However, Riemannian manifolds exhibit intricate local geometric structures. Approximating using Euclidean space leads to an inaccurate portrayal of the manifold's local structure. And *we could not determine the precise distance metric in the latent space that is best suited for approximating geodesics on the special Euclidean group*. 2) Moreover, the convergence process of existing methods tends to be relatively protracted.
- We propose a novel similarity learning–based paradigm for single-view 2D/3D registration. By learning similarity metrics in a non-Euclidean hyperspherical space with bi-invariant properties, the optimization landscape becomes smoother and less ambiguous, enabling a larger capture range and more reliable convergence. In addition, a differentiable Levenberg–Marquardt optimizer is introduced to accelerate and stabilize the registration process. Extensive experiments on pelvic, spine, and clinical datasets demonstrate state-of-the-art accuracy and robustness.

Methods and Materials

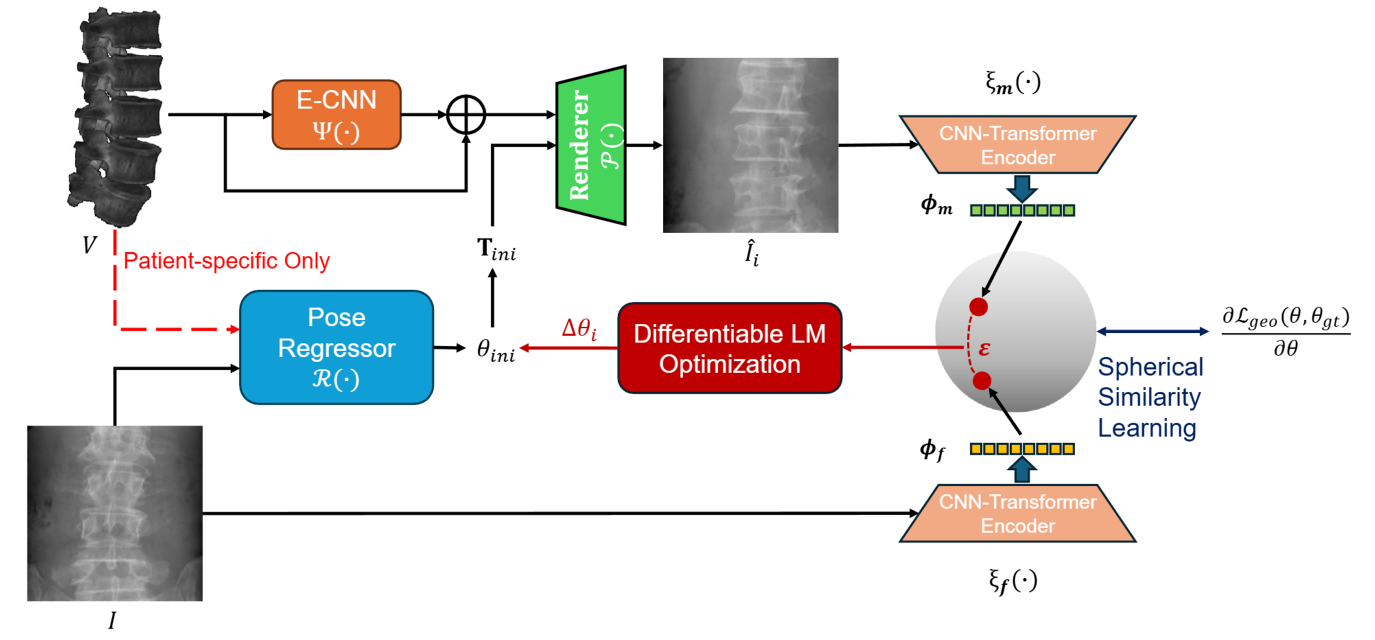


Figure 1. **Overview of the proposed framework.** We first employ a regressor to initialize the pose and then refine it using differentiable Levenberg-Marquardt optimization based on spherical similarity learning. Spherical similarity learning consists of two main components: extracting image feature representations using CNN-Transformer encoders $\xi(\cdot)$ and projecting these embeddings into hypersphere space, where the geodesic distance between them is computed as a measure of deep similarity. During training, we enforce the gradient of this deep similarity with respect to θ to approximate the gradient of the geodesic distance between θ and the ground truth θ_{gt} in SE(3).

- **Step 1 – Pose Initialization**: A lightweight pose regressor predicts an initial transformation $\theta_{ini} \in SE(3)$ from the X-ray image. A coarse alignment reduces the search space for subsequent optimization.
- **Step 2 – Equivariant DRR Rendering:** CT volume is processed by SE(3)-equivariant CNN to extract transformation-consistent features, then uses differentiable renderer to generate DRRs.
- **Step 3 – Spherical Similarity Learning**: We encode the X-ray and DRR using a CNN–Transformer encoder and map the learned features onto a hypersphere. Similarity is defined as spherical geodesic distance *1) preserves non-Euclidean geometry; 2) avoids local distortion from Euclidean metrics; 3)yields smoother and more contrastive similarity landscapes*. To further improve stability, pose gradients are approximated in SO(4) space, *producing: 1) consistent distances 2) reduced ambiguity 3)larger capture range.*
- **Step 4 – Differentiable LM Optimization**: Instead of gradient descent, we employ differentiable Levenberg–Marquardt optimization: 1) second-order updates; 2)faster convergence 3) fewer iterationsmore stable near optimum. The pose is iteratively refined by minimizing the learned spherical similarity.

◆ **Key Innovations: 1. Geometry-aware similarity learning in hyperspherical space. 2. Bi-invariant SO(4) pose embedding for stable gradients. 3. Differentiable LM for fast&robust convergence.**

Results

- We evaluate our method on three public datasets (**pelvic, spine, and clinical CBCT/angiography data**), covering both patient-specific and patient-agnostic registration scenarios. Performance is assessed using **mean target registration error (mTRE)** and **sub-millimeter success rate (SMSR)**, and compared against representative learning-based and optimization-based baselines.

	Method	SMSR	Median (mm)	Percentile (mm) 75%	Percentile (mm) 95%	Run Time
DeepFluoro	PSSS-reg [64]	56.0%	0.93	2.51	5.57	12.7 s
	PoseNet [3]	4.3%	16.6	22.0	29.2	**0.1 s**
	DFLNet [22]	36.6%	3.20	7.29	13.1	1.0 s
	SCR-reg [54]	33.3%	4.70	9.59	12.8	1.1 s
	DiffPose [18]	83.1%	0.60	0.89	1.47	5.3 s
	Ours-$\mathfrak{se}(3)$	82.8%	0.60	0.89	1.77	5.6 s
	Ours-SO(4)	**86.1%**	**0.51**	**0.85**	**1.42**	6.2 s
Ljubljana	PSSS-reg [64]	40.0%	2.48	5.87	11.3	15.3 s
	PoseNet [3]	0%	23.3	26.2	29.2	<**0.1 s**
	DiffPose [18]	80.0%	0.63	0.94	1.78	6.0 s
	Ours-$\mathfrak{se}(3)$	**85.0%**	0.57	**0.85**	1.77	6.6 s
	Ours-SO(4)	**85.0%**	**0.55**	**0.85**	**1.35**	6.5 s
CTSpine1k	PSSS-reg [64]	31.4%	4.57	9.75	15.8	16.2 s
	PoseNet [3]	9.8%	11.7	17.2	24.5	< **0.1 s**
	DFLNet [22]	28.4%	4.80	10.9	18.1	1.3 s
	SCR-reg [54]	22.2%	7.08	12.6	19.7	1.3 s
	DiffPose [18]	66.4%	0.77	1.51	3.39	7.3 s
	Ours-$\mathfrak{se}(3)$	76.5%	0.65	0.97	2.11	6.6 s
	Ours-SO(4)	**80.6%**	**0.59**	**0.93**	**1.83**	6.6 s

Method	SMSR	Median (mm)	Percentile (mm) 75%	Percentile (mm) 95%	Run Time
BOBYQA	18.5%	5.01	8.02	32.4	22.3 s
ProST-m [15]	37.6%	3.03	7.36	13.6	18.7 s
ProST-t [16]	46.3%	2.03	9.56	20.4	13.2 s
SOPI [8]	43.8%	1.99	6.28	12.5	14.2 s
CDreg [7]	50.1%	0.99	7.72	17.4	**10.1 s**
Ours-$\mathfrak{se}(3)$	53.1%	0.94	5.01	**11.4**	11.7 s
Ours-SO(4)	**55.5%**	**0.90**	**4.85**	11.9	12.5 s

	SMSR ↑	mTRE (mm) ↓
Ours-$\mathfrak{se}(3)$	53.1%	3.2 ± 3.9
Ours-SO(4)	**55.5%**	**2.1 ± 4.1**
Hyperbolic similarity	51.3%	3.5 ± 4.3
Euclidean similarity	52.8%	3.2 ± 3.9
w/o E-CNN	48.2%	5.3 ± 6.7
w/ 3D CNN	49.9%	3.6 ± 4.4

Key Findings

- **Best overall performance across all datasets** (pelvic, spine, clinical).
- Up to **+14% SMSR improvement** over prior methods on CTSpine1K.
- **Lower mTRE** and tighter error distribution (75% / 95% percentiles).
- **Faster and more stable convergence** with differentiable LM.
- Bi-invariant SO(4) metric consistently outperforms se(3) parameterization.

Generalization Performance

- **Best accuracy** under patient-agnostic setting.
- **+5% SMSR** over strongest baseline.
- More robust to large initialization errors.
- Comparable runtime to learning-based methods.

Ablation Study

- Spherical similarity outperforms Euclidean and hyperbolic metrics.
- SO(4) embedding further reduces mTRE.
- Equivariant CNN improves feature stability.
- All components contribute to final performance.

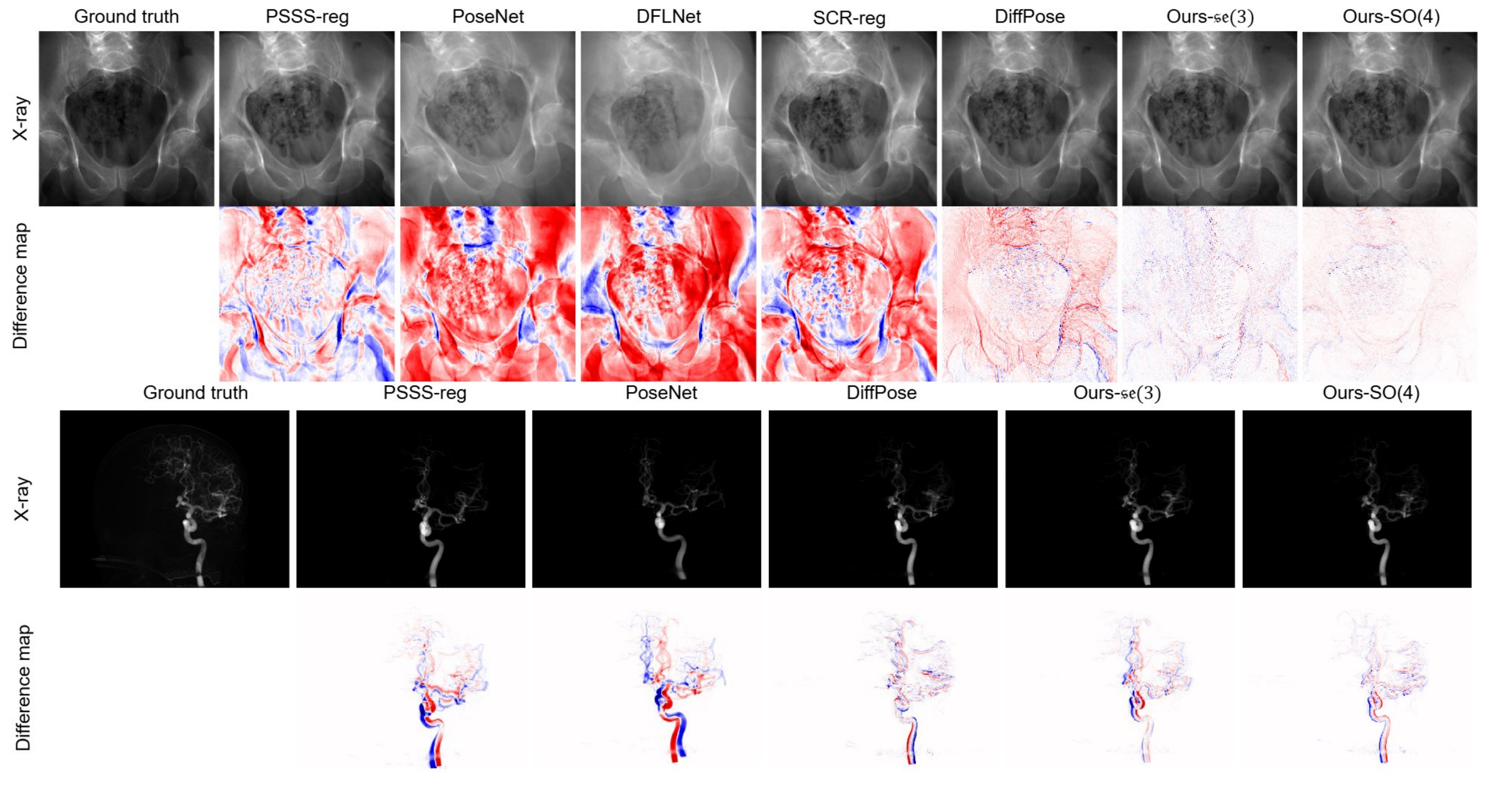
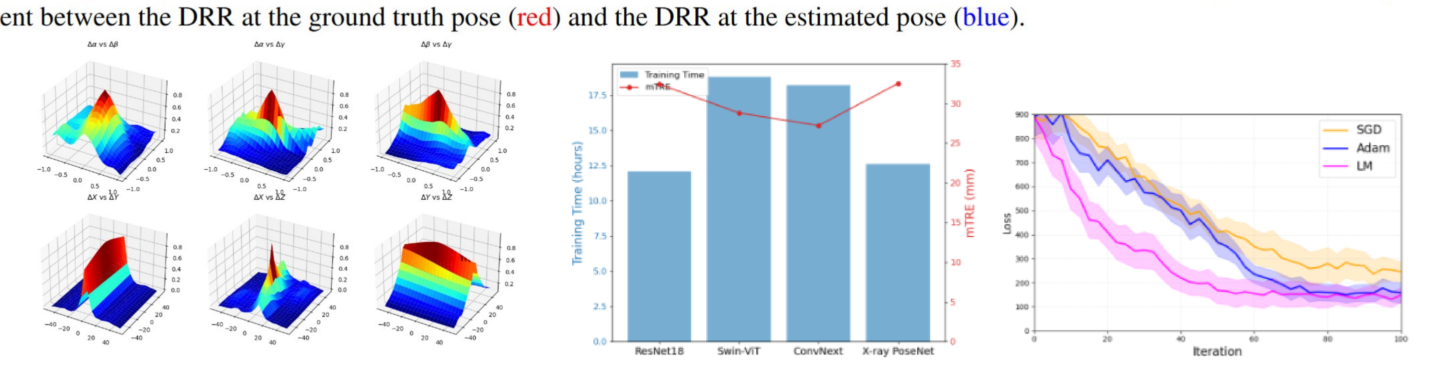


Figure 2. **Qualitative results.** We visualize the registration results of the proposed method and the baselines on the DeepFluoro and Ljubljana dataset under the patient-specific scenario. In each example, the **top** row shows the ground truth X-ray image alongside the corresponding DRR generated from the pose estimated by each method. The **bottom** row displays the difference map, demonstrating the alignment between the DRR at the ground truth pose (red) and the DRR at the estimated pose (blue).



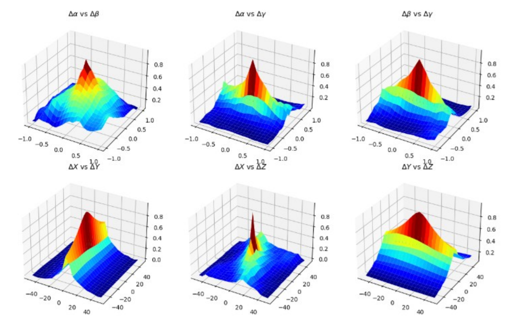


Figure 4. **Left:** comparison of training time and mTRE for position regressors with different backbone architectures.The training time reports the time when the standard deviation of the model's loss function is less than 10e-4 in the last ten epochs. **Right:** comparison of the convergence speed of the proposed framework using different gradient-based optimization methods in the inference phase.

Conclusions

- **Proposed a geometry-aware spherical similarity learning framework for single-view 2D/3D registration**
- **Introduced bi-invariant SO(4) embedding + differentiable LM for smoother and faster optimization**
- **Achieved state-of-the-art SMSR and lowest mTRE across pelvic, spine, and clinical datasets**
- **Enables robust, accurate, and radiation-efficient intraoperative registration**

Contact

Minheng Chen
Department of Computer Science and Engineering, University of Texas at Arlington
Email: mxc2442@mavs.uta.edu
Website: https://m1nhengchen.github.io

References

1. Gao, Cong, et al. "Generalizing spatial transformers to projective geometry with applications to 2D/3D registration." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer International Publishing, 2020.
2. Gao, Cong, et al. "A fully differentiable framework for 2D/3D registration and the projective spatial transformers." IEEE transactions on medical imaging 43.1 (2023): 275-285.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com