RayEmb: Arbitrary Landmark Detection in X-Ray Images Using Ray Embedding Subspace

Pragyan Shrestha, Chun Xie, Yuichi Yoshii, and Itaru Kitahara
University of Tsukuba, Tokyo Medical University

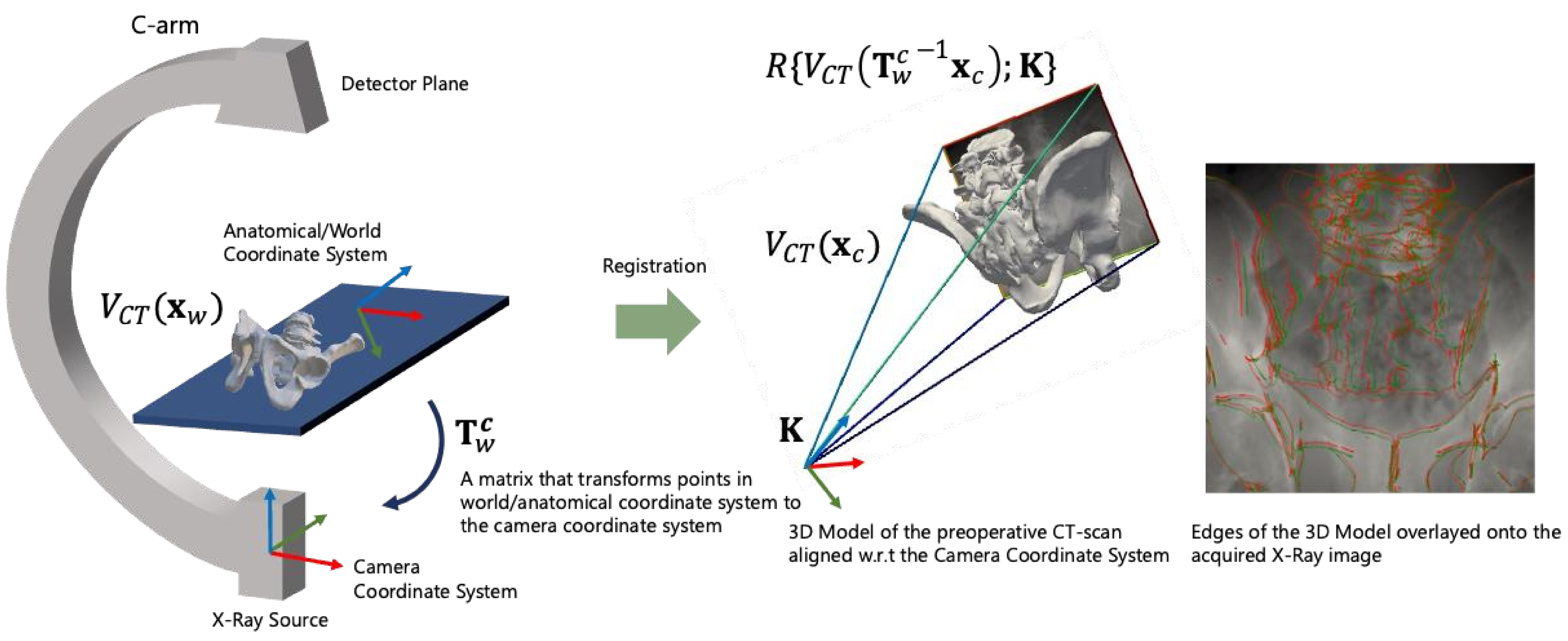
筑波大学 University of Tsukuba

東京医科大学 TOKYO MEDICAL UNIVERSITY

ACCV HANOI VIETNAM 2024 DEC 8-12

Visit our project page for interactive visualizations and code !

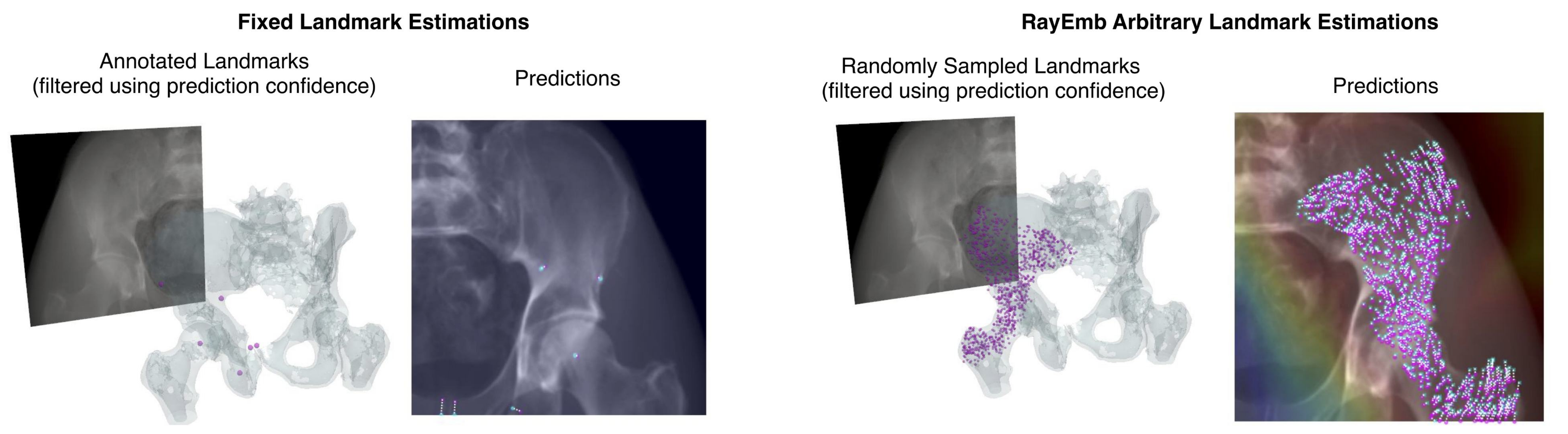
1 BACKGROUND



2D-3D Registration

The problem of estimating the **camera extrinsics / camera pose** of the imaging system, given 3D model and 2D projection Image. It is a crucial technique for precise navigation and alignment during **orthopedic surgeries.**

2 MOTIVATION



Fixed Landmark Estimation

Previous methods estimated 2D landmarks of **pre-defined fixed 3D landmarks**. In images with heavy truncation effect, the landmark estimates are not reliable for registration.

Arbitrary Landmark Estimation

We propose to estimate the 2D landmarks of **arbitrary 3D landmarks** inside the volume to mitigate the issue of non-visible landmarks.

3 METHOD OVERVIEW

Key Idea

Represent a **3D point** in a volume as a **subspace** spanned by **ray embeddings** associated with rays intersecting at the point.

Train an **image encoder** such that the ray embeddings of intersecting rays are mapped close to subspace spanned by a subset of known intersecting ray embeddings.

Subspace Representation of a 3D Landmark

Let $\mathbf{e}(\mathbf{x}) = E(I_q(\mathbf{x});\mathbf{w})$ be the ray embedding of the query input image and $\mathbf{e}'_t = E(I'_t(\mathbf{x}'_t);\mathbf{w}), t \in \{1,2,3,\ldots,N\}$ be the ray embeddings of the template images.

Stack the template ray embeddings to form a matrix. $\mathbf{F} = (\mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_N)$

Calculate projection matrix onto the subspace. $\mathbf{P} = \mathbf{F}\mathbf{F}^+ = \mathbf{U}\Sigma\Sigma^+\mathbf{U}^\mathbf{T}$

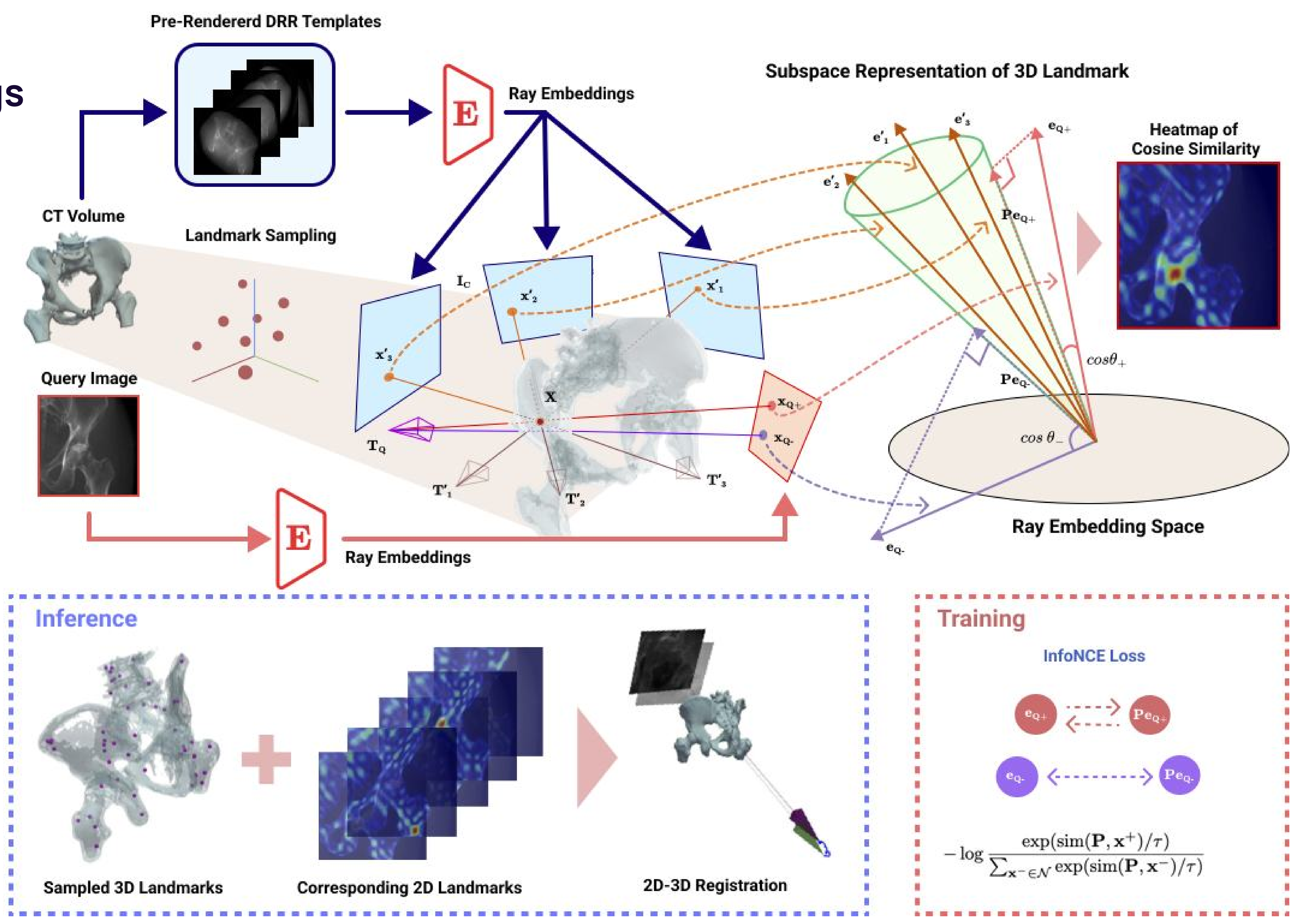
Corresponding 2D Landmark Estimation

Calculate closeness of a ray embedding to the subspace using its projection

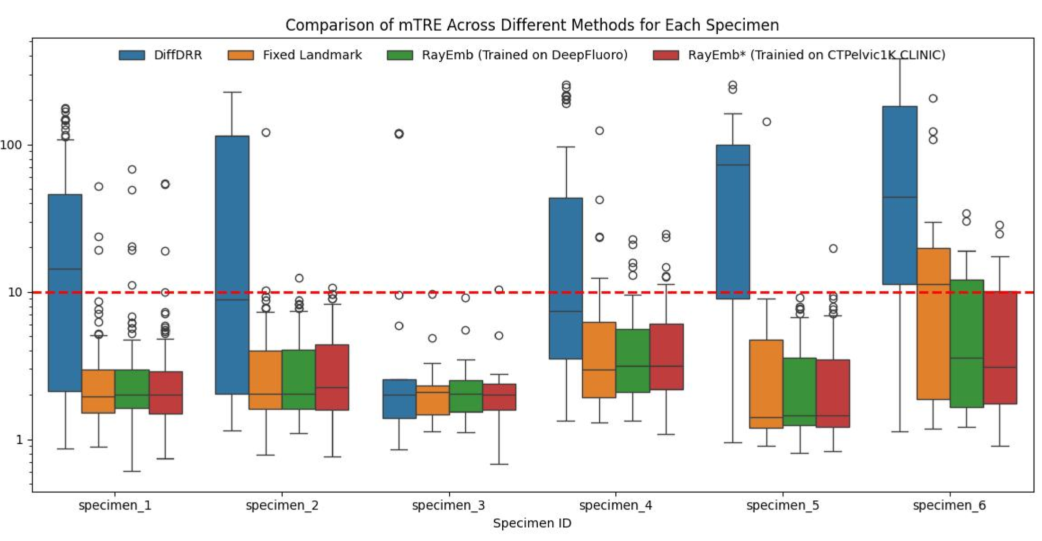
$$\mathrm{sim}(\mathbf{P},\mathbf{x}) = \frac{\mathbf{e}^\mathbf{T}(\mathbf{x})\mathbf{P}\mathbf{e}(\mathbf{x})}{|\mathbf{e}^\mathbf{T}(\mathbf{x})||\mathbf{P}\mathbf{e}(\mathbf{x})|}$$

Ray embedding with the largest similarity is considered as the corresponding 2D landmark

$$\hat{\mathbf{x}} = \mathrm{argmax}\ \mathrm{sim}(\mathbf{P},\mathbf{x})$$



4 QUANTITATIVE RESULTS



Specimen	Method	25^{th}	50^{th}	95^{th}	GFR@10	GFR@5
1	DiffDRR	2.125	14.259	147.349	54.054	63.964
	Fixed Landmark	1.520	**1.960**	6.701	**2.703**	**9.009**
	RayEmb	1.625	2.010	**6.520**	4.505	9.910
	RayEmb*	**1.502**	2.011	6.535	**2.703**	10.811
2	DiffDRR	2.049	8.879	203.835	48.077	59.615
	Fixed Landmark	**1.602**	**2.027**	**7.660**	1.923	19.231
	RayEmb	1.604	2.049	7.772	**0.962**	**17.308**
	RayEmb*	1.578	2.264	8.281	**0.962**	20.192
3	DiffDRR	**1.402**	2.007	119.635	12.500	20.833
	Fixed Landmark	1.474	2.100	**4.628**	**0.000**	**4.167**
	RayEmb	1.549	2.053	5.239	**0.000**	8.333
	RayEmb*	1.583	**2.019**	4.737	4.167	8.333
4	DiffDRR	3.529	7.402	216.490	41.667	66.667
	Fixed Landmark	**1.937**	**2.965**	23.770	14.583	35.417
	RayEmb	2.094	3.157	15.375	**10.417**	**29.167**
	RayEmb*	2.186	3.167	**14.103**	12.500	33.333
5	DiffDRR	9.062	73.597	140.134	74.545	76.364
	Fixed Landmark	**1.190**	**1.414**	8.172	1.818	25.455
	RayEmb	1.248	1.447	**7.633**	**0.000**	**18.182**
	RayEmb*	1.211	1.452	8.313	1.818	21.818
6	DiffDRR	11.263	43.853	375.553	75.000	79.167
	Fixed Landmark	1.875	11.250	121.944	50.000	62.500
	RayEmb	**1.647**	3.580	28.496	33.333	**45.833**
	RayEmb*	1.754	**3.090**	**23.832**	**25.000**	**45.833**

Datasets

DeepFuoro[1] - CT scans of six specimens with 366 real registered X-ray images. All real X-ray images were used for evaluating RayEmb*.

CTPelvic1K CLINIC[2] - 103 CT volumes with varying resolutions and diverse range of subjects were used for simulating X-ray images using DiffDRR.

Baselines

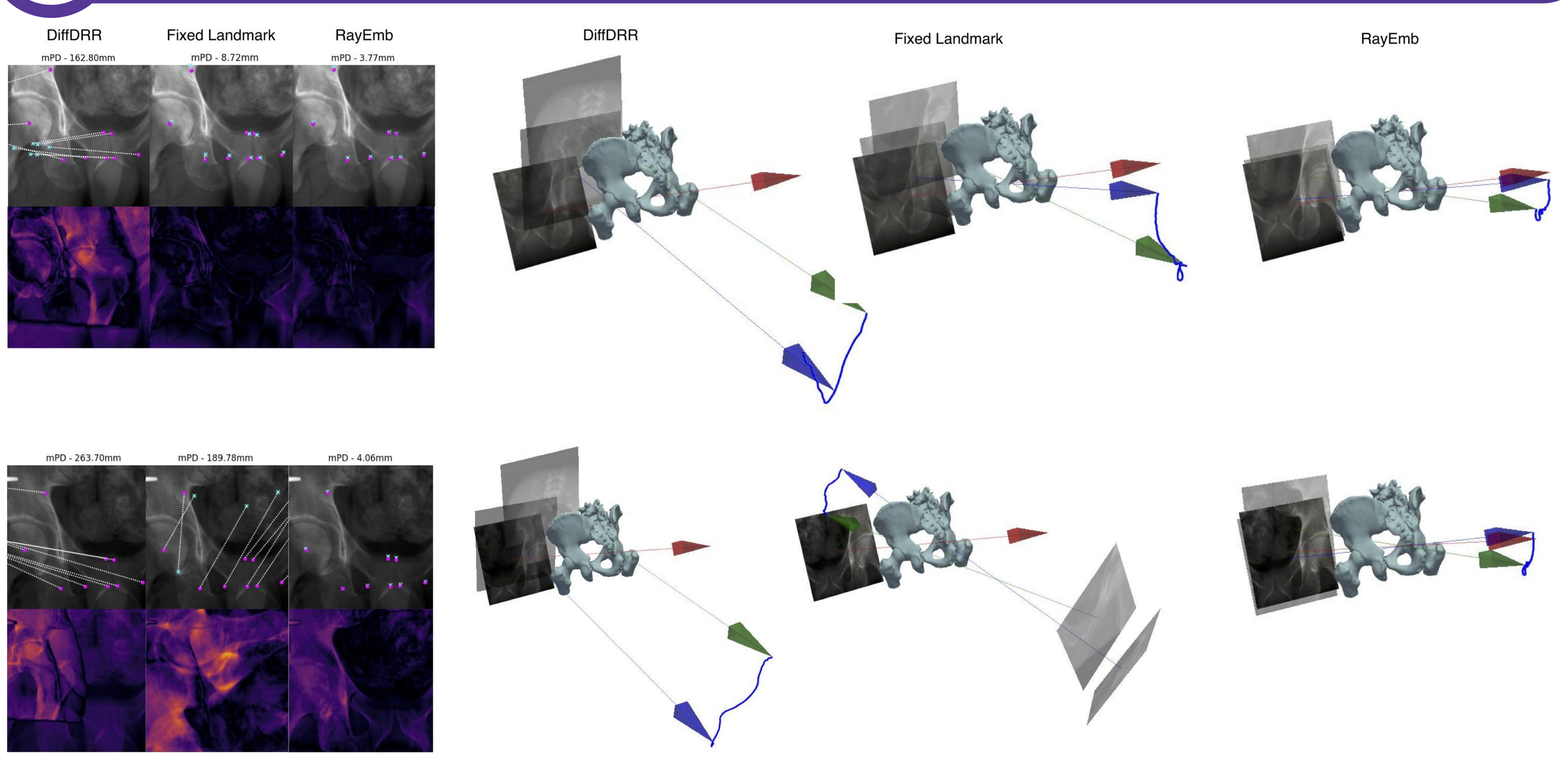
DiffDRR[3] - Gradient based optimization of normalized cross correlation between the X-ray image and rendered image.

Fixed Landmark[1] - Estimates heatmaps of 2D landmarks and runs PnP with RANSAC for registration.

Results

Both, Fixed landmark Estimation and Rayemb maintained median **mTRE below 10mm across all specimens**. However, rayemb demonstrated **lower mTRE and Failure rates on specimen 6**, which is the difficult test case in terms of patient and pose variability.

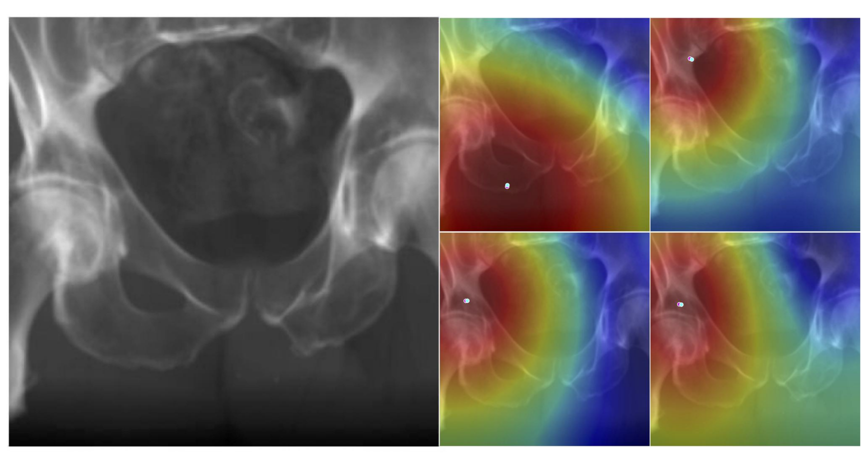
5 QUALITATIVE RESULTS



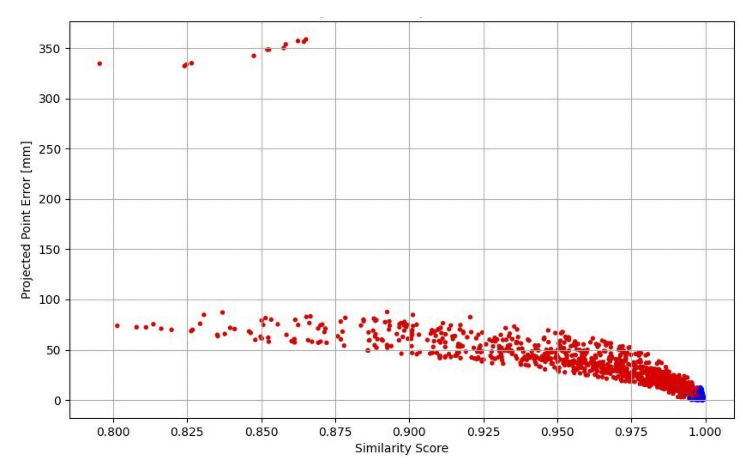
Visualizing Registration Results

Figure above illustrates the reprojected landmarks in X-ray image (left) and registered camera poses (right) where initial registration result (green), optimized result (blue) and ground truth (red) cameras are drawn for each method.

Estimated Heatmaps



Similarity Scores and Projection Error



References

[1] Grupp, R.B., Unberath, M., Gao, C., Hegeman, R.A., Murphy, R.J., Alexander, C.P., Otake, Y., McArthur, B.A., Armand, M., Taylor, R.H.: Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration. Int. J. Comput. Assist. Radiol. Surg. 15(5), 759–769 (May 2020)

[2] Liu, P., Han, H., Du, Y., Zhu, H., Li, Y., Gu, F., Xiao, H., Li, J., Zhao, C., Xiao, L., Wu, X., & Zhou, S. K. (2021). Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. International journal of computer assisted radiology and surgery, 16(5), 749–756. https://doi.org/10.1007/s11548-021-02363-8 (Mar 2018)

[3] Gopalakrishnan, V., & Golland, P. (2022, September). Fast auto-differentiable digitally reconstructed radiographs for solving inverse problems in intraoperative imaging. In Workshop on Clinical Image-Based Procedures (pp. 1-11). Cham: Springer Nature Switzerland.