

# Revisiting Pose Sensitivity in Splat-based Computed Tomography under Sparse-view Reconstruction

Kiseok Choi

Hyeongjun Cho

Inchul Kim

Min H. Kim

KAIST

{kschoi;hjcho;ickim;mhkim}@vclab.kaist.ac.kr

## Abstract

X-ray computed tomography (CT) reconstructs volumetric representations of objects from projection images obtained by transmitting X-rays through a target. Recent splat-based tomography, which represents a volume as a continuous distribution of 3D Gaussians, has demonstrated both high reconstruction quality and fast convergence in cone-beam sparse-view CT. However, when deployed in real CT systems with limited and non-uniform view distributions, we observe distinctive streak and strip artifacts that are far more pronounced than in conventional reconstruction methods. Through detailed analysis, we show that these artifacts primarily originate from pose inaccuracies in the acquisition geometry rather than from view sparsity itself. We revisit pose sensitivity in the splatting formulation and derive a stable gradient-based framework that jointly refines geometric parameters during reconstruction. Our study not only identifies how pose perturbations propagate through the differentiable projection operator but also reveals why splat-based CT is particularly vulnerable to geometric misalignment. The resulting formulation remains lightweight and easily integrable into existing pipelines while substantially improving reconstruction fidelity under real-world sparse-view conditions.

## 1. Introduction

X-ray computed tomography (CT) reconstructs the internal structure of an object by measuring the attenuation of X-rays transmitted through it [15]. It plays an essential role in both medical diagnostics and industrial inspection, providing non-destructive access to volumetric structures. Among various system geometries such as parallel-beam, fan-beam, and cone-beam configurations [12], cone-beam CT (CBCT) has become the most widely adopted due to its ability to reconstruct high-quality 3D volumes with fewer projections. The Feldkamp–Davis–Kress (FDK) algorithm [8] remains

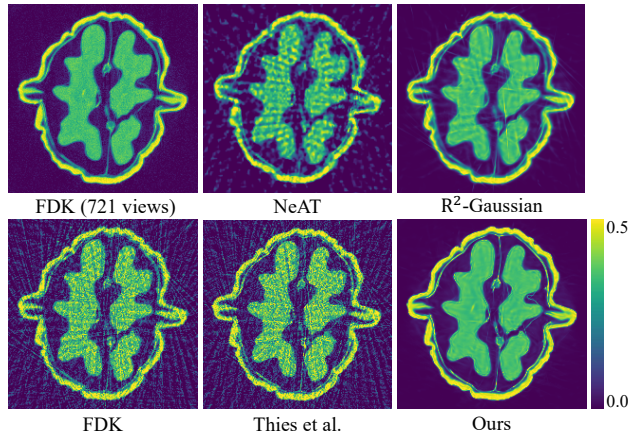


Figure 1. Reconstruction of a real walnut from cone-beam CT using FDK [8] (721 views), joint reconstruction-calibration methods (NeAT [30], Thies et al. [39]),  $R^2$ -Gaussian [48], and our method (others use 75 sparse views). FDK suffers from severe radial artifacts and blur. NeAT reduces noise but introduces directional streaks. Thies et al. improve edges but leave stripe artifacts.  $R^2$ -Gaussian shows needle artifacts from pose errors. Our self-calibrating splat-based method suppresses these artifacts while preserving fine details, highlighting the importance of geometric calibration in real sparse-view CT.

a standard solution for fast reconstruction, while iterative optimization techniques [1, 34] improve quality at the cost of computation.

Recent advances in differentiable rendering and implicit representation have transformed CT reconstruction into an optimization problem on continuous volumetric fields. Methods based on neural implicit functions [4, 30, 47] and, more recently, splat-based representations [48] have demonstrated superior performance in sparse-view reconstruction. By modeling the attenuation field as a set of anisotropic 3D Gaussians, splat-based tomography achieves both high-quality reconstruction and rapid convergence. However, despite its success on synthetic data, we observe that applying this approach to real CT acquisitions introduces distinctive streak and strip artifacts (Figure 3(a)). This discrepancy indicates that the degradation is not solely

due to data sparsity, but arises from underlying geometric inconsistencies in the acquisition process.

Through detailed analysis, we identify geometric calibration errors—particularly inaccuracies in camera poses—as the dominant source of these artifacts. This finding suggests that pose sensitivity, rather than view sparsity, fundamentally limits the robustness of splat-based CT. We therefore revisit the formulation of pose estimation in the splatting pipeline, examining how geometric perturbations propagate through differentiable projection operators and affect reconstruction stability. Our analysis leads to a refined gradient-based framework that allows joint optimization of volumetric and geometric parameters, ensuring stability against small pose deviations.

In summary, this work provides both empirical and theoretical insights into pose sensitivity in splat-based CT and offers a practical self-calibrating formulation that mitigates artifacts under sparse-view conditions. Our main contributions are as follows:

- We conduct the first systematic analysis of pose sensitivity in splat-based CT reconstruction, revealing that geometric misalignment, rather than data sparsity, is the dominant source of artifacts.
- We reformulate the gradient-based pose optimization in splatting, deriving a stable and differentiable calibration framework that jointly refines geometry and volume.
- We present an unbiased simulation framework for generating cone-beam datasets with controlled pose perturbations, enabling reproducible evaluation of geometric sensitivity.

## 2. Related Work

This section reviews recent advances in cone-beam CT reconstruction.

**Cone-Beam CT Reconstruction.** Cone-beam reconstruction algorithms in X-ray CT can be categorized by the way the volume is represented. Voxel-based algorithms construct a cubic grid in 3D space and estimate the value of each element. A representative method is the FDK algorithm [8], which leverages the Fourier slice theorem, resulting in fast and reliable performance that has made it a standard algorithm. Iterative reconstruction methods [1, 34], on the other hand, use a ray-projection model based on the Beer–Lambert law and iteratively optimize the voxel values to fit the model. Advances in computer graphics and neural networks have enabled implicit representation-based reconstruction methods [4, 30, 46, 47]. These approaches represent the volume using neural networks instead of explicit voxel grids. These methods perform well even under sparse-view conditions; however, due to the dense modeling of a volume using neural networks, their time complexity is generally significantly higher.  $R^2$ -Gaussian [48] introduces a new approach to CT reconstruction using Gaussian splat-

ting [17], where the volume is represented as a combination of 3D Gaussian primitives [50]. This approach shows fast and high-quality reconstruction.

**Artifact Analysis.** CT artifacts arise in various forms depending on their sources and have been extensively studied [2]. For example, ring artifacts appear as circular bands and are typically caused by defective detector pixels [6, 21, 32, 41]. Metal artifacts arise from the polychromatic nature of X-rays [16, 26, 42]. Sparse-view artifacts manifest as directional streaks due to insufficient projection data [5, 18, 20, 23, 40]. Calibration artifacts appear as blurring or subtle streaks [10, 30, 31, 37–39, 43], originating from inaccuracies in system geometry. Recently, neural networks have been actively used to classify artifacts and identify their causes [14, 24, 29]. However, works in this category take a slice from the reconstructed volume to produce heat maps of each artifact, so they do not account for these in 3D, and only partial 2D features from slices can be leveraged. Additionally, the performance of these methods is significantly constrained due to the limited access to real artifact datasets.

**Geometric Calibration.** Precise calibration of a geometry configuration is crucial to the quality of the CT reconstruction. This is a challenging problem, particularly in rotating systems, which can result in discrepancies between the actual and ideal geometries. Calibration methods can be broadly categorized into *offline* and *online* approaches [9]. The offline methods [13, 44, 45] use specially manufactured phantoms of known size to estimate accurate geometry. Although feature extraction from pre-scanned phantoms can yield precise calibration parameters, these methods cannot accommodate real-time geometric perturbations during scanning. Moreover, they require additional scanning time and are primarily used for initial system setup. Online approaches [19, 25, 27], in contrast, calibrate based on the actual target object. The calibration parameters are jointly optimized during the reconstruction process, thereby improving quality and eliminating the need for an additional scan. These approaches calculate a loss based on reconstruction quality and compute gradients with respect to camera pose parameters to iteratively reduce the loss by pose errors. Thies et al. [39] perform FDK-based reconstruction while correcting head motion, but the FDK limits the output quality and is dependent on a pre-trained network to measure the quality of reconstruction. Gao et al. [10] reduce pose errors using a pre-trained projection network with segmented object volumes and captured projections, but their method requires ground-truth segmentation as prior information. Wu et al. [43] optimize an implicit representation network while correcting camera poses, but their approach supports only the fan-beam geometry and does not provide publicly available code. Rückert et al. [30] employ an adaptive octree structure for an efficient implicit repre-

sensation network with pose correction. In our method, we employ a splat-based method [48] that offers fast, robust, and fully differentiable reconstruction. We further study the geometry parameters in [48] that support volumetric X-ray projections with various poses, which differ from those of the splatting method in the novel view synthesis [17].

### 3. Analysis

In the splat baseline work [48], the results from the real datasets in the sparse-view setting show streaky artifacts (Figure 3(a)), even though there were no metallic objects that induced anomalies in the scene. We study this problem by designing experiments shown in Figure 2. We first obtain a pseudo ground-truth volume by reconstructing with FDK using dense 721 projections, under the assumption that FDK can produce a sufficiently accurate reconstruction when enough projections are available (Step 1 in Figure 2). Generally, the root mean squared error (RMSE) scales as

$$\text{RMSE} \propto \frac{1}{\sqrt{N}}, \quad (1)$$

where  $N$  denotes the number of projections under unbiased beam geometry errors. The mathematical derivation supporting this assumption is provided in the supplemental material. Then, we create projection images from the reconstructed volume, which consists of 75 views, serving as a sparse-view set of projections. We run the splat reconstruction [48] using the regenerated 75 projections to obtain another volume. Notably, in the volume from the recreated projections, the streaky artifacts are significantly suppressed (Figure 3(b)) compared with those from the 75 projections from the original real projections (Figure 3(a)). According to this experiment, we find that the streaky artifacts do not mainly come from the sparsity of input projections. To further investigate the cause, we compare the error of the estimated projections from the splat reconstruction between the synthesized dataset and the real dataset (Figure 4). The difference between the ground truth and the estimated projections using the synthetic data is uniformly distributed (Figure 4(c)); however, with the real data, it shows a *directional bias* around edges (Figure 4(f)). From this observation, we conclude that the artifacts from the real dataset, when using the splat reconstruction, mainly originate from geometric errors.

### 4. Method

**Projection Formulation.** We first restate the forward projection in the splatting formulation to examine how small pose perturbations affect the rendered intensity. Unlike the voxel-based Beer–Lambert model, the Gaussian representation introduces anisotropic weighting along the ray direction [48], making the reconstructed intensity inherently sen-

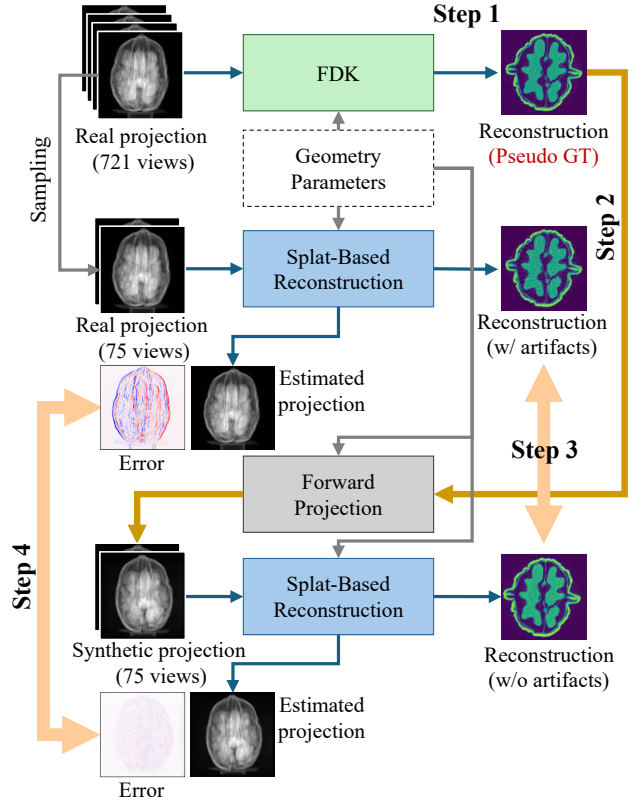


Figure 2. The pipeline of our artifact analysis experiment. (Step 1) A pseudo ground truth (GT) volume is reconstructed using 721-view projections from the real dataset by the FDK [8]. (Step 2) We synthesize the 75 projection images from the pseudo-GT volume. (Step 3) We compare the results from the splat reconstruction [48] using both real and synthetic projections. (Step 4) We compare the error of the estimated projections from the splat reconstruction between the real dataset and the synthesized dataset to identify the source of the error.

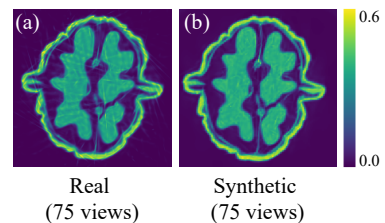


Figure 3. Comparison between the reconstructions from real 75-view projections and synthetic 75-view projections (Figure 2). The artifacts are observed only in the reconstruction from real projections, indicating the sparse input is not the main cause of the streaky artifacts.

sitive to geometric misalignment. Formally, the attenuation field is modeled as:

$$\mu(\mathbf{x}) = \sum_i^M \rho_i \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{p}_i)^T \Sigma_i^{-1}(\mathbf{x} - \mathbf{p}_i)\right), \quad (2)$$

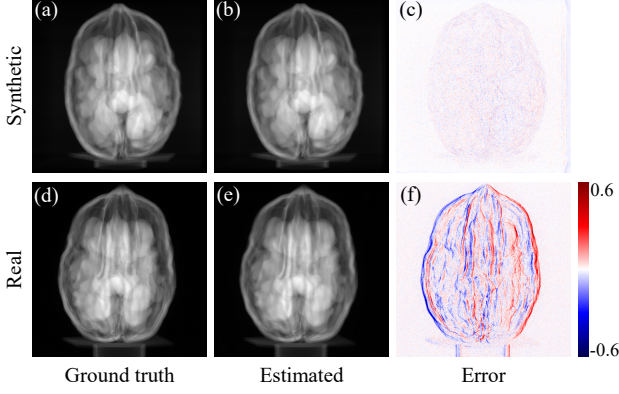


Figure 4. Comparison of projection images and error maps between the real and synthetic datasets with 75 projections. (a) and (b) show projections from the synthetic dataset and the estimated one from the splat reconstruction [48]. (d) and (e) show projections from the real dataset and the estimated one from the splat method. The difference between the ground truth and the estimated projections using the synthetic data is uniformly distributed (c); however, with the real data, it shows a *directional bias* around edges (f), revealing the inaccuracy of camera poses.

where  $\mathbf{x}$  is the spatial query point,  $M$  is the number of Gaussians,  $\rho_i$  is the density of the  $i$ -th Gaussian, and  $\mathbf{p}_i \in \mathbb{R}^{3 \times 1}$ ,  $\Sigma_i \in \mathbb{R}^{3 \times 3}$  denote its center and covariance, respectively. The forward projection of a ray is then computed by integrating the contribution of all Gaussians along the ray path:

$$I(\hat{\mathbf{x}}_k) = \sum_i^M \sqrt{\frac{2\pi|\check{\Sigma}_{i,k}|}{|\hat{\Sigma}_{i,k}|}} \rho_i \exp\left(-\frac{1}{2}(\hat{\mathbf{x}}_k - \hat{\mathbf{p}}_{i,k})^T \hat{\Sigma}_{i,k}^{-1} (\hat{\mathbf{x}}_k - \hat{\mathbf{p}}_{i,k})\right), \quad (3)$$

where  $\hat{\mathbf{x}}_k \in \mathbb{R}^{2 \times 1}$  is the projected query point,  $\check{\Sigma}_{i,k} \in \mathbb{R}^{3 \times 3}$  is the ray-space covariance,  $\hat{\Sigma}_{i,k} \in \mathbb{R}^{2 \times 2}$  is the projected Gaussian covariance, and  $\hat{\mathbf{p}}_{i,k} \in \mathbb{R}^{2 \times 1}$  is the center coordinate of the projected Gaussian.  $k$  is the index of the projection image. The parameters  $\rho_i$ ,  $\mathbf{p}_i$ , and  $\Sigma_i$  are optimized to reduce the discrepancy between the computed projections from Equation (3) and the input projection images. After optimization, the final attenuation volume is obtained by summing up the Gaussian components at query points such as voxel grids.

**Pose Parameterization.** To ensure stable optimization near the nominal geometry, we parameterize pose perturbations as small increments relative to the initial configuration. This incremental formulation mitigates gradient explosion under small-angle approximation and allows fine-grained refinement of rotation and translation during joint optimization. We express the rigid body motion of each camera using a rotation matrix  $\mathbf{W}_k \in \mathbb{R}^{3 \times 3}$  and a translation vector  $\mathbf{t}_k \in \mathbb{R}^{3 \times 1}$ , then the  $i$ -th Gaussian primitive (Equation (2)) is transformed as:

$$\tilde{\mathbf{p}}_{i,k} = \mathbf{W}_k \mathbf{p}_i + \mathbf{t}_k, \quad \tilde{\Sigma}_{i,k} = \mathbf{W}_k \Sigma_i \mathbf{W}_k^T, \quad (4)$$

where  $\tilde{\mathbf{p}}_{i,k}$  and  $\tilde{\Sigma}_{i,k}$  are the transformed Gaussian's coordinates and covariance matrix in 3D. The transformed Gaussian is then projected onto the 2D detector as below:

$$\hat{\mathbf{p}}_{i,k} = \psi(\tilde{\mathbf{p}}_{i,k}), \quad \hat{\Sigma}_{i,k} = \left(\mathbf{J}_{i,k} \tilde{\Sigma}_{i,k} \mathbf{J}_{i,k}^T\right)_{\{0,1\},\{0,1\}}, \quad (5)$$

where  $\psi(\cdot) : \mathbb{R}^{3 \times 1} \rightarrow \mathbb{R}^{2 \times 1}$  is the non-linear projection function,  $(\cdot)_{\{0,1\},\{0,1\}}$  is an operator that crops the first  $2 \times 2$  submatrix from a matrix, and  $\mathbf{J}_{i,k}$  is the Jacobian of projection function with respect to the camera-space position:

$$\mathbf{J}_{i,k} = \frac{\partial \phi(\tilde{\mathbf{p}}_{i,k})}{\partial \tilde{\mathbf{p}}_{i,k}} \in \mathbb{R}^{3 \times 3}, \quad (6)$$

$$\phi(\tilde{\mathbf{p}}_{i,k}) = \begin{pmatrix} \hat{\mathbf{p}}_{i,k} \\ \|\tilde{\mathbf{p}}_{i,k}\| \end{pmatrix} \in \mathbb{R}^{3 \times 1}$$

where  $\phi(\cdot)$  is the mapping function from the camera-space to ray-space used in the volumetric EWA splatting [50]. Note that in the original Gaussian Splatting [17], this mapping function does not include the last component, so it becomes 2D. On the other hand, in the CT splatting the third component in  $\phi(\cdot)$  is included, since we need 3D information of the mapping function to compute the ray integral of the Gaussian primitives located in the real distance to simulate the X-ray attenuation energy.

To optimize the pose parameters, we model the motion using the 4D quaternion for rotation and the 3D vector for translation, following prior works [7, 22, 33, 49]. Unlike conventional CT calibration methods, our approach optimizes these parameters simultaneously with the volume, allowing for self-calibrating reconstruction. To ensure fine-grained optimization around the initial geometric configuration, we parameterize pose errors relative to the prior camera pose by introducing a motion increment as follows [7]:

$$\Delta \mathbf{q}_k = \mathbf{q}_k - \mathbf{q}_{k,\text{init}} \in \mathbb{R}^{4 \times 1}, \quad \Delta \mathbf{t}_k = \mathbf{t}_k - \mathbf{t}_{k,\text{init}} \in \mathbb{R}^{3 \times 1}, \quad (7)$$

where  $\mathbf{q}_k$  and  $\mathbf{t}_k$  represent the quaternion and translation vectors, while  $\Delta \mathbf{q}_k$  and  $\Delta \mathbf{t}_k$  denote the small pose correction terms optimized during reconstruction. This formulation models a small linearized increment in motion, allowing more accurate pose refinement. The quaternion and translation parameters are converted to the rotation matrix and the translation vector:

$$\mathbf{W}_k = \text{Rot}(\Delta \mathbf{q}_k + \mathbf{q}_{k,\text{init}}), \quad \mathbf{t}_k = \Delta \mathbf{t}_k + \mathbf{t}_{k,\text{init}}, \quad (8)$$

where  $\text{Rot}(\cdot)$  is the rotation mapping function from the quaternion vector. The parameters for our joint calibration and reconstruction method are as follows:

$$\Theta = \{\mathbf{p}_i, \Sigma_i, \rho_i, \Delta \mathbf{q}_k, \Delta \mathbf{t}_k\}. \quad (9)$$

**Optimization.** Our optimization problem is given by:

$$\Theta = \arg \min_{\Theta} \mathcal{L}(I_k, \hat{I}_k), \quad (10)$$

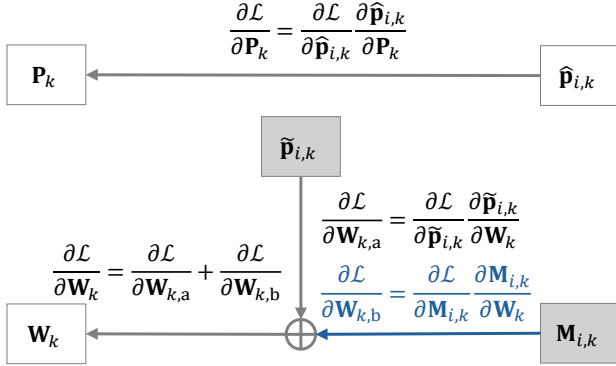


Figure 5. The backward gradient computation pipeline with respect to the intermediate values  $\mathbf{P}_k$  and  $\mathbf{W}_k$ , which are directly dependent on the camera pose parameters.  $\mathbf{P}_k \in \mathbb{R}^{3 \times 4}$  is the perspective projection matrix that transforms 4D homogenous coordinates to the 3D ones.  $\mathbf{M}_{i,k} \in \mathbb{R}^{3 \times 3}$  is the multiplication of  $\mathbf{J}_{i,k}$  and  $\mathbf{W}_k$ . Note that the term  $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{k,b}}$  in blue is different from the calibration method in the RGB splatting [7].

where  $\mathcal{L}$  is the loss function, and  $I_k, \hat{I}_k \in \mathbb{R}^{H \times W}$  denote the given and estimated projection images respectively. Our loss function is defined as:

$$\mathcal{L}(I_k, \hat{I}_k) = \mathcal{L}_{L1}(I_k, \hat{I}_k) + \lambda \mathcal{L}_{\text{SSIM}}(I_k, \hat{I}_k). \quad (11)$$

The objective function combines pixel-wise and structural similarity losses without total variation (TV) regularization. We find that, under our joint calibration framework, the explicit TV term becomes unnecessary and can even suppress essential geometric gradients, thereby reducing the system’s ability to recover fine pose corrections. This observation empirically supports that stability arises naturally from accurate geometric modeling rather than from heuristic smoothness constraints.

Figure 5 illustrates the backward gradient flow through pose-dependent transformations. We explicitly track the Jacobian terms with respect to both the camera projection matrix  $\mathbf{P}_k$  and the rotation matrix  $\mathbf{W}_k$ , allowing us to observe how small geometric perturbations propagate through the differentiable rendering pipeline. This analysis provides insight into the source of instability in prior splat-based CT methods, which often neglect these cross-derivative dependencies.

## 5. Results

### 5.1. Experiment Setup

We implement our method using PyTorch [28] and CUDA [11], running experiments on an Intel Xeon 4214R CPU (2.4 GHz) and an NVIDIA RTX A6000 GPU. For optimization, we adopt the Gaussian learning rates from R<sup>2</sup>-Gaussian [48] while setting the learning rate for camera parameters to  $2e-4$ , which exponentially decays to  $2e-5$  over

30,000 steps. For synthetic data generation, we use TIGRE-v2.3 [3], while real-world evaluations are conducted on the CT dataset from [36]. This setup ensures a controlled evaluation on synthetic data and a realistic assessment of our method’s robustness in real-world scenarios.

### 5.2. Dataset Generation with Geometric Error

To validate our self-calibration framework, we generate a CT dataset with simulated geometric errors. Conventional synthetic datasets typically assume an ideal circular trajectory, where the sensor rotates around a fixed center of rotation. However, mechanical imperfections in real systems introduce deviations from this ideal configuration, resulting in misaligned projections and degraded reconstruction quality. To simulate these real-world effects, we apply perturbations to the camera parameters during projection generation, embedding mechanical misalignment into the dataset. We model the detector geometry as a rigid-body transformation  $\mathbf{T}$  in the special Euclidean group  $\text{SE}(3)$ . To introduce realistic and unbiased perturbations, we treat rotation and translation errors separately. Since translation  $\mathbf{t}$  is linear, we sample it from a zero-mean Gaussian distribution:  $\mathbf{t} \sim \mathcal{N}(0, \sigma_{\text{trans}}^2 \mathbf{I}_{3 \times 3})$ , where  $\sigma_{\text{trans}}^2$  controls the magnitude of translation variation. In contrast, rotation  $\mathbf{R}$  lies in the nonlinear space  $\text{SO}(3)$ , preventing direct additive perturbations. Instead, we apply a logarithmic map to transform it into the tangent space, add Gaussian noise with variance  $\sigma_{\text{rot}}^2$ , and then apply the exponential map to project it back to  $\text{SO}(3)$  [35]. This ensures realistic rotational perturbations while preserving valid transformations. Note that because translations are sampled in voxel space, their physical effect may vary with dataset scaling. By incorporating these calibrated perturbations, our dataset provides an unbiased and realistic testbed for evaluating self-calibration in X-ray CT reconstruction. Mathematical proofs of the unbiasedness of our dataset generation are provided in the supplemental material.

### 5.3. Evaluation

**Quantitative Evaluation.** We evaluate our method on synthetic datasets and present the reconstructed volume results in Figure 6. We compare our approach against traditional tomography methods [8], sparse-view reconstruction methods [4, 48], and the state-of-the-art joint calibration and reconstruction methods [30, 39]. To simulate real-world geometric miscalibrations, we introduce camera rotation and translation noise with standard deviations of 0.03 in the Lie algebra domain and 1.0 in unit length (corresponding to the voxel size of the reconstruction volume), respectively. Our method consistently outperforms all baseline methods both quantitatively and qualitatively, demonstrating robustness under geometric perturbations. Note that our approach exceeds the SOTA method [39] by approximately

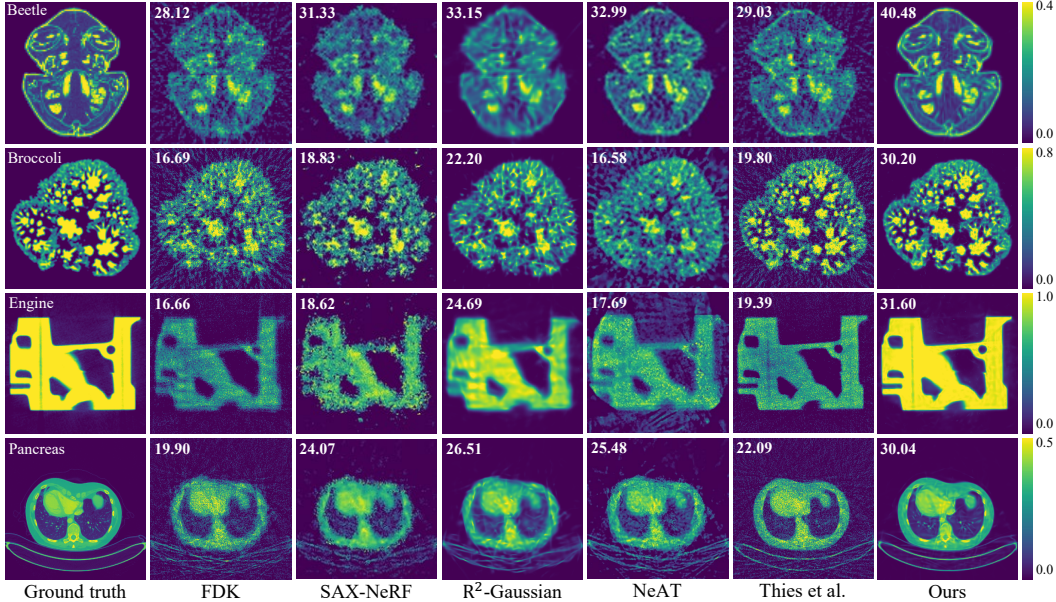


Figure 6. Comparison of reconstructed volume slices on the synthetic dataset with the pose perturbation. We compare our method against existing approaches, including FDK [8], SAX-NeRF [4], R<sup>2</sup>-Gaussian [48], NeAT [30], and Thies et al. [39] across multiple objects. The numbers in the upper-left corners of each image indicate the PSNR values (dB) relative to the respective ground truth volume. Our method achieves higher PSNR values and better visual quality, demonstrating superior robustness against geometric misalignments.

10 dB in PSNR, highlighting its superior reconstruction accuracy. In contrast, existing methods often fail to recover accurate volumes under geometric errors, as they rely on a pre-calibrated hardware setup or lack of self-correction performance. To further validate the effectiveness of our approach, we report quantitative results across different synthetic datasets, with and without geometric perturbations on the baseline splat-based reconstruction method [48] and ours (Table 1). On the noise-free dataset, the baseline and ours indicates similar performance. But on the noisy dataset, ours shows significantly better results compared to the baseline. These results confirm that our method effectively reconstructs volumes while simultaneously handling camera pose errors. In addition to reconstruction accuracy, we evaluate geometric calibration performance in Table 2. We measure camera pose estimation errors using the transformation:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^t & 1 \end{bmatrix} = \mathbf{T}_{\text{gt}}^{-1} \mathbf{T}_{\text{est}}, \quad (12)$$

where  $\mathbf{T}_{\text{gt}}$  and  $\mathbf{T}_{\text{est}}$  represent the ground truth and estimated camera transformations, respectively. We compute the orientation error as  $\theta = \frac{1}{N} \sqrt{\arccos \frac{\text{tr}(\mathbf{R})-1}{2}}$ , and define the translation error as the Euclidean distance between the estimated and ground truth camera positions. Table 2 reports the root mean squared error (RMSE) of  $\theta$  and  $\mathbf{t}$ , demonstrating that our self-calibration framework significantly reduces pose estimation errors. These results highlight the effectiveness of our method in simultaneously optimizing

Table 1. Comparison of the baseline splat-based reconstruction [48] and ours on the noise-free and noisy datasets. Values are reported in PSNR.

Scene	Noise-free dataset		Noisy dataset	
	Baseline	Ours	Baseline	Ours
Chest	35.81	35.68	26.69	30.44
Foot	32.51	32.04	25.46	30.57
Head	40.99	40.92	30.91	34.84
Jaw	37.27	36.77	27.12	32.38
Pancreas	37.91	37.50	26.51	30.04
Beetle	43.18	43.22	33.15	40.48
Bonsai	35.62	34.99	26.61	30.05
Broccoli	36.54	34.70	22.21	30.20
Kingsnake	39.85	39.97	35.86	38.16
Pepper	39.40	38.68	26.51	31.96
Backpack	38.69	38.90	27.88	32.19
Engine	40.25	39.33	24.69	31.60
Mount	38.94	38.08	29.07	31.13
Present	38.06	37.82	28.20	33.84
Teapot	47.81	47.79	36.65	43.43

volume reconstruction and geometric calibration, making it a robust solution for real-world X-ray tomography applications.

**Qualitative Evaluation.** Figure 7 presents the reconstruction results on the real dataset [36] using 75 views. The FDK reconstruction exhibits strong radial artifacts, which arise due to sparse-view sampling. While SART provides improved reconstruction quality over FDK, it still suffers from similar structured artifacts. SAX-NeRF [4] effectively suppresses the radial artifacts, but details are smoothed out due to calibration errors. R<sup>2</sup>-Gaussian [48] reduces anoma-

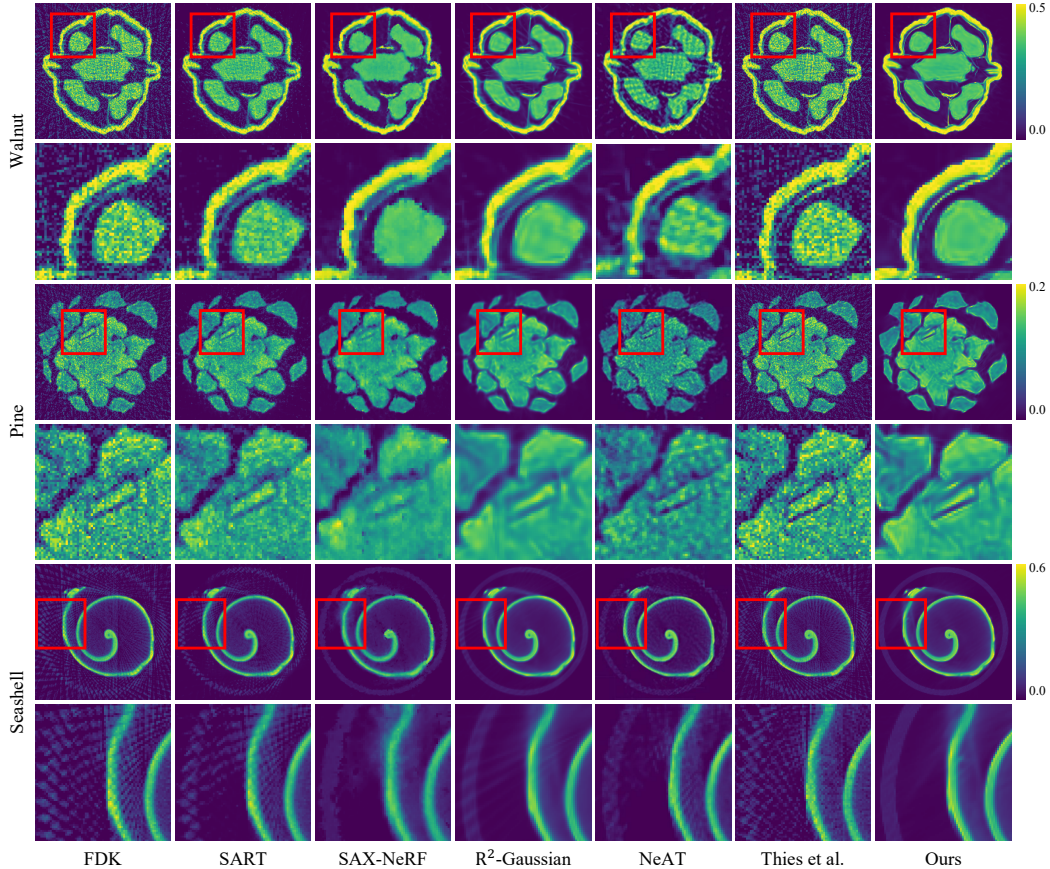


Figure 7. Comparison of reconstructed volume slices on the real dataset [36]. All methods reconstruct the volume using 75 projections. Our approach achieves superior performance, effectively mitigating reconstruction artifacts and preserving fine details, thanks to our self-calibrating geometric correction. The zoomed-in regions (highlighted in red boxes) demonstrate that our method recovers finer structures with fewer artifacts compared to other approaches.

Table 2. We report the root mean squared error (RMSE) of translation and orientation errors after applying our self-calibration framework. The translation error is expressed in arbitrary units (1 AU corresponds to one voxel size of the reconstruction volume), and the orientation error is measured in degrees. Our method outperforms over the state-of-the-art online calibrating reconstruction methods with significantly reduced pose errors.

Scene	Translation (AU)			Orientation (°)		
	NeAT [30]	Thies et al. [39]	Ours	NeAT [30]	Thies et al. [39]	Ours
Chest	1.076	2.347	<b>0.973</b>	2.720	3.919	<b>0.621</b>
Foot	1.768	2.685	<b>1.213</b>	2.867	4.434	<b>0.461</b>
Head	1.254	2.371	<b>0.726</b>	3.019	3.682	<b>0.782</b>
Jaw	1.310	2.616	<b>0.924</b>	2.872	4.295	<b>0.523</b>
Pancreas	1.455	2.587	<b>1.247</b>	3.058	3.649	<b>1.020</b>
Beetle	1.076	2.037	<b>0.809</b>	3.030	3.144	<b>0.194</b>
Bonsai	2.804	3.083	<b>0.732</b>	3.035	3.961	<b>0.747</b>
Broccoli	1.475	2.095	<b>0.671</b>	2.857	5.580	<b>0.314</b>
Kingsnake	1.097	2.025	<b>0.484</b>	2.983	3.214	<b>1.590</b>
Pepper	1.469	2.711	<b>0.397</b>	2.789	4.892	<b>0.629</b>
Backpack	1.570	2.413	<b>0.549</b>	2.894	3.412	<b>0.563</b>
Engine	1.632	2.357	<b>0.598</b>	2.739	5.845	<b>0.484</b>
Mount	1.169	2.768	<b>0.769</b>	2.550	4.313	<b>0.567</b>
Present	1.243	2.546	<b>0.399</b>	3.022	3.784	<b>0.577</b>
Teapot	1.152	2.308	<b>0.397</b>	2.781	3.017	<b>0.339</b>
Mean	1.437	2.463	<b>0.726</b>	2.881	4.076	<b>0.627</b>

lies but remains affected by needle-like artifacts, a consequence of imperfect camera calibration. To mitigate these issues,  $R^2$ -Gaussian employs TV regularization, which suppresses artifacts but over-smooths surfaces, as observed in the fourth column of Figure 7. NeAT [30] produces sharp object boundaries but fails to reconstruct homogeneous regions effectively. Thies et al. [39] show slightly better performance than their baseline method [8], but suffer from severe sparse-view artifacts inherited from the baseline FDK properties. In contrast, our method achieves high-accuracy reconstruction, effectively recovering fine details while suppressing radial artifacts without relying on aggressive smoothing. By jointly optimizing volumetric reconstruction and geometric calibration, our approach ensures artifact-free and detail-preserving reconstructions, demonstrating clear advantages over existing methods.

**Computation Time.** To evaluate the computational efficiency of our method, we compare the average reconstruction time over all synthetic scenes across different approaches, as summarized in Table 4. Our method achieves

Table 3. Reconstruction performance with camera noise varies. We compare the baseline [48] and our method under different camera pose perturbations. Standard deviations  $\sigma_{\text{rot}}$  and  $\sigma_{\text{trans}}$  indicate rotational and translational noise levels. Both methods perform well without noise (first row), but our method shows greater robustness to pose errors, achieving higher PSNR and SSIM in most scenario levels. The best scores for each experiment are highlighted in bold.

Scene	$\sigma_{\text{rot}}$	$\sigma_{\text{trans}}$	Baseline [48]		Ours	
			PSNR (dB)	SSIM	PSNR (dB)	SSIM
Beetle	-	-	43.18	0.995	<b>43.22</b>	<b>0.995</b>
	0.01	0.5	37.28	0.979	<b>41.38</b>	<b>0.992</b>
	0.02	1.0	33.90	0.954	<b>40.65</b>	<b>0.990</b>
	0.05	2.5	31.91	0.930	<b>35.42</b>	<b>0.970</b>
	0.10	5.0	30.80	0.920	<b>32.32</b>	<b>0.939</b>
Head	-	-	<b>40.99</b>	<b>0.983</b>	40.92	0.981
	0.01	0.5	35.29	0.932	<b>37.87</b>	<b>0.954</b>
	0.02	1.0	32.25	0.883	<b>35.63</b>	<b>0.944</b>
	0.05	2.5	28.53	0.808	<b>32.61</b>	<b>0.892</b>
	0.10	5.0	26.41	0.768	<b>28.17</b>	<b>0.805</b>

Table 4. Comparison of mean computation time (in minutes) among joint reconstruction and calibration methods, including the baseline.

Computation time (min)			
Baseline [48]	NeAT [30]	Thies et al. [39]	Ours
23.19	48.35	31.15	<b>20.89</b>

the lowest computation time among joint reconstruction and calibration methods, even outperforming the baseline [48]. The reduction in computation time compared to the baseline mainly arises from the absence of the total variation (TV) regularization term in our formulation.

#### 5.4. Ablation Studies

We evaluate the reconstruction performance of our method against the baseline splat-based algorithm [48] under different levels of pose perturbation and varying numbers of input views. As shown in Table 3, both methods achieve high reconstruction accuracy in the absence of pose errors. However, as the perturbation level increases, the baseline performance rapidly deteriorates, while our method maintains stable results thanks to its self-calibrating design. Table 5 summarizes the results under pose perturbations ( $\sigma_{\text{rot}} = 0.03$ ,  $\sigma_{\text{trans}} = 1.0$ ) across 75, 50, and 25 views. As the number of views decreases, reconstruction quality drops for both methods, yet our approach consistently outperforms the baseline. These results demonstrate the robustness of our joint calibration module in mitigating geometric misalignment while preserving fine structural details.

#### 6. Limitation and Discussion

We further evaluate our method and the baseline method [48] under extremely sparse-view conditions (25 projections) with pose errors. Compared with the baseline (Figure 8(a)), our method still shows significantly fewer

Table 5. Quantitative comparison under geometric perturbations with different numbers of input views. Our method consistently outperforms the baseline [48] across 75, 50, and 25 views on synthetic datasets.

Views	Baseline [48]		Ours	
	PSNR	SSIM	PSNR	SSIM
75	28.50	0.820	<b>33.42</b>	<b>0.906</b>
50	27.67	0.799	<b>31.73</b>	<b>0.879</b>
25	26.44	0.764	<b>29.02</b>	<b>0.810</b>

artifacts with the extremely sparse inputs (Figure 8(b)). However, ours starts to contain the needle-like artifacts due to the absence of the TV regularization. This suggests that

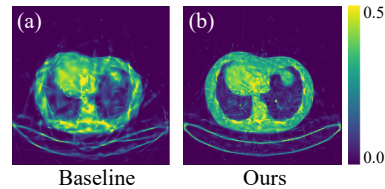


Figure 8. Comparison between the baseline [48] and ours from extremely sparse input projections (25 views) with geometric perturbations. Our method reconstructs the volume with significantly fewer artifacts even under extreme conditions.

while our self-calibration framework consistently mitigates geometric misalignment, additional regularization strategies may be necessary for extreme sparse-view scenarios. Addressing this limitation to enhance reconstruction quality in the extreme sparse inputs is future work.

#### 7. Conclusion

We have presented an analysis method for localizing the root cause of stripe artifacts and a self-calibrating splat-based tomography framework that jointly optimizes volume reconstruction and geometric calibration. By introducing a differentiable loss with respect to camera parameters, our method corrects geometric errors and real-world misalignment in X-ray CT. Experiments demonstrate consistent reconstruction accuracy in the presence of hardware-induced geometric faults, outperforming existing approaches. Our method is robust to pose errors and does not require pre-calibrated hardware. Furthermore, the estimated calibration parameters can be reused in other tomographic reconstruction methods.

#### Acknowledgements

Min H. Kim acknowledges the Samsung Research Funding & Incubation Center (SRFC-IT2402-02), the Korea NRF grant (RS-2024-00357548), the MSIT/IITP of Korea (RS-2022-00155620, RS-2024-00398830, RS-2024-00436680, and 2017-0-00072), the MSIT Advanced GPU Utilization Support Program, and Microsoft Research Asia, with additional support by the KAIST Jang Young Sil Fellow Program.

## References

- [1] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984. PMID: 6548059. 1, 2
- [2] Julia F Barrett and Nicholas Keat. Artifacts in ct: recognition and avoidance. *Radiographics*, 24(6):1679–1691, 2004. 2
- [3] 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. 5
- [4] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *CVPR*, 2024. 1, 2, 5, 6
- [5] Kyle M Champley, Michael B Zellner, Joseph W Tringe, and Harry E Martz Jr. Methods for few-view ct image reconstruction. *arXiv preprint arXiv:2410.07552*, 2024. 2
- [6] Linda CP Croton, Gary Ruben, Kaye S Morgan, David M Paganin, and Marcus J Kitchen. Ring artifact suppression in x-ray computed tomography using a simple, pixel-wise response correction. *Optics express*, 27(10):14231–14245, 2019. 2
- [7] Youming Deng, Wenqi Xian, Guandao Yang, Leonidas Guibas, Gordon Wetzstein, Steve Marschner, and Paul Debevec. Self-calibrating gaussian splatting for large field of view reconstruction. *arXiv preprint arXiv:2502.09563*, 2025. 4, 5
- [8] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984. 1, 2, 3, 5, 6, 7
- [9] Massimiliano Ferrucci, Richard K Leach, Claudiu Giusca, Simone Carmignato, and Wim Dewulf. Towards geometrical calibration of x-ray computed tomography systems—a review. *Measurement Science and Technology*, 26(9):092003, 2015. 2
- [10] Cong Gao, Anqi Feng, Xingtong Liu, Russell H Taylor, Mehran Armand, and Mathias Unberath. A fully differentiable framework for 2d/3d registration and the projective spatial transformers. *IEEE transactions on medical imaging*, 43(1):275–285, 2023. 2
- [11] Design Guide. Cuda c programming guide. *NVIDIA*, July, 29:31, 2013. 5
- [12] Jiang Hsieh. Computed tomography: principles, design, artifacts, and recent advances. 2003. 1
- [13] Jaehong Hwang, Hyeongseok Kim, Taewon Lee, Da in Choi, Taejin Kwon, and Seungryong Cho. Geometry calibration for a dental cone-beam ct system with an offset detector. *Precision Engineering*, 79:264–276, 2023. 2
- [14] Satu I Inkinen, AO Kotiaho, M Hanni, Miika T Nieminen, and Mikael AK Brix. Computed tomography artefact detection using deep learning—towards automated quality assurance. In *Nordic Conference on Digital Health and Wireless Solutions*, pages 17–28. Springer, 2024. 2
- [15] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001. 1
- [16] Willi A Kalender, Robert Hebel, and Johannes Ebersberger. Reduction of ct artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987. 2
- [17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 3, 4
- [18] Byeongjoon Kim, Hyunjung Shim, and Jongduk Baek. A streak artifact reduction algorithm in sparse-view ct using a self-supervised neural representation. *Medical physics*, 49(12):7497–7515, 2022. 2
- [19] Y Kyriakou, R M Lapp, L Hillebrand, D Ertel, and W A Kalender. Simultaneous misalignment correction for approximate circular cone-beam computed tomography. *Physics in Medicine & Biology*, 53(22):6267, 2008. 2
- [20] Anish Lahiri, Gabriel Maliakal, Marc L Klasky, Jeffrey A Fessler, and Saiprasad Ravishankar. Sparse-view cone beam ct reconstruction using data-consistent supervised and adversarial learning from scarce training data. *IEEE Transactions on Computational Imaging*, 9:13–28, 2023. 2
- [21] Xiaokun Liang, Zhicheng Zhang, Tianye Niu, Shaode Yu, Shibin Wu, Zhicheng Li, Huailing Zhang, and Yaoqin Xie. Iterative image-domain ring artifact removal in cone-beam ct. *Physics in Medicine & Biology*, 62(13):5276, 2017. 2
- [22] Changkun Liu, Shuai Chen, Yash Sanjay Bhalgat, HU Siyan, Ming Cheng, Zirui Wang, Victor Adrian Prisacariu, and Tristan Braud. Gs-cpr: Efficient camera pose refinement via 3d gaussian splatting. In *The Thirteenth International Conference on Learning Representations*. 4
- [23] Elena Loli Piccolomini. A comparison of regularization models for few-view ct image reconstruction. *ANNALI DELL'UNIVERSITA' DI FERRARA*, 68(2):385–396, 2022. 2
- [24] Frederic Madesta, Thilo Sentker, Tobias Gauer, and René Werner. Deep learning-based conditional inpainting for restoration of artifact-affected 4d ct images. *Medical Physics*, 51(5):3437–3454, 2024. 2
- [25] Yuanzheng Meng, Hui Gong, and Xiaoquan Yang. Online geometric calibration of cone-beam computed tomography for arbitrary imaging objects. *IEEE Transactions on Medical Imaging*, 32(2):278–288, 2013. 2
- [26] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß. Normalized metal artifact reduction (nmar) in computed tomography. *Medical physics*, 37(10):5482–5493, 2010. 2
- [27] S Oudah, J W Stayman, G J Gang, T Ehtiati, and J H Siewerdsen. Self-calibration of cone-beam ct geometry using 3d–2d image registration. *Physics in Medicine & Biology*, 61(7):2613, 2016. 2
- [28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. 2019. 5
- [29] Prakhar Prakash and Sandeep Dutta. Deep learning-based artifact detection for diagnostic ct images. In *Medical Imaging*

- 2019: *Physics of Medical Imaging*, pages 1055–1060. SPIE, 2019. [2](#)
- [30] Darius Rückert, Yuanhao Wang, Rui Li, Ramzi Idoughi, and Wolfgang Heidrich. Neat: Neural adaptive tomography. *ACM Transactions on Graphics (TOG)*, 41(4):1–13, 2022. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#)
- [31] Darius Rückert, Lorenz Butzhammer, Simon Wittl, Gabriel Herl, Tino Hausotte, and Philipp Kurth. Uncalibrated ct reconstruction for one-shot scanning of arbitrary trajectories. In *13th conference on industrial computed tomography (iCT)*, 2024. [2](#)
- [32] Jakub Šalplachta, Tomáš Zikmund, Marek Zemek, Adam Břínek, Yoshihiro Takeda, Kazuhiko Omote, and Jozef Kaiser. Complete ring artifacts reduction procedure for lab-based x-ray nano ct systems. *Sensors*, 21(1):238, 2021. [2](#)
- [33] Dongbo Shi, Shen Cao, Lubin Fan, Bojian Wu, Jinhui Guo, Renjie Chen, Ligang Liu, and Jieping Ye. No parameters, no problem: 3d gaussian splatting without camera intrinsics and extrinsics. *arXiv preprint arXiv:2502.19800*, 2025. [4](#)
- [34] 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. [1](#), [2](#)
- [35] Joan Sola, Jeremie Deray, and Dinesh Atchuthan. A micro lie theory for state estimation in robotics. *arXiv preprint arXiv:1812.01537*, 2018. [5](#)
- [36] The Finnish Inverse Problems Society. X-ray tomographic datasets, 2024. [5](#), [6](#), [7](#)
- [37] Mareike Thies, Fabian Wagner, Noah Maul, Lukas Folle, Manuela Meier, Maximilian Rohleder, Linda-Sophie Schneider, Laura Pfaff, Mingxuan Gu, Jonas Utz, et al. Gradient-based geometry learning for fan-beam ct reconstruction. *Physics in Medicine & Biology*, 68(20):205004, 2023. [2](#)
- [38] Mareike Thies, Noah Maul, Siyuan Mei, Laura Pfaff, Nastassia Vysotskaya, Mingxuan Gu, Jonas Utz, Dennis Posart, Lukas Folle, Fabian Wagner, et al. Differentiable score-based likelihoods: Learning ct motion compensation from clean images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 253–263. Springer, 2024.
- [39] Mareike Thies, Fabian Wagner, Noah Maul, Haijun Yu, Manuela Goldmann Meier, Linda-Sophie Schneider, Mingxuan Gu, Siyuan Mei, Lukas Folle, Alexander Preuhs, et al. A gradient-based approach to fast and accurate head motion compensation in cone-beam ct. *IEEE Transactions on Medical Imaging*, 2024. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#)
- [40] Yanli Wang, Lianying Chao, Wenqi Shan, Haobo Zhang, Zhiwei Wang, and Qiang Li. Improving the quality of sparse-view cone-beam computed tomography via reconstruction-friendly interpolation network. In *Proceedings of the Asian Conference on Computer Vision*, pages 2255–2269, 2022. [2](#)
- [41] Markus Wedekind, Susana Castillo, and Marcus Magnor. Feature-sensitive ring artifact correction for computed tomography. *Journal of Nondestructive Evaluation*, 42(1):5, 2023. [2](#)
- [42] Qing Wu, Lixuan Chen, Ce Wang, Hongjiang Wei, S Kevin Zhou, Jingyi Yu, and Yuyao Zhang. Unsupervised polychromatic neural representation for CT metal artifact reduction. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. [2](#)
- [43] Qing Wu, Xin Li, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. Joint rigid motion correction and sparse-view ct via self-calibrating neural field. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2023. [2](#)
- [44] Jingyan Xu and Benjamin M. W. Tsui. An analytical geometric calibration method for circular cone-beam geometry. *IEEE Transactions on Medical Imaging*, 32(9):1731–1744, 2013. [2](#)
- [45] Kai Yang, Alexander L. C. Kwan, DeWitt F Miller, and John M Boone. A geometric calibration method for cone beam ct systems. *Medical Physics*, 33(6), 2006. [2](#)
- [46] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021. [2](#)
- [47] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–452. Springer, 2022. [1](#), [2](#)
- [48] Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R<sup>2</sup>-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [8](#)
- [49] Qitao Zhao and Shubham Tulsiani. Sparse-view pose estimation and reconstruction via analysis by generative synthesis. In *NeurIPS*, 2024. [4](#)
- [50] M. Zwicker, H. Pfister, J. van Baar, and M. Gross. Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):223–238, 2002. [2](#), [4](#)