# EA-LDM: Explicit Alignment Latent Diffusion Model for Patient-Specific CT Reconstruction from Bi-planar X-rays

Jiewen Xiao[1], Shuqiong Wu[2], Le Chen[1], Bin Feng[3], Fu Jin[3], and Xin Fan[1(✉)]

[1] Dalian University of Technology, Dalian, China
jiewen.xiao@foxmail.com, chenle@mail.dlut.edu.cn, xin.fan@dlut.edu.cn
[2] The University of Osaka, Osaka, Japan
wu@am.sanken.osaka-u.ac.jp
[3] Chongqing University Cancer Hospital
{fengbin77, jfazj}@126.com

**Abstract.** Reconstructing 3D CT volumes from bi-planar X-rays offers a low-dose, efficient alternative to CBCT for Adaptive Radiotherapy. However, the anatomical overlaps in X-rays and the ill-posed nature of this task often lead to spatial ambiguity and a loss of patient-specific details in existing methods. In this paper, we propose EA-LDM, an Explicit Alignment Latent Diffusion Model framework to address these challenges. First, we introduce an Explicit Alignment (EA) strategy, where an alignment network maps 2D X-rays directly into a pre-trained 3D CT latent space, utilizing volumetric semantic priors to resolve spatial ambiguity. Second, to recover patient-specific details, we propose Patient-Specific Optimization (PSO). Formulated through a bi-level optimization, PSO effectively incorporates priors from the pre-treatment planning CT while maintaining robustness to temporal anatomical changes. Experiments on a collected clinical dataset demonstrate that our method significantly outperforms state-of-the-art approaches in both structural fidelity and perceptual quality.

**Keywords:** CT Reconstruction · Adaptive Radiotherapy · Patient-Specific Priors · Explicit Alignment · Latent Diffusion Models.

## 1 Introduction

Adaptive Radiotherapy (ART) is a cancer treatment using radiation to destroy tumors, delivered over multiple fractions across several weeks. In the standard ART workflow, a planning CT (pCT) is first acquired, on which the clinical target and radiation dose are determined. At each fraction, cone-beam CT (CBCT) is acquired from the onboard kV system for patient setup and plan optimization (e.g., dose optimization) before treatment [27,14]. However, the cumulative radiation dose from frequent CBCTs raises concerns regarding potential risks, such as secondary malignancies [4]. As an alternative, reconstructing 3D CT volumes from low-dose 2D bi-planar X-rays (typically anterior-posterior and lateral

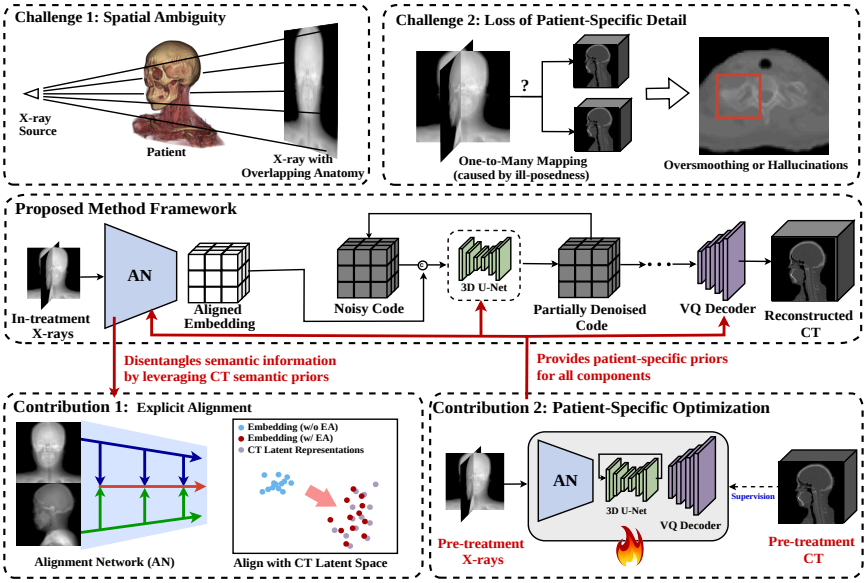


**Fig. 1.** Overview of our EA-LDM. **Top: Challenges.** Illustration of the two main challenges in existing methods. **Middle: Our framework.** Our method uses EA to extract an aligned embedding from X-rays, guiding the LDM backbone to generate a clean latent code, which is then decoded into the final CT. **Bottom: Key contributions.** (1) EA for mitigating the spatial ambiguity; and (2) PSO for recovering patient-specific details.

views) overcomes the limitations of CBCT byreducing the imaging dose by approximately an order of magnitude (1/10 to 1/20), and providing instantaneous acquisition instead of a minute-long gantry rotation [18]. Moreover, going beyond standard image registration which merely warps prior geometry, this approach recovers the critical 3D spatial and tissue density information necessary for both patient setup and plan optimization [2].

Nevertheless, recovering 3D structures from 2D X-rays remains a significant challenge. Early approaches focused on learning a direct mapping from X-rays to CT [13,15,23,29]. Recently, the success of diffusion models and Latent Diffusion Models (LDMs) in sparse-view reconstruction [3,8,28] has inspired their application to bi-planar X-ray-to-CT generation [12,16,22]. While these methods successfully reconstruct bones and large organs, they often fail to accurately render fine anatomical details, leading to oversmoothing or artifacts.

These limitations stem primarily from two factors. First, the severe anatomical overlap intrinsic to X-rays creates substantial spatial ambiguity. Consequently, it is difficult to infer the accurate volumetric distribution, often resulting in ghosting artifacts or incorrect structural topology. Second, due to the ill-posed nature of this problem, previous population-based methods frequently fail to capture patient-specific details [29]. As similar 2D projections can originate from

varying 3D volumes in population data (one-to-many mapping), previous models tend to learn a statistical average or generate structural hallucinations, rather than learning the authentic patient-specific information.

To address these limitations, we propose the EA-LDM (Explicit Alignment Latent Diffusion Model) framework with Patient-Specific Optimization for patient-specific CT reconstruction (Figure 1). To mitigate the spatial ambiguity inherent in X-rays, we introduce Explicit Alignment (EA). Specifically, we align X-ray features with the LDM's pre-trained CT latent space via an alignment network, creating conditional embeddings to guide the LDM. By leveraging 3D semantic priors within the CT latent space, EA disentangles semantic information from the X-ray, effectively alleviating spatial ambiguity and enabling the model to infer anatomically plausible 3D structures. To recover missing patient-specific information, we propose Patient-Specific Optimization (PSO), which builds a patient-specific model by incorporating patient-specific priors derived from the pre-treatment pCT. However, leveraging the pCT is challenging, as temporal anatomical changes (e.g., tumor regression and organ deformation) often render the pre-treatment pCT misaligned with the in-treatment X-rays, preventing the pCT from providing precise guidance. To overcome this, PSO adopts a bi-level optimization (BLO) strategy [21]: the lower-level optimization operates on population-level data, while the upper-level optimization incorporates patient-specific data (pCT). The lower level ensures the model learns generalizable reconstruction patterns from the population, while the outer level injects patient-specific priors from the pCT. Therefore, the model benefits from the high-quality priors of the pCT without being misled by its obsolete anatomy, ensuring robustness to in-treatment anatomical variations.

## 2 Methods

Figure 1 shows our EA-LDM framework. An alignment network first maps anterior-posterior (AP) and lateral (LA) X-rays to a 3D embedding within a pre-trained VQGAN [5] latent space. This embedding is concatenated with noisy latents to condition a 3D U-Net [10] backbone, which progressively denoises the latent code before the VQGAN decoder reconstructs the final CT volume.

### 2.1 Explicit Alignment

We propose Explicit Alignment (EA) where an alignment network produces a 3D X-rays embedding that is aligned with the CT latent space. Since X-rays and CTs of the same patient share a common semantic representation [20], we train the alignment network to directly predict this representation, thereby disentangling anatomical semantics in a manner defined by the CT latent representation and alleviating the spatial ambiguity inherent in 2D inputs.

**Explicit Alignment Training** We train the alignment network in conjunction with the fixed, pre-trained VQGAN decoder $\mathcal{D}_{vq}$ of LDM. The alignment network

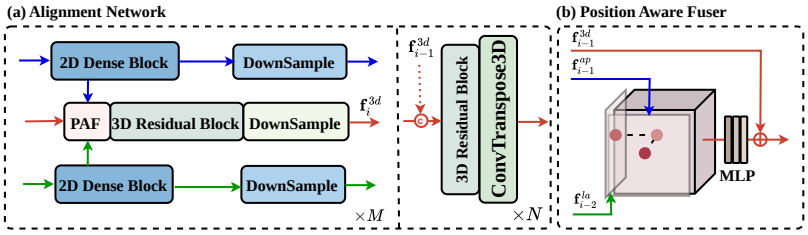


**Fig. 2.** Architecture of the alignment network. (a) Overview of the network structure. Blue, green, and red arrows indicate the data flow of AP, LA, and 3D features, respectively. $f_i^{ap}$ represents the AP feature at stage $i$; $f_i^{la}$ and $f_i^{3d}$ are defined similarly. M and N represent the number of stages of encoder and decoder. (b) Details of the Position-Aware Fuser (PAF).

(AN) takes bi-planar X-rays $(\mathbf{x}^{ap}, \mathbf{x}^{la})$ as input to generate a latent embedding $\hat{\mathbf{z}}_e$, which is then decoded into a 3D CT volume:

$$\hat{\mathbf{z}}_e = \text{AN}(\mathbf{x}^{ap}, \mathbf{x}^{la}), \tag{1}$$

$$\hat{\mathbf{y}}_e = \mathcal{D}_{vq}(\hat{\mathbf{z}}_e). \tag{2}$$

The optimization minimizes the discrepancy between $\hat{\mathbf{y}}_e$ and the ground-truth CT $\mathbf{y}$. We adopt the VQGAN training objective [1], comprising a VQ loss, an L1 reconstruction loss, an LPIPS perceptual loss [30], and a GAN loss. Inspired by [6], we additionally integrate a Multi-scale Normalized Cross-Correlation (mNCC) loss to mitigate artifacts and hallucinations. Specifically, we compute differentiable DRR projections [7] from $\hat{\mathbf{y}}_e$ and calculate a mNCC loss between these projections and the input X-rays. The total loss can be formulated as:

$$\mathcal{L}_e = \lambda_1 \mathcal{L}_{vq} + \lambda_2 \mathcal{L}_{rec} + \lambda_3 \mathcal{L}_{lpips} + \lambda_4 \mathcal{L}_{gan} + \lambda_5 \mathcal{L}_{mNCC} \tag{3}$$

where $\lambda_i$ are weighting hyperparameters. Loss weighting is set to balance the scales. We set $\lambda_3 = 100, \lambda_4 = 0.1$, and others=1.

**Architecture of the Alignment Network** To project the 2D X-rays into 3D embeddings and align them with the 3D CT latent space, we design a multi-branch encoder-decoder network (Figure 2). Structurally, the encoder comprises two parallel 2D branches based on dense blocks [11] for extracting multi-scale features from bi-planar X-rays, and a central 3D branch based on residual blocks [9] that evolves the volumetric features. Subsequently, the decoder utilizes these multi-scale volumetric features to generate the condition embedding $\hat{\mathbf{z}}_e$. We use M=4 encoder stages and N=3 decoder stages.

At each scale, we employ a lifting module Position-Aware Fuser (PAF) inspired by [23]. As shown in Figure 2 (b), PAF queries 2D feature vectors for each 3D voxel through projecting its coordinates onto the bi-planar 2D features. The two vectors associated with each voxel are then fused via an MLP to generate

a 3D feature volume. This interaction strategy based on spatial correspondence allows the network to cross-reference information from orthogonal views at every feature level, thereby correcting spatial misalignments from low-level anatomical textures to high-level structural semantics.

### 2.2 Patient-Specific Optimization

We propose Patient-Specific Optimization (PSO) based on BLO to integrate individual priors while mitigating the influence of obsolete anatomy in pre-treatment pCTs. This approach balances population-level generalization with patient-specific precision. The framework is formulated as a nested optimization problem:

$$\begin{aligned} &\arg\min_{\phi^*} \mathcal{L}_{\text{psp}}(\phi^*; \mathbb{D}_{\text{psp}}) \\ &\text{s.t.} \quad \phi^* = \arg\min_{\phi} \mathcal{L}_{\text{pop}}(\phi; \mathbb{D}_{\text{pop}}). \end{aligned} \tag{4}$$

where the lower-level optimization learns parameters for general anatomical features ($\phi$) from population data $\mathbb{D}_{pop}$ to ensure model generalization, and the upper-level optimization optimizes these parameters ($\phi^*$) using patient-specific pre-treatment data $\mathbb{D}_{psp}$ to capture patient-specific priors.

The training procedure for both levels involves three sequential steps: (1) training the VQGAN to define the CT latent space; (2) training the alignment network to generate 3D X-ray embeddings aligned with this latent space; and (3) training the LDM backbone for the denoising task. For both lower-level and upper-level optimization, we employ consistent loss functions for each respective module. The VQGAN and alignment network are optimized using a unified reconstruction objective (detailed in Equation (3)). The LDM backbone employs the standard diffusion denoising loss, formulated as:

$$\mathcal{L}_d = \mathbb{E}_{\mathbf{z}_0,\epsilon,t}\left[\|\epsilon - \epsilon_\theta(\mathbf{z}_t, \hat{\mathbf{z}}_e, t)\|^2\right], \tag{5}$$

where $\mathbf{z}_t$ denotes the noisy target latent code at timestep $t$ and $\hat{\mathbf{z}}_e$ represents the aligned 3D X-ray embedding. This embedding is concatenated with $\mathbf{z}_t$ along the channel dimension at each timestep to serve as a conditioning signal.

Leveraging the intrinsic similarity between the lower-level and upper-level optimization tasks, we empirically observe that optimal performance can be achieved by first converging on the lower-level task, followed by a brief upper-level optimization with limited steps. This facilitates the rapid construction of distinct patient-specific models for different individuals.

## 3 Experiments

### 3.1 Experiment Settings

Our framework is implemented in PyTorch [19] on four NVIDIA A40 GPUs. All modules are optimized via AdamW [17]. In the lower-level optimization, the

**Table 1.** Quantitative comparison with state-of-the-art methods. We report results both with and without Patient-Specific Optimization (PSO). The best results are highlighted in bold. The symbol † indicates that the method is re-implemented.

| PSO | Method | FLOPs (T) | Infer. Time (s) | PSNR-3D (↑) | PSNR (↑) | SSIM-3D (↑) | SSIM (↑) | LPIPS (↓) |
|---|---|---|---|---|---|---|---|---|
| w/o (×) | X2CT-GAN | **0.312** | **0.10** | 24.9320 | 29.8349 | 0.8833 | 0.8739 | 0.0792 |
| | INRR3CT | 0.680 | 0.20 | **26.0310** | 31.7075 | 0.9001 | 0.8917 | 0.0807 |
| | PerX2CT | 11.401 | 2.62 | 25.8292 | 30.5154 | 0.8984 | 0.8902 | 0.0671 |
| | DiffuX2CT† | 454.556 | 45.87 | 24.2540 | 30.2144 | 0.8710 | 0.8601 | 0.0643 |
| | DIFR3CT† | 21.847 | 1.79 | 25.0178 | 30.7809 | 0.8846 | 0.8756 | 0.0609 |
| | **Ours** | 21.978 | 1.88 | 26.0287 | **31.7087** | **0.9016** | **0.8940** | **0.0531** |
| w/ (✓) | X2CT-GAN | **0.312** | **0.10** | 27.0751 | 33.7918 | 0.9111 | 0.9036 | 0.0468 |
| | INRR3CT | 0.680 | 0.20 | 28.1142 | 32.2475 | 0.9250 | 0.9185 | 0.0692 |
| | PerX2CT | 11.401 | 2.62 | 27.5547 | 33.0548 | 0.9242 | 0.9186 | 0.0501 |
| | DiffuX2CT† | 454.556 | 45.87 | 25.7854 | 32.8269 | 0.8990 | 0.8909 | 0.0475 |
| | DIFR3CT† | 21.847 | 1.79 | 27.0134 | 33.9801 | 0.9172 | 0.9111 | 0.0439 |
| | **Ours** | 21.978 | 1.88 | **28.1725** | **34.9080** | **0.9309** | **0.9259** | **0.0408** |

VQGAN and alignment network are trained with a learning rate of $4 \times 10^{-5}$ for $8 \times 10^4$ steps, while the LDM backbone is trained with a learning rate of $6 \times 10^{-5}$ for $1 \times 10^5$ steps. In the upper-level optimization, the learning rates for all modules are reduced by a factor of 10, and training for $2 \times 10^4$ steps.

For population-level optimization, we utilize the RADCURE dataset [26], containing CT data from 3,346 patients. For patient-specific optimization and evaluation, we collected real-world clinical CTs from 92 patients, where each subject includes a pre-treatment pCT and an in-treatment CT. Corresponding bi-planar X-rays of all CTs are generated via Digitally Reconstructed Radiographs (DRRs) [7].

### 3.2 Comparison with Previous Methods

**Quantitative Comparison** Table 1 benchmarks our framework against SOTA GAN- [29], CNN- [13], NeRF- [23], and diffusion-based [16,22] methods. In the setting without PSO, our method outperforms the leading approach, DIFR3CT, by reducing LPIPS by 12.8% and achieving improvements of 4.0% in PSNR-3D, validating that the EA strategy effectively mitigates spatial ambiguity. With PSO, all methods show substantial gains, demonstrating PSO's generality in capturing patient-specific priors. Our method maintains the lead, surpassing DIFR3CT by 7.1% in LPIPS, 4.3% in PSNR-3D and 1.6% in SSIM. Notably, while the NeRF-based INRR3CT shows competitive pixel-wise metrics, its high LPIPS reflects inferior perceptual fidelity, which is consistent with the over-smoothed results shown in Figure 3.

Although diffusion-based methods typically incur higher computational costs, our approach completes reconstruction within 2 seconds with FLOPs competitive with SOTA diffusion models. In clinical ART workflows, such latency is negligible compared to the time required for patient positioning or CBCT acquisition.

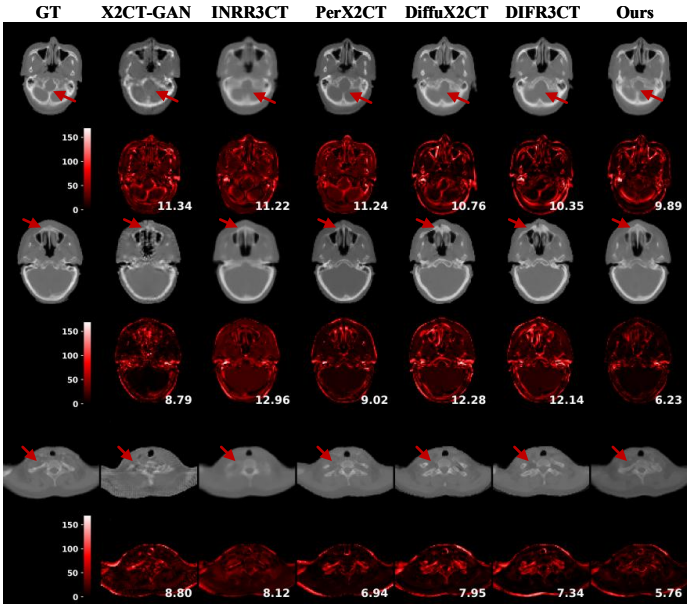


**Fig. 3.** Qualitative comparison of reconstruction results. The visual comparisons include CT slices and corresponding error maps. Red arrows indicate regions with significant improvements. Numbers on the error maps report the slice-wise MAE.

**Qualitative Comparison** Figure 3 visually benchmarks reconstruction quality. Our method achieves superior fidelity, accurately recovering anatomical details while avoiding the oversmoothing seen in INRR3CT and hallucinations in other baselines. The superiority of our method is particularly evident in the complex bony structures (bottom row). While competing methods generate distinct skeletal hallucinations, our method faithfully reconstructs the true geometry. This demonstrates that our EA strategy successfully mitigates spatial ambiguity, while PSO effectively incorporates patient-specific priors to guide correct reconstruction. The error maps further confirm this performance gap. Our method yields the darkest error maps with the lowest reported MAE, thereby validating the robustness of our framework.

### 3.3 Ablation Study

**Effectiveness of EA and PSO** We first evaluate the contributions of EA and PSO (Table 2 left part). Incorporating +EA yields a performance gain of 5.6% in PSNR and 2.7% in SSIM. This suggests that EA suppresses spatial ambiguity, thereby enhancing both numerical precision and structural correctness. Applying +PSO leads to a substantial improvement of 11.9% in PSNR and 5.7% in SSIM, indicating that patient-specific priors are indispensable for accurate anatomical reconstruction. Finally, our full model (Ours) surpasses the baseline by 16.3% in

**Table 2.** Ablation study on our framework. Left: Effectiveness of EA and PSO. Right: Comparison of different alignment strategies (with out PSO).

| Metric | EA & PSO | | | | EA Alternatives | | |
|---|---|---|---|---|---|---|---|
| | **Baseline** | **+EA** | **+PSO** | **Ours** | **Decoupled** | **End-to-end** | **EA** |
| **PSNR** | 30.0174 | 31.7087 | 33.5771 | **34.9080** | 30.0174 | 31.3812 | **31.7087** |
| **SSIM** | 0.8703 | 0.8940 | 0.9195 | **0.9259** | 0.8703 | 0.8843 | **0.8940** |

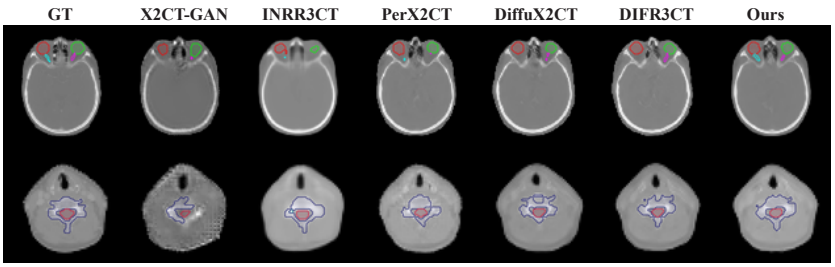


**Fig. 4.** Segmentation results of eyeballs, optic nerves, and cervical spine.

PSNR and 6.4% in SSIM. This confirms their complementarity: EA captures robust semantic information from X-rays, while PSO infers fine anatomical details from patient-specific priors.

**Ablation on EA Strategies** We compare our approach against common embedding paradigms: a Decoupled method (training the embedding network independently, i.e., the Baseline evaluated above) and an End-to-end method (direct joint optimization with LDM backbone). As shown in Table 2 right part, the Decoupled method performs poorly due to a semantic gap with the LDM. Although the End-to-end strategy achieves a relative gain, it remains suboptimal because of spatial ambiguity in input X-rays. Ultimately, our EA strategy yields the highest performance, outperforming baselines. This confirms that leveraging the 3D CT latent space effectively disentangles semantic information, serving as a superior mechanism to mitigate spatial ambiguity.

### 3.4 Segmentation Analysis

To assess clinical utility, we applied TotalSegmentator [25] to both GT and reconstructed volumes on all 92 patients. Figure 4 visualizes the results for eyeballs, optic nerves, and the cervical spine [24]. As shown, our method most closely matches the GT in eyeball morphology and successfully recovers fine optic nerve details. Moreover, it significantly outperforms other methods in reconstructing complex cervical structures, yielding bone and spinal cord shapes highly similar to the GT. This demonstrates the superior anatomical fidelity and clinical potential of our approach.

## 4 Conclusion

In this paper, we present EA-LDM for high-quality patient-specific CT reconstruction from bi-planar X-rays. By introducing Explicit Alignment, our method effectively mitigates the spatial ambiguity inherent in X-rays, minimizing artifacts and topological errors. Furthermore, the proposed Patient-Specific Optimization successfully integrates patient-specific priors from pre-treatment pCT for faithful reconstruction while maintaining robustness to temporal anatomical changes. Extensive experiments on a collected real-world clinical dataset and downstream segmentation tasks demonstrate that our approach outperforms SOTA methods in both perceptual quality and structural fidelity, highlighting its significant clinical potential.

**Acknowledgments.** This work was supported by the Special Project for Performance Incentive Guidance of Scientific Research Institutions in Chongqing Municipality, under the grant entitled "Construction and application of high-quality datasets for spatio-temporal precision and biological optimization of radiotherapy and immunotherapy for lung cancer" (Grant No. CSTB2025JXJL-YFX0003).

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

## References

1. Bai, Q., Liu, T., Liu, Z., Tong, Y., Torigian, D., Udupa, J.: XctDiff: Reconstruction of CT images with consistent anatomical structures from a single radiographic projection image. arXiv preprint arXiv: 2406.04679 (2024)
2. Brock, K.K., Mutic, S., McNutt, T.R., Li, H., Kessler, M.L.: Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM radiation therapy committee task group no. 132. Medical Physics **44**(7), e43–e76 (2017)
3. Chung, H., Ryu, D., Mccann, M.T., Klasky, M.L., Ye, J.C.: Solving 3d inverse problems using pre-trained 2d diffusion models. In: Conference on Computer Vision and Pattern Recognition. pp. 22542–22551. BC, Canada (2023)
4. Ding, G.X., Alaei, P., Curran, B., Flynn, R., Gossman, M., Mackie, T.R., Miften, M., Morin, R., Xu, X.G., Zhu, T.C.: Image guidance doses delivered during radiotherapy: Quantification, management, and reduction: Report of the AAPM therapy physics committee task group 180. Medical Physics **45**(5), e84–e99 (2018)
5. Ge, S., Hayes, T., Yang, H., Yin, X., Pang, G., Jacobs, D., Huang, J., Parikh, D.: Long video generation with time-agnostic VQGAN and time-sensitive transformer. In: European Conference on Computer Vision. pp. 102–118. Tel Aviv, Israel (2022)
6. Gopalakrishnan, V., Dey, N., Golland, P.: Intraoperative 2d/3d image registration via differentiable x-ray rendering. In: Conference on Computer Vision and Pattern Recognition. pp. 11662–11672. WA, USA (2024)
7. 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. Singapore (2022)

8. He, J., Li, B., Yang, G., Liu, Z.: Blaze3DM: Marry triplane representation with diffusion for 3d medical inverse problem solving. arXiv preprint arXiv:2405.15241 (2024)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition. pp. 770–778. Las Vegas, USA (2016)
10. He, Y., Yang, T., Zhang, Y., Shan, Y., Chen, Q.: Latent video diffusion models for high-fidelity long video generation. arXiv preprint arXiv: 2211.13221 (2023)
11. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Conference on Computer Vision and Pattern Recognition. pp. 2261–2269. HI, USA (2017)
12. Jeong, Y.S., Yoo, H.B., Chun, I.Y.: DX2CT: Diffusion model for 3d CT reconstruction from bi or mono-planar 2d X-ray(s). In: International Conference on Acoustics. pp. 1–5. Hyderabad, India (2025)
13. Kyung, D., Jo, K., Choo, J., Lee, J., Choi, E.: Perspective projection-based 3d CT reconstruction from biplanar X-rays. In: International Conference on Acoustics, Speech and Signal Processing. pp. 1–5. RI, Greece (2023)
14. Lavrova, E., Garrett, M.D., Wang, Y.F., Chin, C., Elliston, C., Savacool, M., Price, M., Kachnic, L.A., Horowitz, D.P.: Adaptive radiation therapy: a review of CT-based techniques. Radiology: Imaging Cancer **5**(4), e230011 (2023)
15. Lin, Y., Luo, Z., Zhao, W., Li, X.: Learning deep intensity field for extremely sparse-view CBCT reconstruction. In: Medical Image Computing and Computer Assisted Intervention. pp. 13–23. BC, Canada (2023)
16. Liu, X., Qiao, Z., Liu, R., Li, H., Zhang, J., Zhen, X., Qian, Z., Zhang, B.: DiffuX2CT: Diffusion learning to reconstruct CT images from biplanar x-rays. In: European Conference on Computer Vision. pp. 458–476. Milan, Italy (2024)
17. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations. LA, USA (2019)
18. Murphy, M.J., Balter, J., Balter, S., BenComo Jr., J.A., Das, I.J., Jiang, S.B., Ma, C.M., Olivera, G.H., Rodebaugh, R.F., Ruchala, K.J., Shirato, H., Yin, F.F.: The management of imaging dose during image-guided radiotherapy: Report of the aapm task group 75. Medical Physics **34**(10), 4041–4063 (2007)
19. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: PyTorch: An imperative style, high-performance deep learning library. In: Neural Information Processing Systems. pp. 8024–8035. BC, Canada (2019)
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**, 880–888 (2019)
21. Sinha, A., Malo, P., Deb, K.: A review on bilevel optimization: From classical to evolutionary approaches and applications. IEEE Transactions on Evolutionary Computation **22**(2), 276–295 (2018)
22. Sun, Y., Baroudi, H., Netherton, T., Court, L., Mawlawi, O., Veeraraghavan, A., Balakrishnan, G.: DIFR3CT: Latent diffusion for probabilistic 3d CT reconstruction from few planar X-rays. arXiv preprint arXiv: 2408.15118 (2024)
23. Sun, Y., Netherton, T.J., Court, L.E., Veeraraghavan, A., Balakrishnan, G.: CT reconstruction from few planar X-rays with application towards low-resource radiotherapy. In: Deep Generative Models. pp. 225–234. BC, Canada (2023)
24. Walter, A., Hoegen-Saßmannshausen, P., Stanic, G., Rodrigues, J.P., Adeberg, S., Jäkel, O., Frank, M., Giske, K.: Segmentation of 71 anatomical structures necessary

for the evaluation of guideline-conforming clinical target volumes in head and neck cancers. Cancers **16**(2) (2024)
25. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., et al.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023)
26. Welch, M.L., Kim, S., Hope, A.J., Huang, S.H., Lu, Z., Marsilla, J., Kazmierski, M., Rey-McIntyre, K., Patel, T., O'Sullivan, B., et al.: Radcure: An open-source head and neck cancer ct dataset for clinical radiation therapy insights. Medical Physics **51**(4), 3101–3109 (2024)
27. Yan, D., Vicini, F., Wong, J., Martinez, A.: Adaptive radiation therapy. Physics in Medicine & Biology **42**(1), 123–132 (1997)
28. Yang, L., Huang, J., Yang, G., Zhang, D.: CT-SDM: A sampling diffusion model for sparse-view CT reconstruction across various sampling rates. IEEE Transactions on Medical Imaging **44**(6), 2581–2593 (2025)
29. 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: Conference on Computer Vision and Pattern Recognition. pp. 10619–10628. CA, USA (2019)
30. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Conference on Computer Vision and Pattern Recognition. pp. 586–595. UT, USA (2018)