ALFA: Biplanar X-ray Reconstruction via Anatomy-Latent Field Adaptation

Tianqi Yu[1], Chenhe Du[1], Jie Wen[1], Hongjiang Wei[2], and Yuyao Zhang[1]

[1] School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[2] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
zhangyy8@ashanghaitech.edu.com

Abstract. Biplanar X-ray 3D reconstruction offers a low-dose alternative to CT but remains a severely ill-posed inverse problem. Existing supervised end-to-end models often overfit to fixed geometries, suffering catastrophic performance degradation when acquisition parameters, such as Source-to-Image Distance (SID), change. We present ALFA, a novel decoupled paradigm that separates anatomical prior learning from test-time physical adaptation. First, ALFA constructs a resolution-agnostic latent space using a shared Implicit Neural Representation (INR) to encode diverse subject-specific morphologies into a continuous manifold. Reconstruction is then achieved via geometry-aware latent optimization, using a differentiable forward model to align projections under true imaging geometry. Unlike traditional black-box mappings, ALFA faithfully recovers patient-specific pathologies, such as mandibular deformities, by retrieving morphological traits from the manifold rather than reverting to population averages. Evaluations demonstrate superior zero-shot generalization across diverse SIDs and robust stability under clinical non-idealities, including scattering and noise. ALFA provides a flexible, 'plug-and-play' foundation for high-fidelity 3D clinical assessment in heterogeneous imaging environments.

Keywords: Biplanar X-ray Reconstruction · Implicit Neural Representation.

1 Introduction

Three-dimensional (3D) anatomical reconstruction from biplanar X-ray imaging aims to recover CT-like volumetric information while preserving the low radiation dose and accessibility of conventional radiography. By leveraging orthogonal posteroanterior (PA) and lateral (LAT) projections, biplanar reconstruction has emerged as a promising alternative to CT for orthopedic assessment, surgical planning, and longitudinal monitoring[14, 11, 16, 1]. However, recovering accurate 3D anatomy from only two projections is a severely ill-posed inverse problem due to massive information loss[2, 13].

A clinically reliable reconstruction must satisfy two key requirements: *anatomical consistency*, ensuring realistic morphology aligned with human anatomy, and *geometric consistency*, guaranteeing that forward projections match real measurements under true imaging geometry. Existing supervised end-to-end mappings[10, 8, 6, 19, 3] typically act as geometry-coupled black boxes, entangling anatomical priors with specific imaging parameters (e.g., SID). Consequently, they suffer from catastrophic performance degradation across varying clinical settings[17] and often produce "mean-shape" hallucinations that obscure critical patient-specific pathologies, such as mandibular deformities essential for orthodontic diagnosis.

Implicit Neural Representations (INRs)[21] model volumes as continuous, resolution-agnostic functions[22, 15, 9, 23]. While effective for sparse-view CT[20, 25], INR regularization alone is insufficient for the extreme sparsity of biplanar imaging. Hybrid approaches[5, 4, 12, 7] improve quality but remain computationally intensive. Notably, existing benchmarks often operate on fixed low-resolution grids (e.g., $\leq 64^3$), whereas the continuous nature of INRs offers a more flexible foundation for clinical-grade detail recovery.

To address these bottlenecks, we propose ALFA, a two-stage framework that introduces a paradigm shift by decoupling anatomical prior learning from physics-driven geometric adaptation. ALFA first constructs a continuous anatomical manifold via supervised auto-decoding on 3D CT volumes. In the second stage, ALFA achieves reconstruction through physics-guided latent optimization using a differentiable forward model. By freezing the pretrained decoder, ALFA acts as a "plug-and-play" solution that retrieves the optimal latent code satisfying both the learned anatomical manifold and the observed physical constraints, ensuring the recovery of patient-specific traits across unseen scanner configurations without retraining.

Our contributions are summarized as follows:

- We introduce a decoupled reconstruction paradigm that separates geometry agnostic prior learning from test-time physical adaptation. We experimentally demonstrate its -shot generalization capability across diverse, unseen imaging geometries (SID 800/1200mm).
- We leverage an INR-based latent space to faithfully recover patient-specific pathologies (e.g., severe underbite and overbite). We show that ALFA preserves these critical diagnostic features where traditional end-to-end methods revert to population averages.
- We conduct a rigorous stress test under simulated clinical non-idealities, proving that ALFA maintains exceptional robustness against scattering, Poisson noise, and extrinsic geometric misalignments through its constrained optimization on a pre-trained manifold.

2 Method

2.1 Problem Formulation

In CT imaging, X-ray attenuation follows Lambert–Beer's law. After logarithmic transformation, the detected intensity I along a ray $\mathbf{r}$ simplifies to a line integral

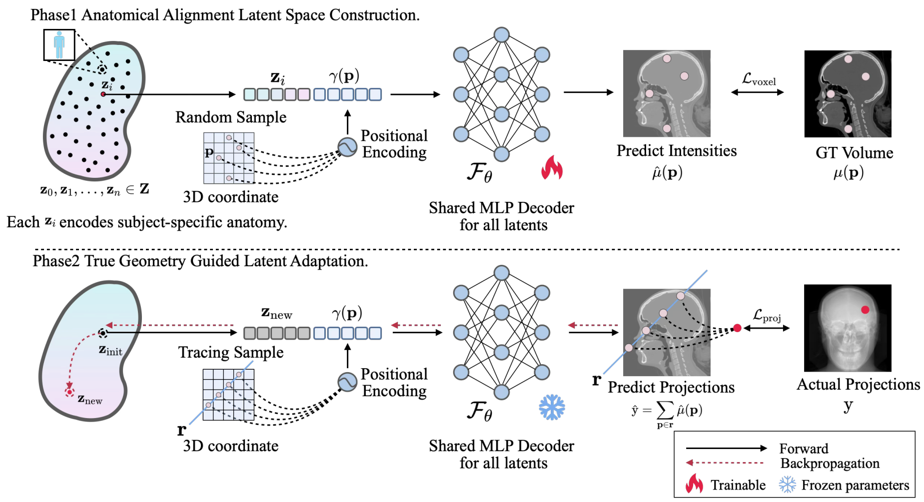


Fig. 1. Overview of ALFA. **Phase 1:** Joint optimization of latent codes $\mathbf{z}_i$ and an INR decoder $\mathcal{F}_\theta$ constructs a geometry-agnostic anatomical prior. **Phase 2:** With $\mathcal{F}_\theta$ frozen, a new latent code is iteratively optimized via differentiable forward projection to match the measured biplanar X-rays, and then decoded into the final 3D volume.

of the linear attenuation coefficient (LAC) field $\mu(\mathbf{p})$:

$$\mathbf{y} = \int_{\mathbf{r}} \mu(\mathbf{p})\, \mathrm{d}\mathbf{p} + \mathbf{n} \approx \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the discretized volume and $\mathbf{A}$ is the projection operator [2]. In biplanar imaging, we observe only two orthogonal projections, $\mathbf{y}_{\text{bip}} = [\mathbf{A}_{\text{PA}}^\top, \mathbf{A}_{\text{LAT}}^\top]^\top \mathbf{x} + \mathbf{n}$. This reconstruction is severely underdetermined, requiring strong anatomical priors to recover the missing volumetric information.

2.2 Spatially Continuous Anatomical Latent Space

We employ Implicit Neural Representations (INRs) to model the attenuation field as a continuous function $\mathcal{F}_\theta$. To represent population-level variability, each subject i is embedded into a latent space via a vector $\mathbf{z}_i$. The LAC at any coordinate $\mathbf{p}$ is queried as:

$$\hat{\mu}_i(\mathbf{p}) = \mathcal{F}_\theta(\gamma(\mathbf{p}), \mathbf{z}_i), \tag{2}$$

where $\gamma(\cdot)$ denotes Fourier positional encoding [21]. During the first stage, we jointly optimize the shared decoder θ and subject-specific latents $\{\mathbf{z}_i\}$ by minimizing the voxel-wise loss $\mathcal{L}_{\text{voxel}} = \mathbb{E}_{\mathbf{p}}[|\hat{\mu}_i(\mathbf{p}) - \mu_i(\mathbf{p})|]$ over CT volumes.

This joint optimization forces the decoder $\mathcal{F}_\theta$ to learn a universal anatomical prior, while $\mathbf{z}_i$ captures individual morphological variations. Under the manifold assumption, the latent space effectively parameterizes a low-dimensional function space $v_i(\mathbf{p}) = \Phi(\mathbf{p}; \mathbf{z}_i)$, where the inherent structural similarity in medical images provides implicit regularization for the learned representation.

2.3 Geometry-Guided Latent Adaptation

For reconstruction from new biplanar X-rays, the pretrained decoder $\mathcal{F}_\theta$ is frozen to act as a strict anatomical constraint. We solve the inverse problem by optimizing only the latent vector $\mathbf{z}$ to align with the observed geometry. For a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, the projected pixel $\hat{y}$ is calculated via differentiable rendering:

$$\hat{y} = \sum_j \mathcal{F}_\theta(\mathbf{r}(t_j), \mathbf{z})\Delta t_j. \tag{3}$$

The latent code is adapted by minimizing the projection error

$$\mathcal{L}_{\text{proj}} = \sum_{k \in \{\text{PA, LAT}\}} \|\hat{\mathbf{y}}_k - \mathbf{y}_k\|_1. \tag{4}$$

The optimized latent $\mathbf{z}^*$ is then queried across the coordinate grid to produce a 3D volume $\hat{\mu}^*(\mathbf{p}) = \mathcal{F}_\theta(\gamma(\mathbf{p}), \mathbf{z}^*)$. This process ensures that the reconstruction is simultaneously consistent with the learned anatomical manifold and the specific imaging geometry.

3 Experiments

3.1 Compared Methods and Metrics

We evaluate ALFA against two representative baselines: X2CT-GAN[24] and PerX2CT[8]. Reconstruction quality is quantitatively assessed using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM).

3.2 Datasets and Experimental Setup

We utilize a clinical dataset of 800 head CT scans (split into 600/100/100 for training/test/validation) featuring diverse orthodontic phenotypes, including severe mandibular deformities (e.g., underbite and overbite). Biplanar X-ray data were simulated at 128^3 resolution.

Notably, this resolution is chosen to establish a rigorous benchmark against state-of-the-art supervised methods, specifically X2CT-GAN (128^3) and PerX2CT (typically $\leq 64^3$). While the current evaluation is performed on a discrete grid for comparative purposes, our INR-based framework is inherently resolution-agnostic, allowing for continuous coordinate querying and potential scaling to higher-fidelity reconstructions. More importantly, ALFA only utilizes 600 unpaired CT volumes, while supervised baselines require paired X-ray/CT data.

To validate ALFA's plug-and-play adaptability across varying clinical Source-to-Image Distances (SIDs), we evaluate performance under three distinct training-testing scenarios:

S1 **Identical Conditions**: Models are trained and tested on the same parallel-beam geometry to establish a baseline for raw reconstruction fidelity.

S2 **Zero-shot Cross-Geometry**: Supervised baselines trained exclusively on parallel-beams are tested on unseen cone-beam geometries (SID: 800/1200mm). This assesses ALFA's geometry-agnostic nature against the inherent sensitivity of end-to-end mappings to domain shifts.

S3 **Mixed-Condition Robustness**: To ensure a fair comparison, baselines are also trained on a heterogeneous dataset containing all projection geometries (parallel, SID 800/1200mm). This benchmarks ALFA's physics-guided adaptation against models with prior exposure to all imaging setups.

3.3 Implementation Details

Our framework employs 8-band Fourier positional encoding and a 128D latent vector follow DeepSDF[18]. The decoder is an MLP with 8 hidden layers (512 units each). For phase 2 inference, the latent code is optimized for 32 epochs using the Adam optimizer with a learning rate of 10^{-2}.

Experiments were conducted on a single NVIDIA GeForce RTX 4090 GPU. Training the shared decoder took 18 hours, while inference via test-time adaptation required only 15 seconds per case. This computational overhead is clinically acceptable, prioritizing anatomical fidelity and zero-shot flexibility over the fixed-geometry speed of end-to-end black boxes.

4 Results

4.1 Identical and Zero-shot Cross-Geometry Comparison

As illustrated in Fig. 2, baseline methods fail to resolve critical diagnostic features, often producing "mean-shape" hallucinations that obscure patient-specific mandibular deformities. In contrast, ALFA achieves superior anatomical fidelity, accurately recovering complex sagittal relationships (e.g., underbite and deep overbite) essential for orthodontic planning. This suggests that ALFA's latent optimization effectively constrains the reconstruction within a clinically valid anatomical manifold, retrieving specific pathological traits from the pre-trained prior rather than reverting to population averages.

Quantitatively, Tab. 1 highlights the robustness of ALFA across varying geometries. While supervised baselines perform competitively under identical settings (S1), they suffer catastrophic performance degradation in zero-shot evaluations (S2), with PSNR dropping from $\sim$25 dB to $\sim$15 dB. This failure stems from their reliance on geometry-coupled direct mappings. Conversely, by incorporating a physical forward model, ALFA demonstrates geometry-agnostic generalization, maintaining near-identical performance across both parallel and cone-beam geometries without retraining.

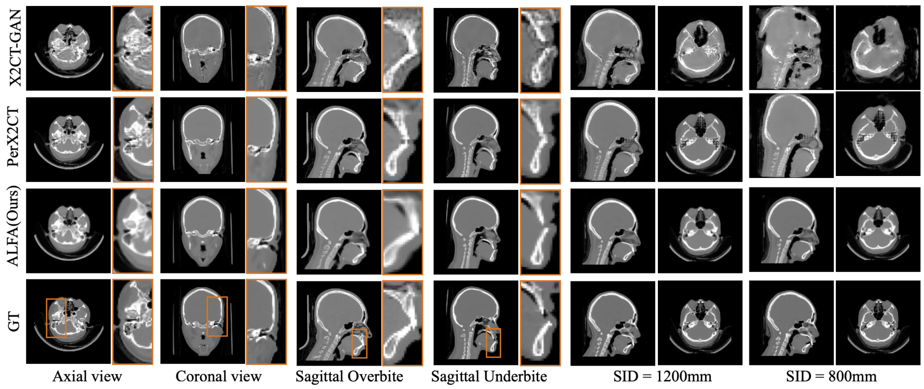


Fig. 2. Qualitative comparison under idealized fixed-geometry (S1) and zero-shot cross-geometry (S2) conditions (SID: 800/1200mm). ALFA preserves patient-specific pathologies where baselines produce blurred averages.

Table 1. Quantitative comparison across scenarios S1, S2, and S3. S1: Identical conditions; S2: Zero-shot generalization (Avg. of 800/1200mm); S3: Mixed-condition training. PSNR is measured in dB. Best results are in bold.

Method	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑
	S1: Identical		**S2: Zero-shot (Avg.)**	
X2CT-GAN[24]	$0.867 \pm .030$	25.16 ± 0.93	$0.494 \pm .037$	15.46 ± 0.30
PerX2CT[8]	$0.851 \pm .034$	23.96 ± 1.05	$0.511 \pm .038$	15.12 ± 0.55
ALFA (Ours)	$\mathbf{0.891 \pm .036}$	$\mathbf{27.21 \pm 1.14}$	$\mathbf{0.891 \pm .035}$	$\mathbf{27.17 \pm 1.21}$
	S3: Cone-800mm		**S3: Cone-1200mm**	
X2CT-GAN[24]	$0.814 \pm .035$	23.09 ± 0.92	$0.817 \pm .040$	23.20 ± 1.15
PerX2CT[8]	$0.823 \pm .032$	22.84 ± 0.81	$0.825 \pm .033$	22.89 ± 0.87
ALFA (Ours)	$\mathbf{0.887 \pm .037}$	$\mathbf{26.95 \pm 1.31}$	$\mathbf{0.895 \pm .033}$	$\mathbf{27.38 \pm 1.10}$

4.2 Comparison under Mixed-Condition Robustness

As shown in Fig. 3, end-to-end baselines degrade significantly even when trained on mixed-geometry datasets (S3) due to geometric ambiguity. Specifically, direct 2D-to-3D mappings lose injectivity when an identical projection corresponds to different 3D spatial representations across varying SIDs, leading to reconstruction collapse.

ALFA fundamentally circumvents this by decoupling anatomical priors from imaging physics. Rather than learning entangled mappings, ALFA explicitly enforces geometric consistency by integrating an analytical forward model into its test-time adaptation. Quantitatively (Tab. 1), ALFA achieves the highest PSNR/SSIM with minimal variance across all SIDs. This stability proves that ALFA's physics-guided paradigm is inherently robust to geometric domain shifts,

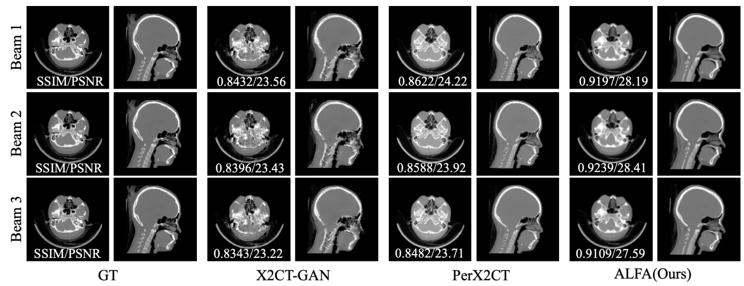


Fig. 3. Reconstruction consistency across diverse acquisition geometries. ALFA maintains stable, geometry-agnostic performance and superior structural fidelity compared to baselines under mixed-condition training.

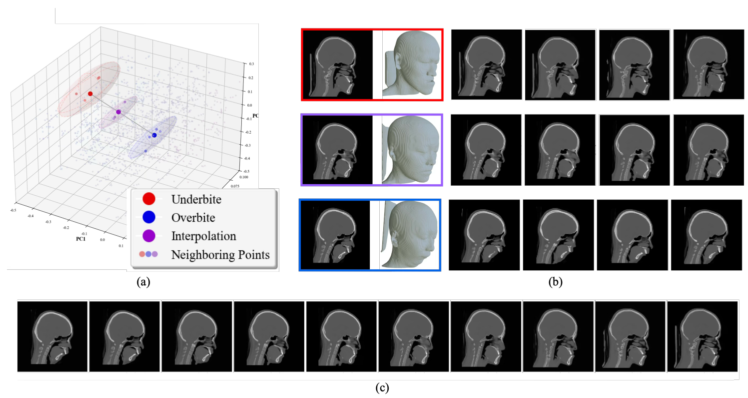


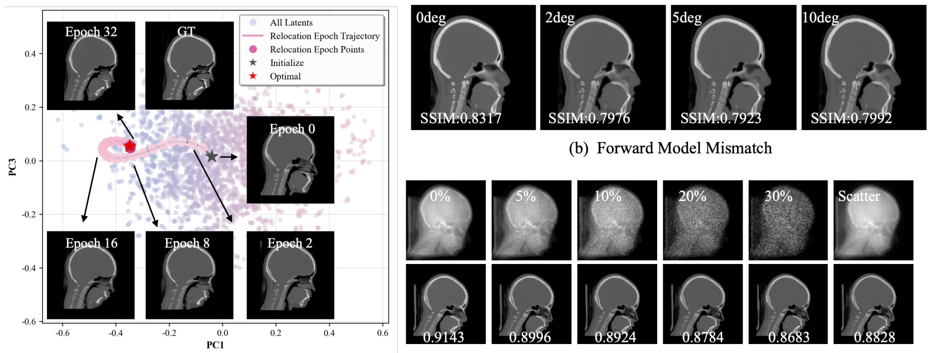
Fig. 4. Visualization of the anatomically consistent latent space. (a) 3D distribution of the latent manifold; (b) Reconstructions from phenotypic centers; (c) Linear interpolation between deep overbite and underbite showing smooth transitions through normal morphology.

providing a reliable "plug-and-play" solution for heterogeneous clinical environments.

5 Discussion

5.1 Latent Space Characterization

PCA was performed on the training set latent codes to characterize the learned anatomical manifold (Fig. 4). By projecting the optimized latents of extreme phenotypes—severe underbite and overbite—into this space, we observe a semantically structured distribution where morphological traits are organized continuously. Linear interpolation between these extremes yields intermediate codes representing normal morphology, confirming that the latent space encodes smooth



(a) Latent space adaptation trajectory

(b) Forward Model Mismatch

(c) Poisson Noise and Scatter

Fig. 5. Analysis of latent space adaptation and robustness. (a) PCA trajectory showing smooth convergence from random initialization; (b–c) Performance stability under forward model mismatch, Poisson noise, and scattering.

and clinically meaningful anatomical transitions (Fig. 4c). Furthermore, sampling latent vectors around these phenotypic centers produces volumes that consistently retain their respective pathological traits. This verifies that the representation is anatomically coherent yet highly expressive, capturing patient-specific deformities rather than merely reverting to a population average.

5.2 Optimization Stability and Robustness Analysis

ALFA exhibits stable convergence during test-time adaptation. As shown by the PCA trajectory (Fig. 5a), latent codes evolve along a smooth, consistent path toward the target representation, even when the search space is expanded by two orders of magnitude. This stability reflects the well-structured, convex-like optimization landscape within the learned anatomical manifold.

To assess clinical viability, we conducted stress tests under three non-idealities: geometric misalignment, Poisson noise, and scattering. Results in Fig. 5b–c show that ALFA maintains high fidelity under significant signal degradation. This robustness is attributed to the inherent constraints of the pre-trained manifold, which regularizes the ill-posed inverse problem by restricting the solution space to valid anatomical configurations. By mitigating artifacts that typically cause unconstrained or end-to-end methods to fail, ALFA demonstrates strong potential for reliable reconstruction in non-idealized clinical environments.

6 Conclusion

We presented ALFA, a decoupled paradigm for biplanar 3D reconstruction that separates anatomical prior learning from physical adaptation. By optimizing within a resolution-agnostic latent space via an analytical forward model, ALFA circumvents geometry-specific overfitting and enables zero-shot generalization across diverse clinical setups without retraining. Crucially, ALFA faithfully preserves patient-specific pathologies and maintains high robustness against clinical noise and geometric misalignment. As a physics-consistent, "plug-and-play" solution, ALFA offers a reliable and flexible foundation for high-fidelity 3D clinical assessment from sparse X-ray views.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 62571328).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arn Roth, T., Jokeit, M., Sutter, R., Vlachopoulos, L., Fucentese, S.F., Carrillo, F., Snedeker, J.G., Esfandiari, H., Fürnstahl, P.: Deep-learning based 3d reconstruction of lower limb bones from biplanar radiographs for preoperative osteotomy planning. International Journal of Computer Assisted Radiology and Surgery **19**(9), 1843–1853 (2024)
2. Buzug, T.M.: Computed tomography: from photon statistics to modern cone-beam ct. Soc Nuclear Med **50**(7) (2009)
3. Chen, C.C., Fang, Y.H.: Using bi-planar x-ray images to reconstruct the spine structure by the convolution neural network. In: International Conference on Biomedical and Health Informatics. pp. 80–85. Springer (2019)
4. Chu, J., Du, C., Lin, X., Zhang, X., Wang, L., Zhang, Y., Wei, H.: Highly accelerated mri via implicit neural representation guided posterior sampling of diffusion models. Medical Image Analysis **100**, 103398 (2025)
5. Du, C., Lin, X., Wu, Q., Tian, X., Su, Y., Luo, Z., Zheng, R., Chen, Y., Wei, H., Zhou, S.K., et al.: Dper: Diffusion prior driven neural representation for limited angle and sparse view ct reconstruction. arXiv preprint arXiv:2404.17890 (2024)
6. Huang, B., Pei, Y.: Generalizable structure-aware inf: Biplanar-view ct reconstruction via disentangled implicit neural field. In: Proceedings of the Asian Conference on Computer Vision. pp. 699–715 (2024)
7. Hui, M., Wei, Z., Zhu, H., Xia, F., Zhou, Y.: Microdiffusion: Implicit representation-guided diffusion for 3d reconstruction from limited 2d microscopy projections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11460–11469 (2024)
8. Kyung, D., Jo, K., Choo, J., Lee, J., Choi, E.: Perspective projection-based 3d ct reconstruction from biplanar x-rays. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
9. Lee, J., Baek, J.: Iterative reconstruction for limited-angle ct using implicit neural representation. Physics in Medicine & Biology **69**(10), 105008 (2024)
10. Liu, K., Ying, Z., Jin, J., Li, D., Huang, P., Wu, W., Chen, Z., Qi, J., Lu, Y., Deng, L., et al.: Swin-x2s: Reconstructing 3d shape from 2d biplanar x-ray with swin transformers. arXiv preprint arXiv:2501.05961 (2025)
11. Loisel, F., Durand, S., Goubier, J.N., Bonnet, X., Rouch, P., Skalli, W.: Three-dimensional reconstruction of the hand from biplanar x-rays: Assessment of accuracy and reliability. Orthopaedics & Traumatology: Surgery & Research **109**(6), 103403 (2023)

12. Lyu, J., Wang, G., Wang, Z., Dong, S., Ding, W., Wang, C.: Diffusion-prior based implicit neural representation for arbitrary-scale cardiac cine mri super-resolution. Information Fusion p. 103510 (2025)
13. Maken, P., Gupta, A.: 2d-to-3d: a review for computational 3d image reconstruction from x-ray images. Archives of Computational Methods in Engineering **30**(1), 85–114 (2023)
14. Melhem, E., Assi, A., El Rachkidi, R., Ghanem, I.: Eos® biplanar x-ray imaging: concept, developments, benefits, and limitations. Journal of children's orthopaedics **10**(1), 1–14 (2016)
15. Molaei, A., Aminimehr, A., Tavakoli, A., Kazerouni, A., Azad, B., Azad, R., Merhof, D.: Implicit neural representation in medical imaging: A comparative survey. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2381–2391 (2023)
16. Nguyen, D.C.T., Benameur, S., Mignotte, M., Lavoie, F.: 3d biplanar reconstruction of lower limbs using nonlinear statistical models. Medical & Biological Engineering & Computing **61**(11), 2877–2894 (2023)
17. Nguyen, V.T.H.: Advances in biplanar X-ray imaging: calibration and 2D/3D registration. Ph.D. thesis, University of Antwerp (2022)
18. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 165–174 (2019)
19. Ratul, M.A.R., Yuan, K., Lee, W.: Ccx-raynet: a class conditioned convolutional neural network for biplanar x-rays to ct volume. In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI). pp. 1655–1659. IEEE (2021)
20. Shi, J., Zhu, J., Pelt, D.M., Batenburg, K.J., Blaschko, M.B.: Implicit neural representations for robust joint sparse-view ct reconstruction. arXiv preprint arXiv:2405.02509 (2024)
21. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
22. Wu, Q., Feng, R., Wei, H., Yu, J., Zhang, Y.: Self-supervised coordinate projection network for sparse-view computed tomography. IEEE Transactions on Computational Imaging **9**, 517–529 (2023)
23. Yang, S., Ding, M., Wu, Y., Li, Z., Zhang, J.: Implicit neural representation for cooperative low-light image enhancement. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12918–12927 (2023)
24. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2ct-gan: reconstructing ct from biplanar x-rays with generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10619–10628 (2019)
25. Zang, G., Idoughi, R., Li, R., Wonka, P., Heidrich, W.: Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1960–1970 (2021)