X-LRM: X-ray Large Reconstruction Model for Extremely Sparse-View Computed Tomography Recovery in One Second

Guofeng Zhang[1,*], Ruyi Zha[2,*], Hao He[3], Yixun Liang[3], Alan Yuille[1] Hongdong Li[2], Yuanhao Cai[1,†]
[1] Johns Hopkins University, [2] Australian National University, [3] HKUST

Abstract

Sparse-view 3D CT reconstruction aims to recover volumetric structures from a limited number of 2D X-ray projections. Existing feedforward methods are constrained by the scarcity of large-scale training datasets and the absence of direct and consistent 3D representations. In this paper, we propose an X-ray Large Reconstruction Model (X-LRM) for extremely sparse-view (<10 views) CT reconstruction. X-LRM consists of two key components: X-former and X-triplane. X-former can handle an arbitrary number of input views using an MLP-based image tokenizer and a Transformer-based encoder. The output tokens are then upsampled into our X-triplane representation, which models the 3D radiodensity as an implicit neural field. To support the training of X-LRM, we introduce Torso-16K, a large-scale dataset comprising over 16K volume-projection pairs of various torso organs. Extensive experiments demonstrate that X-LRM outperforms the state-of-the-art method by 1.5 dB and achieves 27× faster speed with better flexibility. Furthermore, the evaluation of lung segmentation tasks also suggests the practical value of our approach. Our code and dataset will be released at https://github.com/Richard-Guofeng-Zhang/X-LRM.

1. Introduction

Computed Tomography (CT) uses X-rays with penetrating power to reveal internal structures non-invasively. It is widely used in medical imaging for disease diagnosis, treatment planning, and surgical navigation [12, 13, 22, 23]. In particular, CT reconstruction aims to recover the 3D radiodensity of the scanned object given 2D X-ray projections.

Traditional methods [1, 17, 42, 54] usually require hundreds of X-ray projections to yield good reconstruction quality, which exposes significant radiation to patients. Recently, some self-supervised algorithms based on neural radiance field (NeRF) [7, 55] or 3D Gaussian splatting (3DGS) [5, 57] have been designed to reconstruct CT with $\sim$ 50 projections. Yet, these methods usually require a long time ($\sim$15 minutes) for each reconstruction with still relatively high radiation exposure. In this work, we study the extremely sparse-view (<10 views) CT reconstruction in a feedforward manner to inference **in one second**.

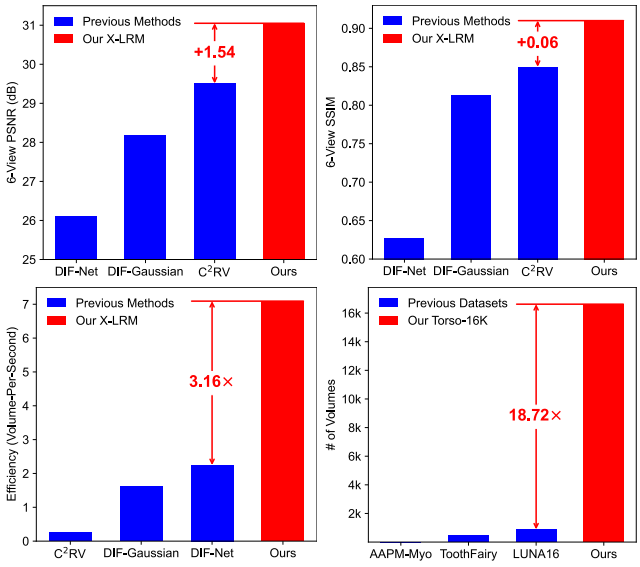

Figure 1. Our X-LRM outperforms previous 3D feedforward methods in quality and efficiency, including DIF-Net [29], DIF-Gaussian [30], and C^2-RV [31]. Our collected CT dataset, Torso-16K, is over 18× larger than previous benchmarks: LUNA16 [40], ToothFairy [11], and AAPM-Myo [35].

Some recent works [24, 29–31, 34] also try to explore this task. However, existing feedforward methods suffer from the following issues. **(i)** They rely on single-organ datasets containing fewer than 1,000 cases [11, 35, 40], which severely lack the diversity and scale required to develop robust and generalizable models. **(ii)** The number of the input projections of existing methods is fixed and cannot be adjusted, which lacks flexibility and limits the application in practice. **(iii)** Previous feedforward methods lack an explicit 3D representation, which limits their ability to model complex spatial structures and hampers performance in sparse-view or large-scale 3D reconstruction settings.

To cope with these problems, we design an X-ray Large Reconstruction Model (X-LRM), for extremely sparse-view

∗ = Equal Contribution. † = Corresponding Author

(< 10 views) CT recovery. X-LRM consists of two parts: X-former and X-triplane. Firstly, X-former uses a multi-layer perception (MLP) based tokenizer to split an arbitrary number of input images into patch tokens. Then X-former adopts a pure Transformer [16] encoder to compute the self-attention among these patch tokens. The output tokens of the X-former are then upsampled and reshaped into our X-triplane representation. The point feature of the X-triplane is fed into an MLP to learn an implicit neural field of the 3D volume radiodensity. To explore the potential of large-scale training, we collect a 3D CT reconstruction dataset, Torso-16K, containing ∼16K volume-projection data pairs. With the proposed techniques and collected dataset, our X-LRM can significantly benefit from large-scale training to boost the reconstruction quality and flexibly handle different numbers of input X-ray projections.

In a nutshell, our contributions can be summarized as:

- We propose X-LRM, a novel feedforward framework for sparse-view CT reconstruction.
- We design a Transformer-based encoder, X-former, to flexibly encode an arbitrary number of input X-ray projections. Besides, we present a new 3D representation, X-triplane, which directly and consistently models the radiodensity in X-ray imaging.
- We collect a large-scale dataset, Torso-16K, containing over 16K samples of 2D X-ray projections and 3D CT volumes. To the best of our knowledge, our Torso-16K is the largest CT reconstruction benchmark and is over 18× larger than the existing largest dataset.
- X-LRM drastically outperforms the SOTA by 1.5 dB PSNR and is 27× faster in inference.

2. Related Works

2.1. Sparse-View CT Reconstruction

We adopt a cone-beam CT (CBCT) setup that acquires multi-view 2D X-ray projections for volumetric reconstruction. Existing sparse-view CT reconstruction approaches can be categorized into optimization-based and prediction-based methods. Optimization-based methods iteratively refine the 3D volume to align with the measured projections. Traditional methods [1, 39, 43] formulate reconstruction as a *maximum a posteriori* problem, while learning-based methods leverage neural representations [5, 6, 41, 55, 57] and diffusion models [9, 10, 27]. Despite their effectiveness, these methods typically require minutes to hours to process a single case, making them impractical for real-time clinical applications. Prediction-based methods, in contrast, utilize neural networks to learn semantic priors from external datasets. Given a test case, they employ pre-trained models for projection extrapolation [2, 18], slice denoising [25, 34, 47], or volume regression [29–31]. While these methods enable rapid inference, they are constrained by the limited capacity of CNN-based models and the scarcity of large-scale training datasets. We cope with these problems by designing X-LRM and collecting Torso-16K.

2.2. Feedforward 3D Reconstruction

Unlike optimization-based methods NeRF [37] or 3D gaussian splatting [26], which take time-consuming optimization phase for shape recovery, feedforward 3D reconstruction aims to learn diverse geometry types (*e.g.*, mesh [48, 52], implicit fields [36, 53] *etc.*) from input images in a forward manner with neural network architectures. Boosting from large-scale 3D datasets like Objaverse-XL [14, 15] and the scalability of Transformer architectures [46], Large Reconstruction Model [21] and its subsequent variants [8, 19, 28, 44, 49, 51, 58] has greatly promoted reconstruction ability and efficiency of current fields. However, due to the data scarcity of CT reconstruction and 3D model, current feedforward CT reconstruction methods often suffer from poor reconstruction quality and generalization. Our goal is to fill these research gaps.

3. Method

The pipeline of our method is shown in Fig. 2. Our X-LRM consists of two parts: X-former and X-triplane, corresponding to Fig. 2 (b) and Fig. 2 (c). X-former begins with an MLP-based image tokenizer. Then a Transformer-based encoder processes an arbitrary number of multi-view image tokens with view-associated ray information into patch-based features. These features are then mapped into triplane tokens through cross-attention in a triplane decoder. We upsample and unpatchify these tokens to our X-triplane representation. Finally, we adopt an MLP to learn an implicit mapping from the 3D point features on the triplane to the corresponding volume radiodensity.

3.1. X-former

As aforementioned, existing feedforward methods struggle with large-scale training and varying numbers of projections, resulting in degraded performance, limited scalability, and reduced flexibility. To address these challenges, we propose X-former, an architecture composed of an MLP-based image tokenizer and a Transformer-based image encoder tailored for variable-view processing.

Image Tokenizer. As shown in Fig. 2 (b), the input to the tokenizer is multi-view X-ray projections $\mathbf{I}_i \in \mathbb{R}^{H\times W\times 1}$ concatenated with the corresponding viewpoint camera conditions $\mathbf{C}_i \in \mathbb{R}^{H\times W\times 6}$. We denote the input at the i-th view as $\mathbf{X}_i = [\mathbf{I}_i, \mathbf{C}_i] \in \mathbb{R}^{H\times W\times 7}$. During training, X-former can take varying numbers of views, denoted as $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$, where m is the count of varying numbers of input views. For a specific V_i, the input is denoted as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{V_i}] \in \mathbb{R}^{V_i\times H\times W\times 7}$, where $V_i \in \mathbf{V}$ can change dynamically during training.

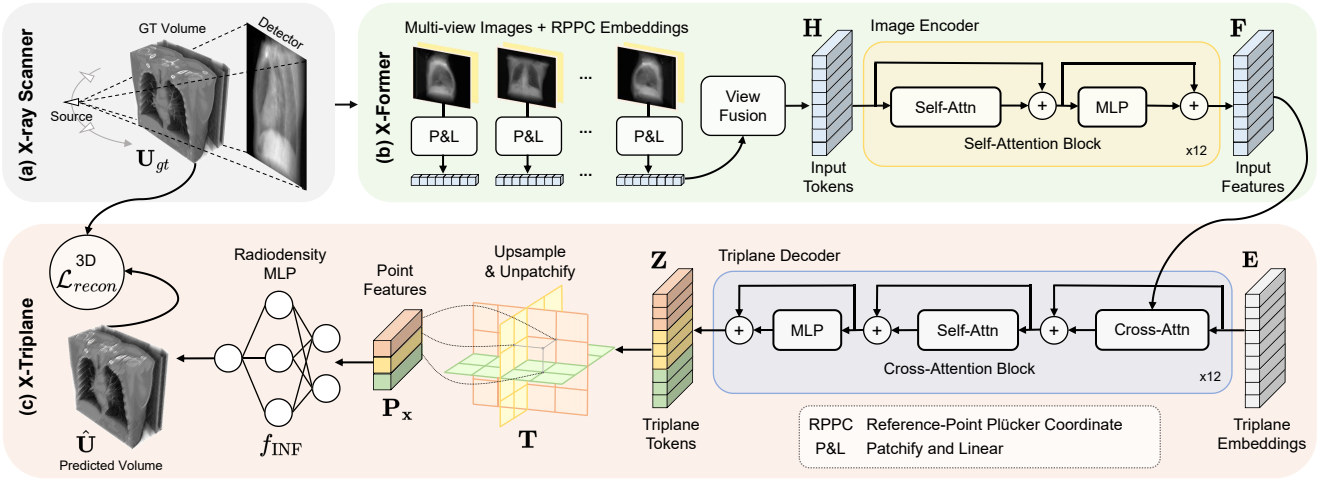

Figure 2. The overall architecture of X-LRM: (a) We collect Torso-16K, the largest CT reconstruction dataset (Sec. 4.1). (b) Our X-Former features an image tokenizer and encoder, designed to process a variable number of input views (Sec. 3.1). (c) Our X-Triplane includes a triplane decoder followed by our implicit neural field, directly predicting the 3D CT volume $\hat{\mathbf{U}}$ (Sec. 3.2).

We adopt the Reference-Point Pl"ucker Coordinate (RPPC) [8] as our camera condition, as it encodes more ray position and relative depth information than standard Pl"ucker coordinates. Thus, we have $\mathbf{C}_i = (\mathbf{o}_i - (\mathbf{o}_i \cdot \mathbf{d}_i)\mathbf{d}_i, \mathbf{d}_i)$, which better captures spatial relationships. Here, $\mathbf{o}_i$ and $\mathbf{d}_i$ denote the origins and directions of pixel-aligned rays at the i-th view.

Subsequently, the tokenizer partitions each input view into non-overlapping patches and projects each patch into a latent space of dimension d_E via an MLP layer. Then we fuse patchified tokens of different views by concatenating them to derive the initial patch-wise tokens $\mathbf{H} \in \mathbb{R}^{n \times d_E}$.

Image Encoder. The feature tokens $\mathbf{H}$ are then encoded by a Transformer-based encoder to produce input feature tokens: $\mathbf{F} \in \mathbb{R}^{n \times d_E}$, where d_E is the hidden dimension of our image encoder. The image encoder consists of N_e self-attention blocks [46], and each block comprises a multi-head self-attention layer and an MLP layer. We add layer normalization [3] before both layers. For the j-th self-attention block, we first split input $\mathbf{H}^j_{in}$ into k_E heads as

$$\mathbf{H}^j_{in} = [\mathbf{H}^j_1, \mathbf{H}^j_2, \ldots, \mathbf{H}^j_{k_E}]. \tag{1}$$

Then for the i-th head, we project input $\mathbf{H}^j_i$ into $\mathbf{Q}^j_i \in \mathbb{R}^{n \times d_{ke}}$, $\mathbf{K}^j_i \in \mathbb{R}^{n \times d_{ke}}$, and $\mathbf{V}^j_i \in \mathbb{R}^{n \times d_{ke}}$ as

$$\mathbf{Q}^j_i = \mathbf{H}^j_i \mathbf{W}_{\mathbf{Q}^j_i}, \ \mathbf{K}^j_i = \mathbf{H}^j_i \mathbf{W}_{\mathbf{K}^j_i}, \ \mathbf{V}^j_i = \mathbf{H}^j_i \mathbf{W}_{\mathbf{V}^j_i}, \tag{2}$$

where $\mathbf{W}_{\mathbf{Q}^j_i}, \mathbf{W}_{\mathbf{K}^j_i}, \mathbf{W}_{\mathbf{V}^j_i} \in \mathbb{R}^{d_E \times d_{ke}}$ are learnable parameters of the fc layers and $d_{ke} = d_E / k_E$. Then the output of i-th head of the j-th self-attention layer $\mathbf{A}^j_i$ is computed as

$$\mathbf{A}^j_i = \text{softmax}\left(\frac{\mathbf{Q}^j_i (\mathbf{K}^j_i)^\top}{\sqrt{d_{ke}}}\right)\mathbf{V}^j_i + \mathbf{H}^j_i. \tag{3}$$

Then k_E heads are concatenated to pass through a fc layer to derive the output of self-attention as

$$\mathbf{H}^j_{mid} = [\mathbf{A}^j_1, \ \mathbf{A}^j_2 \ \ldots \ \mathbf{A}^j_{k_E}] \ \mathbf{W}^j_s. \tag{4}$$

where $\mathbf{W}^j_s \in \mathbb{R}^{d_E \times d_E}$ is the learnable parameter. Then we forward $\mathbf{H}^j_{mid}$ to the MLP layer:

$$\mathbf{H}^j_{out} = \sigma(\mathbf{H}^j_{mid}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 + \mathbf{H}^j_{mid}, \tag{5}$$

where σ is the activation function, and $\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1, \mathbf{b}_2$ are learnable parameters. The output of the last layer of the image encoder is $\mathbf{F} = \mathbf{H}^{N_e}_{out} \in \mathbb{R}^{n \times d_E}$. This process is illustrated in Fig. 2 (b)

Our X-former leverages the inherent flexibility of the transformer architecture, which can naturally process input tokens of different lengths. This allows our model to seamlessly train with different numbers of input views within a single training session, boosting the reconstruction performance and resulting in a unified framework capable of handling diverse multi-view configurations.

3.2. X-triplane

To lift the features from 2D projection into 3D space, we design a Transformer-based decoder to map the 2D patch-wise features $\mathbf{F}$ into 3D triplane tokens $\mathbf{Z} \in \mathbb{R}^{(3\times32\times32)\times d_D}$, where d_D is the hidden dimension of the triplane decoder. $\mathbf{Z}$ is later upsampled and reshaped into our X-triplane representations, which encode 3D information. Then we adopt an MLP to learn an implicit mapping from the 3D point feature on the triplane representation to the corresponding radiodensity.

Triplane Decoder. As shown in Fig. 2 (c), the input of the triplane decoder includes $\mathbf{F}$ and a learnable triplane embeddings $\mathbf{E} \in \mathbb{R}^{(3\times32\times32)\times d_D}$. Our triplane decoder has

Dataset	Body Parts	# of Volumes
AbdomenAtlas v1.0	Abdomen, Chest, Pelvis	5,171
RSNA2023	Abdomen, Pelvis	4,711
AMOS	Abdomen	1,851
PENGWIN	Pelvis	100
TCIA	Abdomen	833
MELA	Chest	1,100
FLARE24 (subset)	Abdomen, Chest	1,868
FUMPE	Chest	35
LNDb	Chest	294
RibFrac	Abdomen, Chest	660
Torso-16K (Ours)	Abdomen, Chest, Pelvis	16,623

Table 1. The statistics of our collected Torso-16K benchmark. It integrates ten public datasets covering major anatomical regions in different clinical applications.

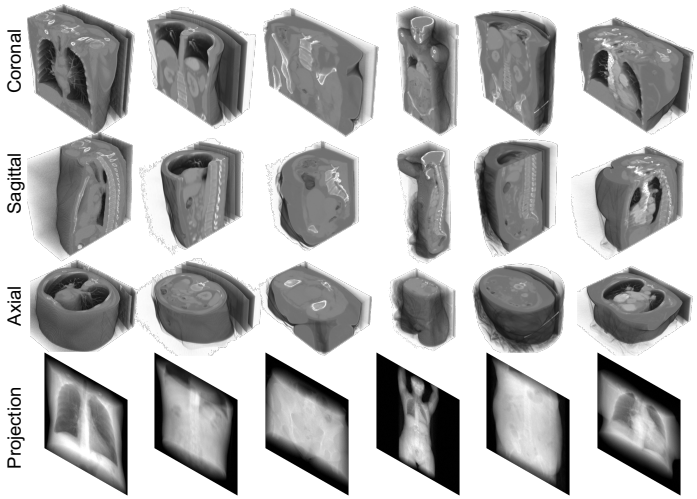

Figure 3. Example CT volumes and corresponding X-ray projections in our Torso-16K dataset.

N_d cross-attention blocks. Each cross-attention block comprises a cross-attention layer, a self-attention layer, and an MLP layer. To guide the reconstruction of the triplane tokens and lift the feature into 3D space, we adopt the cross-attention to extract 2D projection and camera information by querying the input feature $\mathbf{F}$.

Similar to self-attention, for the j-th cross-attention block in our triplane decoder, we first split $\mathbf{F}$ and input triplane embeddings $\mathbf{E}_{in}^{j}$ into k_D heads as

$$\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{k_D}], \mathbf{E}_{in}^{j} = [\mathbf{E}_1^{j}, \mathbf{E}_2^{j}, \dots, \mathbf{E}_{k_D}^{j}]. \quad (6)$$

Then for the i-th cross-attention head, we project $\mathbf{F}_i$ into *query* $\mathbf{Q}_i^{j} \in \mathbb{R}^{n \times d_{kd}}$, and project $\mathbf{E}_i^{j}$ into *key* $\mathbf{K}_i^{j} \in \mathbb{R}^{(3\times32\times32)\times d_{kd}}$ and *value* $\mathbf{V}_i^{j} \in \mathbb{R}^{(3\times32\times32)\times d_{kd}}$ by three fc layers, where $d_{kd} = d_D/k_D$. Then the output of i-th head of the j-th cross-attention layer $\mathbf{B}_i^{j}$ is computed as

$$\mathbf{B}_i^{j} = \text{softmax}\left(\frac{\mathbf{Q}_i^{j}(\mathbf{K}_i^{j})^{\top}}{\sqrt{d_{kd}}}\right)\mathbf{V}_i^{j} + \mathbf{E}_i^{j}. \quad (7)$$

Subsequently, k_D heads are concatenated to pass through an fc layer for the output:

$$\mathbf{E}_{mid}^{j} = [\mathbf{B}_1^{j},\ \mathbf{B}_2^{j}\ \dots\ \mathbf{B}_{k_D}^{j}]\ \mathbf{W}_c^{j}, \quad (8)$$

Similar to previous self-attention (SA) and MLP in Sec. 3.1 we have

$$\mathbf{E}_{out}^{j} = \text{MLP}\big(\text{SA}(\mathbf{E}_{mid}^{j}) + \mathbf{E}_{mid}^{j}\big) + \text{SA}(\mathbf{E}_{mid}^{j}). \quad (9)$$

Finally, the triplane decoder outputs $\mathbf{Z} = \mathbf{E}_{out}^{N_d} \in \mathbb{R}^{(3\times32\times32)\times d_D}$, as illustrated in Fig. 2 (c). $\mathbf{Z}$ is further upsampled by a deconvolution layer and unpatchified to our X-triplane representation $\mathbf{T}$.

Triplane Implicit Neural Field. Our X-triplane $\mathbf{T}$ is composed of three orthogonal feature planes: $\mathbf{T}xy$, $\mathbf{T}yz$, and $\mathbf{T}xz \in \mathbb{R}^{(64\times64)\times d_T}$, where 64×64 refers to the spatial resolution of each plane and d_T is the dimension of the point feature $\mathbf{P}\boldsymbol{x} \in \mathbb{R}^{3\times d_T}$. Then we build an implicit neural field mapping from the position and feature of a 3D point to its radiodensity.

For a given 3D point $\boldsymbol{x} = (x, y, z) \in [-1, 1]^3$ within the unit bounding box (where each coordinate is normalized), we obtain its feature embeddings by projecting it onto three orthogonal plane features $\mathbf{T}_{xy}, \mathbf{T}_{yz}$, and $\mathbf{T}_{xz}$ at $\mathbf{p}_{xy} = (x, y), \mathbf{p}_{yz} = (y, z)$, and $\mathbf{p}_{xz} = (x, z)$. We then apply bilinear interpolation to extract features from each plane. Take the xy-plane $\mathbf{T}_{xy}$ and a point $\mathbf{p}_{xy} = (x, y)$ for instance, the interpolated feature value is computed as

$$\begin{aligned}\mathbf{T}_{xy}(\mathbf{p}_{xy}) = (1-\alpha)\beta\mathbf{T}(x_0, y_1) + \alpha(1-\beta)\mathbf{T}_{xy}(x_1, y_0) \\ + (1-\alpha)(1-\beta)\mathbf{T}_{xy}(x_0, y_0) + \alpha\beta\mathbf{T}_{xy}(x_1, y_1),\end{aligned} \quad (10)$$

where x_0, x_1 and y_0, y_1 are the neighboring points, and the interpolation weights are $\alpha = x - x_0, \beta = y - y_0$. Applying to all three triplanes, we obtain the feature at the point $\boldsymbol{x}$ as

$$\mathbf{P}_{\boldsymbol{x}} = \left(\mathbf{T}_{xy}(\mathbf{p}_{xy}), \mathbf{T}_{yz}(\mathbf{p}_{yz}), \mathbf{T}_{xz}(\mathbf{p}_{xz})\right). \quad (11)$$

As the radiodensity is isotropic and only related to the point property, we adopt an MLP to learn the mapping f_{INF} from the point feature $\mathbf{P}_{\boldsymbol{x}}$ to the radiodensity $\rho_{\boldsymbol{x}}$ as

$$f_{\text{INF}} : \left(\mathbf{T}_{xy}(\mathbf{p}_{xy}), \mathbf{T}_{yz}(\mathbf{p}_{yz}), \mathbf{T}_{xz}(\mathbf{p}_{xz})\right) \rightarrow \rho_{\boldsymbol{x}}. \quad (12)$$

3.3. Training Objective

Existing RGB 3D reconstruction methods mainly adopt 2D rendering training loss to achieve good image recovery

Type	Method	Time (s)↓	6-View		8-View		10-View	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Traditional	FDK	**0.008**	9.51	0.039	10.68	0.047	11.46	0.058
	ASD-POCS	1.385	22.17	0.573	23.40	0.612	24.62	0.667
	SART	1.400	22.61	0.537	23.56	0.548	24.57	0.585
2D Feedforward	FBPConvNet	<u>0.010</u>	26.99	0.704	27.22	0.722	28.05	0.737
	FreeSeed	0.163	28.93	0.841	<u>30.08</u>	0.843	30.17	0.855
3D Feedforward	DIF-Net	0.445	26.10	0.627	26.81	0.663	27.47	0.708
	DIF-Gaussian	0.621	28.19	0.813	28.53	0.820	29.52	0.848
	C^2RV	3.837	<u>29.51</u>	<u>0.850</u>	29.83	<u>0.849</u>	<u>30.96</u>	<u>0.871</u>
	X-LRM (Ours)	0.141	**31.05**	**0.910**	**31.24**	**0.912**	**31.33**	**0.915**

Table 2. Comparison with traditional and feedforward methods on 750 test cases. X-LRM is 1.5 dB better and 27× faster than the best baseline. Best result is in **bold** and second-best is <u>underlined.</u>

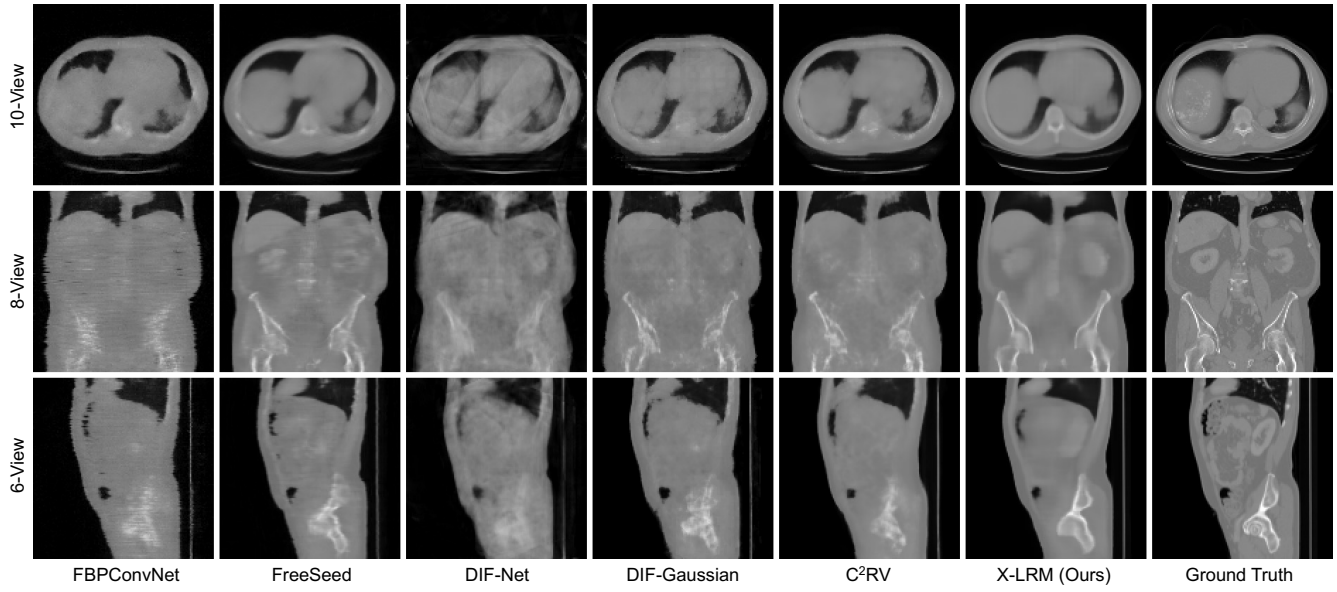

Figure 4. Qualitative results of feedforward methods across multiple anatomical views on the 750-sample test set. From top to bottom: 10-view axial, 8-view coronal, and 6-view sagittal slices.

quality. However, this supervision involves volume rendering that needs to sample many 3D points to compute for each ray, taking a long time and increasing memory cost. Besides, in X-ray imaging, the 3D CT reconstruction is more concerned than the 2D X-ray rendering. Thus, we adopt the more precise 3D reconstruction loss with varying numbers of input views as

$$\mathcal{L}_{recon} = \frac{1}{m} \sum_{V_i \in \mathcal{V}} \left|\left| \hat{\mathbf{U}}_{V_i} - \mathbf{U}_{gt} \right|\right|^2, \tag{13}$$

where $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ represents training settings with different input view numbers V_i. $\hat{\mathbf{U}}_{V_i}$ refers to the CT volume reconstructed by X-LRM given V_i views, and $\mathbf{U}_{gt}$ is the ground-truth CT volume. Such 3D supervision enables better anatomical consistency to view sparsity.

4. Experiments

4.1. Experiment Setup

Datasets. Previous works rely on small datasets [11, 35, 40] (fewer than 1,000 samples), which limits the ability to train generalizable models. To overcome this constraint, we introduce Torso-16K, the largest and most diverse CT reconstruction dataset, comprising 16,623 real-world CT scans from ten public datasets (Sec. 3.1). It covers key anatomical regions in clinical applications, including chest, abdomen, and pelvis. Some examples are shown in Fig. 3. Torso-16K is split into 15,000 / 873 / 750 for training, validation, and testing.

We standardize CT scans by resampling and cropping to a 50^3 cm^3 volume at 128^3 resolution. Radiodensity values are normalized from the Hounsfield unit range [-1000, 1000] to [0,1], ensuring coverage of primary organs of in-

Type	Method	Time↓	6-View		8-View		10-View	
			PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Self-Supervised	NAF	11m	23.86	0.644	24.64	0.654	25.38	0.685
	R^2-Gaussian	6m	20.28	0.528	20.79	0.529	22.09	0.581
	SAX-NeRF	8h	24.08	0.669	24.73	0.674	25.68	0.692
Diffusion Based	DDS	12m	24.42	0.529	25.64	0.570	26.64	0.607
	DiffusionMBIR	11h	26.61	0.734	28.51	0.803	30.05	0.835
3D Feedforward	X-LRM (Ours)	**0.14s**	**30.14**	**0.888**	**30.10**	**0.886**	**30.28**	**0.889**

Table 3. Comparison with self-supervised and diffusion-based methods on 10 test cases. Our X-LRM achieves 3.53 dB higher PSNR and is 2570× faster than the best-performing baseline.

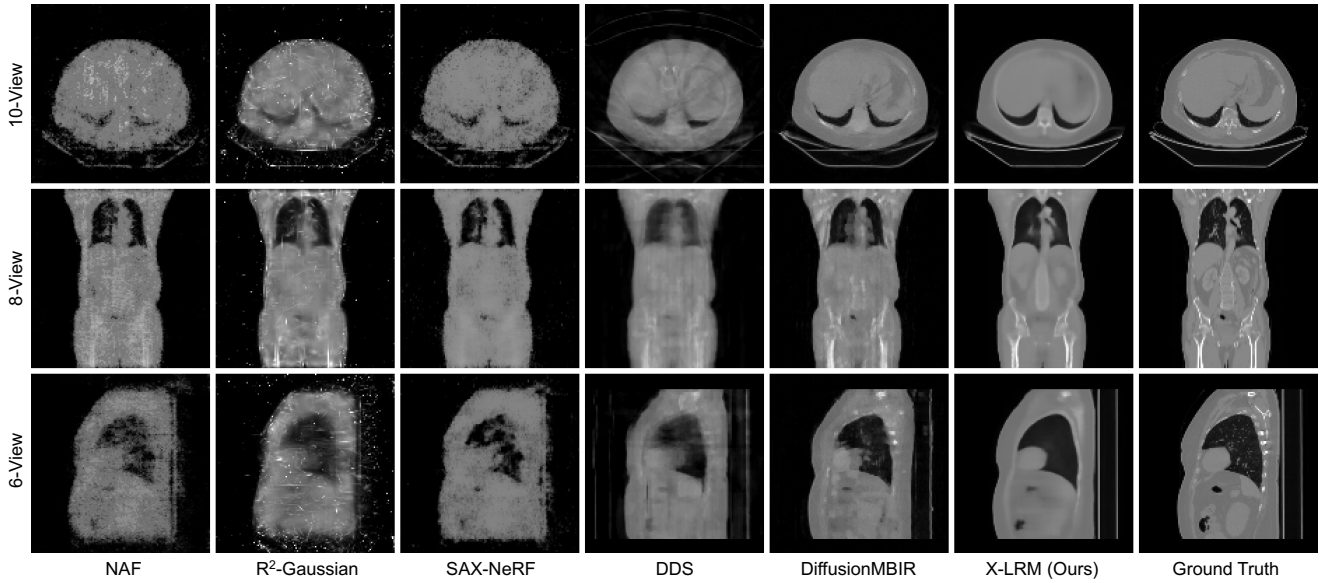

Figure 5. Qualitative results of self-supervised and diffusion-based methods on the 10-sample test set. From top to bottom: 10-view axial, 8-view coronal, and 6-view sagittal slices.

terest. Since most public datasets only provide CT volumes, we render multi-view X-ray projections via TIGRE toolbox [4]. 256^2 resolution projections span the full range 0° ~360° with 3.9^2 mm^2 pixel spacing. To enhance realism, we add Gaussian and Poisson noise to simulate Compton scattering and adopt UCT 960+ scanner [45], with 0.6m source-object and 1.118m source-detector distance.

Implementation Details. We implement X-LRM by PyTorch [38]. X-LRM is trained with the AdamW optimizer [33] ($\beta_1 = 0.9$, $\beta_2 = 0.95$). The initial learning rate is set to 4×10^{-4} and follows a cosine annealing scheduler [32] with a warm-up phase of 3000 iterations. For the network architecture, we utilize a ViT-B/16 transformer encoder, which processes 256×256 inputs to 257 feature tokens at an embedding dimension of $d_E = 384$ with $N_e = 12$ layers. The transformer decoder consists of $N_d = 12$ layers at an output dimension of $d_D = 512$, while the X-triplane has a feature dimension of $d_T = 32$. The MLP used for radiodensity queries has four layers with a hidden dimension of 64.

During training, our model is designed to learn from a set of possible input view counts, $V = \{6, 8, 10\}$. For each epoch, the same instance is processed 3 times, each with a different number of views selected from V. Training is conducted on 8 RTX A5000 GPUs at a per-gpu batch size of 6 for 100 epochs. For evaluation, we adopt the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [50] as the quantitative metrics. Please note that PSNR is measured directly in 3D space and SSIM is computed as the average of 2D SSIM values.

4.2. Comparison with State-of-the-Art Methods

We evaluate our X-LRM model against baseline methods under different numbers of projection views (*i.e.* 6, 8, 10) using the following two different settings:

- **Traditional and feedforward methods**: Traditional methods are directly tested on the 750-sample test set. The 2D and 3D feedforward methods are first trained on

Method	Recon.		Left Lung		Right Lung	
	PSNR	SSIM	DICE	ASD↓	DICE	ASD↓
FDK	9.14	0.03	0.34	43.41	0.26	45.12
SART	21.7	0.51	28.29	13.44	2.92	28.12
ASD-POCS	21.48	0.53	25.35	15.62	2.52	31.84
FBPConvNet	26.02	0.68	93.59	0.65	93.58	0.56
FreeSeed	27.77	0.83	91.01	1.07	90.56	0.82
DIF-Net	24.71	0.55	84.63	1.70	84.78	1.44
DIF-Gaussian	26.84	0.79	92.16	0.83	91.69	0.72
C^2RV	28.24	0.83	91.47	0.88	90.28	0.87
X-LRM (Ours)	**30.59**	**0.92**	**95.21**	**0.49**	**94.63**	**0.48**

Table 4. Traditional and feedforward methods.

Method	Recon.		Left Lung		Right Lung	
	PSNR	SSIM	DICE	ASD↓	DICE	ASD↓
NAF	21.91	0.57	48.54	19.14	50.08	9.49
R^2-Gaussian	18.58	0.45	29.62	26.12	34.76	12.86
SAX-NeRF	21.83	0.59	39.34	27.55	19.87	20.86
DiffusionMBIR	25.31	0.72	93.10	0.74	93.25	0.67
DDS	23.47	0.53	71.04	2.61	69.58	2.60
X-LRM (Ours)	**27.63**	**0.85**	**95.60**	**0.51**	**95.48**	**0.48**

Table 5. Self-supervised and diffusion methods.

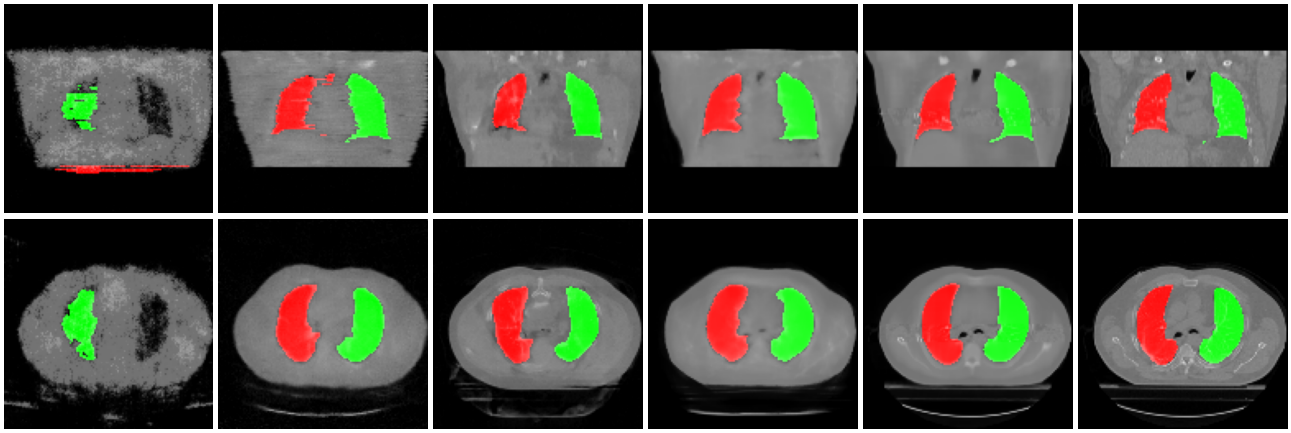

Figure 6. Visual comparison of lung segmentation on 6-view reconstructed CT slices with the recent best self-supervised method NAF [55], 2D feedforward method FreeSeed [34], diffusion-based method DiffusionMBIR [9], and 3D feedforward method C^2RV [31].

the train set and then tested on the 750-sample test set.

- **Self-supervised and diffusion-based methods**: We use a subset of 10 samples selected from the 750-sample test set, ensuring all 10 sub-datasets in Sec. 3.1 are covered. We test on this small dataset due to the long inference time of these methods.

Quantitative Results. Firstly, we compare X-LRM with three traditional methods (FDK [17], SART [1], and ASD-POCS [42]) and five feedforward methods (FBPConvNet [24], FreeSeed [34], DIF-Net [30], DIF-Gaussian [30], and C^2RV [31]). The results are reported in Tab. 2. **(i)** When reconstructing CT volumes from 6, 8, and 10 X-ray projection views, our X-LRM surpasses the SOTA 2D feedforward method, FreeSeed, by 2.12, 1.16, and 1.16 dB in PSNR. Compared to the SOTA 3D feedforward method, C^2RV, X-LRM improves the performance by 1.54, 1.41, and 0.37 dB in PSNR, while enjoying over 27× faster inference speed. **(ii)** Unlike previous feedforward methods, X-LRM enjoys better flexibility as it can efficiently reconstruct CT volume with different numbers of input views without training separate models.

Secondly, we compare with three self-supervised methods (NAF [55], R^2-Gaussian [56], and SAX-NeRF [7]) and two diffusion-based methods (DiffusionMBIR [9] and DDS [10]). The quantitative results are listed in Tab. 3. Our X-LRM achieves the best performance and fastest inference speed. Compared with the second-best method, DiffusionMBIR, our X-LRM is 3.53 dB higher in PSNR. Compared with the second-fastest method, R^2-Gaussian, our method is over 2570× faster in inference.

Qualitative Results. The qualitative results are depicted in Fig. 4 (compared with feedforward methods) and Fig. 5 (compared with self-supervised and diffusion-based methods). As observed from the reconstructed slices, all baseline methods struggle with generating high-quality reconstructions, particularly in sparser-view scenarios. Both feedforward and optimization-based approaches exhibit noticeable blurriness and lack of fine details, leading to incomplete anatomical structures and texture inconsistencies. Structural elements, such as lung regions and organ boundaries, appear unclear, often blending into surrounding areas due to the loss of high-frequency details.

Method	Base Model	+ X-Triplane	+ X-former
PSNR	13.09	28.76	**31.33**
SSIM	0.42	0.84	**0.92**

Table 6. Break-down ablation towards higher performance by adding the components of X-LRM. The ablation study is conducted under the 10-view CT reconstruction setting.

Noisy parameters			PSNR	SSIM(10^{-2})
Angles	DSO	DSD		
-	-	-	31.05 (-0.00)	91.04 (-0.00)
$\pm 0.5°$	-	-	30.93 (-0.12)	90.95 (-0.09)
$\pm 1°$			30.62 (-0.43)	90.71 (-0.33)
-	$\pm 2mm$	-	30.85 (-0.20)	90.89 (-0.15)
	$\pm 3mm$		30.67 (-0.38)	90.73 (-0.31)
-	-	$\pm 2mm$	30.99 (-0.06)	90.99 (-0.05)
		$\pm 3mm$	30.93 (-0.12)	90.95 (-0.09)

Table 7. Ablation study of X-LRM's robustness to noisy X-ray scanner parameters under a 6-view CT reconstruction setting.

In contrast, our X-LRM yields visually sharper reconstructions with well-defined textures and more coherent anatomical structures. Across different view settings, it preserves fine-grained details while maintaining spatial smoothness. Our method consistently reconstructs realistic features with minimal artifacts, demonstrating high-quality performance in sparse-view CT reconstruction.

Application in Segmentation. We evaluate the reconstructed CT volumes using medical segmentation. We employ the LungMask toolkit [20] to segment the left and right lung from CT reconstructions produced by various methods and compare the results against the ground-truth segmentation obtained from the original CT scans. Specifically, we evaluate lung test data from the 750-test set and 10-test set, testing the corresponding baseline methods and reporting reconstruction performance (PSNR and SSIM) alongside lung segmentation accuracy (DICE and ASD) for 6-view reconstructed volumes. As shown in Tab. 4 and Tab. 5, X-LRM achieves superior reconstruction quality, surpassing C^2RV by and DiffusionMBIR by 2.35 and 2.32 dB in PSNR. Additionally, the higher DICE scores and lower ASD values on both the left and right lung indicate that the 3D segmentation on the CT volume reconstructed by X-LRM has a larger overlap and smaller boundary discrepancies with the segmentation mask on the ground-truth CT volume. Fig. 6 shows the visual comparison with four kinds of recent best methods. Both quantitative and qualitative results demonstrate the ability of X-LRM to preserve anatomical structures more accurately and maintain precise shape consistency, surpassing other methods in both reconstruction fidelity and segmentation alignment.

4.3. Ablation Study

Ablation studies evaluate the effectiveness of the proposed modifications compared to the standard LRM, including X-former and X-triplane. Additionally, we assess the robustness of X-LRM under varying noisy scanning parameters, such as viewing angles, DSD, and DSO. The breakdown study is performed under 6,8,10-view setting, and the robustness analysis is conducted under 6-view setting.

Break-down Ablation. We adopt the Open-LRM [21] as the base model to study the effect of each component of X-LRM towards higher performance. The results of the 10-view reconstruction are reported in Tab. 6. The base model only achieves poor results of 12.33 dB in PSNR on average. After applying our X-triplane and X-former, the model gains by 15.53 and 3.34 dB in PSNR on average. These results validate the effectiveness of our proposed methods.

Robustness Analysis. We conduct a robustness analysis under varying noisy scanning parameters, including viewing angles, source-to-origin distance (DSO), and source-to-detector distance (DSD). The introduced noise follows a uniform distribution, modeled as $\eta \sim \mathcal{U}(-\epsilon, +\epsilon)$. With this noise, the projection images change but the model processes them as if captured under perfect conditions. Tab. 7 shows that X-LRM remains robust to noises of scanning parameters. Viewing angle shifts of $\pm 0.5°$ (PSNR -0.12 dB, SSIM -0.0009) and $\pm 1°$ results (PSNR -0.43 dB, SSIM -0.0033) have minimal impact. Noises in DSO and DSD only introduce minor effects, demonstrating the reliability of X-LRM under different real-world possible noises.

5. Conclusion

In this paper, we collect the largest dataset, Torso-16K, to enable large-scale training for CT reconstruction. Torso-16K is over 18× larger than the existing largest benchmark. We propose X-LRM, a Transformer-based feed-forward framework consisting of X-former and X-triplane. X-former employs a tokenizer and Transformer backbone to flexibly encode an arbitrary number of input views, enabling X-LRM to reconstruct CT volumes without retraining. X-triplane decodes image tokens into a triplane representation and learns a neural implicit function to model 3D radiodensity. Experiments show that X-LRM surpasses the SOTA 3D feedforward method by 1.5 dB while achieving 27× faster speed, with its application in medical segmentation further highlighting its practical value.

Acknowledgement

This work was supported by the Lustgarten Foundation for Pancreatic Cancer Research, the Patrick J. McGovern Foundation Award, and the National Institutes of Health (NIH) under Award Number R01EB037669.

References

[1] Anders H Andersen and Avinash C Kak. Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. *Ultrasonic imaging*, 1984. 1, 2, 7

[2] Rushil Anirudh, Hyojin Kim, Jayaraman J Thiagarajan, K Aditya Mohan, Kyle Champley, and Timo Bremer. Lose the views: Limited angle ct reconstruction via implicit sinogram completion. In *CVPR*, 2018. 2

[3] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016. 3

[4] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. Tigre: a matlab-gpu toolbox for cbct image reconstruction. *Biomedical Physics & Engineering Express*, 2016. 6

[5] Yuanhao Cai, Yixun Liang, Jiahao Wang, Angtian Wang, Yulun Zhang, Xiaokang Yang, Zongwei Zhou, and Alan Yuille. Radiative gaussian splatting for efficient x-ray novel view synthesis. In *ECCV*, 2024. 1, 2

[6] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *CVPR*, 2024. 2

[7] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *CVPR*, 2024. 1, 7

[8] Yuanhao Cai, He Zhang, Kai Zhang, Yixun Liang, Mengwei Ren, Fujun Luan, Qing Liu, Soo Ye Kim, Jianming Zhang, Zhifei Zhang, et al. Baking gaussian splatting into diffusion denoiser for fast and scalable single-stage image-to-3d generation. *arXiv preprint arXiv:2411.14384*, 2024. 2, 3

[9] Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. In *CVPR*, 2023. 2, 7

[10] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *ICLR*, 2024. 2, 7

[11] Marco Cipriano, Stefano Allegretti, Federico Bolelli, Mattia Di Bartolomeo, Federico Pollastri, Arrigo Pellacani, Paolo Minafra, Alexandre Anesi, and Costantino Grana. Deep segmentation of the mandibular canal: a new 3d annotated dataset of cbct volumes. *Ieee Access*, 2022. 1, 5

[12] Allan Macleod Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of applied physics*, 1963. 1

[13] Allan Macleod Cormack. Representation of a function by its line integrals, with some radiological applications. ii. *Journal of Applied Physics*, 1964. 1

[14] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. Objaverse-xl: A universe of 10m+ 3d objects. In *NeurIPS*, 2023. 2

[15] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *CVPR*, 2023. 2

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 2

[17] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Josa a*, 1984. 1, 7

[18] Muhammad Usman Ghani and W Clem Karl. Deep learning-based sinogram completion for low-dose ct. In *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*, 2018. 2

[19] Hao He, Yixun Liang, Luozhou Wang, Yuanhao Cai, Xinli Xu, Hao-Xiang Guo, Xiang Wen, and Yingcong Chen. Lucidfusion: Generating 3d gaussians with arbitrary unposed images. *arXiv preprint arXiv:2410.15636*, 2024. 2

[20] Johannes Hofmanninger, Forian Prayer, Jeanny Pan, Sebastian Röhrich, Helmut Prosch, and Georg Langs. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology Experimental*, 2020. 8

[21] Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. In *ICLR*, 2024. 2, 8

[22] Godfrey N Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British journal of radiology*, 1973. 1

[23] Godfrey N Hounsfield. Computed medical imaging. *Science*, 1980. 1

[24] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *TIP*, 2017. 1, 7

[25] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 2017. 2

[26] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 2023. 2

[27] Suhyeon Lee, Hyungjin Chung, Minyoung Park, Jonghyuk Park, Wi-Sun Ryu, and Jong Chul Ye. Improving 3d imaging with pre-trained perpendicular 2d diffusion models. In *ICCV*, 2023. 2

[28] Jiahao Li, Hao Tan, Kai Zhang, Zexiang Xu, Fujun Luan, Yinghao Xu, Yicong Hong, Kalyan Sunkavalli, Greg Shakhnarovich, and Sai Bi. Instant3d: Fast text-to-3d with sparse-view generation and large reconstruction model. In *ICLR*, 2024. 2

[29] Yiqun Lin, Zhongjin Luo, Wei Zhao, and Xiaomeng Li. Learning deep intensity field for extremely sparse-view cbct reconstruction. In *MICCAI*, 2023. 1, 2

[30] Yiqun Lin, Hualiang Wang, Jixiang Chen, and Xiaomeng Li. Learning 3d gaussians for extremely sparse-view cone-beam ct reconstruction. In *MICCAI*, 2024. 1, 7

[31] Yiqun Lin, Jiewen Yang, Hualiang Wang, Xinpeng Ding, Wei Zhao, and Xiaomeng Li. C^ 2rv: Cross-regional and

cross-view learning for sparse-view cbct reconstruction. In *CVPR*, 2024. 1, 2, 7

[32] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. 6

[33] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[34] Chenglong Ma, Zilong Li, Junping Zhang, Yi Zhang, and Hongming Shan. Freeseed: Frequency-band-aware and self-guided network for sparse-view ct reconstruction. In *MICCAI*, 2023. 1, 2, 7

[35] Cynthia McCollough. Tu-fg-207a-04: overview of the low dose ct grand challenge. *Medical physics*, 2016. 1, 5

[36] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, 2019. 2

[37] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 2021. 2

[38] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019. 6

[39] Ken Sauer and Charles Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 1993. 2

[40] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas Van Den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical image analysis*, 2017. 1, 5

[41] Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 2

[42] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 2008. 1, 7

[43] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 2008. 2

[44] Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. Triposr: Fast 3d object reconstruction from a single image. *arXiv preprint arXiv:2403.02151*, 2024. 2

[45] United Imaging Healthcare. Uct 960+. https://eu.united-imaging.com/en/product-service/products/ct/uct-960, 2023. 6

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 2, 3

[47] Ce Wang, Kun Shang, Haimiao Zhang, Qian Li, and S Kevin Zhou. Dudotrans: dual-domain transformer for sparse-view ct reconstruction. In *International Workshop on Machine Learning for Medical Image Reconstruction*, 2022. 2

[48] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, 2018. 2

[49] Peng Wang, Hao Tan, Sai Bi, Yinghao Xu, Fujun Luan, Kalyan Sunkavalli, Wenping Wang, Zexiang Xu, and Kai Zhang. Pf-lrm: Pose-free large reconstruction model for joint pose and shape prediction. *arXiv preprint arXiv:2311.12024*, 2023. 2

[50] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncell. Image quality assessment: from error visibility to structural similarity. *TIP*, 2004. 6

[51] Xinyue Wei, Kai Zhang, Sai Bi, Hao Tan, Fujun Luan, Valentin Deschaintre, Kalyan Sunkavalli, Hao Su, and Zexiang Xu. Meshlrm: Large reconstruction model for high-quality meshes. *arXiv preprint arXiv:2404.12385*, 2024. 2

[52] Rundi Wu, Yixin Zhuang, Kai Xu, Hao Zhang, and Baoquan Chen. Pq-net: A generative part seq2seq network for 3d shapes. In *CVPR*, 2020. 2

[53] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. In *NeurIPS*, 2019. 2

[54] Lifeng Yu, Yu Zou, Emil Y Sidky, Charles A Pelizzari, Peter Munro, and Xiaochuan Pan. Region of interest reconstruction from truncated data in circular cone-beam ct. *TMI*, 2006. 1

[55] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *MICCAI*, 2022. 1, 2, 7

[56] Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R2-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. In *NeurIPS*, 2024. 7

[57] Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R^2-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. In *NeurIPS*, 2024. 1, 2

[58] Kai Zhang, Sai Bi, Hao Tan, Yuanbo Xiangli, Nanxuan Zhao, Kalyan Sunkavalli, and Zexiang Xu. Gs-lrm: Large reconstruction model for 3d gaussian splatting. In *ECCV*, 2024. 2