

Patient-Agnostic Synthetic Pretraining for Efficient Patient-Specific Intraoperative 2D/3D Registration

Minheng Chen^{1,2} and Youyong Kong^{2*}

¹Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, 76013, TX, USA.

^{2*}School of Computer Science and Engineering, Southeast University, Nanjing, 210096, Jiangsu, China.

*Corresponding author(s). E-mail(s): kongyouyong@seu.edu.cn;
Contributing authors: mxc2442@mavs.uta.edu;

Abstract

Intraoperative 2D/3D registration aligns preoperative CT volumes with intraoperative X-ray or fluoroscopic images and is essential for image-guided interventions. Recent learning-based and differentiable registration methods have shown promising accuracy, especially in patient-specific settings where abundant digitally reconstructed radiographs (DRRs) can be synthesized from the target CT. However, training a separate patient-specific model from scratch for every new patient is computationally inefficient and limits practical deployment. In this work, we propose an efficient patient-specific 2D/3D registration framework based on patient-agnostic synthetic pretraining and spherical similarity learning. The model is first pretrained on synthetic DRRs generated from multiple CT volumes to learn transferable pose-sensitive representations, and is then adapted to a new patient using only a limited number of synthetic projections from the target CT. To improve synthetic-to-real robustness without requiring anatomical labels, we introduce a segmentation-free domain randomization strategy that perturbs image intensity, projection physics, field-of-view, occlusion, and fluoroscopic artifacts. The adapted model provides an initial pose estimate, which is further refined using spherical similarity learning and differentiable Levenberg-Marquardt optimization. Experiments on multiple anatomical datasets evaluate whether patient-agnostic synthetic pretraining can improve the efficiency of patient-specific registration, with particular focus on the trade-off between adaptation cost and registration accuracy. The results demonstrate that patient-agnostic synthetic pretraining can significantly reduce patient-specific training requirements while preserving accurate intraoperative 2D/3D registration.

Keywords: 2D/3D registration, Synthetic pretraining, Patient-specific adaptation, Similarity learning

1 Introduction

Intraoperative 2D/3D registration aims to align intraoperative radiographic images, such as X-ray or fluoroscopy, with preoperative 3D volumes, such as CT. It is essential for image-guided interventions, where accurate spatial correspondence

is required for surgical navigation, planning, and instrument or implant localization. Single-view 2D/3D registration is clinically attractive because it can simplify the imaging workflow and reduce radiation exposure, but it is also highly challenging due to projection ambiguity, limited anatomical visibility, and sensitivity to initialization.

Conventional methods usually solve this problem by generating digitally reconstructed radiographs (DRRs) from CT volumes and optimizing the rigid transformation that maximizes image similarity between the DRR and the intraoperative X-ray [1–8]. Hand-crafted similarity metrics, such as normalized cross-correlation (NCC), mutual information (MI), and local feature descriptors, are commonly combined with derivative-free optimizers such as BOBYQA or CMA-ES [9–11]. Although these methods do not require annotated training data, their similarity landscapes are often highly non-convex outside a small capture range, leading to local minima and long optimization times.

Learning-based methods have been proposed to improve registration efficiency and robustness. Regression-based approaches directly predict the pose from paired X-ray/DRR images or 3D–2D inputs [12–15], while landmark-based methods rely on anatomical keypoints followed by geometric pose estimation [16–19]. However, direct pose regression can generalize poorly to unseen patients or imaging conditions, and landmark-based methods require reliable anatomical landmark annotations, which are difficult to obtain in clinical fluoroscopy. These limitations motivate self-supervised synthetic learning strategies, where DRRs generated from CT volumes provide pose-supervised training pairs without manual annotation [20–22]. Recently, differentiable rendering and similarity learning have provided a promising alternative between conventional optimization and direct regression. Gao *et al.* [23, 24] proposed learning a deep similarity metric whose gradient approximates pose-space geodesic directions on $SE(3)$, thereby improving the capture range of 2D/3D registration. Subsequent studies further explored correlation-driven feature decomposition and embedded feature similarity optimization [25, 26]. Nevertheless, two important challenges remain. First, most learned similarity methods represent pose-induced image changes in Euclidean feature spaces, although rigid transformations naturally lie on non-Euclidean manifolds. Second, high-accuracy patient-specific registration often requires training a separate model from scratch for each new patient, which is computationally expensive and limits clinical scalability.

Patient-specific and patient-agnostic training represent different trade-offs. Patient-specific

methods can synthesize abundant DRRs from the target CT and achieve high accuracy [22, 27], but require costly training for every new patient. Patient-agnostic models can be trained once across multiple subjects and reused on unseen patients [23, 25], but their accuracy is often limited by anatomical variability and the synthetic-to-real domain gap. This raises a natural question: *can patient-agnostic synthetic pretraining provide an effective initialization for efficient patient-specific 2D/3D registration?*

In this work, we propose an efficient patient-specific 2D/3D registration framework based on patient-agnostic synthetic pretraining and spherical similarity learning. Instead of training a full patient-specific model from scratch, we first pre-train the registration network using synthetic DRRs generated from multiple CT volumes, and then adapt the pretrained model to a new patient using a limited number of synthetic projections from the target CT. To improve synthetic-to-real robustness without requiring anatomical labels, we introduce a segmentation-free domain randomization strategy that perturbs image appearance, projection geometry, field-of-view truncation, occlusion, and fluoroscopic artifacts.

Building on our previous spherical similarity learning framework [28], image features are projected onto a hyperspherical manifold to provide a geometrically more consistent similarity measure. We further use a bi-invariant $SO(4)$ pose-gradient formulation and differentiable Levenberg–Marquardt optimization for iterative pose refinement. Inspired by recent advances in parameter-efficient fine-tuning of vision foundation models, we systematically evaluate different adaptation settings, including the amount of patient-specific synthetic data and the model components selected for fine-tuning, to analyze the trade-off between registration accuracy and adaptation cost. The main contributions of this work are summarized as follows:

- We propose a patient-agnostic synthetic pretraining and patient-specific adaptation strategy for efficient intraoperative 2D/3D registration, reducing the need to train a separate model from scratch for every patient.
- We introduce a segmentation-free domain randomization scheme for synthetic pretraining,

covering image appearance, projection geometry, field-of-view variation, occlusion, and fluoroscopic artifacts without requiring CT or X-ray segmentation.

- We build upon spherical similarity learning and an $SO(4)$ -based pose-gradient formulation to provide a geometrically consistent similarity landscape for differentiable pose refinement.
- We systematically evaluate the trade-off between patient-specific adaptation cost and registration accuracy, including adaptation data size, trainable modules, training time, and the effect of domain randomization.

2 Related Work

2.1 Learning-based 2D/3D Registration

Learning-based methods have been widely studied for 2D/3D registration because they can improve computational efficiency and enlarge the capture range compared with purely optimization-based pipelines. Existing methods can be broadly grouped into direct regression, landmark-based estimation, and hybrid two-stage frameworks. Regression-based methods directly predict the 3D pose from a 2D radiograph, a 3D volume, or paired 2D/3D inputs [14, 15, 29, 30]. These methods enable fast feed-forward pose estimation, which is attractive for intraoperative navigation and robotic-assisted interventions. However, direct regression often lacks the precision required for high-accuracy alignment, especially under single-view ambiguity, appearance variation, or synthetic-to-real domain shift. Landmark-based methods estimate the transformation from anatomical landmarks, local descriptors, or point-to-plane correspondences [16, 17, 31–33]. They typically formulate registration as a geometric pose estimation problem, such as perspective-n-point (PnP) [34], or use point-to-plane solvers for 2D/3D alignment [21, 35–37]. While these methods can be accurate when reliable correspondences are available, their performance depends strongly on the visibility and localization accuracy of anatomical structures [9]. In clinical fluoroscopy, landmarks may be occluded, truncated,

or degraded by low contrast and surgical instruments, often requiring robust estimation strategies such as RANSAC [38].

Hybrid two-stage methods combine learning-based initialization with iterative refinement [22, 26, 27, 39, 40]. A neural network first predicts an initial pose, which is then refined using differentiable rendering, learned similarity optimization, or derivative-free search [12, 41]. This strategy is effective in patient-specific scenarios, where synthetic projections from a fixed preoperative CT can provide dense training supervision. However, training a separate patient-specific model from scratch is computationally costly, whereas patient-agnostic models are reusable but usually less accurate due to inter-patient anatomical variability. This trade-off motivates the patient-agnostic pre-training and patient-specific adaptation strategy studied in this work.

2.2 Synthetic Learning and Domain Randomization

A major challenge in learning-based 2D/3D registration is the lack of real X-ray/CT pairs with accurate ground-truth poses. Synthetic learning addresses this limitation by generating DRRs from CT volumes under known projection geometries, thereby providing pose-supervised training data without manual annotation [20, 22, 42]. Since pose labels are directly determined by the rendering process, DRR-based learning is particularly suitable for 2D/3D registration. Synthetic data have also become an important strategy for improving generalization in learning-based medical imaging [43–45]. Representative Synth-based methods, including SynthSeg [46–48], SynthStrip [49], and SynthMorph [50, 51], use randomized synthetic images to generalize across imaging protocols, contrasts, resolutions, and populations. These studies suggest that carefully designed synthetic distributions can help models learn task-relevant anatomical and geometric representations that are less dependent on a specific acquisition setting.

For intraoperative 2D/3D registration, however, models must bridge the gap between synthetic DRRs and real fluoroscopic images [52, 53]. DRRs generated from simplified projection models may not reproduce scatter, detector response, truncation, surgical tools, or patient-positioning effects. Domain randomization has

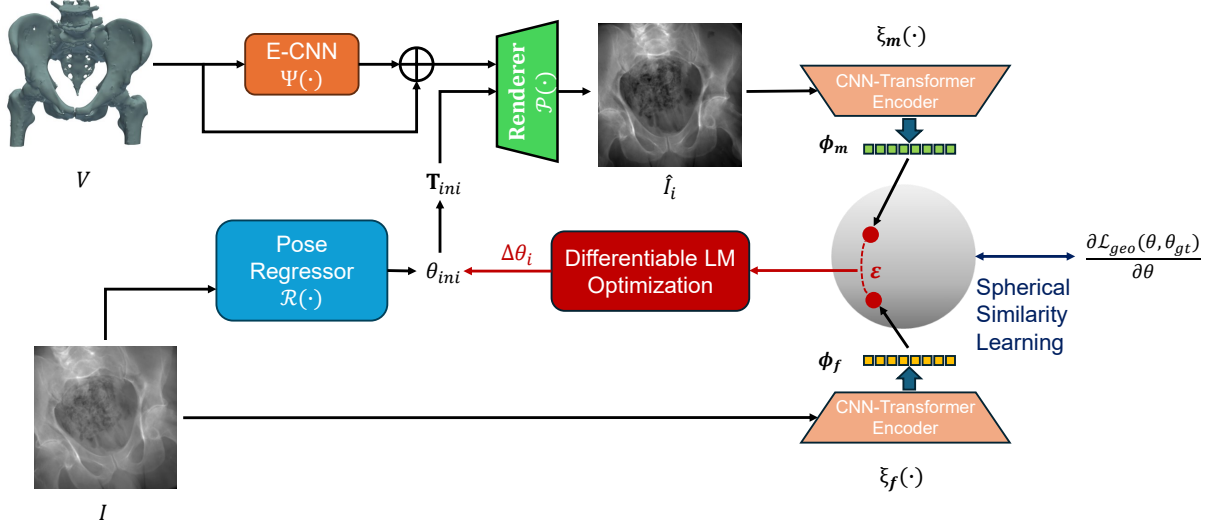


Fig. 1 Overview of the proposed framework. We first employ a regressor to initialize the pose and then refine it using differentiable Levenberg-Marquardt optimization based on spherical similarity learning. Spherical similarity learning consists of two main components: extracting image feature representations 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).

therefore been used to diversify synthetic training data and improve robustness to real imaging conditions [54, 55]. However, perturbations in 2D/3D registration must be carefully controlled, since they should improve appearance invariance without destroying pose-sensitive anatomical information. Synthetic DRR learning can be applied in both patient-agnostic and patient-specific settings. Patient-agnostic models are reusable across unseen patients but are often limited by anatomical variability and the synthetic-to-real gap. Patient-specific models can learn more accurate anatomy-dependent representations from the target CT [22, 40], but training or adapting a separate model for each patient introduces additional computational cost.

In this work, we study whether patient-agnostic synthetic pretraining can reduce the cost of patient-specific 2D/3D registration. We first learn transferable pose-sensitive representations from DRRs generated across multiple CT volumes, and then adapt the pretrained model to a new patient using limited synthetic projections from the target CT. Unlike segmentation-driven synthetic frameworks such as SynthSeg or SynthMorph, our setting requires neither anatomical

labels nor segmentation masks, making the proposed strategy annotation-free and segmentation-free.

2.3 Similarity Learning for Image Registration

Image registration relies critically on the choice of image similarity measure. Traditional metrics, such as sum of squared differences (SSD), normalized cross-correlation (NCC), and mutual information (MI), have been widely used in medical image registration [6]. However, their similarity landscapes are often non-convex over a large pose range, which can trap iterative optimization in local minima. This issue is especially challenging in single-view 2D/3D registration, where projection ambiguity and the X-ray/DRR appearance gap are substantial.

Similarity learning addresses this problem by learning task-specific feature representations and similarity functions from data [25, 56, 57]. Learned metrics can provide more informative gradients for optimization and improve robustness to local appearance variations [58–61]. In 2D/3D registration, Gu *et al.* [62] introduced similarity

learning through Riemannian pose-gradient estimation, and Gao *et al.* [23, 24] extended this idea into a fully differentiable framework. Subsequent work further improved learned similarity optimization through correlation-driven feature decomposition and parameter-specific initialization [25, 26]. Most existing similarity learning frameworks represent image features in Euclidean latent spaces, although rigid transformations and pose-induced image changes naturally lie on non-Euclidean manifolds. To provide a more geometrically consistent similarity landscape, our method builds on spherical similarity learning and a bi-invariant $SO(4)$ pose-gradient formulation. Combined with patient-agnostic synthetic pretraining and patient-specific adaptation, this learned similarity metric can be efficiently transferred to new patients while maintaining sensitivity to anatomical misalignment.

3 Method

3.1 Problem Formulation

Given a preoperative CT volume V and an intraoperative X-ray image I^x , 2D/3D registration aims to estimate the rigid transformation that aligns the CT volume with the X-ray image. Let $\mathcal{P}(\cdot)$ denote a differentiable projector that generates a digitally reconstructed radiograph (DRR) from the CT volume under a given pose. The rigid transformation is represented as:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \in SE(3) \quad (1)$$

where $\mathbf{R} \in SO(3)$ is the rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is the translation vector. The registration objective is formulated as:

$$\mathbf{T}^* = \arg \min_{\mathbf{T} \in SE(3)} \mathcal{S}(I^x, \mathcal{P}(\mathbf{T}) \circ V) \quad (2)$$

where $\mathcal{S}(\cdot, \cdot)$ measures the dissimilarity between the intraoperative X-ray image and the rendered DRR.

3.2 Network Architecture

The overall framework follows our conference version [28] as shown in Fig. 1. Given an input X-ray image and a CT volume, a pose regressor $\mathcal{R}(\cdot)$ first

predicts an initial pose parameter $\theta_{\text{ini}} \in \mathfrak{se}(3)$. The predicted parameter is mapped to $SE(3)$ through the exponential map:

$$\mathbf{T}_{\text{ini}} = \text{Exp}(\theta_{\text{ini}}) \quad (3)$$

The initial transformation is then used to render an initial DRR from the CT volume.

To improve the representation of the projected volume, we employ an equivariant CNN module [63, 64] $\Psi(\cdot)$ to extract transformation-aware 3D features. The enhanced volume is obtained by residual addition:

$$\hat{V} = V + \Psi(V), \quad (4)$$

and the corresponding moving DRR is generated as:

$$\hat{I}_m(\mathbf{T}) = \mathcal{P}(\mathbf{T}) \circ \hat{V} \quad (5)$$

The fixed image and moving DRR are then encoded by two feature extractors, ξ_f and ξ_m , which share the same architecture:

$$\phi_f = \xi_f(I_f), \quad \phi_m(\mathbf{T}) = \xi_m(\hat{I}_m(\mathbf{T})) \quad (6)$$

The learned feature maps are used to compute the spherical similarity described below.

3.3 Spherical Similarity Learning

To obtain a geometrically consistent similarity metric, we embed the learned feature vectors onto a hyperspherical manifold. Given feature maps ϕ_f and ϕ_m , we apply a spherical exponential mapping along the channel dimension:

$$\Phi_f = \text{EXP}(\phi_f), \quad \Phi_m = \text{EXP}(\phi_m) \quad (7)$$

The spherical discrepancy between the fixed and moving images is defined as:

$$\varepsilon(\mathbf{T}) = \sum_u (1 - \Phi_m(u, \mathbf{T})^\top \Phi_f(u)) \quad (8)$$

where u indexes spatial feature locations. A smaller value of $\varepsilon(\mathbf{T})$ indicates better alignment between the fixed image and the moving DRR.

Following our previous work, the similarity network is trained using a double-backward strategy. Instead of directly supervising the absolute similarity value, we align the gradient of the

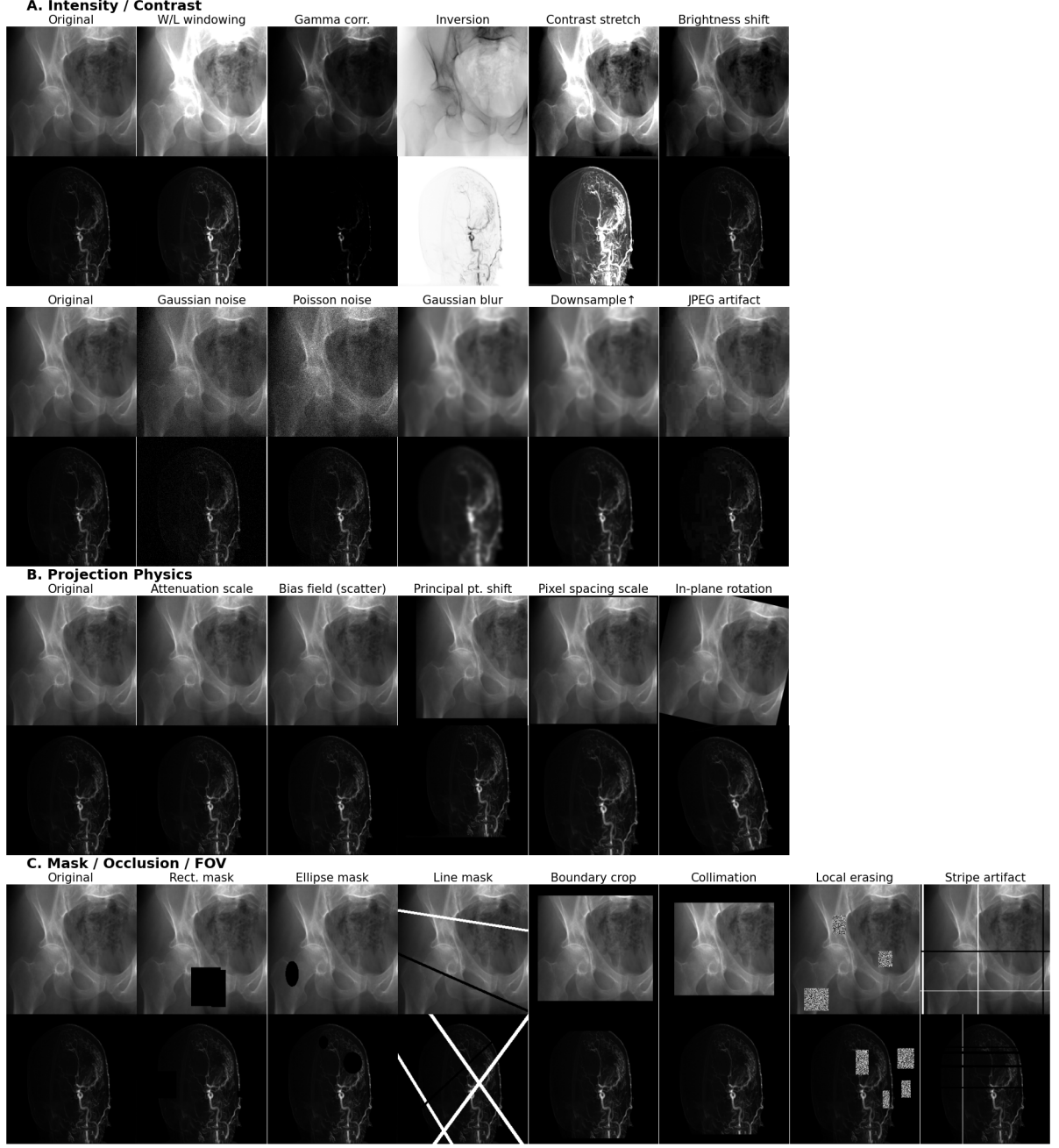


Fig. 2 Illustration of the proposed segmentation-free domain randomization. The augmentations are grouped into intensity/contrast perturbations, projection-physics variations, and mask & occlusion & FOV transformations, covering common synthetic-to-real differences between DRRs and intraoperative X-ray images.

learned similarity with the gradient of a pose-space geodesic distance:

$$\nabla_{\theta} \mathcal{L}_{\text{net}}(\theta) \approx \nabla_{\theta} \mathcal{L}_{\text{geo}}(\theta, \theta_{\text{gt}}) \quad (9)$$

where $\mathcal{L}_{\text{net}} = \varepsilon(\mathbf{T})$. To provide more geometrically consistent supervision, the pose is embedded into a bi-invariant $SO(4)$ representation, and the

geodesic discrepancy is computed in this space as in our conference version [28].

3.4 Segmentation-Free Domain Randomization

To improve the transferability of synthetic DRRs to real intraoperative X-ray images, we introduce a segmentation-free domain randomization strategy. Unlike segmentation-based synthetic frameworks, the proposed augmentation does not require CT segmentation, X-ray segmentation, anatomical labels, or landmark masks. It operates directly on DRR images and contains three operation pools: intensity and contrast randomization $\mathcal{G}_I$, projection-physics randomization $\mathcal{G}_P$, and mask, occlusion, and field-of-view randomization $\mathcal{G}_M$. The operations used for randomization are illustrated in Fig. 2. The intensity and contrast pool $\mathcal{G}_I$ simulates photometric differences between DRRs and real X-ray images, including window-level adjustment, gamma correction, intensity inversion, contrast stretching, brightness shifts, Gaussian and Poisson noise, Gaussian blur, downsampling-upsampling, and JPEG compression. The projection-physics pool $\mathcal{G}_P$ approximates acquisition-related variations, including attenuation scaling, smooth multiplicative bias fields, principal-point shifts, pixel-spacing or zoom changes, and in-plane detector rotations. The mask, occlusion, and field-of-view pool $\mathcal{G}_M$ simulates content mismatch commonly observed in fluoroscopy, including rectangular and elliptical masks, line-shaped occlusions, boundary cropping, collimation masks, local erasing, and stripe artifacts.

Let $\mathcal{A}(\cdot)$ denote the stochastic augmentation operator. During synthetic training, the fixed DRR generated from the ground-truth pose is transformed into a pseudo-X-ray image:

$$\tilde{I}_{\text{gt}} = \mathcal{A}(\mathcal{P}(\mathbf{T}_{\text{gt}}) \circ V) \quad (10)$$

The moving DRR remains differentially rendered from the perturbed pose:

$$I_m(\mathbf{T}) = \mathcal{P}(\mathbf{T}) \circ V \quad (11)$$

This design encourages robustness to appearance variation while preserving differentiability with respect to pose parameters. For each operation

pool $g \in \{I, P, M\}$, two distinct operations are sampled without replacement and applied in random order:

$$\mathcal{A}_g^* = A_g^{(2)} \circ A_g^{(1)}, \quad A_g^{(1)}, A_g^{(2)} \in \mathcal{G}_g \quad (12)$$

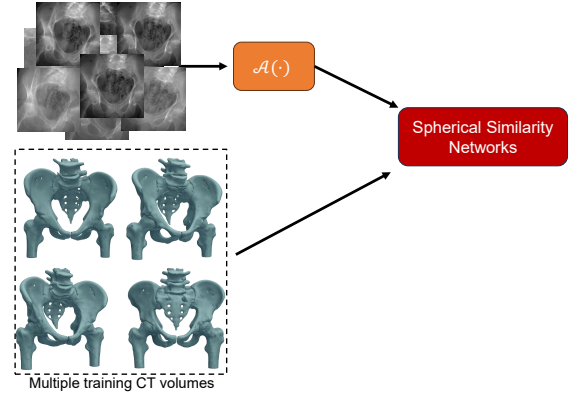
The final compound augmentation is defined as

$$\mathcal{A} = (\mathcal{A}_M^*)^{z_M} \circ (\mathcal{A}_P^*)^{z_P} \circ (\mathcal{A}_I^*)^{z_I} \quad (13)$$

where $z_g \sim \text{Bernoulli}(p_g)$, and $p_I = p_P = p_M = 0.5$. When $z_g = 0$, the corresponding operation pool is replaced by the identity mapping.

3.5 Patient-Agnostic Pretraining and Patient-Specific Adaptation

(a) Patient-Agnostic Pretraining



(b) Patient-Specific Adaptation

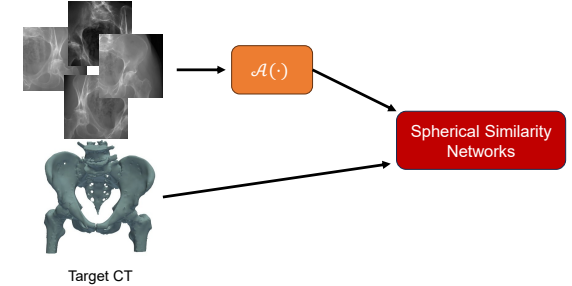


Fig. 3 Overview of the proposed pretraining–adaptation strategy. (a) Patient-agnostic pretraining learns spherical similarity networks from synthetic DRRs generated from multiple training CT volumes with segmentation-free domain randomization $\mathcal{A}(\cdot)$. (b) Patient-specific adaptation fine-tunes the pretrained model using limited randomized synthetic projections from the target CT.

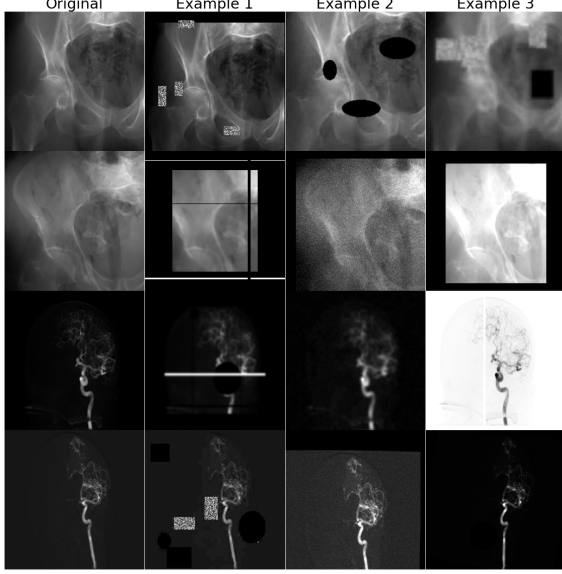


Fig. 4 Examples of segmentation-free domain randomization. The first column shows the original projection, and the remaining columns show randomly augmented examples with intensity, physics, occlusion, and FOV perturbations.

The proposed training strategy consists of patient-agnostic synthetic pretraining followed by patient-specific adaptation. In the pretraining stage, synthetic DRR pairs are generated from multiple training CT volumes. For each sample, a ground-truth pose $\mathbf{T}_{\text{gt}}$ is sampled within an anatomy-specific pose range, and a perturbed pose $\mathbf{T}$ is sampled around it to generate the moving DRR. The fixed image is randomized by $\mathcal{A}(\cdot)$, while the moving image is rendered from the perturbed pose. Given a new patient, the pretrained model is fine-tuned using a limited number of synthetic DRRs generated only from the target CT volume V_{new} . The adaptation set is written as:

$$\mathcal{D}_{\text{adapt}} = \left\{ \left(\tilde{I}_{\text{gt}}^k, I_m^k, \mathbf{T}^k, \mathbf{T}_{\text{gt}}^k \right) \right\}_{k=1}^K \quad (14)$$

where K is the number of patient-specific synthetic samples. During adaptation, we use a modified augmentation operator $\mathcal{A}_{\text{adapt}}$, which follows the same compound augmentation strategy but excludes the mask, occlusion, and field-of-view pool $\mathcal{G}_M$. This avoids artificial removal of patient-specific anatomical structures while retaining mild appearance and acquisition variability. Compared

with patient-specific training from random initialization, the proposed strategy starts from a patient-agnostic pretrained representation and therefore requires fewer synthetic samples and fewer training iterations. Motivated by recent advances in parameter-efficient fine-tuning, we evaluate different adaptation settings by varying both the amount of patient-specific synthetic data and the set of trainable modules.

3.6 Training and Inference

As shown in Fig. 3, the overall training strategy consists of two stages: patient-agnostic synthetic pretraining and patient-specific adaptation. In the first stage, the model is pretrained using synthetic DRRs generated from multiple training CT volumes. For each training sample, a ground-truth pose is sampled within the anatomy-specific pose range, and a perturbed pose is used to generate the moving DRR. The pose regressor is optimized by minimizing the geodesic discrepancy between the predicted initial pose and the ground-truth pose, while the spherical similarity network is trained using the double-backward gradient alignment strategy described above. In the second stage, the pretrained model is adapted to a new patient using a limited number of synthetic DRRs generated only from the target CT volume. The same pose-supervised training objectives are used during adaptation, but only selected modules are fine-tuned depending on the adaptation setting. This strategy transfers the patient-agnostic pose-sensitive representation to the target anatomy while avoiding training the entire patient-specific model from random initialization.

During inference, the adapted regressor first predicts an initial pose $\mathbf{T}_{\text{ini}}$. The pose is then refined by minimizing the learned spherical discrepancy:

$$\mathbf{T}^* = \arg \min_{\mathbf{T}} \varepsilon(\mathbf{T}) \quad (15)$$

We use differentiable Levenberg-Marquardt optimization for iterative refinement. At iteration i , the pose increment is computed as:

$$\Delta \theta_i = (J_i^\top W J_i + \lambda I)^{-1} J_i^\top W \mathbf{r}(\theta_{i-1}) \quad (16)$$

where J_i is the Jacobian of the residual vector $\mathbf{r}(\theta_{i-1})$, W is the weighting matrix, and λ is the damping parameter. The pose is updated by left

Table 1 Comparison of the proposed pretrained fine-tuning method with existing patient-specific 2D/3D registration baselines on DeepFluoro and Ljubljana. Our pretrained FT method adapts a patient-agnostic pretrained model to each target patient using synthetic projections from the target CT. SMSR denotes the sub-millimeter success rate, defined as the percentage of test cases with mTRE < 1 mm. Median, 75th percentile, and 95th percentile mTREs are reported. For accuracy metrics, the best results are **bolded**, and the second-best results are underlined. Run time is reported for reference and is not used for ranking.

Dataset	Method	SMSR	Median (mm)	Percentile (mm)		Run Time
				75%	95%	
DeepFluoro	PSSS-reg [27]	56.0%	0.93	2.51	5.57	12.7 s
	PoseNet [65]	4.3%	16.6	22.0	29.2	0.1 s
	DFLNet [66]	36.6%	3.20	7.29	13.1	1.0 s
	SCR-reg [32]	33.3%	4.70	9.59	12.8	1.1 s
	DiffPose [22]	83.1%	0.60	<u>0.89</u>	<u>1.47</u>	5.3 s
	Ours (from scratch)	86.1%	0.51	0.85	1.42	6.2 s
	Ours (pretrained FT)	<u>84.7%</u>	<u>0.55</u>	<u>0.89</u>	1.58	6.3 s
Ljubljana	PSSS-reg [27]	40.0%	2.48	5.87	11.3	15.3 s
	PoseNet [65]	0%	23.3	26.2	29.2	< 0.1 s
	DiffPose [22]	<u>80.0%</u>	0.63	<u>0.94</u>	1.78	6.0 s
	Ours (from scratch)	85.0%	<u>0.55</u>	0.85	<u>1.35</u>	6.5 s
	Ours (pretrained FT)	85.0%	0.52	0.85	1.33	6.4 s

multiplication:

$$\mathbf{T}_i = \text{Exp}(\Delta\theta_i)\mathbf{T}_{i-1} \quad (17)$$

The final pose after LM refinement is used as the registration result.

4 Experiments and Results

4.1 Datasets and Evaluation Metrics

We evaluate the proposed method on two publicly available datasets covering different anatomical regions and imaging scenarios.

- **DeepFluoro:** The DeepFluoro dataset contains six pelvic CT scans and 366 real fluoroscopic X-ray images [66]. The dataset provides the calibrated intrinsic matrix of the C-arm imaging system and the ground-truth extrinsic matrix for each X-ray image. In addition, 14 manually annotated anatomical landmarks are available for each CT scan, enabling quantitative evaluation of target registration error.
- **Ljubljana:** The Ljubljana dataset [67] contains ten clinical 3D cone-beam computed tomography (CBCT) subtracted angiography scans from patients undergoing cerebral endovascular

treatment. Each CBCT scan is associated with two X-ray images, and calibrated intrinsic and extrinsic matrices are provided for registration evaluation.

Following prior studies [20, 22, 68], we pre-process the generated DRRs to reduce the intensity discrepancy between simulated projections and real X-ray images. X-ray imaging records the attenuation of photons after traversing tissue, whereas DRR rendering based on line integration often represents accumulated absorption along each ray. As a result, high-density structures may appear with opposite contrast in DRRs and real X-ray images. To make the grayscale distribution of DRRs generated by Siddon’s ray-tracing method [69] more consistent with real X-ray images, we apply the following intensity inversion:

$$\tilde{I} = 1 - \frac{\log(1 + I)}{\log(1 + I_0)} \quad (18)$$

where I denotes the rendered DRR intensity and I_0 is the maximum intensity used for normalization. The logarithmic transformation improves numerical stability and enhances contrast consistency between DRRs and X-ray images. The evaluation metrics used in our experiments are as follows:

Table 2 Main comparison of patient-specific adaptation strategies on DeepFluoro and Ljubljana. Speed-up is computed relative to patient-specific training from scratch on each dataset.

Dataset	Strategy	Pretraining	Fine-tuning amount	Training time (h)	Speed-up vs. scratch	SMSR $\uparrow$	Median / 95% mTRE (mm) $\downarrow$
DeepFluoro	From scratch	No	Full	20.70	1.0 $\times$	86.1%	0.51 / 1.42
	Patient-agnostic only	Yes	0	0	–	19.6%	4.50 / 22.63
	Pretrain + 10% FT	Yes	10%	0.04	517.5 $\times$	42.4%	2.78 / 15.86
	Pretrain + 25% FT	Yes	25%	0.10	207.0 $\times$	60.9%	0.92 / 8.94
	Pretrain + 50% FT	Yes	50%	0.22	94.1 $\times$	74.3%	0.75 / 1.67
	Pretrain + full FT	Yes	Full	0.47	44.0 $\times$	84.7%	0.55 / 1.58
Ljubljana	From scratch	No	Full	20.90	1.0 $\times$	85.0%	0.55 / 1.35
	Patient-agnostic only	Yes	0	0	–	23.4%	4.26 / 19.33
	Pretrain + 10% FT	Yes	10%	0.06	348.3 $\times$	70.0%	0.81 / 2.73
	Pretrain + 25% FT	Yes	25%	0.13	160.8 $\times$	80.0%	0.63 / 1.86
	Pretrain + 50% FT	Yes	50%	0.25	83.6 $\times$	81.0%	0.59 / 1.59
	Pretrain + full FT	Yes	Full	0.48	43.5 $\times$	85.0%	0.52 / 1.33

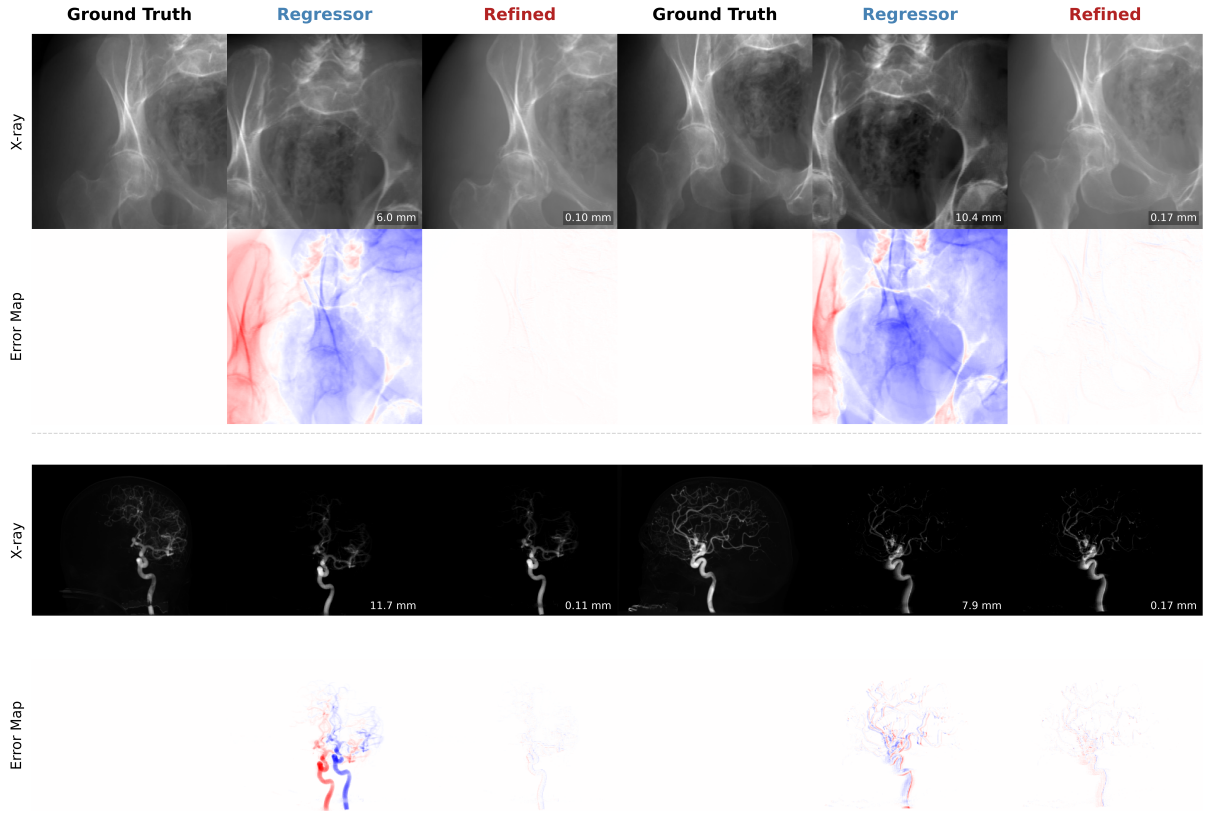


Fig. 5 Qualitative registration results on DeepFluoro and Ljubljana. For each case, we show the ground-truth X-ray, the DRR generated from the regressor initialization, and the DRR after differentiable LM refinement. The corresponding error maps show that the refinement step substantially reduces the residual misalignment from the initial pose estimate.

- **Mean target registration error (mTRE):**
mTRE measures the average Euclidean distance

between corresponding target points transformed by the estimated pose and the ground-truth pose. For DeepFluoro, we use the provided

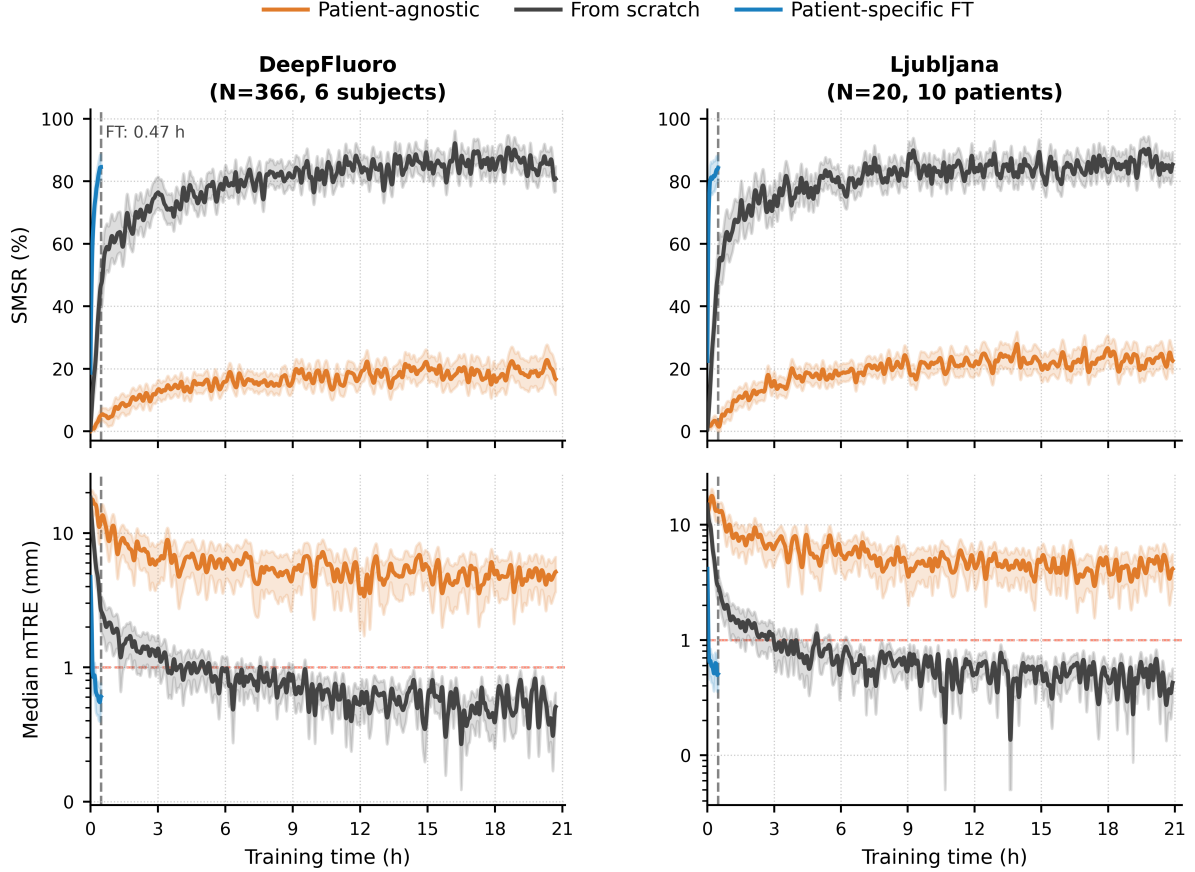


Fig. 6 Training efficiency comparison between patient-specific training from scratch and patient-agnostic pretraining with patient-specific fine-tuning. Curves show SMSR and median mTRE as a function of training time on DeepFluoro and Ljubljana. Patient-agnostic direct transfer provides limited accuracy, whereas patient-specific fine-tuning rapidly reaches near from-scratch performance within approximately 0.5 h. The red dashed line indicates the 1 mm mTRE threshold for sub-millimeter registration.

anatomical landmarks. For Ljubljana, where manual landmarks are not available, we compute the error using sampled 3D points from the reconstructed anatomical structure, following the calibrated ground-truth projection geometry. We report the median, 75th percentile, and 95th percentile mTRE.

- **Sub-millimeter success rate (SMSR):** SMSR is defined as the percentage of test cases with mTRE below 1 mm. This metric reflects the proportion of registrations that achieve sub-millimeter accuracy, which is important for high-precision image-guided interventions.
- **Training time:** Since the main goal of this work is to reduce patient-specific adaptation cost, we also report the time required

for patient-specific fine-tuning under different adaptation settings.

4.2 Training and Implementation Details

All models were implemented in PyTorch and trained on an NVIDIA A6000 GPU. For patient-specific evaluation, we adopted a leave-one-out protocol. For each target patient, the patient-agnostic model was pretrained using CT volumes from all remaining subjects, ensuring that the target CT was excluded from pretraining. The pre-trained model was then adapted using synthetic DRRs generated only from the target patient’s CT volume. We follow the pose parameter setting

in [40], for both patient-agnostic pretraining and patient-specific adaptation, synthetic DRR pairs were generated by sampling ground-truth poses within anatomy-specific pose ranges and then perturbing them to obtain moving projections. For the DeepFluoro dataset, rotations were sampled from $\alpha \in [-45^\circ, 45^\circ]$, $\beta \in [-45^\circ, 45^\circ]$, and $\gamma \in [-15^\circ, 15^\circ]$, while translations were sampled from $x \in [-150, 150]$ mm, $y \in [-1000, -450]$ mm, and $z \in [-150, 150]$ mm. For the Ljubljana dataset, rotations were sampled from $\alpha \in [-45^\circ, 90^\circ]$, $\beta \in [-5^\circ, 5^\circ]$, and $\gamma \in [-5^\circ, 5^\circ]$, while translations were sampled from $x \in [-25, 25]$ mm, $y \in [700, 800]$ mm, and $z \in [-25, 25]$ mm.

During patient-agnostic pretraining, the fixed projection was transformed into a pseudo-X-ray image using the proposed segmentation-free domain randomization, while the moving projection was rendered from the perturbed pose to preserve differentiability with respect to pose parameters [41]. Unlike the patient-agnostic setting in our conference version [28], where the pose regressor was replaced by RTPiv3 [26] to incorporate CT spatial information and trained with 750k regressor samples and 300k similarity-learning samples, we keep the same regressor architecture as in the patient-specific setting in this study. This design provides a controlled comparison between patient-specific training from scratch and pre-trained fine-tuning, ensuring that performance differences mainly reflect the effect of synthetic pretraining and adaptation rather than changes in the regressor architecture. For consistency in training-data scale across experiments, we also reduced the number of self-supervised training samples to 500k for the pose regressor and 200k for the similarity network.

We used the Adam optimizer with an initial learning rate of 1×10^{-3} , weight decay of 1×10^{-3} , and a cyclic learning-rate scheduler. For patient-specific adaptation, the pretrained model was fine-tuned using synthetic DRRs generated only from the target patient’s CT volume. Unless otherwise specified, we used 20k patient-specific synthetic samples for adaptation and fine-tuned the pose regressor and similarity encoder while keeping the remaining modules fixed. The adaptation stage used a learning rate of 1×10^{-3} and a batch size of 4. To evaluate adaptation efficiency, we varied the number of patient-specific synthetic samples and

compared the results with patient-specific training from random initialization.

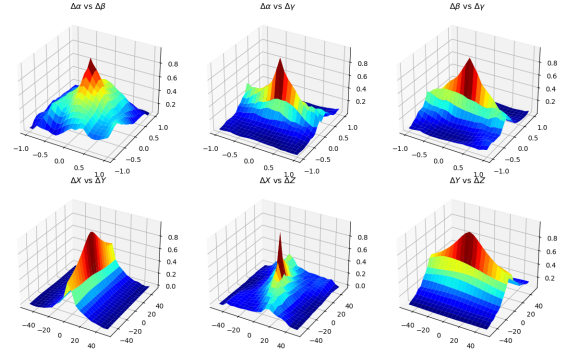


Fig. 7 Visualization of the proposed spherical deep similarity landscape. For clearer visualization, the deep similarity values are first normalized to the range $[0, 1]$, and then transformed by computing $1 - \epsilon$, effectively inverting the scale to enhance contrast in the display.

At inference time, the adapted pose regressor first predicted an initial pose, which was then refined using differentiable Levenberg–Marquardt optimization. The LM optimizer was run for 150 iterations with the learning rate initialized to 5×10^{-3} . A step learning-rate decay with a factor of 0.9 was applied every 25 iterations. The termination criterion was defined such that the standard deviation of the learned similarity over the last ten iterations was less than 1×10^{-2} . The final pose after LM refinement was used for quantitative evaluation. All input DRRs and X-ray images were resized to 256×256 and normalized to $[0, 1]$. The learned deep similarity is visualized in Fig. 7

4.3 Efficient Patient-Specific Adaptation

The comparison results between the proposed method and other patient-specific baseline methods are presented in Table 1, the baselines follow the same setting as we introduced in [28]. Table 2 compares different patient-specific adaptation strategies on DeepFluoro and Ljubljana, while Fig. 6 further shows the corresponding training dynamics. The patient-specific model trained from scratch serves as the upper-bound baseline,

Table 3 Ablation study of fine-tuned modules during patient-specific adaptation. All settings use the same number of patient-specific synthetic DRRs.

Dataset	Regressor	Similarity network	E-CNN module	SMSR $\uparrow$	Median mTRE (mm) $\downarrow$	Training time (min)
DeepFluoro	✓	×	×	68.5%	0.86	7.4
	×	✓	×	76.8%	0.68	18.7
	✓	✓	×	83.9%	0.57	25.6
	✓	✓	✓	84.7%	0.55	28.3
Ljubljana	✓	×	×	71.0%	0.78	7.8
	×	✓	×	79.5%	0.63	19.2
	✓	✓	×	84.0%	0.54	25.1
	✓	✓	✓	85.0%	0.52	28.9

whereas the patient-agnostic-only model evaluates direct transfer without target-patient adaptation. Direct deployment of the patient-agnostic model leads to a clear performance drop, achieving only 19.6% SMSR on DeepFluoro and 23.4% on Ljubljana. This indicates that although patient-agnostic pretraining learns reusable representations, patient-specific adaptation remains necessary to account for target anatomy and dataset-specific appearance differences. Patient-specific fine-tuning rapidly recovers registration accuracy. On DeepFluoro, SMSR increases from 19.6% without adaptation to 74.3% with 50% fine-tuning, and reaches 84.7% with full fine-tuning, close to the from-scratch result of 86.1%. At the same time, training time is reduced from 20.70 hours to 0.47 hours, corresponding to an approximately 44-fold speed-up. On Ljubljana, 10% fine-tuning already improves SMSR from 23.4% to 70.0%, and full fine-tuning matches the from-scratch SMSR of 85.0% while reducing training time from 20.90 hours to 0.48 hours.

As shown in Fig. 6, training from scratch improves gradually over more than 20 hours, whereas patient-specific fine-tuning reaches near from-scratch performance within a short adaptation period. These results demonstrate that patient-agnostic synthetic pretraining provides an effective initialization for patient-specific 2D/3D registration, converting a costly from-scratch training process into a lightweight adaptation problem while preserving comparable registration accuracy.

4.4 Adaptation Module Analysis

Motivated by recent advances in parameter-efficient fine-tuning of vision foundation models [70–72], we investigate multiple patient-specific adaptation strategies by the set of modules updated during fine-tuning. This allows us to analyze the trade-off between registration accuracy and adaptation cost. Table 3 analyzes which modules should be updated during patient-specific adaptation. Fine-tuning only the pose regressor already improves performance compared with direct patient-agnostic transfer, reaching 68.5% SMSR on DeepFluoro and 71.0% on Ljubljana. This indicates that adapting the initialization network to the target anatomy is beneficial. However, fine-tuning only the similarity network achieves higher accuracy, with SMSR increasing to 76.8% on DeepFluoro and 79.5% on Ljubljana, suggesting that the learned similarity landscape also requires patient-specific adaptation. Updating both the pose regressor and the similarity network provides the best trade-off between accuracy and efficiency. On DeepFluoro, this setting achieves 83.9% SMSR and a median mTRE of 0.57 mm, which is close to full fine-tuning with 84.7% SMSR and 0.55 mm median mTRE. Similarly, on Ljubljana, fine-tuning the regressor and similarity network reaches 84.0% SMSR and 0.54 mm median mTRE, while full fine-tuning achieves 85.0% SMSR and 0.52 mm median mTRE. The additional adaptation of the E-CNN module brings only marginal improvement but increases training time.

These results suggest that most of the patient-specific adaptation benefit comes from updating the pose initialization and similarity estimation components. Therefore, fine-tuning the regressor and similarity network while keeping the volume module fixed offers an efficient adaptation strategy with near full fine-tuning performance.

4.5 Effect of Segmentation-Free Domain Randomization

Table 4 evaluates the effect of different domain randomization strategies during patient-agnostic pretraining. We compare five settings: no domain randomization (No DR), the randomization strategy used in our conference version (Previous DR), intensity and contrast randomization only ($\mathcal{G}_I$), intensity/contrast plus projection-physics randomization ($\mathcal{G}_I + \mathcal{G}_P$), and the proposed full segmentation-free domain randomization strategy. The proposed strategy further includes the mask, occlusion, and field-of-view pool $\mathcal{G}_M$ during the synthetic pretraining process, in addition to $\mathcal{G}_I$ and $\mathcal{G}_P$. Examples of the proposed segmentation-

Table 4 Ablation of domain randomization strategies during patient-agnostic pretraining. Median and 95% denote mTRE statistics.

Setting	SMSR $\uparrow$	Median $\downarrow$	95% $\downarrow$
<i>DeepFluoro</i>			
No DR	76.2%	0.70	2.18
Previous DR [28]	82.7%	0.59	1.70
$\mathcal{G}_I$	79.6%	0.64	1.91
$\mathcal{G}_I + \mathcal{G}_P$	83.2%	0.58	1.65
$\mathcal{G}_I + \mathcal{G}_P + \mathcal{G}_M$	84.7%	0.55	1.58
<i>Ljubljana</i>			
No DR	77.0%	0.70	2.20
Previous DR [28]	83.0%	0.56	1.55
$\mathcal{G}_I$	80.0%	0.62	1.84
$\mathcal{G}_I + \mathcal{G}_P$	83.8%	0.55	1.47
$\mathcal{G}_I + \mathcal{G}_P + \mathcal{G}_M$	85.0%	0.52	1.33

free domain randomization are provided in Fig. 4. Without domain randomization, the pretrained model achieves 76.2% SMSR on DeepFluoro and 77.0% on Ljubljana, indicating that synthetic pretraining alone is insufficient to fully bridge the DRR-to-X-ray appearance gap. Using only intensity and contrast randomization improves

performance to 79.6% and 80.0% SMSR, respectively, showing that photometric perturbations help reduce sensitivity to fixed DRR intensity distributions. Adding projection-physics randomization further improves the results to 83.2% on DeepFluoro and 83.8% on Ljubljana, suggesting that acquisition-related variations such as detector shifts, zoom changes, and bias fields are important for improving synthetic-to-real transfer.

Compared with the previous domain randomization strategy, the proposed full randomization achieves consistently better performance. On DeepFluoro, it improves SMSR from 82.7% to 84.7% and reduces the 95th percentile mTRE from 1.70 mm to 1.58 mm. On Ljubljana, it improves SMSR from 83.0% to 85.0% and reduces the 95th percentile mTRE from 1.55 mm to 1.33 mm. These results indicate that the mask/occlusion/-FOV perturbations in $\mathcal{G}_M$ provide complementary robustness beyond intensity and projection-physics randomization for pretraining. Overall, the proposed segmentation-free domain randomization better diversifies the synthetic pretraining distribution and improves patient-specific adaptation without requiring anatomical segmentation masks or landmark annotations.

5 Conclusion

In this work, we proposed an efficient patient-specific 2D/3D registration framework based on patient-agnostic synthetic pretraining and patient-specific adaptation. The model first learns transferable pose-sensitive representations from synthetic DRRs generated across multiple CT volumes, and is then adapted to a target patient using a limited number of patient-specific synthetic projections. We further introduced a segmentation-free domain randomization strategy to improve synthetic-to-real robustness without requiring anatomical labels or segmentation masks. Experiments on DeepFluoro and Ljubljana showed that direct patient-agnostic transfer is insufficient, while patient-specific fine-tuning rapidly recovers registration accuracy. Compared with training from scratch, the proposed pretraining-adaptation strategy achieved comparable accuracy while reducing patient-specific training time by more than 40 times. Ablation studies further demonstrated the benefits of fine-tuning the pose regressor and similarity network, as well as the

effectiveness of the proposed domain randomization. Future work will validate the method on larger multi-center clinical datasets and investigate faster adaptation strategies for real-time intraoperative deployment. Overall, our results suggest that patient-agnostic synthetic pretraining can transform patient-specific 2D/3D registration from a costly from-scratch training problem into an efficient adaptation problem.

Data availability. We used two publicly available datasets for 2D/3D registration: DeepFluoro (<https://github.com/rg2/DeepFluoroLabeling-IPCAI2020>) and Ljubljana (<https://lit.fe.uni-lj.si/en/research/resources/3D-2D-GS-CA>).

Author contribution. M. C.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review and editing. Y. K.: Supervision, Resources, Project administration, Writing – review and editing.

Declarations

Conflict of interest The authors declare no competing interests.

References

- [1] Grupp, R.B., Armand, M., Taylor, R.H.: Patch-based image similarity for intraoperative 2d/3d pelvis registration during periacetabular osteotomy. In: International Workshop on Computer-Assisted and Robotic Endoscopy, pp. 153–163 (2018). Springer
- [2] Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L., *et al.*: A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE transactions on medical imaging* **17**(4), 586–595 (1998)
- [3] Knaan, D., Joskowicz, L.: Effective intensity-based 2d/3d rigid registration between fluoroscopic x-ray and ct. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 351–358 (2003). Springer
- [4] Frysck, R., Pfeiffer, T., Rose, G.: A novel approach to 2d/3d registration of x-ray images using grangeat’s relation. *Medical Image Analysis* **67**, 101815 (2021)
- [5] De Silva, T., Uneri, A., Ketcha, M., Reaung-amornrat, S., Kleinszig, G., Vogt, S., Aygun, N., Lo, S., Wolinsky, J., Siewerdsen, J.: 3d–2d image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch. *Physics in Medicine & Biology* **61**(8), 3009 (2016)
- [6] Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3d/2d registration methods for image-guided interventions. *Medical image analysis* **16**(3), 642–661 (2012)
- [7] Chen, M., Zhang, Z.: Introducing learning rate adaptation cma-es into rigid intraoperative 2d/3d registration for spinal surgery. In: 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), pp. 1–4 (2025). IEEE
- [8] Chen, M., Li, T., Zhang, Z., Kong, Y.: An optimization-based baseline for rigid 2d/3d registration applied to spine surgical navigation using cma-es. *arXiv preprint arXiv:2402.05642* (2024)
- [9] Toews, M., Wells, W.M.: Phantomless auto-calibration and online calibration assessment for a tracked freehand 2-d ultrasound probe. *IEEE transactions on medical imaging* **37**(1), 262–272 (2017)
- [10] Powell, M.J., *et al.*: The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge, Cambridge **26**, 26–46 (2009)
- [11] Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation* **9**(2), 159–195 (2001)
- [12] Li, T., Chen, M., Li, M., Li, C., Kong, Y.: Automatic x-ray to ct registration using

- embedding reconstruction and lite cross-attention. *Medical Physics* **52**(7), 17896 (2025)
- [13] Miao, S., Wang, Z.J., Liao, R.: A cnn regression approach for real-time 2d/3d registration. *IEEE transactions on medical imaging* **35**(5), 1352–1363 (2016)
- [14] Leroy, A., Cafaro, A., Gessain, G., Champagnac, A., Grégoire, V., Deutsch, E., Lepetit, V., Paragios, N.: Structuregnet: Structure-guided multimodal 2d-3d registration. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 771–780 (2023). Springer
- [15] Zhao, H., Zhu, W., Deng, X., Zhang, G., Zou, W., *et al.*: Automatic 2d/3d spine registration based on two-step transformer with semantic attention and adaptive multi-dimensional loss function. *Biomedical Signal Processing and Control* **95**, 106384 (2024)
- [16] Esteban, J., Grimm, M., Unberath, M., Zahnd, G., Navab, N.: Towards fully automatic x-ray to ct registration. In: *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 631–639 (2019). Springer
- [17] Grimm, M., Esteban, J., Unberath, M., Navab, N.: Pose-dependent weights and domain randomization for fully automatic x-ray to ct registration. *IEEE transactions on medical imaging* **40**(9), 2221–2232 (2021)
- [18] Brandstätter, S., Seeböck, P., Fürböck, C., Pochepnia, S., Prosch, H., Langs, G.: Rigid single-slice-in-volume registration via rotation-equivariant 2d/3d feature matching. In: *International Workshop on Biomedical Image Registration*, pp. 280–294 (2024). Springer
- [19] Markova, V., Ronchetti, M., Wein, W., Zettinig, O., Prevost, R.: Global multi-modal 2d/3d registration via local descriptors learning. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 269–279 (2022). Springer
- [20] Unberath, M., Zaech, J.-N., Lee, S.C., Bier, B., Fotouhi, J., Armand, M., Navab, N.: Deepdr—a catalyst for machine learning in fluoroscopy-guided procedures. In: *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 98–106 (2018). Springer
- [21] Jaganathan, S., Kukla, M., Wang, J., Shetty, K., Maier, A.: Self-supervised 2d/3d registration for x-ray to ct image fusion. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2788–2798 (2023)
- [22] Gopalakrishnan, V., Dey, N., Golland, P.: Intraoperative 2d/3d image registration via differentiable x-ray rendering. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11662–11672 (2024)
- [23] Gao, C., Liu, X., Gu, W., Killeen, B., Armand, M., Taylor, R., Unberath, M.: Generalizing spatial transformers to projective geometry with applications to 2d/3d registration. In: *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 329–339 (2020). Springer
- [24] Gao, C., Feng, A., Liu, X., Taylor, R.H., Armand, M., Unberath, M.: A fully differentiable framework for 2d/3d registration and the projective spatial transformers. *IEEE transactions on medical imaging* (2023)
- [25] Chen, M., Zhang, Z., Gu, S., Ge, Z., Kong, Y.: Fully differentiable correlation-driven 2d/3d registration for x-ray to ct image fusion. In: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5 (2024)
- [26] Chen, M., Zhang, Z., Gu, S., Kong, Y.: Embedded feature similarity optimization with specific parameter initialization for 2d/3d medical image registration. In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1521–1525 (2024). IEEE
- [27] Zhang, B., Faghihroohi, S., Azampour, M.F.,

- Liu, S., Ghotbi, R., Schunkert, H., Navab, N.: A patient-specific self-supervised model for automatic x-ray/ct registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 515–524 (2023). Springer
- [28] Chen, M., Kong, Y.: Intraoperative 2d/3d registration via spherical similarity learning and differentiable levenberg-marquardt optimization. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 7239–7249 (2026)
- [29] Guo, H., Xu, X., Xu, S., Wood, B.J., Yan, P.: End-to-end ultrasound frame to volume registration. In: International Conference on Medical Image Computing and Computer Assisted Intervention, pp. 56–65 (2021). Springer
- [30] Miao, S., Wang, Z.J., Zheng, Y., Liao, R.: Real-time 2d/3d registration via cnn regression. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1430–1434 (2016). IEEE
- [31] Liao, H., Lin, W.-A., Zhang, J., Zhang, J., Luo, J., Zhou, S.K.: Multiview 2d/3d rigid registration via a point-of-interest network for tracking and triangulation. In: In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12638–12647 (2019)
- [32] Shrestha, P., Xie, C., Shishido, H., Yoshii, Y., Kitahara, I.: X-ray to ct rigid registration using scene coordinate regression. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 781–790 (2023). Springer
- [33] Shrestha, P., Xie, C., Yoshii, Y., Kitahara, I.: Rayemb: Arbitrary landmark detection in x-ray images using ray embedding subspace. In: Proceedings of the Asian Conference on Computer Vision, pp. 665–681 (2024)
- [34] Li, S., Xu, C., Xie, M.: A robust $O(n)$ solution to the perspective-n-point problem. *IEEE transactions on pattern analysis and machine intelligence* **34**(7), 1444–1450 (2012)
- [35] Jaganathan, S., Wang, J., Borsdorf, A., Shetty, K., Maier, A.: Deep iterative 2d/3d registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 383–392 (2021). Springer
- [36] Schaffert, R., Wang, J., Fischer, P., Borsdorf, A., Maier, A.: Learning an attention model for robust 2-d/3-d registration using point-to-plane correspondences. *IEEE transactions on medical imaging* **39**(10), 3159–3174 (2020)
- [37] Wang, J., Schaffert, R., Borsdorf, A., Heigl, B., Huang, X., Hornegger, J., Maier, A.: Dynamic 2-d/3-d rigid registration framework using point-to-plane correspondence model. *IEEE transactions on medical imaging* **36**(9), 1939–1954 (2017)
- [38] Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**(6), 381–395 (1981)
- [39] Downs, C., Sluijs, P., Cornelissen, S.A., Nijenhuis, F.t., Zwam, W.H.v., Gopalakrishnan, V., Zhang, X., Su, R., Walsum, T.v.: Improving automatic cerebral 3d-2d cta-dsa registration. *International Journal of Computer Assisted Radiology and Surgery*, 1–10 (2025)
- [40] Gopalakrishnan, V., Dey, N., Chlorogianis, D.-D., Abumoussa, A., Larson, A.M., Orbach, D.B., Frisken, S., Golland, P.: Rapid patient-specific neural networks for intraoperative X-ray to volume registration (2025). <https://arxiv.org/abs/2503.16309>
- [41] Gopalakrishnan, V., Golland, P.: Fast auto-differentiable digitally reconstructed radiographs for solving inverse problems in intraoperative imaging. In: Workshop on Clinical Image-Based Procedures, pp. 1–11 (2022). Springer
- [42] Toth, D., Cimen, S., Ceccaldi, P., Kurzendorfer, T., Rhode, K., Mountney, P.: Training deep networks on domain randomized synthetic x-ray data for cardiac interventions. In:

- International Conference on Medical Imaging with Deep Learning, pp. 468–482 (2019). PMLR
- [43] Gopinath, K., Hoopes, A., Alexander, D.C., Arnold, S.E., Balbastre, Y., Billot, B., Casamitjana, A., Cheng, Y., Chua, R.Y.Z., Edlow, B.L., *et al.*: Synthetic data in generalizable, learning-based neuroimaging. *Imaging Neuroscience* **2**, 2 (2024)
 - [44] Dey, N., Billot, B., Wong, H.E., Wang, C., Ren, M., Grant, E., Dalca, A.V., Golland, P.: Learning general-purpose biomedical volume representations using randomized synthesis. In: The Thirteenth International Conference on Learning Representations (2025)
 - [45] Fu, J., Zheng, Y., Dey, N., Ferreira, D., Moreno, R.: Synthesizing individualized aging brains in health and disease with generative models and parallel transport. *Medical Image Analysis* **105**, 103669 (2025)
 - [46] Billot, B., Greve, D.N., Van Leemput, K., Fischl, B., Iglesias, J.E., Dalca, A.V.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. *Medical Image Analysis* **86**, 102789 (2023)
 - [47] Billot, B., Magdamo, C., Cheng, Y., Arnold, S.E., Das, S., Iglesias, J.E.: Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain mri datasets. *Proceedings of the National Academy of Sciences* **120**(9), 2216399120 (2023)
 - [48] Laso, P., Cerri, S., Sorby-Adams, A., Guo, J., Mateen, F., Goebel, P., Wu, J., Liu, P., Li, H.B., Young, S.I., *et al.*: Quantifying white matter hyperintensity and brain volumes in heterogeneous clinical and low-field portable mri. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1–5 (2024). IEEE
 - [49] Hoopes, A., Mora, J.S., Dalca, A.V., Fischl, B., Hoffmann, M.: Synthstrip: skull-stripping for any brain image. *NeuroImage* **260**, 119474 (2022)
 - [50] Hoffmann, M., Billot, B., Greve, D.N., Iglesias, J.E., Fischl, B., Dalca, A.V.: Synthmorph: learning contrast-invariant registration without acquired images. *IEEE Transactions on Medical Imaging* **41**(3), 543–558 (2022)
 - [51] Hoffmann, M., Hoopes, A., Greve, D.N., Fischl, B., Dalca, A.V.: Anatomy-aware and acquisition-agnostic joint registration with synthmorph. *Imaging Neuroscience* **2**, 1–33 (2024)
 - [52] Unberath, M., Zaech, J.-N., Gao, C., Bier, B., Goldmann, F., Lee, S.C., Fotouhi, J., Taylor, R., Armand, M., Navab, N.: Enabling machine learning in x-ray-based procedures via realistic simulation of image formation. *International journal of computer assisted radiology and surgery* **14**(9), 1517–1528 (2019)
 - [53] Gao, C., Killeen, B.D., Hu, Y., Grupp, R.B., Taylor, R.H., Armand, M., Unberath, M.: Synthetic data accelerates the development of generalizable learning-based algorithms for x-ray image analysis. *Nature Machine Intelligence* **5**(3), 294–308 (2023)
 - [54] Huang, Z., Li, H., Shao, S., Zhu, H., Hu, H., Cheng, Z., Wang, J., Kevin Zhou, S.: Pele scores: pelvic x-ray landmark detection with pelvis extraction and enhancement. *International Journal of Computer Assisted Radiology and Surgery* **19**(5), 939–950 (2024)
 - [55] Zakharov, S., Ambrus, R., Guizilini, V., Kehl, W., Gaidon, A.: Photo-realistic neural domain randomization. In: European Conference on Computer Vision, pp. 310–327 (2022). Springer
 - [56] Qin, C., Shi, B., Liao, R., Mansi, T., Rueckert, D., Kamen, A.: Unsupervised deformable registration for multi-modal images via disentangled representations. In: International Conference on Information Processing in Medical Imaging, pp. 249–261 (2019). Springer
 - [57] Grzech, D., Azampour, M.F., Glocker, B.,

- Schnabel, J., Navab, N., Kainz, B., Le Folgoc, L.: A variational bayesian method for similarity learning in non-rigid image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 119–128 (2022)
- [58] Ronchetti, M., Wein, W., Navab, N., Zettinig, O., Prevost, R.: Disa: Differentiable similarity approximation for universal multimodal registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 761–770 (2023). Springer
- [59] Grzech, D., Folgoc, L.L., Azampour, M.F., Vlontzos, A., Glocker, B., Navab, N., Schnabel, J., Kainz, B.: Unsupervised similarity learning for image registration with energy-based models. In: International Workshop on Biomedical Image Registration, pp. 229–240 (2024). Springer
- [60] Sideri-Lampretsa, V., Zimmer, V.A., Qiu, H., Kaissis, G., Rueckert, D.: Mad: Modality agnostic distance measure for image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 147–156 (2023). Springer
- [61] Mok, T.C., Li, Z., Bai, Y., Zhang, J., Liu, W., Zhou, Y.-J., Yan, K., Jin, D., Shi, Y., Yin, X., *et al.*: Modality-agnostic structural image representation learning for deformable multimodality medical image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11215–11225 (2024)
- [62] Gu, W., Gao, C., Grupp, R., Fotouhi, J., Unberath, M.: Extended capture range of rigid 2d/3d registration by estimating riemannian pose gradients. In: Machine Learning in Medical Imaging: 11th International Workshop (2020). Springer
- [63] Billot, B., Dey, N., Moyer, D., Hoffmann, M., Turk, E.A., Gagoski, B., Grant, P.E., Golland, P.: Se (3)-equivariant and noise-invariant 3d rigid motion tracking in brain mri. *IEEE transactions on medical imaging* (2024)
- [64] Moyer, D., Abaci Turk, E., Grant, P.E., Wells, W.M., Golland, P.: Equivariant filters for efficient tracking in 3d imaging. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24, pp. 193–202 (2021). Springer
- [65] Bui, M., Albarqouni, S., Schrapp, M., Navab, N., Ilic, S.: X-ray posenet: 6 dof pose estimation for mobile x-ray devices. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1036–1044 (2017). IEEE
- [66] Grupp, R.B., Unberath, M., Gao, C., Hege-man, 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. *International journal of computer assisted radiology and surgery* **15**, 759–769 (2020)
- [67] Mitrović, U., Špiclin, Z., Likar, B., Pernuš, F.: 3d-2d registration of cerebral angiograms: A method and evaluation on clinical images. *IEEE Transactions on Medical Imaging* **32**(8), 1550–1563 (2013)
- [68] Momeni, M., Gopalakrishnan, V., Dey, N., Golland, P., Frisken, S.: Differentiable voxel-based x-ray rendering improves sparse-view 3d cbct reconstruction. *Machine Learning and the Physical Sciences, NeurIPS 2024* (2024)
- [69] Siddon, R.L.: Fast calculation of the exact radiological path for a three-dimensional ct array. *Medical physics* **12**(2), 252–255 (1985)
- [70] Kong, Y., Li, H., Zhang, S., Chen, M.: Pro-tuning: Prototype tuning of foundation models for volumetric medical image segmentation. *IEEE Transactions on Biomedical Engineering* (2026)
- [71] Lin, X., Xiang, Y., Yu, L., Yan, Z.: Beyond

adapting sam: Towards end-to-end ultrasound image segmentation via auto prompting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 24–34 (2024). Springer

- [72] Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., *et al.*: Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. *Medical Image Analysis* **98**, 103310 (2024)