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

Anonymous CVPR submission

Paper ID 4997

Abstract

X-ray is essential in medical diagnostics for visualizing internal body structures, yet its use is strictly regulated due to the health risks posed by radiation exposure. To mitigate these risks, recent research has explored generating novel views from sparse input to minimize radiation doses, employing methods like NeRF and 3D Gaussian Splatting. However, these approaches primarily adhere to principles for visible light imaging, failing to account for the distinct characteristics of X-ray imaging. In this paper, we propose a novel 3D representation, named ***X-Field****, specifically designed to align with the intrinsic characteristics of X-ray imaging. In lieu of the continuous, view-dependent representations used in visible light, X-Field models X-ray representation as discrete and view-independent, rooted in the physical property of energy absorption rate. To capture such property, we employ ellipsoids with uniform energy absorption rates, effectively representing complex material distributions in internal structures. Our method further empowers a hybrid progressive initialization strategy, leveraging structural priors from CT imaging, and optimizes via a material-based approach that dynamically adjusts to local variations in material composition. Experimental results demonstrate that X-Field achieves state-of-the-art visual fidelity in reconstructing human organs and other objects, highlighting its potential to transform medical imaging by enhancing safety and diagnostic precision.*

1. Introduction

X-rays are indispensable in clinical diagnosis. Conventional 3D X-ray reconstruction techniques, like Computed Tomography (CT) [2, 15, 21, 26], typically require hundreds of X-ray images to accurately reconstruct anatomical structures [9, 28, 49]. However, acquiring such an extensive number of images results in prolonged exposure to ionizing radiation, which presents substantial health risks to patients. To mitigate this, X-ray novel view synthesis has been developed [3, 31, 36, 61]. This technique aims to reconstruct the 3D X-ray volume from only a sparse set of 2D X-ray projections, significantly reducing radiation dose while maintaining high reconstruction quality.

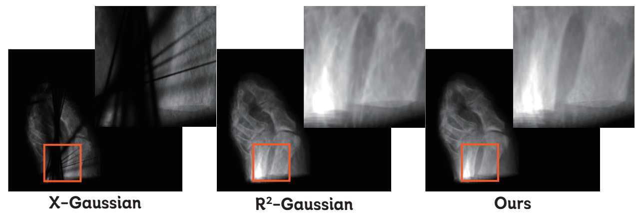

Figure 1. Visual comparison of novel view synthesis on X-ray images. Current methods [11, 62] adapt 3D representations designed for visible light fields, causing needle-like artifacts. Our method, grounded in the physical properties of X-ray field, delivers accurate reconstructions with fine details and clear material structures.

Existing sparse reconstruction methods [11, 12, 61, 62] primarily uses two types of representations: Neural Radiance Fields (NeRF) [39] and 3D Gaussian Splatting (3DGS) [32]. While NeRF-based methods [12, 61] achieve high-quality rendering, their reliance on MLPs to encode the entire X-ray field results in significant inefficiencies. Consequently, recent studies have shifted towards 3DGS [32] for X-ray field representation. With explicit parametrization and a highly parallelized pipeline, these methods provide faster reconstruction and enable real-time rendering. Building on this foundation, advanced 3DGS-based approaches [11, 62] address issues related to grayscale values and integration biases, achieving promising improvements in X-ray reconstruction.

Despite notable advancements, current methods primarily adapt principles from visible light imaging, which is rooted in natural scene reconstruction, to X-ray imaging. However, crucial distinctions between visible light and X-ray imaging are often overlooked. In visible light imaging, rays interact with surfaces, producing complex patterns of reflection and refraction that result in a dense spatial distribution of rays. This distribution can be efficiently modeled as a light field [23, 34, 41], where each point in the field is influenced by omnidirectional light rays, giving it an inherently continuous and view-dependent nature. For example, the color of a pixel at a given point within the field depends on the rays directed toward that viewpoint. As the viewing angle shifts, the collection of rays affecting that point also

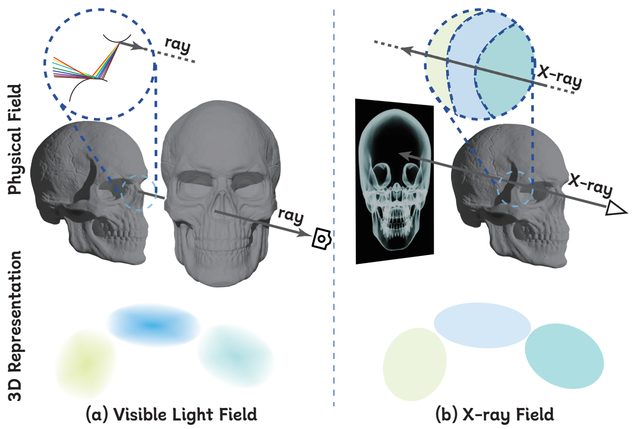

Figure 2. **Comparison of visible light and X-ray fields.** (a) In a visible light field, each pixel's color is formed by the accumulation of omnidirectional light rays due to multiple reflections and refractions, resulting in a continuous field. This makes it well-suited for representation with ellipsoids following a Gaussian distribution. (b) In the X-ray field, passing through various materials, X-rays produce a projection without overlapping paths, leading to a discrete representation. This characteristic can be better captured using ellipsoids with homogeneous properties.

changes, highlighting the view-dependence of visible light fields [1]. Given these characteristics, continuous and view-dependent models are suitable for representing visible light fields. In line with this, 3DGS [32] employs ellipsoids with anisotropic Gaussian distributions to model such fields.

In X-ray imaging [10], the high penetration ability of X-ray beams allows them to traverse deep material layers with minimal scattering, effectively creating approximately straight-line paths. As a result, for a given viewpoint, the intensity of a pixel corresponding to a specific field point is largely determined by a single X-ray ray passing through that point, which implies a fundamentally discrete field in X-ray imaging. Unlike the continuous, view-dependent nature of the visible light field, the X-ray field is inherently discrete and view-independent. Practically, this field captures the energy absorption rate at each point [2, 21, 26], depending solely on the material present at that location, making it independent of the viewing angle. Figure 2 visually compares the visible light and X-ray fields, showcasing their distinct representations.

In this paper, we present a novel 3D representation, **X-Field**, specifically designed around the physical properties unique to X-ray imaging. To capture the distinct characteristics of X-ray fields, we introduce a **Physically Grounded Ellipsoid Representation**, which models material distribution and physical attributes within internal structures. This ellipsoid-based approach enables an efficient representation of spatial structures while accurately modeling material distribution. Each point within our model is characterized by its energy absorption rate, assuming uniformity across each ellipsoid—where all points share the same absorption rate [2, 26]. When a pixel overlaps an ellipsoid, we compute the cumulative energy absorption along the X-ray path through the ellipsoid, yielding a more precise cumulative measurement. In overlapping regions, only one ellipsoid contributes to the field due to the discrete nature of X-rays, implemented via a first-pass precedence strategy. To optimize ellipsoid density, we introduce **Material-Based Optimization**, which increases ellipsoid density in complex regions and along material boundaries. Additionally, we integrate **Hybrid Progressive Initialization**, drawing on structural priors from CT techniques to further enhance the reconstruction performance.

To validate X-Field, we evaluate its performance on both human organ and object datasets, demonstrating substantial improvements over state-of-the-art methods. Even with sparse input views, X-Field consistently achieves robust, high-quality reconstructions using minimal data. Notably, in the human organ dataset with only 10 input views, X-Field attains a 0.45 dB gain in PSNR and a 2.44 reduction in LPIPS compared to the strong baseline R^2-Gaussian [62].

2. Related Work

2.1. 3D Representation in Medical Imaging

Traditional 3D medical imaging relies heavily on voxel-based representations [4, 14, 40], but these approaches become impractical at high resolutions due to excessive computation and storage costs. Later, NeRF-based works [5, 38, 39, 52, 59] used implicit functions to represent color and density fields, achieving high-quality results in novel view synthesis (NVS) tasks. Many works such as [12, 61] have adapted the NeRF rendering pipeline to better fit X-ray reconstruction. However, NeRF-based methods, which use MLPs to store spatial parameters, suffer from significant redundancy in parameters. Moreover, for those methods, querying MLPs and ray tracing cannot be parallelized, resulting in low training and rendering efficiency.

To address these limitations, 3D Gaussian Splatting (3DGS) [32] and its successors [22, 29, 35, 56, 58, 65] introduced an explicit representation structure, enabling parallel rasterization rendering and substantially improving parameter storage and rendering efficiency. Building on these advancements, several works [11, 62] have modified the 3DGS pipeline to render X-ray projections, advancing the reconstruction of X-ray fields. However, these methods are fundamentally tailored to the continuous and view-dependent nature of visible light, and their minor adaptations fail to align with the discrete and view-independent characteristics of X-rays. This mismatch often results in reconstruction artifacts. In contrast, our approach is explicitly designed for the physical properties of X-ray imaging, ensuring more accurate and artifact-free reconstructions.

CVPR #4997

2.2. X-ray Imaging and Reconstruction

X-rays are widely used across various domains due to their strong penetrative ability, making them essential tools such as in medical diagnosis [15, 16, 19, 26, 27, 45], biology [18, 33, 37], industrial inspection [17, 50], and security screening [53]. Traditional X-ray imaging [30] generates 2D projection images by passing X-rays through an object and capturing the attenuated radiation on a detector. Due to the minimal scattering of X-rays [10], the intensity of a pixel is determined by the path of a single X-ray traversing the object, more precisely by the energy absorption rate along this path [2, 21, 26]. This process can be modeled as a X-ray field that describes the energy absorption rate at each spatial point [2, 21, 26].

To reconstruct X-ray field for obtaining internal structural and material information, there has esisted some traditional methods can be divided into two categories: analytical methods [21, 60] and optimization-based techniques [2, 42, 48]. However, these methods typically require hundreds of X-ray images, leading to increased radiation exposure. Recent advances in deep learning [11, 12, 43, 61, 62] have enabled 3D X-ray field reconstruction using fewer 2D projections. However, existing methods do not fully utilize the structural priors from traditional X-ray reconstruction [8, 21, 25, 30], leading to challenges in achieving high-quality results. In contrast, our approach leverages prior knowledge from conventional CT methods to enhance reconstruction quality.

3. Background

3D Gaussian Splatting. 3DGS [32] is a technique for representing and rendering 3D scenes using Gaussian primitives within a differentiable volume splatting framework. This method explicitly parameterizes the Gaussian primitive with a covariance matrix $\mathbf{\Sigma}_{\text{3D}} \in \mathbb{R}^{3\times 3}$ and its center position $\mathbf{p_c} \in \mathbb{R}^3$. The Gaussian function $\mathcal{G}_{\text{3D}}$ at a given position $\mathbf{p}$ is defined as:

$$\mathcal{G}_{\text{3D}}(\mathbf{p}) = \exp\left(-\frac{1}{2}(\mathbf{p}-\mathbf{p_c})^\top \mathbf{\Sigma}_{\text{3D}}^{-1}(\mathbf{p}-\mathbf{p_c})\right). \tag{1}$$

To render an image, the 3D Gaussians are transformed into camera coordinates using the world-to-camera matrix $\mathbf{W}$ and then projected onto the image plane via a local affine transformation $\mathbf{J}$. This yields the transformed covariance matrix: $\mathbf{\Sigma}'_{\text{3D}} = \mathbf{J}\mathbf{W}\mathbf{\Sigma}_{\text{3D}}\mathbf{W}^\top\mathbf{J}^\top$. Discarding the third row and column of $\mathbf{\Sigma}'_{\text{3D}}$, we obtain the corresponding 2D Gaussian $\mathcal{G}_{\text{2D}}$ with covariance matrix $\mathbf{\Sigma}'_{\text{2D}}$. The rendered pixel color at position $\mathbf{x}$ is then computed by volumetric alpha blending across all Gaussians:

$$\mathbf{c}(\mathbf{x}) = \sum_{k=1}^{K} c_k \alpha_k \mathcal{G}_{\text{2D}}^k(\mathbf{x}) \prod_{j=1}^{k-1}\left(1-\alpha_j \mathcal{G}_{\text{2D}}^j(\mathbf{x})\right), \tag{2}$$

where K is the number of Gaussian primitives, α_k is the opacity value, and c_k is the view-dependent color. These attributes are optimized by a photometric loss.

X-ray Physical Field. X-ray physical field describes the attenuation properties of a given space, where each position $\mathbf{x} \in \mathbb{R}^3$ is characterized by an attenuation coefficient $\sigma(\mathbf{x}) \in \mathbb{R}^+$ [6, 26, 30]. From this perspective, X-ray imaging measures the cumulative attenuation of X-rays as they pass through an object. Then, a projection image $I \in \mathbb{R}^{H\times W}$ can be generated by capturing the remaining intensity along each X-ray path after attenuation. Mathematically, let $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d} \in \mathbb{R}^3$ denote an X-ray path, with initial intensity I_0 and path bounds t_0 and t_n. The resulting raw pixel value $I'(\mathbf{r})$ corresponds to the remaining intensity after attenuation. This intensity can be derived from the Beer-Lambert law [30], expressed as:

$$I'(\mathbf{r}) = I_0 \exp\left(-\int_{t_0}^{t_n} \sigma(\mathbf{r}(t))\, dt\right). \tag{3}$$

In practice, tomography formulates the raw data in logarithmic space for computational simplicity:

$$I(\mathbf{r}) = \log I_0 - \log I'(\mathbf{r}) = \int_{t_0}^{t_n} \sigma(\mathbf{r}(t))\, dt. \tag{4}$$

Here, each pixel value $I(\mathbf{r})$ reflects the accumulated attenuation along the X-ray path. The objective of tomographic reconstruction is to recover the 3D distribution of the attenuation coefficient $\sigma(\mathbf{x})$, producing a discrete volumetric representation based on projections from multiple angles. This approach facilitates detailed and comprehensive visualization of an object's internal structure, which is invaluable for applications in medical diagnosis.

4. The Proposed X-Field

In this section, we introduce Physically Grounded Ellipsoid Representation in Section 4.1. This includes the formulation of Material-Adaptive Ellipsoids, an algorithm for calculating Segment Lengths with Intersections, and Physically Faithful Overlap Filtering to ensure accurate pixel-ellipsoid associations. Additionally, we propose Hybrid Progressive Initialization in Section 4.2 and Material-Based Optimization in Section 4.3.

4.1. Physically Grounded Ellipsoid Representation

Material-Adaptive Ellipsoids. In Section 3, the attribute of each point $\mathbf{x} \in \mathbb{R}^3$ in the X-ray field is defined as the attenuation coefficient $\sigma(\mathbf{x})$. Therefore, instead of RGB in the case of visible light, we aim to reconstruct an attenuation coefficient field featured by σ. We use ellipsoids $\{\mathbf{E}_i \mid i = 0, 1, \ldots, n-1\}$ to model regions in space where the attenuation coefficient is non-negative, where n is the

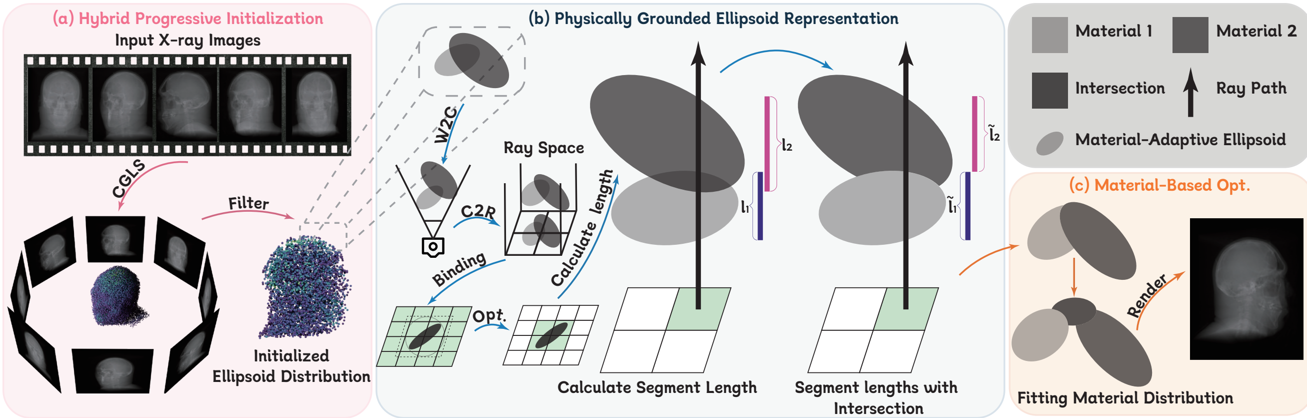

Figure 3. **Overview of the proposed X-Field Pipeline.** **(a) Hybrid Progressive Initialization.** First, we use a small number of X-ray images as input to obtain a rough X-ray field by Conjugate Gradient Least Squares (CGLS) [8, 25, 30]. Based on the CGLS reconstruction, we employ Total Variation (TV) regularization [44] to further reduce noise. Finally, a threshold filtering process is used to obtain the Initialized Ellipsoid Distribution. **(b) Physically Grounded Ellipsoid Representation.** Adhere to the X-ray imaging principle, we first transform the ellipsoids from the world coordinate system to the camera coordinate system, and then to the ray space. In ray space, we establish the association between pixels and ellipsoids, optimized via our designed Physically Faithful Overlap Filtering for precise binding. Then, we derive an explicit form for segment length calculation and further update these lengths to account for the complex intersection among ellipsoids. **(c) Material-Based Optimization.** Beyond geometry, our optimization enables splitting ellipsoids to capture the material distribution essential for high-quality rendering.

total number of ellipsoids. Note that the interior of each $\mathbf{E}_i$ is not modeled as Gaussian distributions. Instead, we assume a homogeneous material structure, where any point $\mathbf{x}$ inside $\mathbf{E}_i$ has the same physical properties. Thus, for any $\mathbf{x} \in \mathbf{E}_i$, we define $\sigma(\mathbf{x}) = \sigma_i$, where σ_i is the uniform attenuation coefficient associated with $\mathbf{E}_i$. Suppose that the X-ray path $\mathbf{r}$ passes through all n ellipsoids, we can derive the accumulated attenuation and corresponding pixel value using Eq. (4) as follows:

$$\begin{aligned} I(\mathbf{r}) &= \int_{t_0}^{t_n} \sigma(\mathbf{r}(t))\, dt \\ &= \int_{t_0}^{t_1} \sigma(\mathbf{r}(t))\, dt + \cdots + \int_{t_{n-1}}^{t_n} \sigma(\mathbf{r}(t))\, dt \\ &= \sigma_0 \int_{t_0}^{t_1} dt + \cdots + \sigma_{n-1} \int_{t_{n-1}}^{t_n} dt \\ &= \sigma_0 l_0 + \sigma_1 l_1 + \cdots + \sigma_{n-1} l_{n-1}, \end{aligned} \tag{5}$$

where $l_i = t_{i+1} - t_i$ denotes the length of the path segment passing through the i-th ellipsoid, and σ_i is its corresponding attenuation coefficient. Therefore, to accurately measure the accumulated attenuation, it is necessary to obtain the segment length l_i for the ray $\mathbf{r}$ and each ellipsoid in $\{\mathbf{E}_i\}$. We then derive an explicit form of l_i, which enables precise and efficient calculation.

Explicit Form of Segment Lengths. We compute the segment length l_i for the i-th ellipsoid $\mathbf{E}_i$ along the view direction $\mathbf{d}$. As shown in Figure 3(b), we first project the ellipsoid $\mathbf{E}_i$ onto the NDC space along $\mathbf{d}$, resulting in a 2D ellipse. Mathematically, the ray path passing through the center of this ellipse reaches the maximal segment length:

$$l_{\text{max}} = \frac{2}{\sqrt{\mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}}}. \tag{6}$$

Here, $\mathbf{\Sigma}_{\text{3D}}$ is the covariance matrix of ellipsoid $\mathbf{E}_i$. Note that $\mathbf{d}$ is the directional vector pointing from the center of the ellipse to the center of the ellipsoid. Leveraging this result, for any other point $\mathbf{u}$ inside the ellipse, we can calculate the corresponding segment length l_i by:

$$l_i = l_{\text{max}} \times \sqrt{1 - \left(\frac{C - B^2}{A}\right)}, \text{ where} \tag{7}$$

$$A = \mathbf{d}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}, \quad B = \mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{d}, \quad C = \mathbf{a}^\top \mathbf{\Sigma}_{\text{3D}}^{-1} \mathbf{a}. \tag{8}$$

Here, $\mathbf{a} = \mathbf{u} - \mathbf{p_c}$ is the displacement from the center of the ellipsoid $\mathbf{p_c}$ to the point $\mathbf{u}$ inside the ellipse. The detailed derivation of l_{max} and l_i is provided in the supplementary materials. Figure 3(b) illustrates the calculated l_i.

Segment Lengths with Intersections. We have derived an explicit form of the segment length for individual ellipsoids. However, when ellipsoids intersect, additional handling is required to account for the overlap between regions. Figure 3(b) illustrates such complex situations. When two ellipsoids intersect, we argue that the intersecting region should not be simultaneously attributed to both ellipsoids, given our assumption of material homogeneity within each ellipsoid. Instead, the overlapping region ought to exclusively belong to one of the intersecting ellipsoids. For optimization simplicity, we adopt a first-pass precedence strategy:

CVPR #4997

Algorithm 1 Compute Segment Lengths with Intersections

Input: $(z_0, z_1, \dots, z_{n-1})$: sorted depths of ellipsoids $\{\mathbf{E}_i\}$
$(l_0, l_1, \dots, l_{n-1})$: segment lengths for individual ellipsoids without considering intersections
Output: Updated segment lengths $\tilde{l}_0, \tilde{l}_1, \dots, \tilde{l}_{n-1}$ and effective regions

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**
      Update the valid region of ellipsoid $\mathbf{E}_i$ as $[z_i + \frac{1}{2}l_i - \tilde{l}_i, z_i + \frac{1}{2}l_i]$
      $z \leftarrow z_i, l \leftarrow l_i$
    end if
  end if
end for

Given the ordered set of ellipsoids $\{\mathbf{E}_i \mid i = 0, 1, \dots, n-1\}$ along the ray path $\mathbf{r}$, where the ellipsoid indices are arranged by increasing depth. Suppose that two ellipsoids $\mathbf{E}_i$ and $\mathbf{E}_{i+1}$ have an intersecting region e_i. We assign this overlap to $\mathbf{E}_i$ as it appears first along the ray, *i.e.*, $\mathbf{E}_i$ has a smaller depth. Consequently, the effective region of $\mathbf{E}_{i+1}$ becomes $\mathbf{E}_{i+1} - e_i$ in this situation.

The first-pass precedence strategy impacts the computation of segment length l_i, requiring updates to account for the adjusted regions. We denote the updated segment lengths by $\tilde{l}_i$ and implement the updating using an efficient algorithm, as elaborated in Algorithm 1. Therefore, the final accumulated attenuation is then given by:

$$I(\mathbf{r}) = \sum_{i=0}^{n-1} \sigma_i \tilde{l}_i. \tag{9}$$

Physically Faithful Overlap Filtering. Before applying the established algorithm for segment length calculation, we first need to identify which ellipsoids are related to each pixel. As illustrated in Figure 4(a), existing methods [11, 32, 62] determine the association between ellipsoids and pixels by finding the circumscribed circle of the projected ellipsoid and subsequently calculating its Axis-Aligned Bounding Box (AABB) [51]. Then, if the AABB overlaps with a pixel, a relation is assumed. However, we observed that many pixels, such as the red ones in Figure 4, do not actually overlap with the ellipse but are still mistakenly considered related, resulting in physically incorrect associations. Therefore, to ensure that each pixel exclusively considers the ellipsoids with actual intersections along the ray, we propose to first determine the Oriented Bounding Box (OBB) [24] and then remove all the redundant regions to resolve the physical unfaithfulness.

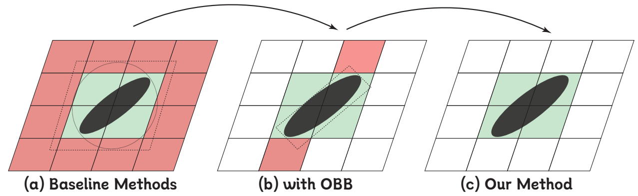

Figure 4. **Illustration of Pixel-Ellipsoid Association.** (a) The baseline method based on AABB [32, 51], resulting in incorrect associations in red pixels. (b) The pixels indicated by OBB [24]. (c) **Our Physically Faithful Overlap Filtering**, which further strictly removes redundant pixels to ensure physical faithfulness.

4.2. Hybrid Progressive Initialization

Before training, we initialize the parameters within each ellipsoid $\mathbf{E}_i$. Specifically, the attenuation coefficient σ_i and the covariance matrix $\mathbf{\Sigma}_{3\text{D}}$ are randomly initialized. Moreover, to estimate the initial positions of ellipsoid centers, SfM-based methods [46, 47] and Dust3R-like methods [54, 57, 63] are widely adopted in conventional 3D reconstruction. However, these methods face significant challenges in the X-ray physical field. Structure-from-Motion (SfM) identifies which pixels across multiple images belong to the same 3D spatial point by matching pixel features [46, 47]. However, in X-ray imaging, even if pixels in different images correspond to the same spatial position, the variations in ray paths can cause significant differences in the remaining energy intensity, resulting in mismatched features. Thus, SfM fails to provide reliable initialization in the X-ray context. Similarly, Dust3R-like methods [54, 57, 63] aggregate color point clouds from multiple views to form a more complete point cloud. However, these methods are designed to predict a discrete map of a color field, which is incompatible with the X-ray domain where we need to model an attenuation coefficient field. Consequently, they cannot directly be used to initialize the attenuation map required for our task.

To overcome these issues, we introduce Hybrid Progressive Initialization tailored for the X-ray physical field, which is based on traditional CT volume reconstruction. This initialization approach comprises three main steps:

(a) CGLS Reconstruction. We begin with the Conjugate Gradient Least Squares (CGLS) method [8, 25, 30] to reconstruct the projection data. CGLS is an iterative optimization technique designed to minimize reconstruction errors, offering robust performance even with sparse or noisy

CVPR
#4997

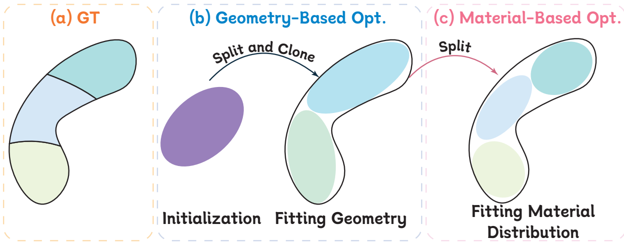

Figure 5. **Illustration of Adaptive Optimization Strategy.** **(a) Ground-truth Geometry and Material Distribution**, with different colors indicating distinct materials. **(b) Geometry-based optimization** [32], which fits the ellipsoids closely to the object geometry. **(c) Our material-based optimization**, which further refines the ellipsoids to capture the material distribution.

data. By iterative refinement, CGLS effectively captures the basic structure and shape of the target object.

(b) TV Regularization. Building on the CGLS reconstruction, we apply Total Variation (TV) regularization [44] to further suppress noise. TV regularization is effective in smoothing out minor fluctuations while preserving edges and structural features, resulting in a cleaner image. By reducing noise and enhancing uniformity in the density distribution, TV regularization improves the quality of the point cloud, aiding subsequent processing steps.

(c) Threshold Filtering. Following TV regularization, we perform threshold filtering on the reconstruction results to eliminate low-density noise, where all points below the threshold are discarded. This effectively reduces the low-intensity noise points scattered around the object, helping to exclude background noise and concentrate the point cloud.

4.3. Material-Based Optimization

X-Gaussian [11] and R^2-Gaussian [62] adopt the optimization strategy of 3DGS [32], which involves splitting and cloning Gaussian ellipsoids in regions with poor geometry fitting and pruning in areas with extremely low opacity. However, this geometry-based optimization strategy fails to accurately model the material distribution. As shown in Figure 5(a), consider a ground-truth object composed of three distinct materials (represented by different colors). While the geometry is successfully learned after applying the 3DGS optimization strategy, it only captures two intermediate material types, failing to accurately represent all three materials (as illustrated in Figure 5(b)).

We hypothesize that this issue stems from the continuous fluctuation of the ellipsoid's attenuation coefficient, preventing it from achieving convergence. To address this, we propose a material-based optimization approach that accounts for the need to accurately fit different materials in the X-ray physical field. As demonstrated in Figure 5(c), when the gradient of the attenuation coefficient exceeds a predefined threshold, we split the ellipsoid into two smaller ones, each scaled down by a factor of 1.6, following empirical guidelines established in prior studies [11, 32, 62]. By aligning with the specific physical characteristics of the X-ray imaging field, we achieve a more natural and accurate modeling of material distribution.

5. Experiment

5.1. Dataset Settings

We conduct experiments on on two sets of datasets including Human Organs [11] (containing five scenes: chest, foot, head, jaw, and pancreas) to test the performance of models in the medical domain, and Daily Objects [11] (including three scenes: bonsai, teapots, engine) to further evaluate the generalization ability. Following R^2-Gaussian [62], we adopt the tomography toolbox TIGRE [7] to capture projections in the range of $0° \sim 180°$ with ponton scatter and electric noise. For highly sparse-view novel view synthesis, 5 and 10 views are used for training and 50 samples 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.

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

Baselines. We compare X-Field with state-of-the-art 3D X-ray reconstruction methods, including TensoRF [13], NeAT [43], NAF [61], SAX-NeRF [12], X-Gaussian [11], and R^2-Gaussian [62]. 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 novel view synthesis, and R^2-Gaussian extends this for CT reconstruction by introducing voxelization. We also evaluate traditional methods, including FDK [21] and SART [2], which are commonly used for CT generation. The novel view images are obtained by leveraging TIGRE for rendering.

Metrics. We adopt peak signal-to-noise ratio (PSNR) [20] to assess the quality of rendered images, structural similarity index measure (SSIM) [55] to measure consistency between predicted images and ground-truths, and Learned Perceptual Image Patch Similarity (LPIPS) [64] to analyze the perceptual quality in high-level feature space. For clarity, we report LPIPS as LPIPS* = LPIPS $\times 10^3$ instead.

Discussion on 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-view (10 views and 5 views) X-Rays reconstruction. Note that we report quantitative results with the mean results of scenes from the same setting, and scene-wise results are presented

	Human Organ 10-views [62]			Object 10-views [62]			Human Organ 5-views [62]			Object 5-views [62]		
Method	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓	PSNR↑	SSIM↑	LPIPS*↓
Traditional Methods												
FDK [21]	12.29	0.678	293.4	16.48	0.718	258.3	8.19	0.620	312.8	14.37	0.691	284.1
SART [2]	13.28	0.693	283.2	17.74	0.726	248.9	9.29	0.632	302.6	15.74	0.661	294.2
Deep Learning-based Methods												
TensoRF [13]	16.57	0.929	183.8	24.27	0.947	152.3	12.29	0.893	188.0	18.23	0.921	211.6
NeAT [43]	16.27	0.936	184.6	25.19	0.958	154.6	11.02	0.889	189.1	17.35	0.916	210.5
NAF [61]	16.93	0.926	194.7	25.48	0.950	152.2	11.22	0.892	196.3	17.05	0.921	209.1
SAX-NeRF [12]	18.26	0.946	185.7	27.44	0.981	148.2	13.22	0.912	190.4	19.05	0.949	205.3
X-Gaussian [11]	16.92	0.948	129.5	22.95	0.983	78.54	15.19	0.928	177.84	19.46	0.958	107.34
R^2-Gaussian [62]	34.64	0.957	86.57	41.24	0.983	40.34	31.18	0.958	108.42	34.78	0.966	82.46
Ours	35.09	0.960	84.13	40.35	0.988	41.21	31.94	0.962	96.24	34.82	0.967	79.15

Table 1. **Results of Quantitative Comparison (§ 5.2).** We compare our X-Field with: (a) Traditional X-ray reconstruction method: FDK[21], SART[2]. (b) Deep Learning-based methods: TensoRF [13], NeAT [43], NAF [61], SAX-NeRF [12], X-Gaussian [11], and R^2-Gaussian [62]. We report LPIPS* = LPIPS $\times 10^3$. We mark out best and second best method for all metrics.

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 Ray Length	27.48	0.875	87.83
Ours	35.03	0.953	72.12

Table 2. **Ablation on the Proposed Components (§ 5.3).**

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

Table 3. **Ablation on the Initialization Methods (§ 5.3).**

#View	PSNR ↑	SSIM ↑	LPIPS* ↓
5	31.61	0.933	92.23
10	35.11	0.959	83.72
15	38.31	0.968	82.91
25	41.71	0.979	75.26
50	42.61	0.993	61.19

Table 4. **Ablation on the Number of Input Views (§ 5.3).**

in the supplementary material. X-Field demonstrates superior performance in reconstructing X-ray novel views across most scenarios, consistently surpassing the state-of-the-art R^2-Gaussian in all metrics. Under the object reconstruction setting, simpler objects and a higher number of views (10 views) are considered. In this scenario, X-Field achieves superior SSIM and competitive PSNR and LPIPS scores compared to R^2-Gaussian. This demonstrates its effectiveness in relatively straightforward cases.

Discussion on Qualitative Results. Figure 6 presents visual comparisons between X-Field and multiple state-of-the-art 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. Take the foot scene for example, R^2-Gaussian introduces black linear artifacts in the bone region, where our method generates smoother textures. In summary, our method is able to effectively mitigates blurry, line, and wave-pattern artifacts while maintaining smoothness in homogeneous areas and on object surfaces.

5.3. 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.

Component Analysis. Table 2 evaluates the effects of individual components on reconstruction performance. We observe that Material Optimization and Overlap Filter have limited impact on reconstruction quality, focusing instead on improving rendering efficiency and aligning with X-ray physical properties. Removing the Intersection Module leads to substantial performance drops, with a PSNR of 1.19, and an SSIM of 0.024, underscoring its importance in preserving structural integrity.

Ray Length is a fundamental component of our ellipsoid representation, 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 drop in PSNR and an 11.71 increase in LPIPS. These results underscore the critical importance of Ray Length in enabling X-Field to achieve accurate and high-quality reconstructions.

Initialization Analysis We compare the proposed hybrid initialization with random initialization and FDK [21]. Table 3 shows that both FDK and our hybrid initialization strategy outperform random initialization. While FDK achieves a slightly lower LPIPS, it also results in lower PSNR and SSIM, likely due to blurrier images reducing high-level feature similarity. In contrast, our method improves both PSNR and SSIM, demonstrating its effectiveness in enhancing reconstruction quality.

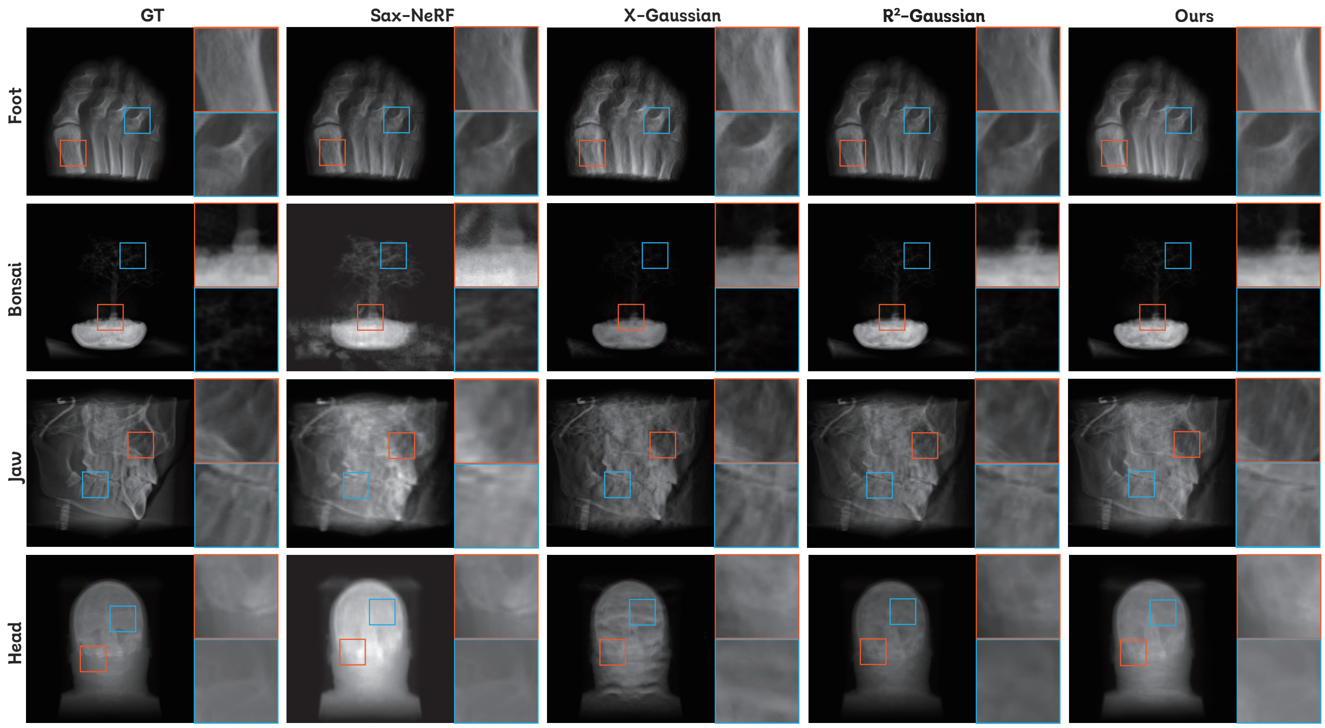

Figure 6. **Qualitative Comparison (§ 5.2).** We present visual examples of reconstructed images across four cases trained with 10 views. Our results demonstrate superior visual quality, richer details, and fewer spatial artifacts. Please zoom in for a closer examination.

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 4, with the increase of the input view numbers, the performance is consistently enhanced. When using 50 views as input, X-Field achieves a performance that is comparable to SOTA works specifically designed for sparse view X-Ray reconstruction [11, 12]. We further show in Figure 7 that, with the increase in the input view number, the reconstructed X-ray images exhibit smoother and clearer bone textures.

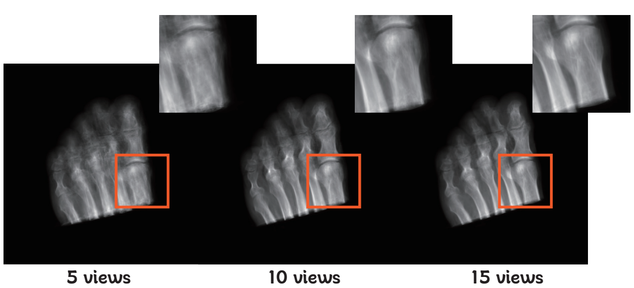

Figure 7. **Comparison of Different Input View Numbers (§ 5.3).**

6. Conclusion and Discussion

Conclusion. This paper presents X-Field, a novel physics-based 3D representation designed for highly sparse-view X-ray reconstruction. We identified and addressed previously overlooked differences between mainstream 3D representations and the physical properties of X-ray, which arise from the fundamental distinctions between visible light and X-ray imaging principles. To overcome these limitations, we redesigned the ellipsoid representation to specifically resolve these conflicts. Moreover, we developed an optimization strategy based on the X-ray distribution. To further improve convergence efficiency, we leveraged traditional medical knowledge to inform the initialization process. Our proposed X-Field significantly surpasses state-of-the-art methods in terms of reconstruction quality, demonstrating its potential for medical applications. More importantly, we provide insights into the design of representations tailored for X-ray imaging, which can also be generalized to other tasks such as reconstruction of translucent objects.

Discussion. Our work, X-Field, does not address all challenges in highly sparse-view X-ray reconstruction. Many existing 3D sparse-view reconstruction methods utilize structural priors and spatial information from pre-trained large models, such as depth estimation, video/image diffusion, or semantic segmentation models. These priors offer valuable constraints and strong guidance, facilitating improved detail recovery and faster convergence. Furthermore, traditional medical structure priors, which could help supervise the internal structure of the X-ray field, remain unexplored in our approach. These areas warrant further investigation in future work.

References

[1] Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, pages 3–20. MIT Press, 1991. 2

[2] 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. 1, 2, 3, 6, 7

[3] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[4] John Ashburner and Karl J. Friston. Voxel-based morphometry—the methods. *NeuroImage*, 11(6):805–821, 2000. 2

[5] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. *arXiv preprint arXiv:2206.05836*, 2022. 2

[6] Jacob Beutel, Harold L. Kundel, and Richard L. Van Metter. *Handbook of Medical Imaging, Volume 1: Physics and Psychophysics*. SPIE Press, Bellingham, WA, 2000. 3

[7] 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. 6

[8] Ake Bjorck and Tommy Elfving. A conjugate gradient method for solving linear least squares problems. *BIT Numerical Mathematics*, 19(2):202–220, 1979. 3, 4, 5

[9] 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. 1

[10] Stewart C. Bushong. *Radiologic Science for Technologists: Physics, Biology, and Protection*. Elsevier Health Sciences, 11th edition, 2017. 2, 3

[11] 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, 3, 5, 6, 7, 8

[12] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *CVPR*, 2024. 1, 2, 3, 6, 7, 8

[13] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision (ECCV)*, 2022. 6, 7

[14] Ozgun Cicek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 424–432, 2016. 2

[15] Allan M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34(9):2722–2727, 1963. 1, 3

[16] 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. 3

[17] 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. 3

[18] 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. 3

[19] 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. 3

[20] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43(12):2959–2965, 1995. 6

[21] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *JOSA A*, 1(6):612–619, 1984. 1, 2, 3, 6, 7

[22] Yuan Gao, Chiyu Jiang, and Yann LeCun. Gaussian splatting for dynamic scene representation. *arXiv preprint arXiv:2308.02114*, 2023. 2

[23] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pages 43–54, 1996. 1

[24] Stefan Gottschalk, Ming C Lin, and Dinesh Manocha. Obbtree: A hierarchical structure for rapid interference detection. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 171–180. ACM, 1996. 5

[25] 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. 3, 4, 5

[26] Godfrey N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British Journal of Radiology*, 46(552):1016–1022, 1973. 1, 2, 3

[27] Godfrey N. Hounsfield. Computed medical imaging. *Science*, 210(4479):22–28, 1980. 3

[28] 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. 1

[29] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers '24*, page 1–11. ACM, 2024. 2

[30] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001. 3, 4, 5

[31] Yoni Kasten, Daniel Doktofsky, and Ilya Kovler. End-to-end convolutional neural network for 3d reconstruction of knee bones from bi-planar x-ray images. In *MICCAI Workshops (MICCAIW)*, 2020. 1

[32] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4), 2023. 1, 2, 3, 5, 6

[33] Timo Kiljunen, Touko Kaasalainen, Anni Suomalainen, and Mika Kortesniemi. Dental cone beam ct: A review. *Physica Medica*, 31(8):844–860, 2015. 3

[34] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '96)*, pages 31–42, 1996. 1

[35] Mingwei Li, Jiachen Tao, Zongxin Yang, and Yi Yang. Human101: Training 100+fps human gaussians in 100s from 1 view, 2023. 2

[36] Yiqun Lin, Zhongjin Luo, Wei Zhao, and Xiaomeng Li. Learning deep intensity field for extremely sparse-view cbct reconstruction. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2023. 1

[37] Vladan Lučić, Friedrich Förster, and Wolfgang Baumeister. Structural studies by electron tomography: from cells to molecules. *Annual Review of Biochemistry*, 74:833–865, 2005. 3

[38] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained scene representation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7206–7215, 2021. 2

[39] 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. 1, 2

[40] Frank Natterer. *The Mathematics of Computerized Tomography*. Society for Industrial and Applied Mathematics, 2001. 2

[41] Ren Ng, Marc Levoy, Mathieu Bredif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CTSR 2005-02, Stanford University, 2005. 1

[42] 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. 3

[43] 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. 3, 6, 7

[44] 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. 4, 6

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

[46] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 5

[47] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016. 5

[48] 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. 3

[49] 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. 1

[50] W. Sun, S. B. Brown, and R. K. Leach. An overview of industrial x-ray computed tomography. 2012. 3

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

[52] Daniel Verbin, Peter Hedman, Pratul P. Srinivasan, Ben Mildenhall, Jonathan T. Barron, Todd Zickler, and Phillip Isola. Ref-nerf: Structured view-dependent appearance for neural radiance fields. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[53] 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. 3

[54] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy, 2024. 5

[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. 6

[56] Yunzhi Yan, Haotong Lin, Chenxu Zhou, Weijie Wang, Haiyang Sun, Kun Zhan, Xianpeng Lang, Xiaowei Zhou, and Sida Peng. Street gaussians for modeling dynamic urban scenes. In *ECCV*, 2024. 2

[57] 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. 5

[58] Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, and Xinggang Wang. Gaussiandreamer: Fast generation from text to 3d gaussians by bridging 2d and 3d diffusion models, 2024. 2

[59] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. *arXiv preprint arXiv:2103.14024*, 2021. 2

[60] 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. 3

[61] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: Neural attenuation fields for sparse-view cbct reconstruction. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2022. 1, 2, 3, 6, 7

[62] 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. 1, 2, 3, 5, 6, 7

[63] 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. 5

[64] 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. 6

[65] Hongyu Zhou, Jiahao Shao, Lu Xu, Dongfeng Bai, Weichao Qiu, Bingbing Liu, Yue Wang, Andreas Geiger, and Yiyi Liao. Hugs: Holistic urban 3d scene understanding via gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21336–21345, 2024. 2