From Sparsity to Geometry: Spatial Modeling for 3D Reconstruction from Biplanar Bone X-rays

Nuo'er Long[1 [0009−0003−8719−7722]], Songkang Liu[1,3 [0009−0003−8241−9736]], Cheng Chen[4,1 [0000−0002−6040−6833]], Tao Tan[5 [0000−0001−5403−0887]], Michael K. T. To[2,3 [0000−0001−6853−0591]], and Peikai Chen[1,2,3 ✉[0000−0003−1880−0893]]

[1] AIBD Lab, The University of Hong Kong - Shenzhen Hospital, Shenzhen, China
[2] Dept Orthopedics, HKU - Shenzhen Hospital, Shenzhen, China
[3] Shenzhen Clinical Research Center for Rare Diseases, Shenzhen, China
[4] The University of Hong Kong, Hong Kong, China
[5] Macao Polytechnic University, Macao, China
✉ pkchen@hku-szh.org

Abstract. Three-dimensional (3D) reconstruction from biplanar X-rays has high clinical values and remains technically challenging. Existing methods emphasize larger datasets and stronger networks, with less focus on spatial modeling. This usually works well for X-ray inputs with dense structural cues (e.g., chest). However, for inputs with sparse foreground (e.g., long-bone), the impact of geometry unawareness might be significant and introducing such constraints may be helpful. Here, we introduce ExpGS (Explicit Gaussian Splatting), a spatial modeling framework that uses a fixed virtual biplanar geometry to eliminate per-case camera metadata at inference. By preserving non-degenerate anteroposterior/lateral ray constraints within a consistent frustum-intersection volume, it reduces background interference and stabilizes reconstruction. ExpGS reduces to canonical and implicit modeling as geometric constraints relax. Experiments were performed on two CT datasets with distinct content-sparsities, Femur (long bones) and LIDC (chest), whereby synthetic X-rays and matched virtual geometry were obtained from CT projections for training and testing. Results show that spatial modeling confers only minor benefits to information-dense LIDC but substantially improves reconstruction stability (SSIM 86.13%) on the sparse Femur dataset, comparing both with other methods and with the two ablated forms of ExpGS. Our work establishes the importance of geometric constraints for content-sparse planar inputs such as long bones. Implementation is available at: https://github.com/HKUSZH/ExpGS.

Keywords: Biplanar X-rays · 3D bone reconstruction · Spatial modeling · Information sparsity.

1 Introduction

Accurate 3D bone reconstruction is highly desirable in orthopedics [17]. 3D reconstruction from planar radiographics (and not CTs) has further dosage and

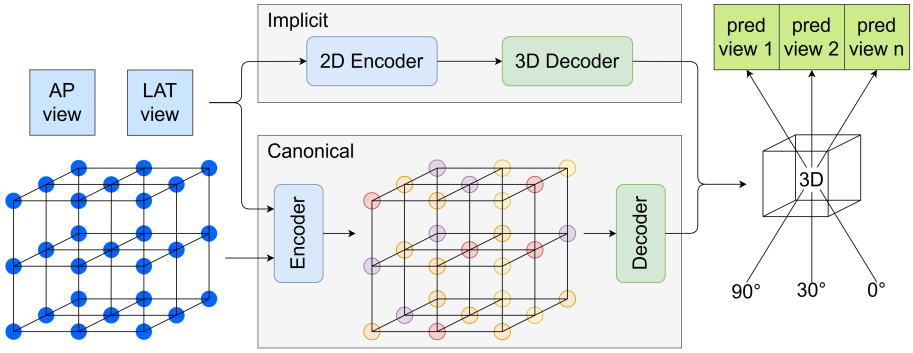

Fig. 1. Illustration of two representative spatial modeling paradigms for 3D reconstruction from biplanar X-rays. (AP: anteroposterior; LAT: lateral)

cost benefits. Consequently, stable 3D bone reconstruction using biplanar radiographics has long been an important research topic in the field [10, 7].

Unfortunately, the problem is mathematically intractable unless strong priors are introduced through large samples. Thus, existing studies often prioritize performance improvements via novel network architectures or strong statistical priors [26, 4]. However, the information density of biplanar inputs may also affect reconstruction quality. For example, large areas of long bones X-rays tend to be dark backgrounds, so apart from the training sample size, proper spatial modeling strategies as additional constraints are helpful in reconstruction [24, 25, 29]. In contrast, when planar inputs contain dense and continuous structural cues (such as chest images), stable reconstruction can still be achieved with weak or implicit spatial constraints. To date, a comprehensive assessment of spatial modeling in inputs with sparse contents remains elusive.

Here, we introduce ExpGS, an explicit spatial modeling framework to study and validate geometry mechanisms in biplanar 3D reconstruction. Unlike methods that require case-specific camera intrinsics/extrinsics and near/far planes, ExpGS adopts a fixed virtual anteroposterior/lateral perspective configuration in a normalized coordinate system. Our design preserves rich spatial constraints from two-view perspective projection while removing dependency on externally provided camera metadata during inference. The recovered non-degenerate ray geometry constrains reconstruction to a physically consistent effective volume, alleviating depth degeneration and suppressing background interference.

In summary, our contributions are three-fold. First, we formalize when biplanar reconstruction is effectively constrained, and show that the benefit of explicit geometry is governed by 2D structural sparsity rather than data scale alone. Second, we build a controlled comparison framework with explicit, canonical, and implicit spatial modeling under a shared backbone/training setting, enabling attribution of performance gaps to spatial assumptions. Third, we propose ExpGS, a fixed-geometry frustum-based model that improves stability on sparse femur data while remaining competitive on information-dense chest data.

2 Related Work

Multiple strategies exist for 3D reconstruction from biplanar X-rays, each with fundamental differences in their spatial modeling assumptions. We categorize them into two broad paradigms, the implicit and the canonical approaches, as illustrated in Fig. 1.

Implicit Spatial Modeling include statistical shape models (SSMs)[15, 2], voxel-based regression networks (e.g., U-Net)[14, 6, 20], and generative models such as VAEs [21], GANs [27], and diffusion models [23]. These methods treat the mapping from 2D projections to 3D volumes as a high-dimensional regression or generation problem. These models excel with sufficient training data or stable shape distributions, but underperform in cases without explicit ray-based geometric constraints, such as biplanar X-rays.

Canonical Spatial Modeling includes recent methods such as NeRF[9, 13, 19] and 3D Gaussian Splatting (3DGS) [31, 12, 3] that have demonstrated strong geometric representation capabilities under multi-view perspective imaging. However, in the X-ray setting, these methods are typically adapted under parallel-beam or weak-perspective assumptions, approximating 3D structure through initialized control points or implicit volumetric representations. In biplanar imaging with only two transmission views, such canonical or degenerate projection assumptions may lead to substantial information loss along the depth direction.

Recent calibrated biplanar or sparse-view methods (e.g., SPIDER[28], SSR-KD[11]) explicitly leverage known projection geometry and structural priors,

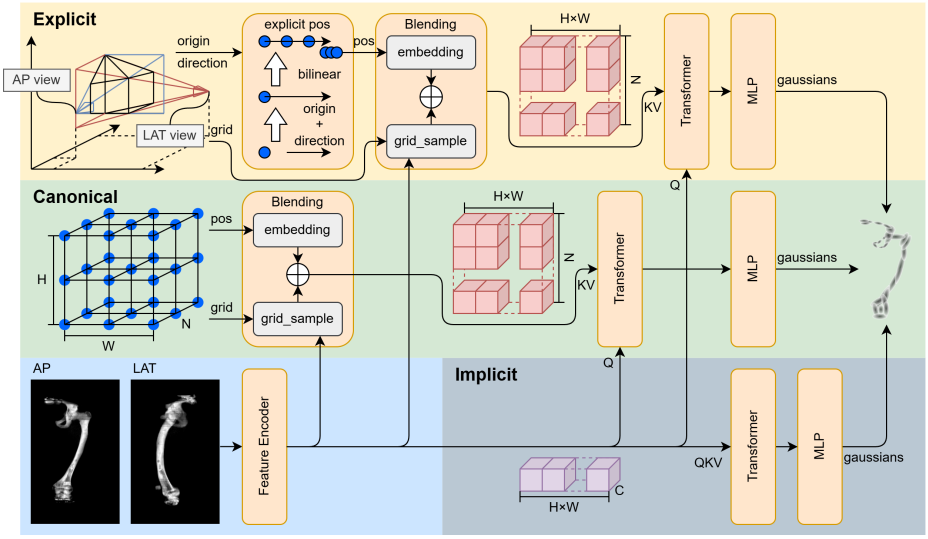

Fig. 2. The flows of ExpGS and its two ablated forms: from most spatially constrained (full ExpGS, top), to canonical (middle), and least constrained (implicit, bottom).

aligning with the explicit modeling category. In parallel, geometry-aware splatting pipelines (e.g., IXGS[5]) demonstrate strong reconstruction under calibrated multi-view C-arm setups. These works complement our framing by highlighting when explicit geometry is crucial; ExpGS focuses on fixed virtual AP/LAT geometry without per-case camera metadata.

Limitations of Existing Spatial Modeling: When projections cannot form non-degenerate view sections, 3D structures cannot be uniquely constrained, forcing reliance on statistical priors. In fact, the quality of reconstruction depends not only on the amount of data but also on whether the model preserves non-degenerate ray geometry. More data may improve visual consistency, but it cannot overcome the fundamental geometric blurring caused by insufficient spatial constraints.

3 Methods

As shown in Fig. 2, we design three reconstruction schemes for ExpGS: (1) explicit view-frustum (full ExpGS), (2) degenerate canonical-space, and (3) implicit spatial modelings. The latter two represent gradual spatial relaxations and increasing conceptual ablations. All three schemes share the same network backbone, inputs, and optimization objectives, enabling a direct attribution of reconstruction differences to spatial modeling strategies only.

3.1 Explicit View-Frustum Spatial Modeling

This full ExpGS modeling (top, Fig. 2) enforces non-degenerate geometric constraints. AP and LAT image inputs are first encoded into 2D feature maps $F \in \mathbb{R}^{H\times W\times C}$. For each spatial position (u, v), a 3D ray is parameterized as

$$\mathbf{r}_{u,v}(t) = \mathbf{o} + t\mathbf{d}_{u,v}, \tag{1}$$

where $\mathbf{o}$ is the virtual camera center and $\mathbf{d}_{u,v}$ the back-projected unit direction. We use a fixed virtual camera layout: orthogonal AP/LAT views, constant focal length, and centered principal point in a normalized canonical coordinate system. This canonicalization anchors the global scale and pose gauge during training. Under this setup, the induced frustums C_{front} and C_{side} define a fixed geometrically consistent region $\Omega = C_{front} \cap C_{side}$ within predefined normalized depth bounds. Practically, only samples whose depth lies in valid intervals of both views are kept, yielding a shared effective support for AP/LAT feature interaction. Along each ray, we uniformly sample N points within Ω:

$$\mathbf{X}_{u,v,n} = \mathbf{o} + t_n\mathbf{d}_{u,v}, \quad (n = 1, \ldots, N). \tag{2}$$

The sampled 3D coordinates are mapped through a positional embedding module to obtain a spatial encoding $S = embedding(X) \in \mathbb{R}^{H\times W\times N\times C}$, which explicitly encodes the spatial structure jointly determined by the biplanar view-frustum geometry. The 2D feature map is then expanded along the depth dimension so that each (u, v) corresponds to N spatial samples:

$$F_{3D} = L(F), \quad (L : \mathbb{R}^{H\times W\times C} \to \mathbb{R}^{H\times W\times N\times C}). \tag{3}$$

We perform cross-attention by using the original 2D features F as Query and the geometry-aware spatial features $F_{3D}+S$ as Key and Value: $Z = T(F,\, F_{3D}+S,\, F_{3D}+S)$. AP and LAT tokens are concatenated before attention and jointly updated in the same transformer blocks, so fusion is driven by shared geometry-aware keys/values in Ω. The resulting representation Z is fed to an MLP to predict continuous 3D Gaussian primitives [8]. Rendering follows an efficient perspective Gaussian splatting pipeline with alpha compositing.

3.2 Degenerate Canonical and Implicit Spatial Modeling

To analyze progressively relaxed spatial assumptions, we further construct two increasingly ablated forms of ExpGS below.

Degenerate canonical modeling (middle, Fig. 2) performs sampling in a predefined canonical 3D space, where depth axes are assumed to be parallel and aligned with the 2D feature grid. This setting can be regarded as a degenerate case of Sec. 3.1, in which view-frustum rays collapse into parallel rays and depth scales are no longer coupled across spatial locations. The resulting spatial embedding $S_{can} \in \mathbb{R}^{H\times W\times N\times C}$ is optionally added to the expanded feature F_{3D}, yielding $Z = T(F, F_{3D}+S_{can}, F_{3D}+S_{can})$, where all other network components remain identical to the explicit scheme.

Implicit modeling (bottom, Fig. 2) further removes explicit spatial structure by directly feeding 2D features into the Transformer, where queries, keys, and values are all derived from F. In this case, feature interactions are driven solely by self-attention without any geometric or spatial encoding.

3.3 Training Objective

Given AP/LAT conditional inputs and a sampled target-view set $\mathcal{V}$, the loss is

$$\mathcal{L}_{2D} = \frac{1}{|\mathcal{V}|}\sum_{v\in\mathcal{V}} \left\| I_v^{pred} - I_v^{gt} \right\|_2^2, \tag{4}$$

where I_v^{pred} and I_v^{gt} denote predicted and ground-truth projections at view v. No direct 3D voxel loss is used. This objective applies to all three schemes.

4 Experiments

4.1 Data and Experimental Setup

We evaluate our method on an in-house long-bone CT dataset (**Femur**) and a public chest CT dataset (**LIDC-IDRI** [1]), representing limited- and sufficient-data regimes, respectively. **Femur** contains 34 volumes (26/8 train/test), while

Table 1. Quantitative comparisons of different spatial modeling paradigms on Femur and LIDC-IDRI datasets, averaged over 72 rendered views.

		Explicit	Canonical		Implicit	
		ExpGS(Ours)	DIF-Gaussian	PerX2CT	DiT	X2CT-GAN
Femur	SSIM↑	**0.8613**	0.5656	0.5947	<u>0.6983</u>	0.6162
	LPIPS↓	**0.1496**	0.3406	0.3295	<u>0.2400</u>	0.3986
	PSNR↑	**21.6132**	<u>16.3335</u>	15.9359	15.7463	15.1413
LIDC-IDRI	SSIM↑	**0.6599**	<u>0.6514</u>	0.6132	0.6346	0.6276
	LPIPS↓	0.2602	<u>0.2130</u>	**0.1763**	0.2175	0.2543
	PSNR↑	<u>21.0379</u>	**22.2947**	17.1806	18.0224	16.9447

LIDC-IDRI includes 1,012 volumes (912/100). For data-scale analysis, LIDC-IDRI is further subsampled at 100%, 50%, 20%, 10%, and 5%. Compared with Femur, LIDC-IDRI provides higher effective projection information density due to richer anatomical structures.

For both datasets, 2D projections are synthesized from CT volumes using a cone-beam point-camera frustum. AP and LAT views are used as conditional inputs, while additional views sampled from a circular orbit provide multi-view supervision during training. Each CT provides 72 DRRs at 5° intervals, with four target views randomly sampled per iteration. All projections are generated at a fixed resolution of 256×256. Reconstruction performance is evaluated over 72 rendered views, using AP/LAT projections only as inputs. This avoids trivial 2D-to-2D copying.

Implementation Details. A ResNet-50 encoder extracts 2D features from AP/LAT views, followed by a Transformer (8 heads, d=512) for cross-view fusion and spatial modeling. The fused representation is mapped to 3D Gaussian parameters $\{\mu, \Sigma, h, \alpha\}$ via a lightweight MLP, forming a continuous Gaussian field for differentiable projection rendering. Training is supervised using an MSE-based multi-view 2D projection consistency loss without direct 3D voxel supervision. All models are optimized using Adam (lr=$1e^{-4}$, batch=4).

Protocol Alignment and Comparison. All baseline methods are retrained under the same AP/LAT-conditioned reconstruction setting, with identical projection geometry, data splits, input resolution, optimizer, and supervision design, enabling controlled comparison of different spatial modeling assumptions.

4.2 Stability of Explicit Spatial Modeling

We first compare explicit spatial modeling (our method, ExpGS) with Canonical modeling (PerX2CT[9] and DIF-Gaussian[12]) and implicit modeling (X2CT-GAN[27] and DiT[16]) approaches. Quantitative evaluations are reported in Table 1, and representative reconstruction results in Fig. 3.

On the Femur dataset, explicit view-frustum-based modeling reconstructs anatomically coherent 3D femur structures with reasonable proportions, whereas canonical methods yield blurred coarse outlines and implicit methods collapse toward averaged shapes, failing to preserve subject-specific geometry. On the

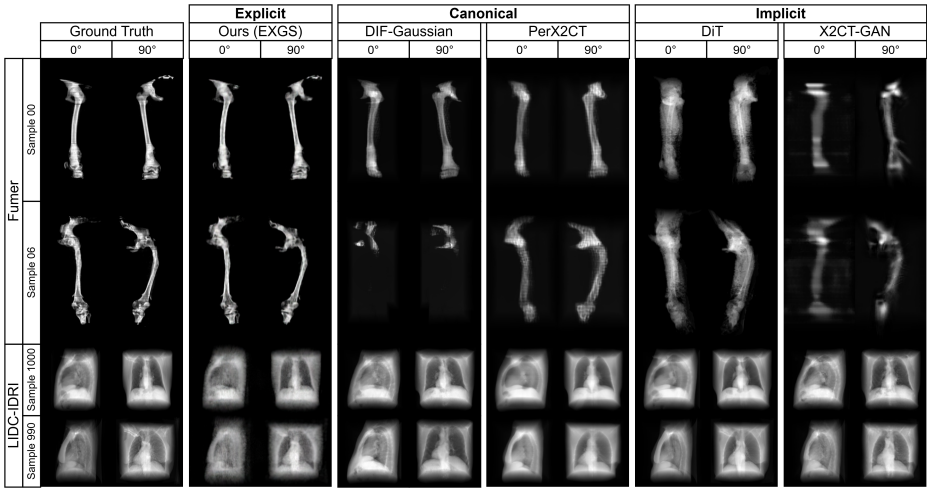

Fig. 3. Qualitative visualization of 3D reconstruction results using different open-source methods.

LIDC dataset, all methods produce visually clear and structurally continuous reconstructions. ExpGS achieves the strongest performance on structure-oriented metrics (SSIM[22]), while photometric metrics (LPIPS[30]/PSNR[18]) remain competitive but not consistently optimal, indicating a structure–pixel trade-off in information-dense settings. Overall, the results indicate that dense imaging enables stable 3D reconstruction even without explicit ray constraints, while sparse, cluttered settings require non-degenerate spatial modeling. This supports the conclusion that explicit geometry is most beneficial in sparse-cue regimes under the current CT-synthesized protocol.

4.3 Impact of Dataset Scale on Reconstruction Performance

We next assess how sample size affects reconstruction, using various subsets of the larger LIDC dataset (Methods). As shown in Fig. 4, from 5% $\sim$ 100%, all methods exhibit only minor variations in SSIM, LPIPS, and PSNR. Increasing sample size mainly introduces slight appearance-level gains, while the core reconstruction behavior remains stable, confirming that sample size matters marginally. This together with the visual contrast between Femur (more background) and LIDC (more foreground) support our sparsity-to-geometry motivation.

4.4 Ablation on Degraded Spatial Modeling Assumptions

Lastly, we conduct systematic ablation studies under the framework set out in Fig. 2, wherein ExpGS and its two conceptually ablated forms are implemented. Results of two representative samples are shown in Fig. 5. It shows explicit

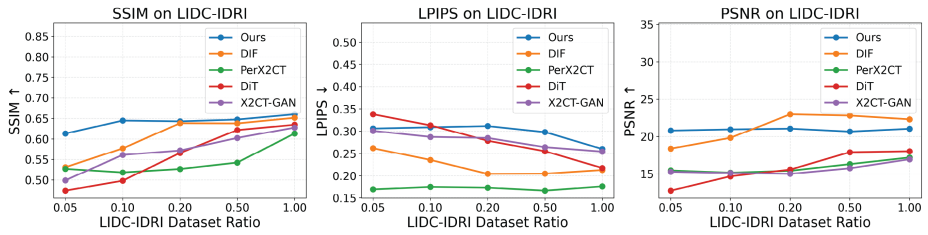

Fig. 4. Reconstruction performance on the LIDC dataset under different training sample sizes for various spatial modeling paradigms.

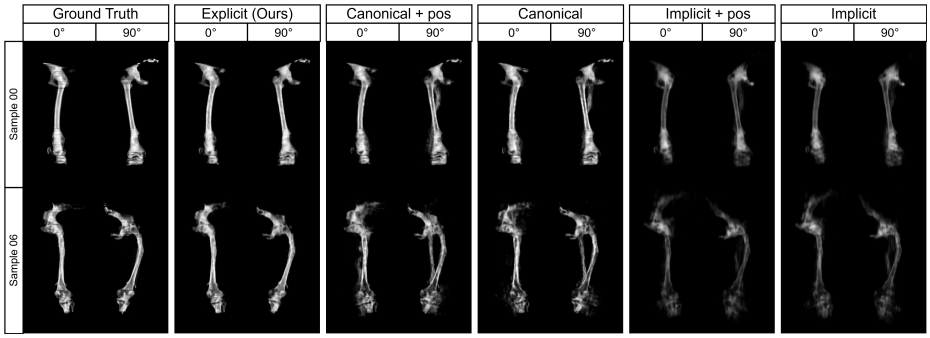

Fig. 5. Reconstruction performances under different spatial modeling assumptions.

spatial modeling (Panel 2) enables coherent and geometrically consistent 3D bone structure reconstruction. When the spatial model degrades to a canonical representation (Panels 3–4), the model remains effective for structures close to the dominant data distribution, but increasingly relies on statistical priors for anatomically complex cases, leading to superposition with averaged shapes. With fully implicit spatial modeling (Panels 5–6), the reconstruction becomes almost entirely driven by data statistics.

4.5 Limitations and Protocol Scope

Several limitations are noted. Synchronous biplanar projections from CT-scans and geometry-matched AP/LAT views were used for training, which might have generalization issues to asynchronous real-world conditions. The geometric effect of the frustum intersection is not explicitly separated from potential region-of-interest constraints. Rendering relies on an efficient splatting formulation, and evaluation uses projection-space metrics rather than standardized voxel-level metrics. Direct voxelizing of 3DGS would require extra resolution and thresholding choices. Moreover, the private Femur dataset is also relatively small.

5 Conclusion

This work studies spatial modeling for biplanar X-ray–based 3D bone reconstruction from a 2D information density perspective. We propose ExpGS, which enforces non-degenerate ray constraints through a fixed virtual AP/LAT frustum without requiring camera metadata at inference. Experiments on Femur and LIDC show clear benefits under sparse-view, matched-geometry settings. Future work will focus on physics-consistent rendering, robustness to camera perturbations, real-radiograph validation, and standardized 3D evaluation metrics.

Acknowledgments. This work was supported by Shenzhen Sci & Tech Program (JCYJ20250604180803005), SZ Sanming (SZSM202311022), SZ Clin Res Ctr for Rare Diseases (LCYSSQ20220823091402005), and HKU-SZH Transl Med Ctr (HZQSWS-KCCYB-2024055). PKC thanks the Shenzhen Peacock Plan (No.20210830100C) and Futian Talent Program. Computes were provided by AIBD Lab's DeepBay platform.

Disclosure of Interests. The authors declare no competing interests.

References

1. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical physics **38**(2), 915–931 (2011)
2. Baka, N., Kaptein, B.L., de Bruijne, M., van Walsum, T., Giphart, J., Niessen, W.J., Lelieveldt, B.P.: 2d–3d shape reconstruction of the distal femur from stereo x-ray imaging using statistical shape models. Medical image analysis **15**(6), 840–850 (2011)
3. Charatan, D., Li, S.L., Tagliasacchi, A., Sitzmann, V.: pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19457–19467 (2024)
4. Di, J., Lin, J., Zhong, L., Qian, K., Qin, Y.: Review of sparse-view or limited-angle ct reconstruction based on deep learning. Laser & Optoelectronics Progress **60**(8), 0811002 (2023)
5. Jecklin, S., Massalimova, A., Zha, R., Calvet, L., Laux, C.J., Farshad, M., Fürnstahl, P.: Ixgs-intraoperative 3d reconstruction from sparse, arbitrarily posed real x-rays (2025), https://arxiv.org/abs/2504.14699
6. Jiang, Y.: Mfct-gan: multi-information network to reconstruct ct volumes for security screening. Journal of Intelligent Manufacturing and Special Equipment **3**(1), 17–30 (2022)
7. Kasten, Y., Doktofsky, D., Kovler, I.: End-to-end convolutional neural network for 3d reconstruction of knee bones from bi-planar x-ray images. In: International Workshop on Machine Learning for Medical Image Reconstruction. pp. 123–133. Springer (2020)
8. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)

9. Kyung, D., Jo, K., Choo, J., Lee, J., Choi, E.: Perspective projection-based 3d ct reconstruction from biplanar x-rays. In: 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
10. Laporte, S., Skalli, W., De Guise, J., Lavaste, F., Mitton, D.: A biplanar reconstruction method based on 2d and 3d contours: application to the distal femur. Computer Methods in Biomechanics & Biomedical Engineering **6**(1), 1–6 (2003)
11. Lin, Y., Sun, H., Li, Y., Aslam, R., Tse, L.F., Cheng, T., Chui, C.S., Yau, W.F., Meur, V.R.L., Amangeldy, M., Cho, K., Ye, Y., Zou, J., Zhao, W., Li, X.: Real-time reconstruction of 3d bone models via very-low-dose protocols (2026), https://arxiv.org/abs/2508.13947
12. Lin, Y., Wang, H., Chen, J., Li, X.: Learning 3d gaussians for extremely sparse-view cone-beam ct reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 425–435. Springer (2024)
13. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
14. Moura, D.C., Boisvert, J., Barbosa, J.G., Labelle, H., Tavares, J.M.R.: Fast 3d reconstruction of the spine from biplanar radiographs using a deformable articulated model. Medical engineering & physics **33**(8), 924–933 (2011)
15. Nolte, D., Xie, S., Bull, A.M.: 3d shape reconstruction of the femur from planar x-ray images using statistical shape and appearance models. BioMedical Engineering OnLine **22**(1), 30 (2023)
16. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
17. Quijano, S., Serrurier, A., Aubert, B., Laporte, S., Thoreux, P., Skalli, W.: Three-dimensional reconstruction of the lower limb from biplanar calibrated radiographs. Medical engineering & physics **35**(12), 1703–1712 (2013)
18. Rosenfeld, A.: Digital picture processing. Academic press (1976)
19. Rückert, D., Wang, Y., Li, R., Idoughi, R., Heidrich, W.: Neat: Neural adaptive tomography. ACM Transactions on Graphics (TOG) **41**(4), 1–13 (2022)
20. Shen, L., Zhao, W., Xing, L.: Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. Nature biomedical engineering **3**(11), 880–888 (2019)
21. Wang, H., Wang, L., Wang, Z., Ma, L., Luo, Y.: Ssc-vae: Structured sparse coding based variational autoencoder for detail preserved image reconstruction. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 7665–7673 (2025)
22. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
23. Webber, G., Reader, A.J.: Diffusion models for medical image reconstruction. BJR| Artificial Intelligence **1**(1), ubae013 (2024)
24. Wu, W., Guo, X., Chen, Y., Wang, S., Chen, J.: Deep embedding-attention-refinement for sparse-view ct reconstruction. IEEE Transactions on Instrumentation and Measurement **72**, 1–11 (2022)
25. Wu, W., Hu, D., Niu, C., Yu, H., Vardhanabhuti, V., Wang, G.: Drone: Dual-domain residual-based optimization network for sparse-view ct reconstruction. IEEE Transactions on Medical Imaging **40**(11), 3002–3014 (2021)
26. Yang, L., Ge, R., Feng, S., Zhang, D.: Learning projection views for sparse-view ct reconstruction. In: Proceedings of the 30th ACM international conference on multimedia. pp. 2645–2653 (2022)

27. 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)
28. Yu, T., Tian, X., Yang, J., He, D., Yu, J., Wang, X., Zhang, Y.: Spider: Structure-preferential implicit deep network for biplanar x-ray reconstruction (2025), https://arxiv.org/abs/2507.04684
29. Zha, R., Zhang, Y., Li, H.: Naf: neural attenuation fields for sparse-view cbct reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 442–452. Springer (2022)
30. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
31. Zhou, S., Chang, H., Jiang, S., Fan, Z., Zhu, Z., Xu, D., Chari, P., You, S., Wang, Z., Kadambi, A.: Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21676–21685 (2024)