X-Field: A Physically Informed Representation for 3D X-ray Reconstruction

Feiran Wang[1*] **Jiachen Tao**[1*] **Junyi Wu**[1*] **Haoxuan Wang**[1] **Bin Duan**[2]
Kai Wang[3] **Zongxin Yang**[4] **Yan Yan**[1†]

[1]University of Illinois Chicago [2]University of Michigan
[3]National University of Singapore [4]Harvard Medical School

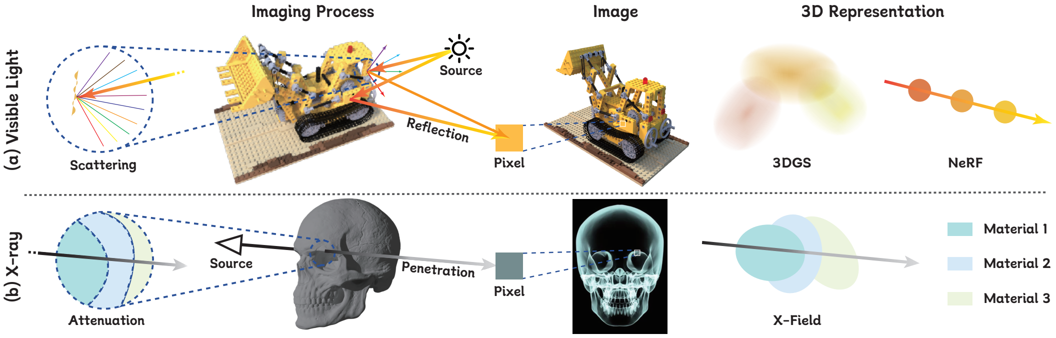

Figure 1: **Comparison of Imaging Processes and Corresponding 3D Representations.** (a) Visible light interacts with surfaces mainly through scattering and reflection. 3D representations of NeRF [1] and 3DGS [2] model this process by accumulating directional light rays. (b) Unlike previous approaches, our model represents internal structures using homogeneous ellipsoids to capture X-ray attenuation and penetration properties.

Abstract

X-ray imaging is indispensable in medical diagnostics, yet its use is tightly regulated due to radiation exposure. Recent research borrows representations from the 3D reconstruction area to complete two tasks with reduced radiation dose: X-ray Novel View Synthesis (NVS) and Computed Tomography (CT) reconstruction. However, these representations fail to fully capture the penetration and attenuation properties of X-ray imaging as they originate from visible light imaging. In this paper, we introduce **X-Field**, a 3D representation informed in the physics of X-ray imaging. First, we employ homogeneous 3D ellipsoids with distinct attenuation coefficients to accurately model diverse materials within internal structures. Second, we introduce an efficient path-partitioning algorithm that resolves the intricate intersection of ellipsoids to compute cumulative attenuation along an X-ray path. We further propose a hybrid progressive initialization to refine the geometric accuracy of X-Field and incorporate material-based optimization to enhance model fitting along material boundaries. Experiments show that X-Field achieves superior visual fidelity on both real-world human organ and synthetic object datasets, outperforming state-of-the-art methods in X-ray NVS and CT Reconstruction. Our code is available on the project page: https://github.com/Brack-Wang/X-Field.

1 Introduction

X-ray imaging is a cornerstone of Computed Tomography (CT) reconstruction [3, 4, 5, 6], providing critical insights into internal structures for clinical diagnostics. X-rays undergo progressive attenuation

---

[*] Equal Contribution [†] Corresponding Author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

while penetrating materials until the residual energy reaches the detector, forming an X-ray projection. Traditional CT reconstruction relies on hundreds of X-ray projections acquired from various angles to recover a density field [7, 8, 9]. However, acquiring such a vast number of projections exposes patients to high doses of ionizing radiation, posing health risks [10, 11]. Consequently, X-ray reconstruction has gained increasing attention [12, 13, 14, 15, 16, 17], aiming to synthesize novel X-ray views from a sparse set of input projections and facilitate the reconstruction of high-quality CT volumes.

Recent advances in 3D reconstruction [2, 1, 18] have laid the groundwork for X-ray reconstruction [14, 15, 16]. Yet, current 3D reconstruction techniques were originally designed for visible light imaging, as illustrated in Figure 1(a). When visible light rays interact with a surface, despite minor absorption, most wavelengths are *reflected* or *scattered* into multiple directions [19]. As a result, light rays from different directions can accumulate to form a single pixel, producing varying appearances of the same surface depending on the viewpoint. Neural Radiance Fields (NeRF) [1] employ deep networks to implicitly encode directional light information at each sampled point, conditioned by spatial positions and view directions to model appearance variation. Similarly, 3D Gaussian Splatting (3DGS) [2] projects anisotropic ellipsoids onto the image plane and accumulates the projected Gaussian splats to determine the pixel colors. Both methods reconstruct scenes considering the accumulation of multi-directional rays, thus well-suited for visible light imaging.

However, X-ray imaging operates on fundamentally different principles from visible light imaging, as depicted in Figure 1(b). First, X-rays are a form of high-energy electromagnetic radiation that is more capable of *penetrating* objects [20], whereas visible light, emitted from natural sources, has lower energy and is primarily reflected by surfaces. Second, X-rays progressively *attenuate* as they pass through an object in a nearly straight line before reaching the detector on the opposite side. Third, the pixel intensity is primarily determined by energy absorption from the corresponding X-ray, rather than by the combined contributions of visible light from multiple directions.

Therefore, X-ray imaging reveals materials composition with distinct densities distributed throughout the object's interior structure. As X-rays penetrate materials, their energy attenuates proportionally to both material density and path length through each substance. Recent X-ray reconstruction methods build upon 3DGS and achieve efficient imaging. For instance, X-Gaussian [15] replaces the spherical harmonics function with an intensity function to capture X-ray grayscale values, while R^2-Gaussian [16] extends this by correcting integration bias to enhance CT reconstruction quality. Yet, these methods represent each material as a Gaussian ellipsoid, depicted as 3DGS representation in Figure 1, with maximum density at the center that gradually diminishes toward the periphery, and render through ellipsoid superposition, without accounting for the critical factor of penetration length through each material. To faithfully adhere to the principles of X-ray imaging, X-ray reconstruction demands a representation that accurately models the radiological density across various materials.

In this paper, we present **X-Field**, a physically informed ellipsoid representation specifically designed to model X-ray attenuation as it penetrates materials during imaging. To differentiate various materials, X-Field represents an object's structure using ellipsoids with distinct attenuation coefficients, shown as X-Field representation in Figure 1. Then, to track the distance an X-ray travels through each material, we derive an explicit form for segment length.

For precise pixel intensity calculation, we express the attenuation integral along the ray path as a finite Riemann sum of the attenuation coefficients and segment lengths. Yet, during optimization, ellipsoids with different attenuation coefficients may overlap. To ensure that each position in 3D space corresponds to a single material, we further design a first-pass precedence strategy, resolving ellipsoid ambiguities in overlapping regions. Furthermore, we propose Material-Based Optimization, which adaptively refines ellipsoid placement by splitting them along material boundaries guided by the local attenuation gradient.

We evaluate X-Field on X-ray reconstruction across various modalities, including real-world human organ datasets and simulated object datasets. Furthermore, we compare X-Field with existing methods in sparse-view CT reconstruction based on Novel View Synthesis (NVS) results. Experiments demonstrate that X-Field outperforms state-of-the-art (SOTA) methods in both NVS and CT reconstruction. Notably, our method achieves optimal results within 10 minutes using only 10 input X-ray projections, delivering substantial PSNR improvements of 2.44 dB for X-ray NVS and 3.98 dB for CT reconstruction over SOTA methods.

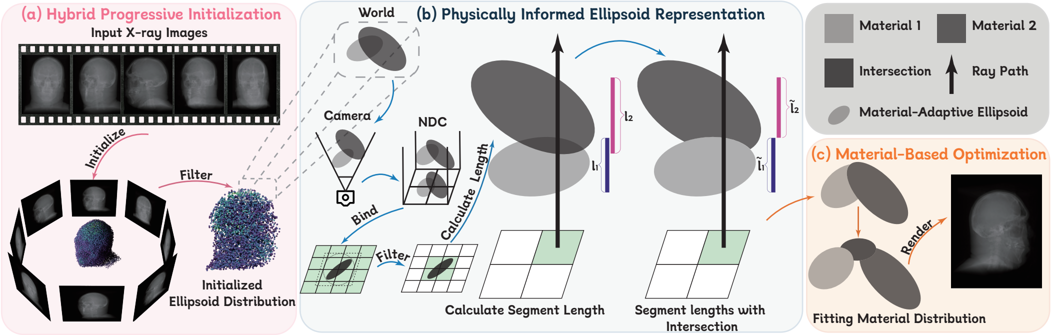

Figure 2: **Overview of X-Field.** **(a) Hybrid Progressive Initialization.** We begin with X-ray images to construct a coarse initialization using combined iterative methods. **(b) Physically Grounded Ellipsoid Representation.** We transform initialized ellipsoids into NDC space and associate them with pixels. We then compute the attenuation integral along the ray, considering segment length and ellipsoid intersections. **(c) Material-Based Optimization.** Our optimization captures detailed material boundaries for high-quality rendering.

2 Related Work

X-ray Novel View Synthesis.

3D X-ray reconstruction consists of two tasks: X-ray novel view synthesis and CT reconstruction [14]. X-ray NVS [14, 15, 13, 16] is essential for medical diagnostics, as it reconstructs images from sparse-view inputs, significantly reducing patient exposure to X-ray radiation. Motivated by the fact that visible light and X-rays are both electromagnetic waves, two representation methods originally designed for visible light have been adopted for X-ray reconstruction: Neural Radiance Fields (NeRF) [1], which employ multi-layer perceptrons for implicit encoding, and 3D Gaussian Splatting (3DGS) [2], which uses Gaussian ellipsoids for explicit representation. NeRF-based methods, such as NAF [13], improve efficiency by incorporating hash tables, while SAX-NeRF [14] introduces a transformer architecture to better model 3D structural dependencies.

3DGS-based approaches, such as X-Gaussian[15], replace the spherical harmonics color in 3DGS with a response function, addressing differences between visible light color and X-ray intensity. R^2-Gaussian[16] further corrects integration bias for more accurate density retrieval. However, neither representation has a comprehensive consideration of the attenuation and penetration properties of the X-ray. Instead, we introduce attenuation-adaptive ellipsoids with distinct attenuation coefficients and consider the X-ray penetration length within the materials, ensuring more accurate reconstructions.

Computed Tomography Reconstruction. CT reconstruction is of vital importance for domains such as medical diagnosis [6, 21, 22, 5, 23, 24], biology [25, 26, 27], industrial inspection [28, 29], and security screening [30]. Early CT reconstruction methods [20] can be categorized into analytical approaches [3, 31] and iterative techniques [4, 32, 33]. Analytical methods such as FDK [3] perform a filtered back-projection of acquired projections. Iterative methods such as SART [4] and CGLS [34] reconstruct CT images by iteratively refining the solution through algebraic projection corrections and solving a least-squares optimization problem, respectively. However, these methods typically require hundreds to thousands of X-ray images, leading to increased radiation exposure.

Fortunately, X-ray NVS offers an effective alternative to CT reconstruction. By generating diverse novel views from sparse X-ray inputs [15, 16, 14, 17, 35], it enables reconstruction of high-quality CT volumes from limited data. We further reduce the number of required X-ray images to as few as five while achieving superior performance both qualitatively and quantitatively compared to other baselines, highlighting its practicality for real-world applications.

3 Method

In this section, we introduce our Physically Informed Ellipsoid Representation in Sec. 3.1, including (1) the formulation of X-ray Physical Field, (2) the Attenuation-Adaptive Ellipsoids, (3) the algorithms for calculating segment lengths with intersections, and (4) the overlap filtering method. Then, we propose Hybrid Progressive Initialization in Sec. 3.2 and Material-Based Optimization in Sec. 3.3.

3.1 Physically Informed Ellipsoid Representation

X-ray Physical Field. X-ray imaging quantifies the cumulative attenuation of X-rays as they traverse an object, governed by the material-dependent absorption properties of the medium. To formally describe this process, we define the X-ray Physical Field, where each spatial position $\mathbf{x} \in \mathbb{R}^3$ is characterized by the local energy absorption rate $\sigma(\mathbf{x}) \in \mathbb{R}^+$ of X-ray [20, 5, 36]. When X-rays with an initial intensity I_0 propagate through the X-ray Physical Field, their energy is progressively attenuated by the materials they traverse. Ultimately, the remaining intensity forms a projection image $I \in \mathbb{R}^{H\times W}$. Mathematically, we represent an X-ray path as $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d} \in \mathbb{R}^3$, where $\mathbf{o}$ denotes the X-ray source position, $\mathbf{d}$ is the unit view direction vector, and t varies between the entry t_0 and exit t_n points of the object. According to the Beer-Lambert law [37], the X-ray intensity $I'(\mathbf{r})$ after attenuation is given by:

$$I'(\mathbf{r}) = I_0 \exp\left(-\int_{t_0}^{t_n} \sigma(\mathbf{r}(t))\, dt\right). \tag{1}$$

In practice, the raw data is typically processed in logarithmic space for computational efficiency [20]:

$$I(\mathbf{r}) = \log I_0 - \log I'(\mathbf{r}) = \int_{t_0}^{t_n} \sigma(\mathbf{r}(t))\, dt. \tag{2}$$

Thus, each pixel intensity $I(\mathbf{r})$ in the projection image aggregates the material absorption along the X-ray path, providing the foundation for reconstructing internal structures.

Attenuation-Adaptive Ellipsoids. Inspired by the infinitesimal method [38], we represent the spatial distribution of materials using a set of N ellipsoids $\{\mathbf{E}_i\}$. Each ellipsoid $\mathbf{E}_i$ encodes material properties and contains a center position $\mathbf{p_c}$ and covariance matrix $\mathbf{\Sigma}_{\text{3D}}$. To maintain physical consistency with X-ray imaging, we further assign a distinct, non-negative attenuation coefficient σ_i to characterize the local absorption behavior. Given an X-ray path $\mathbf{r}$ that traverses multiple ellipsoids, the corresponding pixel intensity is determined by the accumulated attenuation along the path:

$$\begin{aligned} I(\mathbf{r}) = \int_{t_0}^{t_n} \sigma(\mathbf{r}(t))\, dt &= \int_{t_0}^{t_1} \sigma(\mathbf{r}(t))\, dt + \cdots + \int_{t_{n-1}}^{t_n} \sigma(\mathbf{r}(t))\, dt \\ &= \sigma_0 \int_{t_0}^{t_1} dt + \cdots + \sigma_{n-1} \int_{t_{n-1}}^{t_n} dt \\ &= \sigma_0 l_0 + \sigma_1 l_1 + \cdots + \sigma_{n-1} l_{n-1}, \end{aligned} \tag{3}$$

where $l_i = t_{i+1} - t_i \geq 0$ represents the segment length of the X-ray path within ellipsoid $\mathbf{E}_i$. Therefore, accurately measuring the accumulated attenuation requires precise determination of each segment length l_i. In the following, we derive an explicit formulation for l_i.

Explicit Form of Segment Lengths. Our objective is to compute the segment length l_i for ellipsoid $\mathbf{E}_i$ along the view direction $\mathbf{d}$, a process illustrated in Figure 2(b). To achieve this, we first transform $\mathbf{E}_i$ from world to camera coordinates and then into Normalized Device Coordinates (NDC) space, aligning X-rays with the coordinate axes for computational simplicity. To establish a reference for general segment length computation, we define the maximum segment length l_{max} as a pivot, which corresponds to the ray path passing through the center of $\mathbf{E}_i$:

$$l_{\text{max}} = \frac{2}{\sqrt{\mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}}}. \tag{4}$$

Given l_{max}, we then compute the segment length l_i for an arbitrary point $\mathbf{u}$ inside the projected ellipse of $\mathbf{E}_i$ in NDC space. The relationship is formulated as:

$$l_i = l_{\text{max}} \times \sqrt{1 - \left(\frac{C - B^2}{A}\right)}, \text{ where} \tag{5}$$

$$A = \mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}, \quad B = \mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}, \quad C = \mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{a}. \tag{6}$$

Here, $\mathbf{a} = \mathbf{u} - \mathbf{p_c}$ is the displacement vector from the center of the ellipsoid $\mathbf{p_c}$ to the point $\mathbf{u}$. The full derivation of l_{max} and l_i is provided in the supplementary materials.

Segment Lengths with Intersections. As ellipsoids adjust their positions and sizes during optimization, intersections may occur, leading to ambiguities in the attenuation rates within overlapping regions. To ensure that each location corresponds to a single dominant material, resolving these complex overlaps is crucial to maintaining a physically consistent representation.

Algorithm 1 Segment Length Correction with Intersections

```
Input: $(z_0, z_1, \dots, z_{n-1})$: Sorted depths
    $(l_0, l_1, \dots, l_{n-1})$: Initial segment lengths
Output: Updated segment lengths $\tilde{l}_0, \tilde{l}_1, \dots, \tilde{l}_{n-1}$
for $i = 0$ to $n - 1$ do
    if $i == 0$ then
        $\tilde{l}_0 \leftarrow l_0$
        $z \leftarrow z_0, l \leftarrow l_0$
    else
        if $z_i < z + \frac{1}{2}l$ then
            $\tilde{l}_i \leftarrow \max(0, (z_i + \frac{1}{2}l_i) - (z + \frac{1}{2}l))$
        else
            $\tilde{l}_i \leftarrow \min(l_i, (\frac{1}{2}l_i + z_i) - (z + \frac{1}{2}l))$
        end if
        if $\tilde{l}_i \neq 0$ then
            $[z_i + \frac{1}{2}l_i - \tilde{l}_i, z_i + \frac{1}{2}l_i]$ ; $z \leftarrow z_i, l \leftarrow l_i$
        end if
    end if
end for
```

For optimization stability, we adopt a simple yet effective first-pass precedence strategy. Given two intersecting ellipsoids along the X-ray path, we first sort them in ascending depth order, denoted by $\mathbf{E}_i$ and $\mathbf{E}_{i+1}$, where $\mathbf{E}_i$ appears first along the ray. The overlapping region e_i is then exclusively assigned to $\mathbf{E}_i$ by adjusting the effective region of $\mathbf{E}_{i+1}$ to $\mathbf{E}_{i+1} - e_i$. This modification directly impacts the segment length computation, requiring additional updates to ensure consistency in attenuation accumulation, as shown in Figure 2(b). To systematically handle different intersection scenarios, we introduce an efficient correction function f, as outlined in Algorithm 1. The final accumulated attenuation along the X-ray path is then expressed as:

$$I(\mathbf{r}) = \sum_{i=0}^{n-1} \sigma_i \cdot f\left(\frac{2}{\sqrt{A}} \times \sqrt{1 - \left(\frac{C - B^2}{A}\right)}\right), \tag{7}$$

Physically Faithful Overlap Filtering. While resolving segment length inconsistencies ensures a physically valid attenuation model along the X-ray path, accurately associating ellipsoids with their corresponding pixels is equally critical for maintaining consistency in the projected space. Current methods often rely on coarse bounding-box approximations, leading to unintended pixel assignments that compromise reconstruction fidelity. To address this, we introduce a refined Oriented Bounding Box (OBB) strategy that establishes precise pixel-ellipsoid associations. As illustrated in Figure 3, existing approaches [2, 15, 16] typically rely on Axis-Aligned Bounding Boxes (AABB) [39], which approximate the projected ellipse using a circumscribed circle. However, this method includes extraneous pixels (red region) that do not truly overlap with the projected ellipse, leading to physically inconsistent assignments. In contrast, we first determine the OBB via eigenvalue decomposition, aligning it with the principal axes of the projected ellipse. Then, we filter out irrelevant pixels while preserving only the valid tiles (green region) that are closely related to the actual projected ellipse. By eliminating spurious pixel association, our method enhances the physical consistency and numerical stability of the reconstruction process.

3.2 Hybrid Progressive Initialization

It is widely acknowledged that initializing point clouds with detailed geometric information accelerates model convergence [2, 15, 40]. However, conventional initialization techniques fail in the context of X-ray imaging. Specifically, SfM-based methods [41, 2] estimate point clouds by feature matching across multiple images. While in X-ray imaging, variations in ray paths significantly alter the remaining energy intensity, leading to feature mismatches and unreliable correspondences. Similarly, point cloud regression methods such as Dust3R [42, 43, 44] directly infer 3D structures from images but fail in X-ray reconstruction due to domain shifts, as they are trained on natural images and do not account for X-ray-specific attenuation properties.

To address this challenge, we design a Hybrid Initialization strategy that progressively refines the geometric structure through a sequence of iterative methods. We compare the results in Figure 4. Our process begins with Conjugate Gradient Least Squares (CGLS) [45, 34, 20], which efficiently

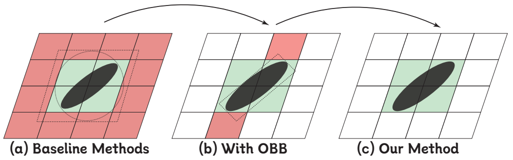

Figure 3: **Pixel-Ellipse Association.** (a) AABB [39] results in incorrect associations. (b) OBB [47] pixels. (c) Ours keeps only aligned pixels.

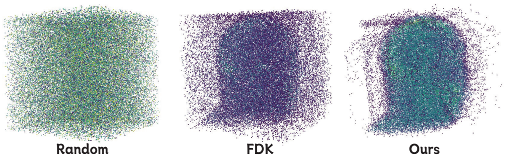

Figure 4: **Initialization Comparisons.** Ours produces informative geometry prior, compared to Random of X-Gaussian [15] and FDK of R^2-Gaussian [16].

provides a coarse global estimate. Next, we refine this estimate using Simultaneous Algebraic Reconstruction Technique (SART) [4], leveraging its capability to enhance local details and correct early-stage inconsistencies. Finally, we incorporate Total Variation (TV) regularization [46] into the initialization process to suppress noise and artifacts, preserving critical structural edges. For each 3D point obtained from our initialization, we define an ellipsoid centered at that position, assigning it a random covariance matrix $\mathbf{\Sigma}_{\text{3D}}$ and an attenuation coefficient σ_i. We subsequently devise an optimization method for these parameters for better modeling the detailed material distribution.

3.3 Material-Based Optimization

Mainstream optimization strategies [2, 15, 16, 40] involve splitting and cloning in regions with poor structural fidelity while pruning with negligible opacity. However, this geometry-based optimization fails to capture the material composition along materials' edge. As depicted in Figure 5(a), the ground-truth object consists of three distinct materials, represented by different colors. While the geometry-based optimization successfully reconstructs the overall shape, it captures only two intermediate material types, as shown in Figure 5(b), highlighting the limitations of purely geometry-centric approaches in faithfully modeling material compositions.

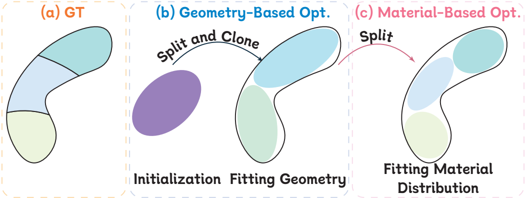

Figure 5: **Illustration of Adaptive Optimization Strategy.** **(a) Ground-truth Material Distribution** with different colors indicating distinct materials. **(b) Geometry-Based Optimization** [2], which fits the ellipsoids closely to the object geometry. **(c) Our Material-Based Optimization**, which further refines the ellipsoids to capture the material distribution.

The challenge arises from the continuous fluctuation of an ellipsoid's attenuation coefficient, which hinders stable convergence and prevents accurate material differentiation. To address this, we propose a material-based optimization strategy that explicitly accounts for the heterogeneous material distributions inherent in the X-ray Physical Field. Our approach is designed to identify material boundaries through local density estimation. Concretely, we randomly sample a subset of ellipsoids and compute the k-nearest neighbors [48], averaging their distances and analyzing the attenuation gradient. Regions with high density and steep gradients indicate material transitions, necessitating a finer adjustment. Therefore, to enhance fidelity, we selectively split ellipsoids in these complex regions, scaling each by a factor of 1.6, which follows empirical heuristics in prior studies [2, 15, 16]. As presented in Figure 5(c), our strategy significantly improves material separation, faithfully capturing the internal composition of objects.

4 Experiment

4.1 Experiment Settings

Dataset. Following the literature convention [14, 15], we conduct experiments on a large-scale X3D dataset containing 15 scenes with two collections: Human Organs, derived from real-world medical datasets, to evaluate model performance in the medical domain; and Daily Objects, generated from synthetic datasets, to assess generalization ability. Specifically, chest scans are sourced from LIDC-IDRI [49], pancreas CT scans from Pancreas-CT [50]. The remaining objects are obtained from VOLVIS [51] and the open scientific visualization dataset [52]. We adopt the tomography toolbox TIGRE [53] to capture projections from CT volumes in the range of $0^\circ \sim 180^\circ$ with minor scatter and electronic noise. For highly sparse-view novel view synthesis, 5 and 10 views are used for

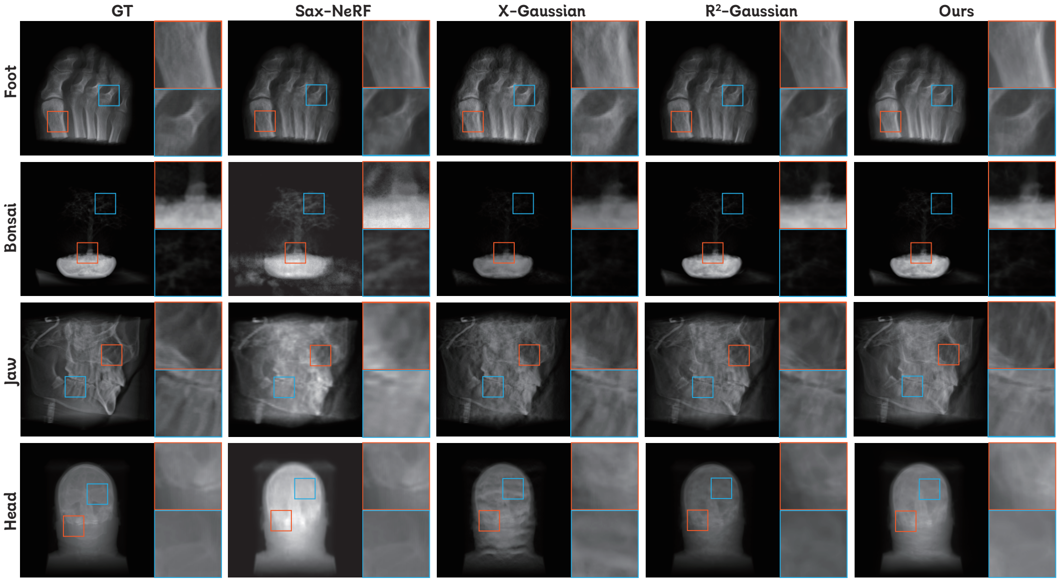

Figure 6: **Qualitative Comparison of NVS.** We present visual examples of reconstructed images across four cases trained with 10 views. Our results demonstrate superior visual quality, richer details, and fewer artifacts.

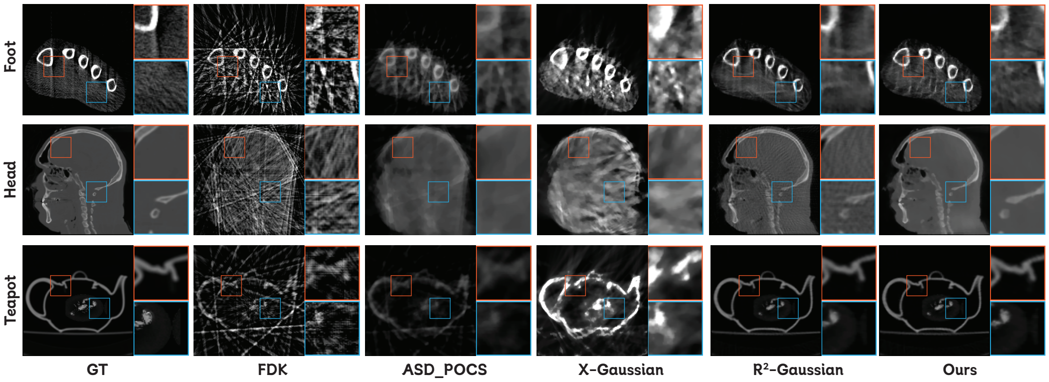

Figure 7: **Qualitative Comparison of CT Reconstruction.** Our method produces clearer textures, more refined anatomical structures, and fewer artifacts, particularly in high-contrast regions such as the cranial cavity.

training and 50 views are used for testing. To further assess model performance and scalability, we generate 50, 25, and 15 views for evaluating the performance under sparse-view synthesis.

Baselines. We compare X-Field with state-of-the-art 3D X-ray reconstruction methods, including TensoRF [12], NeAT [17], NAF [13], SAX-NeRF [14], X-Gaussian [15], and R^2-Gaussian [16]. TensoRF, NeAT, NAF, and SAX-NeRF are NeRF-based methods designed for efficient reconstruction, with SAX-NeRF achieving SOTA performance among them by incorporating a transformer architecture as the model backbone. X-Gaussian and R^2-Gaussian are 3DGS-based methods, where X-Gaussian focuses on NVS, and R^2-Gaussian extends its applicability to CT reconstruction by introducing voxelization. We also compare against traditional CT reconstruction methods including FDK [3] and SART [4]. The novel view images are obtained by leveraging TIGRE for rendering.

Metrics. We adopt peak signal-to-noise ratio (PSNR) [54] to assess the quality of rendered images, structural similarity index measure (SSIM) [55] to measure consistency between predicted images and ground-truth, and Learned Perceptual Image Patch Similarity (LPIPS) [56] to analyze the perceptual quality in high-level feature space. For clarity, we report LPIPS as LPIPS* = LPIPS $\times 10^3$ instead.

Implementation Details. For X-ray novel view synthesis, we evaluate learning-based methods using their official implementations. For CT reconstruction, we follow [17, 14, 15] to synthesize novel-view X-ray images from sparse inputs and reconstruct CT volumes using a total of 100 images. All experiments were conducted on a single RTX 6000 Ada GPU.

	Human Organ 10-views [16]			Daily Object 10-views [16]			Human Organ 5-views [16]			Daily Object 5-views [16]		
Method	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓
Traditional Methods												
FDK [3]	12.35	0.675	291.2	16.52	0.716	259.1	8.15	0.618	310.6	14.42	0.688	283.7
SART [4]	13.23	0.691	284.8	17.69	0.724	247.3	9.31	0.634	303.4	15.68	0.663	293.5
Deep Learning-based Methods												
TensoRF [12]	16.61	0.928	182.5	24.19	0.946	153.4	12.32	0.895	189.6	18.27	0.922	210.8
NeAT [17]	16.22	0.934	185.3	25.15	0.957	155.2	11.08	0.887	188.3	17.29	0.908	211.3
NAF [13]	17.89	0.925	193.2	25.44	0.949	151.9	11.19	0.894	197.1	17.02	0.923	208.5
SAX-NeRF [14]	19.32	0.945	186.4	25.38	0.979	143.6	14.18	0.901	191.2	19.09	0.948	204.9
X-Gaussian [15]	22.88	0.947	130.3	22.91	0.982	79.12	17.23	0.947	176.4	20.31	0.961	108.1
R^2-Gaussian [16]	33.72	0.967	85.97	41.93	0.986	54.31	31.12	0.956	109.7	34.52	0.965	82.46
Ours	35.71	0.980	71.03	42.80	0.983	45.64	32.34	0.963	103.2	37.41	0.970	81.02

Table 1: **Quantitative Comparison of NVS.** We compare our X-Field with: (a) Traditional analytical method: FDK , SART . (b) Deep Learning-based methods: TensoRF , NeAT , NAF , SAX-NeRF , X-Gaussian , and R^2-Gaussian . We report LPIPS* = LPIPS $\times 10^3$. We mark out best and second best method for all metrics.

4.2 Comparison with State-of-the-Art Methods

We present the quantitative results of X-ray NVS and CT reconstruction, discussing the qualitative results across all scenes, highlighting the superior performance of X-Field.

Discussion on NVS Quantitative Results. We compare X-Field with two traditional methods (FDK, SART), three NeRF-based methods (TensoRF, NeAT, NAF), and three SOTA methods (SAX-NeRF, X-Gaussian, and R^2-Gaussian). Table 1 reports the quantitative results of highly sparse views (10 views and 5 views) X-ray NVS. Note that we report quantitative results as the mean results across all scenes under the same setting, and scene-wise results are presented in the supplementary material. X-Field demonstrates superior performance in reconstructing X-ray novel views across most scenarios, surpassing the state-of-the-art R^2-Gaussian in all metrics. Specifically, in the human organ reconstruction setting, X-Field consistently outperforms the state-of-the-art R^2-Gaussian, achieving higher SSIM scores and competitive PSNR and LPIPS values, highlighting its effectiveness for complex reconstruction scenarios.

Discussion on NVS Qualitative Results. Figure 6 presents visual comparisons between X-Field and multiple methods, including SAX-NeRF, X-Gaussian, and R^2-Gaussian. These highly sparse view settings provide limited information, resulting in artifacts of varying severity across all methods. SAX-NeRF reconstructs the overall structure but introduces noticeable blurry artifacts, particularly in the bonsai scene. X-Gaussian produces line and wave-pattern artifacts, which are prominent in the head scene. While R^2-Gaussian performs better than the other baselines, it exhibits flaws in the fine details of the bone structure. For example, R^2-Gaussian introduces black linear artifacts in the bone region in the foot scene, where ours produces smoother textures. In summary, our method is able to effectively mitigate blurry, line, and wave-pattern artifacts while maintaining smoothness in distinct areas and on object surfaces. More visual comparisons are provided in supplementary video.

Discussion on CT Reconstruction Results. Following the tradition [14, 15, 16], we also compare X-Field with traditional learning-free algorithms, FDK [3] and ASD_POCS [57], as well as learning-based methods including X-Gaussian [15] and R^2-Gaussian [16] for sparse-view CT reconstruction. Specifically, we assess the performance of 3DGS-based methods and X-Field by generating novel view images from sparse input projections (5 and 10 views) and reconstructing CT scans using ASD_POCS with a total of 100 views. The quantitative results, presented in Table 2, demonstrate that X-Field consistently achieves the best performance across all scenarios. In particular, when generating novel X-ray projections from 10 input views, X-Field combined with ASD_POCS achieves a PSNR of 33.26 and SSIM of 0.911, a comparable result with the SOTA method, R^2-Gaussian.

Figure 7 shows qualitative results of sparse-view CT reconstruction of foot, head, and teapot scans using 10 input views. Without novel view projections, FDK introduces streak artifacts, while ASD_POCS results in blurred structural details. The NVS of X-Gaussian produces clearer reconstructions but introduces shadow artifacts. In contrast, our method and R^2-Gaussian achieve superior visual quality, with X-Field delivering better results by recovering smoother textures and reducing needle-like artifacts, especially in delicate structures such as the cranial region.

4.3 Ablation Study

To comprehensively assess the performance of X-Field, we evaluate the impact of the proposed components, compare different initialization strategies, and evaluate X-Field under various input view settings from 5 to 50 views.

#Views	5		10	
Method	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FDK	15.20	0.144	19.95	0.257
ASD_POCS	24.89	0.731	27.50	0.787
X-Gaussian	17.57	0.688	18.44	0.537
R^2-Gaussian	26.83	0.804	29.28	0.946
Ours	28.04	0.815	33.26	0.911

Table 2: **CT Reconstruction Comparison**. Ours achieve comparable results under 5 and 10 views.

Method	PSNR ↑	SSIM ↑	LPIPS* ↓
w/o Material Opt.	34.78	0.941	73.45
w/o Overlap Filter	34.59	0.937	74.32
w/o Intersection	33.84	0.929	76.60
w/o Segment Length	27.48	0.875	87.83
Ours	35.03	0.953	72.12

Table 3: **Ablation on Components**. Best performance when all components included.

Initialization	PSNR ↑	SSIM ↑	LPIPS* ↓
Random	37.85	0.966	60.02
FDK [3]	37.96	0.967	59.81
Ours	38.67	0.969	59.95

Table 4: **Initialization Methods.** Our initialization improves reconstruction accuracy.

#Views	PSNR ↑	SSIM ↑	LPIPS* ↓
5	31.61	0.933	92.23
10	35.11	0.959	83.72
25	41.71	0.979	75.26
50	42.61	0.993	61.19

Table 5: **Number of Input Views.** Performance improves as input views increase.

Component Analysis. Table 3 evaluates the effects of individual components on reconstruction performance. We observe that Material-Based Optimization and Overlap Filter have a limited impact on reconstruction quality. Removing the Intersection Module (Sec. 3.1) leads to substantial performance drops, with a PSNR drop of 1.19 dB, and an SSIM decrease of 0.024, underscoring its importance in preserving structural integrity.

Segment Length is a fundamental component of our ellipsoid representation, defined in Sec. 3.1 for capturing the distance each ray traverses within the ellipsoid. Removing it also severely impacts the model's ability to render novel views, leading to significant performance degradation, with a 7.6 dB drop in PSNR and an 11.71 increase in LPIPS. These results underscore the importance of Segment Length in enabling X-Field to achieve high-quality reconstructions.

Initialization Analysis We compare our hybrid initialization with random initialization used in X-Gaussian [15] and FDK [3] introduced in R^2-Guassian [16]. As shown in Figure 4, ours provides a point cloud with clearer boundaries and reduced noise, facilitating initialization. Table 4 shows that both FDK and our hybrid initialization outperform random initialization. While FDK achieves a slightly lower LPIPS, it also results in lower PSNR and SSIM. In contrast, ours improves both PSNR and SSIM, demonstrating its effectiveness in enhancing reconstruction quality.

Input View Number Analysis. To further demonstrate the scalability of X-Field, we conduct experiments to assess the effect of the input view number on reconstruction performance. As shown in Table 5, as the number of input views increases, the performance is consistently enhanced. When using 50 views as input, X-Field achieves performance comparable to work designed for sparse view X-Ray reconstruction [15, 14]. We further show in Figure 8 that, with the increase in the input number, the synthesis X-ray images exhibit smoother and clearer bone textures.

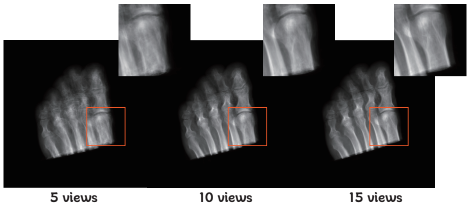

Figure 8: **Input View Numbers.** Five views enable good NVS, and additional input views enhance texture smoothness.

5 Conclusion

We present X-Field, a physically informed representation derived from the penetration and attenuation properties of X-rays for X-ray imaging. To model diverse materials within the internal structures of objects, we introduce attenuation-adaptive ellipsoids with distinct attenuation coefficients. To compute pixel intensity, we design a segment length algorithm that incorporates ellipsoid intersections, enabling accurate estimation of each material's X-ray energy absorption. To improve reconstruction performance, we optimize the ellipsoids along material boundaries and refine the geometric properties during initialization. Our proposed X-Field significantly outperforms state-of-the-art methods in both X-ray NVS and CT reconstruction, demonstrating strong potential for medical applications. Furthermore, we provide insights into the fundamental design of X-ray imaging, which can be extended to other tasks such as the reconstruction of translucent objects.

Acknowledgments: This research is supported by NSF IIS-2525840, CNS-2432534, ECCS-2514574, NIH 1RF1MH133764-01 and Cisco Research unrestricted gift. This article solely reflects opinions and conclusions of authors and not funding agencies.

References

[1] 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. In *ECCV*, 2020.

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023.

[3] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *JOSA A*, 1(6):612–619, 1984.

[4] Anders H Andersen and Avinash C Kak. Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

[5] Godfrey N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British Journal of Radiology*, 46(552):1016–1022, 1973.

[6] Allan M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34(9):2722–2727, 1963.

[7] David J Brenner and Eric J Hall. Computed tomography—an increasing source of radiation exposure. *The New England Journal of Medicine*, 357(22):2277–2284, 2007.

[8] Rebecca Smith-Bindman, Johanna Lipson, Randal Marcus, Kwan Kim, Mahadevappa Mahesh, Robert Gould, Amy Berrington de González, and Diana L Miglioretti. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. *Archives of Internal Medicine*, 169(22):2078–2086, 2009.

[9] Hedvig Hricak, David J Brenner, Steven J Adelstein, Donald P Frush, Eric J Hall, Roger W Howell, Cynthia H McCollough, Fred A Mettler, Mark S Pearce, Oussama H Suleiman, and Richard K Zeman. Managing radiation use in medical imaging: A multifaceted challenge. *Radiology*, 258(3):889–905, 2011.

[10] Hye Jeon Hwang, Joon Beom Seo, Jin Seong Lee, Jae-Woo Song, Song Soo Kim, Hyun Joo Lee, and Chae Hun Lim. Radiation dose reduction of chest ct with iterative reconstruction in image space-part i: studies on image quality using dual source ct. *Korean Journal of Radiology*, 13(6):711–719, 2012.

[11] Amy K Hara, Robert G Paden, Alvin C Silva, Jennifer L Kujak, Holly J Lawder, and William Pavlicek. Iterative reconstruction technique for reducing body radiation dose at ct: feasibility study. *American Journal of Roentgenology*, 193(3):764–771, 2009.

[12] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *ECCV*, 2022.

[13] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: Neural attenuation fields for sparse-view cbct reconstruction. In *MICCAI*, 2022.

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

[15] 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.

[16] 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.

[17] D. Rückert, Y. Wang, R. Li, R. Idoughi, and W. Heidrich. Neat: Neural adaptive tomography. *ACM Transactions on Graphics (TOG)*, 41(4):8–11, 2022.

[18] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022.

[19] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[20] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[21] Allan M. Cormack. Representation of a function by its line integrals, with some radiological applications. ii. *Journal of Applied Physics*, 35(10):2908–2913, 1964.

[22] Idris A. Elbakri and Jeffrey A. Fessler. Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation. *Physics in Medicine & Biology*, 48(15):2453, 2003.

[23] Godfrey N. Hounsfield. Computed medical imaging. *Science*, 210(4479):22–28, 1980.

[24] Ken Sauer and Charles Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Image Processing*, 3(6):600–604, 1993.

[25] Timo Kiljunen, Touko Kaasalainen, Anni Suomalainen, and Mika Kortesniemi. Dental cone beam ct: A review. *Physica Medica*, 31(8):844–860, 2015.

[26] Vladan Lucić, Friedrich Förster, and Wolfgang Baumeister. Structural studies by electron tomography: from cells to molecules. *Annual Review of Biochemistry*, 74:833–865, 2005.

[27] Philip CJ Donoghue, Stefan Bengtson, Xi-ping Dong, Neil J Gostling, Therese Huldtgren, John A Cunningham, Chongyu Yin, Zhao Yue, Fan Peng, and Marco Stampanoni. Synchrotron x-ray tomographic microscopy of fossil embryos. *Nature*, 442(7103):680–683, 2006.

[28] W. Sun, S. B. Brown, and R. K. Leach. An overview of industrial x-ray computed tomography. 2012.

[29] Leonardo De Chiffre, Simone Carmignato, Jean-Pierre Kruth, Robert Schmitt, and Albert Weckenmann. Industrial applications of computed tomography. *CIRP Annals*, 63(2):655–677, 2014.

[30] Claudia Christina Von Bastian, Adrian Schwaninger, and Stefan Michel. Do multi-view x-ray systems improve x-ray image interpretation in airport security screening? 2010.

[31] 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. *IEEE Transactions on Medical Imaging (TMI)*, 25(7):869–875, 2006.

[32] Jinxiao Pan, Tie Zhou, Yan Han, and Ming Jiang. Variable weighted ordered subset image reconstruction algorithm. *International Journal of Biomedical Imaging*, 2006:1–10, 2006.

[33] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777–4807, 2008.

[34] Ake Bjorck and Tommy Elfving. A conjugate gradient method for solving linear least squares problems. *BIT Numerical Mathematics*, 19(2):202–220, 1979.

[35] Yingtai Li, Xueming Fu, Shang Zhao, Ruiyang Jin, and S Kevin Zhou. Sparse-view ct reconstruction with 3d gaussian volumetric representation. *arXiv preprint arXiv:2312.15676*, 2023.

[36] Jacob Beutel, Harold L. Kundel, and Richard L. Van Metter. *Handbook of Medical Imaging, Volume 1: Physics and Psychophysics*. SPIE Press, Bellingham, WA, 2000.

[37] Donald F Swinehart. The beer-lambert law. *Journal of chemical education*, 39(7):333, 1962.

[38] Richard Courant, Fritz John, Albert A Blank, and Alan Solomon. *Introduction to calculus and analysis*, volume 1. Springer, 1965.

[39] David L Toth. Aabb and octree representations for accelerating ray-object intersection tests. *IEEE Computer Graphics and Applications*, 5(10):73–82, 1985.

[40] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pages 1–11, 2024.

[41] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[42] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy, 2024.

[43] Junyi Zhang, Charles Herrmann, Junhwa Hur, Varun Jampani, Trevor Darrell, Forrester Cole, Deqing Sun, and Ming-Hsuan Yang. Monst3r: A simple approach for estimating geometry in the presence of motion. *arXiv preprint arxiv:2410.03825*, 2024.

[44] Chen Yang, Sikuang Li, Jiemin Fang, Ruofan Liang, Lingxi Xie, Xiaopeng Zhang, Wei Shen, and Qi Tian. Gaussianobject: High-quality 3d object reconstruction from four views with gaussian splatting. *ACM Transactions on Graphics*, 2024.

[45] Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[46] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[47] Stefan Gottschalk, Ming C Lin, and Dinesh Manocha. Obbtree: A hierarchical structure for rapid interference detection. In *SIGGRAPH*, pages 171–180. ACM, 1996.

[48] Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

[49] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, 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.

[50] Holger Roth, Amal Farag, Evrim B. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. Data from pancreas-ct, 2016.

[51] Philips. Philips research, hamburg, germany. `https://teem.sourceforge.net/nrrd/volvis/index.html`, 2022. Accessed: 2022-05.

[52] Pavol Klacansky. Open scivis datasets, December 2017.

[53] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. Tigre: a matlab-gpu toolbox for cbct image reconstruction. *Biomedical Physics & Engineering Express*, 2(5):055010, 2016.

[54] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995.

[55] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[56] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

[57] Emil Y Sidky and Xiaochuan Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology*, 53(17):4777, 2008.

[58] Reinhold Baer. *Linear algebra and projective geometry*. Courier Corporation, 2005.

[59] H. Roth, A. Farag, E. B. Turkbey, L. Liu, J. Liu, and R. M. Summers. Data from pancreas-ct (version 2) [data set], 2016.

[60] Pavol Klacansky. Open scivis datasets, December 2017.

[61] Samuli Siltanen et al. FIPS: Open X-ray Tomographic Datasets. Zenodo, 2022.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly present paper's contribution on designing a new 3D representation and corresponding techniques for X-ray Novel View Synthesis and CT Reconstruction. These claims align closely with the Method (Section 3) and Experiment section (Section 4).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Our paper acknowledges that X-Field does not address all challenges in highly sparse-view X-ray reconstruction. Specifically, it does not incorporate structural priors or spatial information from pre-trained large models, nor does it leverage traditional medical structure priors that could further enhance internal structure supervision. These limitations are discussed at appendices.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Our paper provides a complete and rigorous mathematical formulation of our theoretical results, which are derived based on the physical properties of X-ray imaging and the Beer-Lambert law. The main mathematical formulations and their derivations are presented in the Method section, with additional detailed proofs and explanations provided in the Appendix.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We plan to release the code and associated datasets to ensure the reproducibility of our experimental results, including the complete experimental setup and executable code.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: We will organize and release the code after the paper is accepted, but it will not be publicly available at the submission stage. Our experiments are conducted using datasets consistent with those used in prior works, including X-Gaussian and R2-Gaussian. Detailed instructions for data preparation and experimental setup are provided in the supplemental material.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide detailed experimental settings in the appendix due to page constraints in the main paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

Justification: Following the experimental setting of a series of previous works (SAX-NeRF, X-Gaussian, R2-Gaussian), our experiments do not involve testing stochastic processes or randomness.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We report the usage of compute resources details in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our code follows the requirements of NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: We discuss the positive impact of our work in reducing the number of X-ray scans required to achieve high-quality novel X-ray images and further CT reconstruction, thereby minimizing patient radiation exposure. However, we acknowledge a potential negative impact: our method could be misused for unauthorized X-ray image generation, potentially leading to privacy violations or unauthorized medical data reconstruction. To mitigate this, we recommend implementing access controls, watermarking generated images, and ensuring that the application of our method is restricted to authorized and ethical medical use cases.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: Our dataset is sourced from publicly available datasets and follows the same preprocessing procedures as described in previous works.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We properly cite code, data, models used in the paper.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: We do not submit any assets at the submission stage. As previously mentioned, we will organize and release our code with all necessary details, licencse later.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification:Our paper utilizes publicly available datasets and does not involve any experiments or research involving human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in our research does not involve LLMs as important, original components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.

Technical Appendices and Supplementary Material

Overview

A Detailed Derivations for Segment Lengths

In this work, we define the ray space using normalized device coordinates (NDC) [16], providing a convenient framework for deriving explicit segment length formulas. In this space, the segment length l_i for each ellipsoid $\mathbf{E}_i$ is calculated along the view direction $\mathbf{d}$ of a given ray path. Leveraging the orthographic property of NDC space, the 3D ellipsoid $\mathbf{E}_i$ can be projected onto a 2D plane, forming an ellipse. The segment lengths are then constrained within this 2D ellipse, with values outside the ellipse set to zero.

Our goal is to find the relationship between the segment length and the point $\mathbf{u} = (x, y, 0)$ on the ellipse traversed by the ray. Below, we first derive the special value l_{max}, which is the maximum distance a ray travels through the ellipsoid when originating from the center of the ellipse [58]. Then, we use l_{max} as a bridge to express l_i for rays originating from other points on the ellipse.

Consider a ray originating from the center of the ellipse $\mathbf{u_c}$ and traveling along direction $\mathbf{d}$. The modeling of the ray is: $\mathbf{R}(s) = \mathbf{u_c} + s\mathbf{d} \in \mathbb{R}^3$, where s is a scalar parameter. Meanwhile, the ellipsoid centered at $\mathbf{p_c}$ can be described as

$$(\mathbf{p} - \mathbf{p_c})^\top \mathbf{\Sigma}_{\text{3D}}^{-1} (\mathbf{p} - \mathbf{p_c}) = 1. \tag{8}$$

Substituting $\mathbf{p} = \mathbf{R}(s)$ into above equation, we have s as in

$$(\mathbf{u_c} + s\mathbf{d} - \mathbf{p_c})^\top \mathbf{\Sigma}_{\text{3D}}^{-1} (\mathbf{u_c} + s\mathbf{d} - \mathbf{p_c}) = 1. \tag{9}$$

In the NDC space, translating an object does not affect its geometric length. Therefore, without loss of generality, we can assume that $\mathbf{p_c}$ lies on $\mathbf{u_c}$, and then we have:

$$(\mathbf{u_c} + s\mathbf{d} - \mathbf{p_c})^\top \mathbf{\Sigma}_{\text{3D}}^{-1} (\mathbf{u_c} + s\mathbf{d} - \mathbf{p_c}) = s^2 \mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d} = 1. \tag{10}$$

Solving this simplified equation, we can obtain the roots:

$$s = \pm \frac{1}{\sqrt{A}}, \quad \text{where} \quad A = \mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}. \tag{11}$$

The segment length corresponding to $\mathbf{u_c}$ is the difference between the two roots: $l_{\text{max}} = |s_2 - s_1| = \frac{2}{\sqrt{A}}$.

Next, we consider a general point $\mathbf{u}$ on the ellipse plane. The ray originating from $\mathbf{u}$ in the direction $\mathbf{d}$ can be represented by $\mathbf{R}(s) = \mathbf{u} + s\mathbf{d}$. Similarly, we define $\mathbf{a} = \mathbf{u} - \mathbf{p_c}$ and substitute $\mathbf{R}(s)$ into the ellipsoid equation:

$$(\mathbf{a} + s\mathbf{d})^\top \mathbf{\Sigma}_{\text{3D}}^{-1}(\mathbf{a} + s\mathbf{d}) = 1. \tag{12}$$

Expanding the equation results in:

$$s^2\mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1}\mathbf{d} + 2s\mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1}\mathbf{d} + \mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1}\mathbf{a} = 1. \tag{13}$$

For simplicity, we denote $B{=}\mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1}\mathbf{d}$ and $C{=}\mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1}\mathbf{a}$. Thus, the equation becomes $As^2 + 2Bs + C - 1 = 0$, whose roots are:

$$s_{1,2} = \frac{-B \pm \sqrt{B^2 - A(C-1)}}{A}. \tag{14}$$

The segment length can then be derived as:

$$l_i = |s_2 - s_1| = l_{\text{max}} \times \sqrt{1 - \left(\frac{C - B^2}{A}\right)}. \tag{15}$$

B 3DGS Preliminaries

3D Gaussian Splatting (3DGS) [2] represents a scene as a collection of anisotropic Gaussian kernels in 3D space, denoted as $\mathbb{G}^3 = \{G_i^3\}_{i=1}^M$. Each Gaussian is parameterized by its center position, covariance matrix, color, and opacity. By splatting these kernels onto the image plane using a differentiable rasterizer, 3DGS enables efficient view synthesis and supports end-to-end optimization.

Formally, an RGB image $\mathbf{I}_{rgb} \in \mathbb{R}^{H \times W \times 3}$ is rendered from the 3D Gaussians as

$$\mathbf{I}_{rgb} = \mathcal{R}(\mathbb{G}^3) = \mathcal{C}\big(\mathcal{P}(\mathcal{T}(\mathbb{G}^3))\big), \tag{16}$$

where $\mathcal{T}$, $\mathcal{P}$, and $\mathcal{C}$ correspond to the *transformation*, *projection*, and *composition* modules.

Transformation. The set of 3D Gaussians is first transformed into the viewing ray coordinate system: $\mathbb{G}_t^3 = \mathcal{T}(\mathbb{G}^3)$. This step aligns the kernels with the camera rays to reduce computation overhead.

Projection. The transformed Gaussians are then mapped onto the image plane: $\mathbb{G}^2 = \mathcal{P}(\mathbb{G}_t^3)$. The resulting 2D Gaussians retain the opacity and color of their 3D counterparts, while the depth dimension in position and covariance is discarded.

Composition. Finally, an image is synthesized by alpha-compositing the 2D Gaussians in a back-to-front order: $\mathbf{I}_{rgb} = \mathcal{C}(\mathbb{G}^2)$. The differentiable rasterizer $\mathcal{R}$ enables the optimization of the kernel parameters using gradient-based losses.

C Implementation Details

C.1 Dataset

We conduct experiments on the large-scale X3D [14] dataset with multiple modalities across two categories: Human Organ (Chest, Head, Foot, Jaw, Pancreas, Abdomen, aneurism, Leg) and Synthetic Object (Bonsai, Teapot, Engine, Backpack, box, carp, Pelvis). The chest scans are sourced from LIDC-IDRI [49], the pancreas scans from Pancreas-CT [59], and the remaining cases from SciVis [60]. Following prior works [15, 16, 14], we employ the TIGRE tomography toolbox [53] to capture 512×512 projections across a range of 0° to 180°, including both in-distribution and out-of-distribution views.

View Distribution. For training, we uniformly capture 5, 10, 15, 25, and 50 views within 0° to 180°. For testing, we randomly capture 50 views in the same range.

Out-of-Distribution Views. We also evaluate the robustness of our method on out-of-distribution views. To construct these views, we apply a consistent algorithm across all cases. Specifically, for each case, we calculate the Oriented Bounding Box (OBB) [47] of the trained ellipsoid distribution to determine its center and principal axis. Next, we select a direction $\mathbf{d}$ orthogonal to this principal axis.

Starting from the center, we move along $\mathbf{d}$ by a radius r to establish the initial camera position. The camera then rotates 360° around the principal axis to form the view distribution. Due to differences in direction and radius, the resulting new view distribution is entirely distinct from both the training and testing sets. This setup comprehensively evaluates the model's robustness and 3D consistency under unseen views.

C.2 Hyper-parameters

In our experiments, the position learning rate was set to an initial and final value of 0.0002, while the density and scaling learning rates were both initialized and maintained at 0.01 and 0.005, respectively. The structural dissimilarity loss weight (λ_{DSSIM}) was set to 0.25. To refine the density of kernels, a minimum density threshold of 1×10^{-5} was applied. The total number of kernels was capped at 500,000 to balance computational efficiency with accuracy.

D Additional Quantitative Results

D.1 Ablation on 50 views

We evaluate our method under the 50-view setting and compare it with state-of-the-art baselines. As shown in Table 6, our approach achieves the best performance in PSNR and LPIPS while maintaining SSIM comparable to X-Gaussian. These results, together with the 5- and 10-view experiments, demonstrate that our method consistently outperforms existing approaches across different input sparsity levels, from highly sparse to moderately dense scenarios.

Method	SSIM ↑	PSNR ↑	LPIPS* ↓
SAX-NeRF	0.975	28.53	124.10
X-Gaussian	0.997	32.78	70.52
R^2-Gaussian	0.982	42.61	64.38
Ours	0.993	42.65	61.19

Table 6: **Novl View Synthesis with 50 input views.**

D.2 Ablation on Computational Efficiency

We compare the computational efficiency of different approaches as shown in Table 7. Our method achieves training and memory consumption on the same order as 3DGS-based methods, while being significantly more efficient than NeRF-based baselines. In terms of inference, X-Field supports real-time rendering, offering a practical balance between quality and speed for clinical deployment.

	NAF	SAX-NeRF	X-Gaussian	R^2-Gaussian	Ours
Training Time	38m 57s	307m 50s	1m 55s	3m 52s	4m 15s
Inference Speed	2.07 fps	0.36 fps	132 fps	72 fps	45 fps
Memory Usage	24 GB	27 GB	824 MB	968 MB	1456 MB

Table 7: **Computational profile comparison.** Training and inference statistics on RTX 6000 Ada GPU.

D.3 Statistical Significance Tests for Hybrid Initialization

As shown in Table 8, all tests indicate statistically significant improvements of our hybrid initialization over both Random and FDK-based baselines. While the average gains are relatively modest, their consistency across all evaluated scenes demonstrates the reliability of the initialization strategy. This suggests that even small improvements in early-stage initialization can provide a stable advantage for subsequent optimization.

Metric	**Ours vs Random**		**Ours vs FDK**	
	Δ	p-val	Δ	p-val
PSNR ↑	+0.84	0.0002	+0.42	0.0016
SSIM ↑	+0.0071	0.0005	+0.0035	0.0012
LPIPS ↓	–0.53	0.0007	–0.29	0.0013

Table 8: **Statistical significance of hybrid initialization.**

D.4 Component Efficiency Analysis

Table 9 reports runtime and memory for different ablation settings. Removing segment length yields the largest savings but leads to severe quality degradation, as accurate path-dependent attenuation cannot be computed. Excluding material optimization provides moderate efficiency gains but compromises material-aware reconstruction. Overlap filter removal has little effect on training time yet increases memory usage, mainly affecting efficiency. Removing intersection computation offers modest speedup with reduced memory, indicating moderate overhead. Overall, the full model introduces limited additional cost while preserving reconstruction fidelity.

Method	R^2-Gaussian	w/o Material Opt.	w/o Overlap Filter	w/o Intersection	w/o Segment Length	Full
Training Time	3m52s	3m58s	4m12s	4m04s	3m46s	4m15s
Memory Usage	968 MB	1174 MB	1359 MB	1368 MB	945 MB	1456 MB

Table 9: **Training time and memory usage under different ablations.**

D.5 Real-World Data Validation

We evaluate our method on the FIPS dataset [61], which provides real 2D X-ray projections. As shown in Table 10, X-Field consistently outperforms baseline approaches under both 5-view and 10-view settings. Moreover, the performance trends observed on synthetic benchmarks generalize well to real projection data containing scattering, beam hardening, and detector noise, demonstrating that our approach remains robust and effective under practical real-world conditions.

Method	**5-view**			**10-view**		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
SAX-NeRF	22.28	0.963	162.23	25.86	0.978	141.50
X-Gaussian	25.68	0.967	121.62	29.24	0.980	76.94
R^2-Gaussian	36.42	0.970	71.25	38.87	0.984	34.77
Ours	37.56	0.981	70.41	40.26	0.989	34.22

Table 10: **Performance on the FIPS real projection dataset.**

D.6 Quantitative Comparison across Scenes

Tables 11 and 12 present the scene-wise quantitative results for 10-view and 5-view training data, respectively. In the 10-view setting (Table 11), our method achieves the best performance in most scenes across all metrics, including PSNR, SSIM, and LPIPS. Specifically, for challenging scenes such as the pancreas and bonsai, our method consistently outperforms all others, demonstrating its ability to handle complex structures effectively. The R^2-Gaussian method performs well in certain scenes, such as the foot and engine, achieving competitive SSIM and LPIPS scores. However, our method maintains an edge in overall quality, solidifying its superiority in reconstructing finer details.

	Foot			Head			Chest			Jaw		
Method	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓
SAX-NeRF [14]	17.97	0.943	213.3	17.25	0.958	154.6	15.63	0.937	187.2	18.07	0.949	267.9
X-Gaussian [15]	17.86	0.942	120.6	17.20	0.958	136.4	15.71	0.939	132.8	17.93	0.947	227.2
R^2-Gaussian [16]	34.46	0.951	80.73	34.00	0.972	86.25	35.46	0.949	100.4	30.15	0.909	175.3
Ours	35.54	0.959	75.56	35.71	0.975	76.84	34.03	0.944	91.15	29.32	0.910	164.3

	Pancreas			Bonsai			Teapot			Engine		
Method	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓
SAX-NeRF [14]	18.78	0.929	168.5	21.28	0.971	174.5	23.67	0.985	122.8	21.89	0.977	127.2
X-Gaussian [15]	18.77	0.931	121.7	21.66	0.977	91.83	24.24	0.986	66.71	22.14	0.977	72.29
R^2-Gaussian [16]	36.39	0.968	79.36	42.12	0.983	46.85	40.37	0.980	34.03	39.82	0.990	23.44
Ours	36.77	0.969	81.25	43.06	0.987	45.52	38.03	0.988	36.97	41.79	0.993	20.18

Table 11: **Results of Quantitative Comparison on 10 views (§ D.6).** We compare our X-Field on 10 views with Deep Learning-based methods: SAX-NeRF [14], X-Gaussian [15], and R^2-Gaussian [16]. LPIPS* = LPIPS $\times 10^3$. best and second best methods are marked.

In the 5-view scenario (Table 12), the performance gap between methods decreases due to the limited training data. Nonetheless, our method achieves the best results in most scenes, particularly excelling in the chest, bonsai, and teapot, where it surpasses other methods across all metrics. While the R^2-Gaussian method demonstrates strong performance in specific scenes, our approach consistently proves to be more robust. These results underscore our method's adaptability to reduced training data while maintaining state-of-the-art performance.

Method	Foot PSNR↑	SSIM↑	LPIPS*↓	Head PSNR↑	SSIM↑	LPIPS*↓	Chest PSNR↑	SSIM↑	LPIPS*↓	Jaw PSNR↑	SSIM↑	LPIPS*↓
SAX-NeRF [14]	18.21	0.922	234.7	15.23	0.927	192.2	13.67	0.813	237.3	16.16	0.917	322.4
X-Gaussian [15]	18.48	0.925	147.8	15.19	0.929	177.2	13.61	0.824	201.1	16.13	0.918	306.7
R^2-Gaussian [16]	30.32	0.946	117.2	31.25	0.958	108.6	27.65	0.862	169.1	26.43	0.850	232.9
Ours	30.88	0.951	116.2	31.85	0.960	97.53	29.73	0.899	146.4	25.16	0.833	227.8

Method	Pancreas PSNR↑	SSIM↑	LPIPS*↓	Bonsai PSNR↑	SSIM↑	LPIPS*↓	Teapot PSNR↑	SSIM↑	LPIPS*↓	Engine PSNR↑	SSIM↑	LPIPS*↓
SAX-NeRF [14]	18.57	0.924	188.3	19.25	0.953	209.6	21.46	0.975	160.9	21.87	0.960	116.2
X-Gaussian [15]	18.60	0.927	145.5	19.43	0.958	148.2	22.03	0.978	121.8	22.25	0.965	94.87
R^2-Gaussian [16]	30.84	0.937	103.4	34.85	0.966	83.27	27.69	0.945	77.56	31.73	0.968	52.92
Ours	30.99	0.944	100.2	34.73	0.965	82.64	28.30	0.950	76.45	32.13	0.970	52.13

Table 12: **Results of Quantitative Comparison on 5 views (§ D.6).** We compare our X-Field on 5 views with Deep Learning-based methods: SAX-NeRF [14], X-Gaussian [15], and R^2-Gaussian [16]. LPIPS* = LPIPS $\times 10^3$. best and second best methods are marked.

E Additional Qualitative Results

Out-of-Distribution Results To further evaluate X-ray reconstruction capabilities, we rendered 360-degree images around the scene. We compared X-Gaussian, R^2-Gaussian , and our method with viewpoints being out-of-distribution in 5 different scenes, including foot (Figure 9), head (Figure 10), jaw (Figure 11), and teapot (Figure **??**) .

Each figure presents five out-of-distribution views, with rows representing methods and columns corresponding to the same view. In the foot scene (Figure 9), X-Gaussian exhibits black linear artifacts in views 3 and 4. While R^2-Gaussian reconstructs the general foot structure, our method delivers smoother and more visually consistent results. In the head scene (Figure 10), X-Gaussian displays wave-like artifacts at the top of the head, whereas both R^2-Gaussian and our method achieve improved overall reconstructions. Notably, our method captures finer details, showcasing superior reconstruction quality and generalization. Similarly, in the jaw (Figure 11))) scenes, our method produces clearer and more complete reconstructions compared to the other approaches.

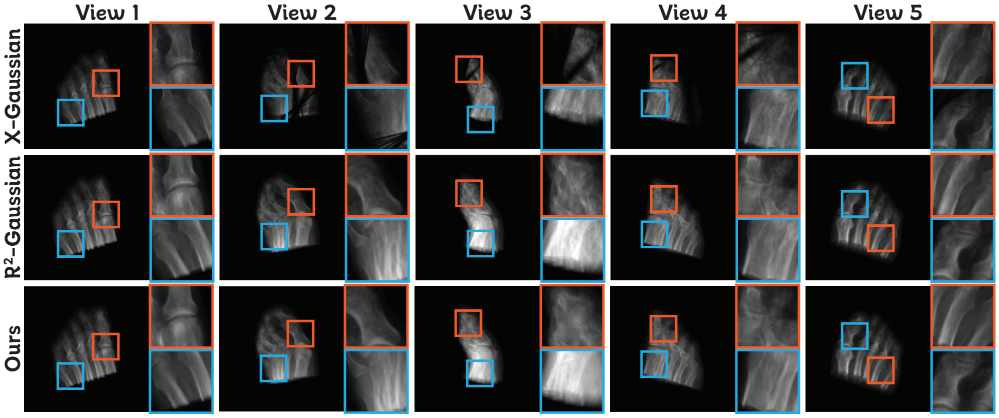

Figure 9: **Qualitative Comparison of Foot on Out-of-Distribution Views.**

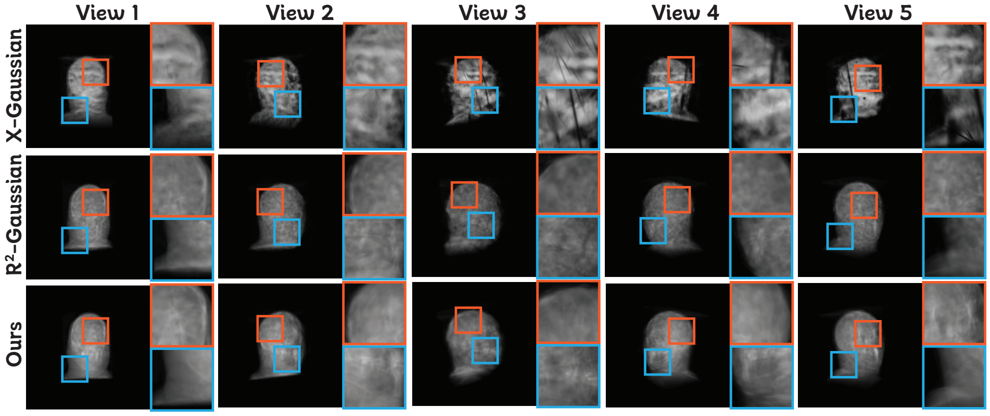

Figure 10: **Qualitative Comparison of Head on Out-of-Distribution Views.**

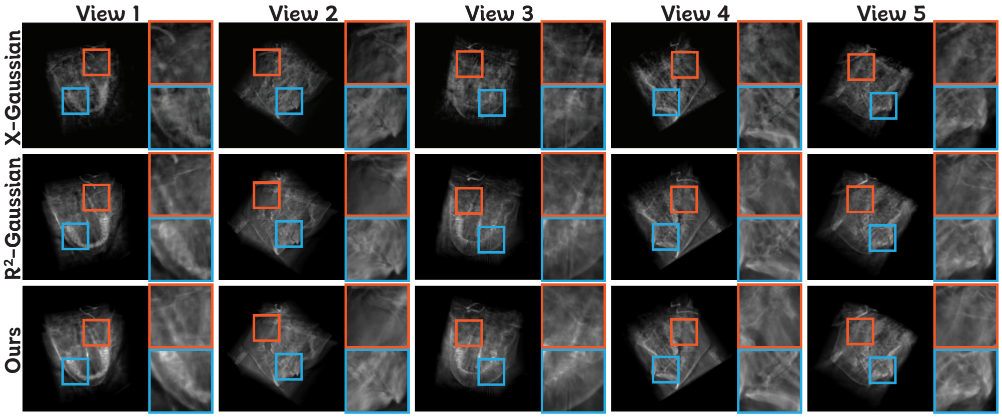

Figure 11: **Qualitative Comparison of Jaw on Out-of-Distribution Views.**

F More Discussions

F.1 Limitations

Despite the significant advancement achieved by X-Field in X-ray sparse-view reconstruction, it is not without limitations:

- **The ellipsoid representation may not be the most efficient.** We use ellipsoids to capture the internal material distribution of objects; however, there could exist a more efficient form that can reduce the number of parameters while also requiring fewer elements to effectively learn the material distribution.
- **Lack of leveraging large model prior knowledge.** Many recent methods for sparse view reconstruction have utilized prior knowledge from large models to guide their learning processes. In our task, we didn't find a suitable large model for X-ray data to introduce into our framework. It is possible that some large-scale medical models could provide prior knowledge to enhance convergence speed and improve overall performance.

F.2 Ethics Considerations

Ethical considerations are critical in developing X-ray reconstruction technologies. Ensuring informed consent for transparent communication about and limitations of X-ray reconstruction methods is essential to respect privacy and prevent misrepresentation. Secure handling and storage of medical imaging data are paramount to safeguard against unauthorized access and misuse. Furthermore, recognizing the potential risks associated with the misuse of advanced imaging technologies, we emphasize the importance of establishing robust ethical guidelines to ensure their responsible application. Our commitment is to uphold the highest ethical standards in all aspects of X-ray reconstruction, protecting the integrity and confidentiality of patient data.