# SAGA: STRUCTURE-AWARE GAUSSIAN ALLOCATION FOR ULTRA-SPARSE-VIEW CT RECONSTRUCTION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Ultra-sparse-view computed tomography (CT) reconstructs a 3D attenuation field from only a few X-ray projections, where severe angular undersampling makes projection consistency insufficient for reliable volumetric recovery. Explicit Gaussian representations provide adaptive volumetric modeling, but their topology evolution is commonly driven by optimization gradients or heuristic criteria that cannot reliably distinguish anatomical structures from sparse-view artifacts and depth-ambiguous responses. We present SAGA, a stage-stabilized structure-aware Gaussian allocation framework for sparse-view CT reconstruction without paired volumetric supervision. SAGA combines a geometry-frequency constrained objective for stabilizing attenuation geometry and preserving structural details with a G+P structure field that integrates volumetric-gradient saliency and residual-backprojection evidence. The resulting structural evidence is converted into per-Gaussian importance scores to guide cloning, splitting, and pruning. A two-stage optimization schedule keeps Gaussian evolution conservative during early geometry formation and progressively activates structure-guided allocation as the reconstruction becomes more reliable. Experiments on five anatomies under 6, 8, and 10-view settings show that SAGA achieves the best average volumetric PSNR and SSIM across all three sampling regimes. Under the challenging 6-view setting, SAGA improves over the explicit Gaussian baseline by 1.44 dB PSNR and 0.0827 SSIM. Ablation studies further validate the contributions of the reconstruction objective, structure-aware allocation, and stage stabilization.

## 1 INTRODUCTION

Computed tomography (CT) reconstructs the three-dimensional distribution of X-ray attenuation coefficients from projections acquired at multiple viewing angles, providing volumetric information for clinical diagnosis and anatomical assessment. Conventional CT relies on sufficiently dense angular sampling to constrain the underlying attenuation field. Reducing the number of projections can lower radiation exposure, shorten acquisition time, and improve scanning efficiency, which is desirable for low-dose imaging, rapid scanning, and geometrically constrained acquisition. Under ultra-sparse-view conditions, however, large portions of angular information are unavailable, substantially weakening the measurement constraints and making reconstruction a severely ill-posed inverse problem.

More specifically, let $\boldsymbol{\mu}$ denote the three-dimensional attenuation volume and $\mathbf{y}$ the measured projections. The CT imaging process can be expressed as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\mu} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{A}$ denotes the projection operator and $\boldsymbol{\varepsilon}$ accounts for measurement noise and model mismatch. With only a few projection angles, $\mathbf{A}$ becomes severely ill-conditioned and admits a large null or near-null space. In particular, there may exist a nonzero perturbation $\Delta\boldsymbol{\mu}$ such that

$$\mathbf{A}\Delta\boldsymbol{\mu} \approx \mathbf{0}. \tag{2}$$

Hence, structurally different attenuation volumes, $\boldsymbol{\mu}$ and $\boldsymbol{\mu} + \Delta\boldsymbol{\mu}$, can produce nearly identical measurements at the acquired views. Noise and modeling errors may also be amplified along weakly constrained directions, resulting in streak artifacts, directional distortions, and structural mislocalization (Kak & Slaney, 2001; Sidky & Pan, 2008; Kudo et al., 2013; Liu et al., 2014; Niu et al.,



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

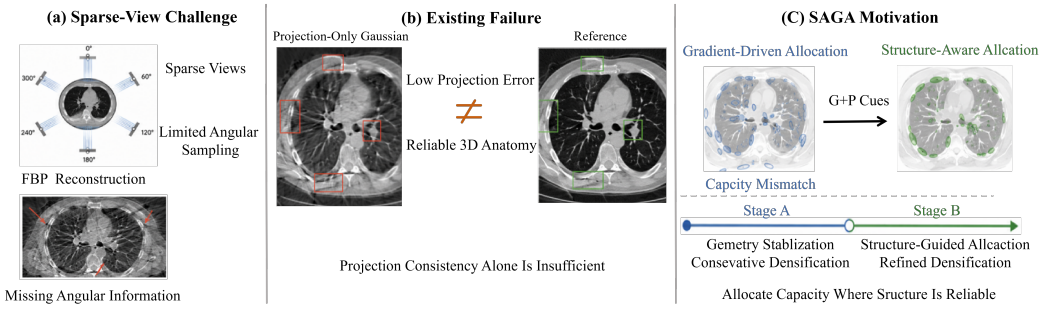


Figure 1: **Motivation of SAGA.** (a) Ultra-sparse angular sampling leaves CT reconstruction severely under-constrained. (b) Projection-consistent Gaussian optimization may still produce unreliable 3D structures, while gradient-driven densification can allocate representation capacity to ambiguous or artifact-dominated regions. (c) SAGA stabilizes the reconstruction before using complementary volumetric and projection-domain cues to guide Gaussian allocation, separating early geometry formation from subsequent structure-guided refinement.

2014). Stronger regularization can reduce these instabilities, but excessive regularization often attenuates thin structures, low-contrast boundaries, and high-frequency anatomical details. A low projection-domain error therefore does not necessarily imply a geometrically accurate 3D volume. Ultra-sparse-view reconstruction must consequently balance consistency with the measured projections and sufficient structural constraints on the recovered attenuation field.

Existing model-based and learning-based approaches address this ambiguity through handcrafted regularization, learned reconstruction priors, sinogram completion, and data-driven inverse mappings (Jin et al., 2017; Zhang et al., 2018; Wang et al., 2018; Han & Ye, 2018; Lee et al., 2019b; Li et al., 2019; Adler & Öktem, 2018; Chen et al., 2018). More recently, neural fields and explicit Gaussian representations have enabled instance-specific reconstruction through differentiable projection and adaptive scene representations (Rückert et al., 2022; Zha et al., 2022; Cai et al., 2024b;a; Zha et al., 2024; Li et al., 2025). Nevertheless, representation flexibility does not remove the ambiguity caused by missing angular measurements. For explicit Gaussian reconstruction, the problem is further coupled with adaptive topology evolution: optimization signals are used not only to update Gaussian parameters but also to determine where primitives should be cloned, split, or removed. Under severe angular undersampling, large responses may arise from genuine anatomical boundaries as well as streaks or depth-ambiguous structures. Allocating additional primitives to such unreliable regions can reinforce early reconstruction errors and make subsequent correction increasingly difficult.

We propose **SAGA**, a stage-stabilized structure-aware Gaussian allocation framework that jointly considers reconstruction stability and representation evolution (Fig. 1). SAGA first initializes an explicit Gaussian attenuation field from an FBP reconstruction and stabilizes its geometry using a geometry-frequency objective. Projection fidelity and volumetric regularization constrain the global attenuation field, while projection-gradient consistency, an FBP-referenced volumetric frequency anchor, and residual suppression provide additional constraints on structural detail. Once the reconstruction becomes sufficiently reliable, volumetric-gradient saliency and residual-backprojection evidence are fused into a G+P structure field. Sampling this field at Gaussian centers yields local importance scores that directly regulate clone, split, and prune operations. A stage-aware schedule keeps topology evolution conservative during early geometry formation and progressively strengthens frequency-related constraints and G+P-guided allocation during later refinement. This design couples reconstruction evidence with adaptive capacity allocation while preventing unstable early structures from prematurely dominating Gaussian evolution.

Our main contributions are:

- We formulate ultra-sparse-view Gaussian CT as a coupled reconstruction–allocation problem, where reconstruction uncertainty affects Gaussian topology evolution and inappropriate allocation can in turn reinforce unreliable structures.

- We introduce geometry-frequency stabilization together with a G+P structure-aware allocation mechanism that combines volumetric-gradient and residual-backprojection evidence to derive per-Gaussian importance for clone, split, and prune operations.
- We develop a stage-aware optimization strategy that separates early geometry formation from subsequent structure-guided refinement, allowing structural cues and detail-oriented constraints to influence Gaussian evolution only after the reconstruction becomes more reliable. Experiments across five anatomies under 6-, 8-, and 10-view settings validate the effectiveness and complementarity of the proposed components.

## 2 Related Work

### 2.1 Sparse-View CT Reconstruction

Analytical methods such as filtered back-projection (FBP) are efficient and physically interpretable, but their accuracy deteriorates rapidly under sparse angular sampling (Kak & Slaney, 2001; Feldkamp et al., 1984; Andersen & Kak, 1984). Iterative and model-based approaches improve robustness by combining projection fidelity with total variation (TV), compressed sensing, higher-order regularization, or edge-preserving priors (Gordon et al., 1970; Rudin et al., 1992; Candès et al., 2006; Donoho, 2006; Sidky & Pan, 2008; Chen et al., 2008; Liu et al., 2014; Niu et al., 2014). These methods remain valuable because they explicitly preserve the CT forward model. Sparse-view reconstruction has also been studied using higher-order regularization and learned completion of missing angular measurements (Liu et al., 2014; Niu et al., 2014; Anirudh et al., 2018). However, ultra-sparse measurements introduce severe *geometric ambiguity* rather than noise alone: multiple volumetric structures can explain the same projections, while strong regularization tends to suppress thin structures and sharp boundaries. This motivates treating reconstruction not only as a value-fitting problem but as a problem of *where to allocate representational capacity*.

### 2.2 Learning-Based CT Reconstruction

Supervised CT reconstruction has been extensively studied through image-domain artifact removal and learned iterative reconstruction (Jin et al., 2017; Chen et al., 2017; Zhang et al., 2018; Adler & Öktem, 2018; Chen et al., 2018; Han & Ye, 2018). Other studies exploit sinogram completion, projection-to-image mappings, and dual- or multi-domain learning to incorporate measurement-domain information (Anirudh et al., 2018; Lee et al., 2019b;a; Li et al., 2019; Pan et al., 2022; Lin et al., 2024; Van der Rauwelaert et al., 2025). Image-domain methods learn to remove artifacts from an initial reconstruction, whereas reconstruction- and dual-domain methods incorporate projections or differentiable imaging operators. Recent model-driven methods further combine explicit imaging models with learned priors to improve reconstruction robustness (Tao et al., 2025). Although such approaches recover visually plausible detail, supervised models typically require paired sparse-view and reference data and often generalize poorly across anatomies, scanners, geometries, and view numbers. Physics-informed, internal-learning, and self-supervised methods reduce this dependency by exploiting only the measured data or imaging model (Ulyanov et al., 2018; Otero Baguer et al., 2020; Sun et al., 2021; Hendriksen et al., 2020; Zang et al., 2021). Nevertheless, projection consistency alone remains under-constrained in the ultra-sparse regime, leaving the allocation ambiguity unresolved.

### 2.3 Neural and Gaussian Tomographic Representations

Implicit neural representations and NeRF-inspired tomography model attenuation as a continuous coordinate function optimized through differentiable projection (Sitzmann et al., 2020; Tancik et al., 2020; Müller et al., 2022; Sun et al., 2021; Rückert et al., 2022; Zha et al., 2022; Yang & Xie, 2024; Cai et al., 2024b). These enable self-supervised reconstruction from sparse measurements but can be costly per case and offer limited direct control over local representational capacity. Explicit Gaussian representations instead provide localized primitives, adaptive density control, and efficient differentiable rendering (Kerbl et al., 2023; Cai et al., 2024a; Zha et al., 2024; Fu et al., 2024; Li et al., 2025; Yu et al., 2025). Recent works have extended Gaussian primitives from radiative X-ray rendering to tomographic reconstruction and dynamic tomography (Cai et al., 2024a; Zha et al., 2024; Fu et al., 2024; Li et al., 2025; Yu et al., 2025). Nevertheless, Gaussian representation alone

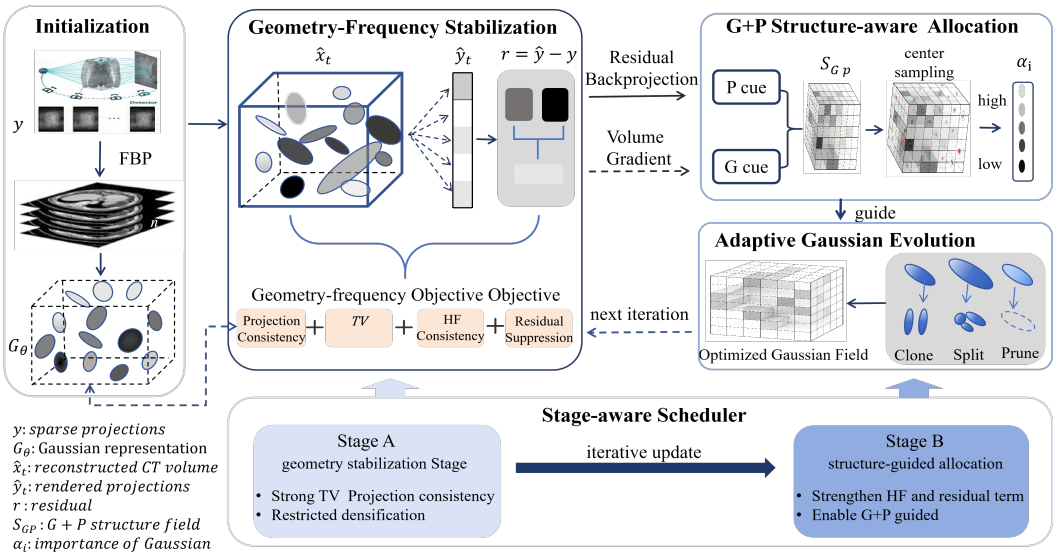


Figure 2: Overview of SAGA. Sparse projections are filtered-backprojected to initialize an explicit Gaussian field. A differentiable projection produces a volumetric-gradient cue ($S_g$) and, via residual backprojection, a projection-consistency cue ($S_p$); the two are fused into a G+P structure field $S_{\mathrm{GP}}$ that estimates per-Gaussian importance $\alpha_i$ and guides adaptive clone/split/prune. A two-stage scheduler separates geometry stabilization (Stage A) from structure-guided allocation (Stage B).

does not resolve sparse-view ambiguity: existing clone, split, and prune operations are commonly driven by gradient magnitude, residuals, or heuristic thresholds, which can allocate primitives to projection-compatible streaks or wrong-depth structures. The question of *what evidence should govern primitive allocation* is therefore largely overlooked.

### 2.4 Structure-Aware Priors and Regularization

TV, higher-order variation, edge-preserving penalties, frequency constraints, and residual- or denoising-based regularizers restrict the solution space of ill-posed reconstruction (Rudin et al., 1992; Sidky & Pan, 2008; Liu et al., 2014; Niu et al., 2014; Romano et al., 2017; Kudo et al., 2013; Hendriksen et al., 2020). Such priors suppress sparse-view artifacts, but overly strong or mismatched assumptions may cause over-smoothing, structural bias, or hallucinated anatomy. A key limitation is that these priors act on reconstruction *values* while adaptive representation growth is left to generic heuristics. In contrast to methods that regularize values but govern capacity by gradient-based rules, our work targets the coupling between objective optimization and representation allocation, using projection-consistent structural evidence rather than an additional learned anatomical prior.

## 3 Method

We propose **SAGA**, a stage-stabilized structure-aware Gaussian allocation framework for ultra-sparse-view CT reconstruction. The key observation underlying SAGA is that sparse-view Gaussian reconstruction involves two tightly coupled problems. First, the reconstructed attenuation field is severely under-constrained by the limited angular measurements, such that a low projection error does not necessarily imply geometrically correct 3D anatomy. Second, the reconstruction itself determines where Gaussian primitives are subsequently cloned, split, or removed. Consequently, unreliable structures formed during early optimization may not only corrupt the current reconstruction, but may also attract additional representation capacity and become progressively harder to correct.

SAGA addresses this coupling by separating *geometry formation* from *structure-guided capacity refinement*. As illustrated in Fig. 2, sparse projections are first used to initialize an explicit Gaussian attenuation field through FBP. The Gaussian representation is then optimized using a geometry-

frequency objective that combines measurement consistency with volumetric and high-frequency regularization. Once the reconstruction becomes sufficiently stable, volumetric-gradient evidence and residual-backprojection evidence are fused into a G+P structure field. Sampling this field at individual Gaussian centers produces local importance scores that guide clone, split, and prune operations. A stage-aware scheduler controls when these structural cues are introduced, preventing uncertain early structures from prematurely driving Gaussian densification.

**Gaussian Initialization and Explicit Reconstruction.** Let $y \in \mathbb{R}^{V\times H\times W}$ denote the measured sparse-view projections acquired under scanning geometry $\Theta$. SAGA represents the unknown attenuation volume using an explicit Gaussian field parameterized by $\theta$. Given a volumetric query domain $\Omega$, the reconstructed attenuation field and its corresponding projections are written as

$$\hat{x}_t = G_{\theta_t}(\Omega), \qquad \hat{y}_t = \mathcal{P}_\Theta(\hat{x}_t), \tag{3}$$

where $G_{\theta_t}$ denotes the Gaussian representation at iteration $t$ and $\mathcal{P}_\Theta$ denotes the differentiable CT forward projector.

Directly optimizing a randomly initialized Gaussian field from only a few projections leads to a highly ambiguous search space. We therefore obtain a coarse reconstruction $x_{\mathrm{FBP}}$ using filtered back-projection (Feldkamp et al., 1984; Kak & Slaney, 2001) and initialize the Gaussian field from this volume. Importantly, $x_{\mathrm{FBP}}$ is not treated as a reconstruction target or additional supervision. Instead, it provides a coarse spatial initialization from which the Gaussian representation is subsequently refined using only the measured sparse projections and the proposed regularization terms. This initialization reduces the burden of recovering global geometry from scratch while retaining the self-supervised nature of the reconstruction.

**Geometry-Frequency Stabilization.** Under ultra-sparse angular sampling, projection consistency alone admits multiple volumetric solutions that can explain the measured views. This ambiguity is particularly problematic for adaptive Gaussian representations because erroneous edges or wrong-depth structures can generate large optimization gradients and subsequently trigger unnecessary densification. SAGA therefore first stabilizes the reconstructed attenuation field before allowing structural evidence to strongly influence Gaussian allocation.

The basic reconstruction objective consists of projection fidelity and volumetric total-variation regularization:

$$\mathcal{L}_{\mathrm{proj}} = \|\hat{y}_t - y\|_2^2, \qquad \mathcal{L}_{\mathrm{tv}} = \|\nabla \hat{x}_t\|_1. \tag{4}$$

The projection term preserves consistency with the measured X-ray data, whereas $\mathcal{L}_{\mathrm{tv}}$ suppresses unstable variations in weakly constrained volumetric regions. This combination follows the established role of TV-type regularization in sparse-view tomography (Rudin et al., 1992; Sidky & Pan, 2008; Liu et al., 2014; Niu et al., 2014). However, excessive reliance on TV tends to attenuate thin structures and sharp anatomical transitions. We therefore complement geometric stabilization with a high-frequency consistency term:

$$\mathcal{L}_{\mathrm{hf}} = \|\nabla \hat{y}_t - \nabla y\|_1 + \eta \|\nabla \hat{x}_t - \nabla x_{\mathrm{FBP}}\|_1, \tag{5}$$

where $\eta$ controls the relative contribution of the volumetric frequency anchor. The first term preserves measured high-frequency structures in the projection domain, while the second discourages the reconstructed volume from drifting arbitrarily away from the coarse structural organization provided by FBP. The FBP volume is detached during optimization and therefore acts only as a fixed structural reference rather than a learnable target.

Sparse-view optimization may additionally leave structured residual patterns that correspond to projection inconsistencies not adequately explained by the current Gaussian field. Let

$$r_t = \hat{y}_t - y. \tag{6}$$

We suppress unstable high-frequency components of the residual through

$$\mathcal{L}_r = \|r_t - \mathcal{B}(r_t)\|_1, \tag{7}$$

where $\mathcal{B}(\cdot)$ denotes the residual smoothing operator. Rather than forcing every residual component to vanish uniformly, $\mathcal{L}_r$ penalizes locally inconsistent residual fluctuations that are particularly susceptible to sparse-view artifacts. Together, $\mathcal{L}_{\mathrm{proj}}$, $\mathcal{L}_{\mathrm{tv}}$, $\mathcal{L}_{\mathrm{hf}}$, and $\mathcal{L}_r$ provide a progressively stabilized reconstruction from which more reliable allocation cues can be extracted.

**G+P Structure-Aware Gaussian Allocation.** Once the coarse volumetric geometry becomes stable, SAGA uses reconstruction evidence to determine where additional Gaussian capacity is needed. Conventional Gaussian densification is typically driven by optimization gradients or fixed thresholds. Under sparse-view CT, however, large gradients may originate from true anatomical boundaries, streak artifacts, or depth-ambiguous structures. Gradient magnitude alone therefore provides insufficient evidence for reliable topology evolution.

We construct two complementary volumetric cues. The first is a geometric cue derived directly from the reconstructed volume:

$$S_g = \mathrm{Norm}\left(\|\nabla \hat{x}_t\|_2\right), \tag{8}$$

where $S_g$ emphasizes spatial locations with strong attenuation transitions. Although these regions are natural candidates for additional representation capacity, $S_g$ alone cannot distinguish genuine anatomical boundaries from artifact-induced gradients.

We therefore introduce a projection cue obtained by backprojecting the current projection discrepancy:

$$S_p = \mathrm{Norm}\left(\mathcal{P}_\Theta^T\left(|r_t|\right)\right), \tag{9}$$

where $\mathcal{P}_\Theta^T$ denotes the adjoint backprojection operator. In contrast to $S_g$, which measures local volumetric saliency, $S_p$ maps measurement-domain disagreement back into the volume and indicates spatial regions that remain insufficiently explained by the current representation. The two cues are therefore complementary: $S_g$ identifies *where structural variation exists*, whereas $S_p$ indicates *where the current representation remains inconsistent with the acquired measurements*.

We combine them into a projection-aware structural field:

$$S_{\mathrm{GP}} = \mathrm{Norm}\left(\alpha_g S_g + \alpha_p S_p\right), \tag{10}$$

where $\alpha_g$ and $\alpha_p$ balance volumetric and projection-domain evidence. For the $i$-th Gaussian centered at $\mu_i$, its local allocation importance is obtained by sampling the structure field at its center:

$$\alpha_i = S_{\mathrm{GP}}(\mu_i). \tag{11}$$

The score $\alpha_i$ provides a direct connection between reconstruction evidence and representation capacity. Gaussians located in structurally informative and measurement-inconsistent regions receive higher priority for refinement, whereas primitives with weak structural support are less likely to consume additional capacity. In this way, the G+P field is not introduced as another reconstruction loss; instead, it explicitly governs *where* Gaussian topology should evolve.

**Adaptive Gaussian Evolution.** The per-Gaussian importance scores are used to regulate topology evolution through clone, split, and prune operations. Cloning increases local representation capacity when additional primitives are required around compact structures, while splitting refines Gaussians whose spatial support is too coarse to model local attenuation variation. Conversely, pruning removes primitives that contribute weakly to the current representation. These operations allow the Gaussian field to redistribute a finite representation budget as optimization proceeds.

A central difficulty is that the reliability of $\alpha_i$ varies throughout training. During early iterations, $\hat{x}_t$ is still strongly affected by FBP artifacts and sparse-view ambiguity, making both volumetric gradients and projection residuals unreliable indicators of anatomical structure. Immediately enabling aggressive structure-guided evolution can therefore create a positive feedback loop: an unstable structure receives additional Gaussians, the increased local capacity reinforces that structure, and later optimization becomes increasingly biased toward the erroneous topology. SAGA avoids this failure mode through stage-stabilized Gaussian evolution.

**Stage-Aware Optimization.** We divide optimization into two functional stages. *Stage A* focuses on global geometry formation. Projection consistency and volumetric regularization dominate this period, while Gaussian densification is kept conservative. The objective is not to recover fine structures immediately, but to establish a sufficiently stable attenuation geometry from which structural evidence can be meaningfully estimated.

*Stage B* progressively introduces frequency preservation, residual suppression, and G+P-guided Gaussian allocation. As the reconstructed volume becomes more reliable, topology evolution is allowed to respond increasingly strongly to the structure field. We define the stage transition using

$$s(\tau) = \mathrm{clip}\left(\frac{\tau - p_a}{p_b - p_a}, 0, 1\right), \qquad \tau = \frac{t}{T}, \tag{12}$$

where $T$ denotes the total number of optimization iterations, and $p_a$ and $p_b$ specify the beginning and completion of the transition. Thus, $s(\tau) = 0$ corresponds to the geometry-dominated regime, whereas $s(\tau) = 1$ corresponds to fully activated structure-aware refinement. The continuous transition avoids an abrupt change in the optimization landscape between the two stages.

The reconstruction loss at iteration $t$ is

$$\mathcal{L}_t = \lambda_{\text{proj}}\mathcal{L}_{\text{proj}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}} + \lambda_{\text{hf}}s(\tau)\mathcal{L}_{\text{hf}} + \lambda_r s(\tau)\mathcal{L}_r. \tag{13}$$

where $\lambda_p$, $\lambda_{\text{tv}}$, $\lambda_{\text{hf}}$, and $\lambda_r$ control the corresponding terms. The stage-dependent functions $s_r(\tau)$ suppress detail-oriented constraints during the unstable early phase and progressively strengthen them as the global geometry converges.

Continuous Gaussian parameters are optimized by gradient descent on $\mathcal{L}_t$, while clone, split, and prune constitute discrete topology updates performed according to the stage state and the local importance scores $\alpha_i$. The updated Gaussian field is then rendered again to produce $\hat{x}_{t+1}$ and $\hat{y}_{t+1}$, from which new residuals and structural cues are computed. SAGA therefore forms a closed reconstruction–allocation loop: the current reconstruction determines where capacity is required, adaptive Gaussian evolution reshapes the representation accordingly, and the stage-aware scheduler determines when this structural evidence is sufficiently reliable to be trusted.

# 4 Experiments

## 4.1 Experimental Setup

**Dataset and Projection Geometry.** We construct sparse-view CT benchmarks from the Open Scientific Visualization Datasets (Klacansky, 2022), covering five anatomical scenes: abdomen, chest, foot, head, and jaw. All volumes are center-cropped or interpolated to $128^3$ and min–max normalized to $[0, 1]$. Sparse projections are generated online using a differentiable cone-beam projector implemented with ODL/ASTRA (van Aarle et al., 2016). We use a circular trajectory around the $z$-axis, a $512 \times 512$ detector, and a $180^\circ$ angular range. Experiments are conducted with 6, 8, and 10 uniformly sampled training views over $[0^\circ, 180^\circ)$, while validation and test views are sampled independently using a fixed random seed.

**Implementation Details.** All methods use the same projections, acquisition geometry, detector resolution, and reconstruction grid. Gaussian-based methods are initialized from FBP, and each case is optimized independently using only the measured sparse projections; reference volumes are used only for evaluation. Unless otherwise specified, we use the chest 6-view setting for ablation studies. Training runs for 15,000 iterations with a batch size of 1, starting from 50,000 Gaussians with a maximum budget of 300,000 and densification every 100 iterations. The loss weights are $\lambda_{\text{proj}} = 1.0$, $\lambda_{\text{tv}} = 0.2$, $\lambda_{\text{hf}} = 0.02$, and $\lambda_{\text{r}} = 0.005$, while the G+P cue weights are $\alpha_g = 0.6$ and $\alpha_p = 0.9$. All experiments are conducted on a single NVIDIA GeForce RTX 4090 GPU, and the checkpoint with the best validation reconstruction quality is reported.

**Compared Methods and Metrics.** We compare SAGA with analytical and iterative methods, including FDK/FBP (Feldkamp et al., 1984; Kak & Slaney, 2001), SART (Andersen & Kak, 1984), and ASD-POCS (Sidky & Pan, 2008); neural representations, including IntraTomo (Zang et al., 2021), NeRF (Mildenhall et al., 2020), TensoRF (Chen et al., 2022), NAF (Zha et al., 2022), and SAX-NeRF (Cai et al., 2024b); and the explicit Gaussian baseline 3DGR (Li et al., 2025). Reconstruction quality is measured by 3D PSNR and SSIM on the normalized $128^3$ volumes. We additionally report PSNR and SSIM for novel-view projection synthesis as an auxiliary measure of projection-domain consistency, while volumetric metrics are used for the primary evaluation.

## 4.2 Main Reconstruction Results

Table 1 summarizes the average 3D reconstruction performance over the five anatomical scenes under 6, 8, and 10-view settings. SAGA achieves the highest average PSNR and SSIM across all three sampling densities, indicating consistent performance across different degrees of angular undersampling. The advantage is most pronounced in the 6-view setting, where SAGA improves the average PSNR from 29.65 dB to 31.09 dB and SSIM from 0.7972 to 0.8799 over the 3DGR baseline. This corresponds to gains of 1.44 dB and 0.0827, respectively, under the most severely under-constrained

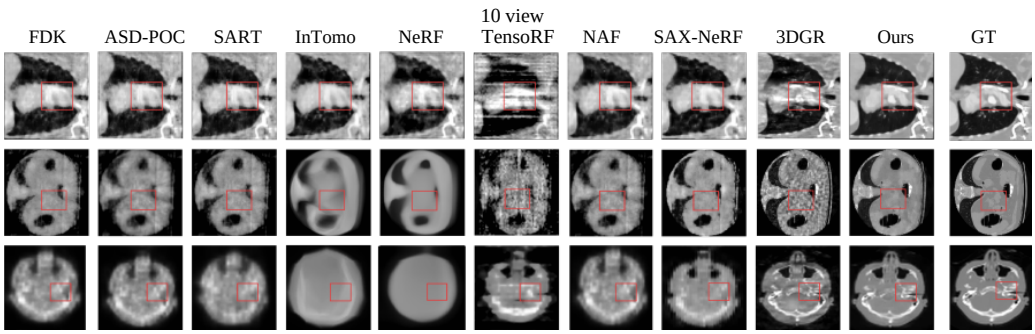


Figure 3: Qualitative comparison of 10-view sparse-view CT reconstruction on chest, abdomen, and head scenes. All methods are evaluated under the same acquisition setting.

Table 1: Average 3D reconstruction performance over five anatomical scenes under 6, 8, and 10-view settings. PSNR and SSIM are averaged over abdomen, chest, foot, head, and jaw. Best results are shown in bold with a light-red background, and second-best results are underlined.

| **Method** | **6-view** | | **8-view** | | **10-view** | |
|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 24.04 | 0.6324 | 25.35 | 0.6756 | 26.48 | 0.7166 |
| ASD-POCS | 24.13 | 0.6352 | 25.53 | 0.6842 | 26.57 | 0.7266 |
| SART | 23.87 | 0.6212 | 25.38 | 0.6848 | 26.11 | 0.7090 |
| InTomo | 22.98 | 0.5912 | 23.57 | 0.6180 | 24.12 | 0.6240 |
| NeRF | 23.45 | 0.6302 | 24.90 | 0.6670 | 25.15 | 0.6638 |
| TensoRF | 21.90 | 0.4294 | 23.07 | 0.4814 | 23.83 | 0.5248 |
| NAF | 24.01 | 0.6254 | 25.44 | 0.6874 | 26.56 | 0.7218 |
| SAX-NeRF | 25.07 | 0.6772 | 26.04 | 0.7094 | 26.94 | 0.7452 |
| 3DGR | <u>29.65</u> | <u>0.7972</u> | <u>32.81</u> | <u>0.8718</u> | <u>33.78</u> | <u>0.8848</u> |
| Ours | **31.09** | **0.8799** | **33.34** | **0.9128** | **34.64** | **0.9306** |

acquisition setting. With 8 and 10 views, the reconstruction problem becomes better constrained and the performance gap narrows, while SAGA retains the highest average volumetric accuracy. These results suggest that the proposed stabilization and allocation strategy is particularly effective when projection measurements provide insufficient constraints on the underlying 3D structure.

Fig. 3 presents qualitative comparisons across representative anatomical scenes. Analytical and iterative methods exhibit visible sparse-view streaks and boundary degradation, while neural-field methods tend to introduce structural smoothing or local inconsistencies. Compared with the explicit Gaussian baseline, SAGA recovers more coherent anatomical boundaries and better preserves local attenuation variations, consistent with the quantitative improvements.

Fig. 4 further compares SAX-NeRF, 3DGR, and SAGA under different view counts. As the number of projections decreases, ambiguity becomes increasingly evident around boundaries, thin structures, and weakly constrained regions. SAGA shows less structural degradation as the acquisition becomes sparser, supporting the role of geometry-frequency stabilization in constraining the reconstruction before structure-aware Gaussian refinement. The G+P allocation mechanism subsequently directs representation capacity toward regions supported jointly by volumetric structure and projection-domain evidence, reducing the tendency to over-refine artifact-dominated regions.

Complete per-anatomy results for the 6, 8, and 10-view settings are reported in Appendix Secs. A.1 and A.2, providing detailed comparisons beyond the cross-anatomy averages presented in the main text.

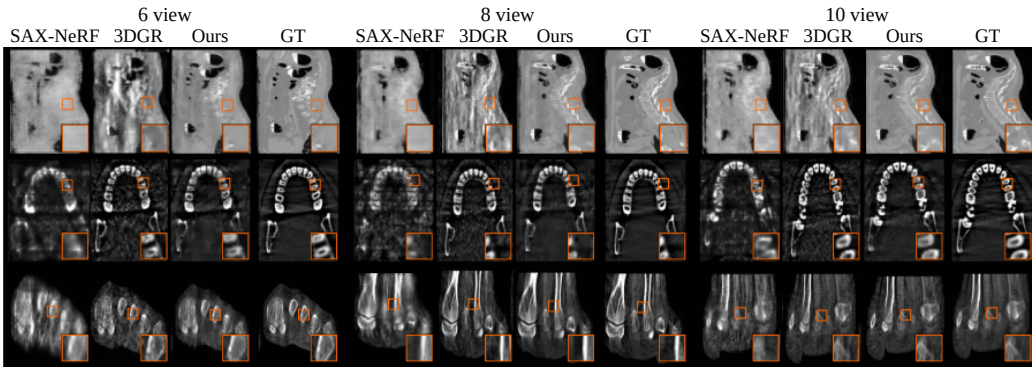


Figure 4: Qualitative comparison of SAX-NeRF, 3DGR, and SAGA on abdomen, jaw, and foot scenes under 6, 8, and 10-view acquisition settings.

Table 2: Main composition ablation on the chest 6-view setting. PSNR and SSIM are computed on the reconstructed volume.

| **Method** | **Obj.** | **G+P** | **Stage** | **PSNR↑** | **SSIM↑** |
|---|---|---|---|---|---|
| baseline | – | – | – | 26.06 | 0.7483 |
| + TV+HF+RED | ✓ | – | – | 30.76 | 0.8956 |
| + G+P allocation | ✓ | ✓ | – | 30.89 | 0.8976 |
| + Two-stage scheduler | ✓ | – | ✓ | 31.09 | 0.8988 |
| Full SAGA | ✓ | ✓ | ✓ | **31.15** | **0.9013** |

### 4.3 Ablation Studies

All ablations are conducted on the chest 6-view setting using the same optimization protocol, and PSNR/SSIM are computed on the reconstructed volume.

**Main Composition.** Table 2 evaluates the three main components of SAGA: the geometry-frequency objective, G+P structure-aware allocation, and stage-aware scheduling. The reconstruction objective provides the dominant improvement over the 3DGR baseline, confirming that projection consistency alone is insufficient under severe angular undersampling. Introducing G+P allocation further improves reconstruction by coupling local Gaussian capacity with volumetric and projection-domain structural evidence. Stage-aware scheduling provides an additional gain by delaying aggressive topology evolution until the reconstruction becomes sufficiently stable. The full model combines these complementary effects and achieves the best performance.

Additional analyses of the allocation scheduler, loss composition, regularization sensitivity, densification strategies, and structure-cue configurations are provided in Appendix Sec. A.3.

## 5 Conclusion

This work presents SAGA, a stage-stabilized, structure-aware Gaussian allocation framework for explicit Gaussian-based sparse-view CT reconstruction. We identify that pure projection-based Gaussian optimization is under-constrained, unable to guarantee reliable volumetric structures from sparse measurements. SAGA resolves this issue via a TV-stabilized, frequency-constrained objective and scheduled structure-guided Gaussian adjustment operations with two-stage stabilization. Experimental results validate that SAGA outperforms vanilla Gaussian methods in volumetric fidelity and structural preservation. The geometry-frequency optimization serves as the core performance booster, while structure-aware allocation and staged stabilization further improve Gaussian placement reliability. This study verifies the efficacy of structure-aware Gaussian optimization for interpretable self-supervised sparse-view CT reconstruction.

## REFERENCES


Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.

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.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6343–6352, 2018. doi: 10.1109/CVPR.2018.00664.

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 *European Conference on Computer Vision*, pp. 283–299, 2024a.

Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view X-ray 3D reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2024b.

Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial radiance fields. In *Computer Vision – ECCV 2022*, volume 13692 of *Lecture Notes in Computer Science*, pp. 333–350, Cham, 2022. Springer. doi: 10.1007/978-3-031-19824-3_20.

Guang-Hong Chen, Jie Tang, and Shuai Leng. Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical Physics*, 35(2):660–663, 2008.

Hu Chen, Yi Zhang, Mannudeep K. Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017.

Hu Chen, Yi Zhang, Yunjin Chen, Jian Zhang, Weihua Zhang, Huan Sun, Yixing Lv, Peixi Liao, Jiliu Zhou, and Ge Wang. LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE Transactions on Medical Imaging*, 37(6):1333–1347, 2018.

David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

Lee A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984.

Xueming Fu, Yingtai Li, Fenghe Tang, Jun Li, Mingyue Zhao, Gao-Jun Teng, and S. Kevin Zhou. 3DGR-CAR: Coronary artery reconstruction from ultra-sparse 2D X-ray views with a 3D Gaussians representation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 14–24, 2024.

Richard Gordon, Robert Bender, and Gabor T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3):471–481, 1970.

Yoseob Han and Jong Chul Ye. Framing U-net via deep convolutional framelets: Application to sparse-view CT. *IEEE Transactions on Medical Imaging*, 37(6):1418–1429, 2018. doi: 10.1109/TMI.2018.2823768.

Allard A. Hendriksen, Daniël M. Pelt, and K. Joost Batenburg. Noise2Inverse: Self-supervised deep convolutional denoising for tomography. *IEEE Transactions on Computational Imaging*, 6:1320–1335, 2020.

Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9): 4509–4522, 2017.

Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023.

Pavol Klacansky. Open scientific visualization datasets, 2022. URL `https://klacansky.com/open-scivis-datasets/`.

Hiroyuki Kudo, Takuya Suzuki, and Essam A. Rashed. Image reconstruction for sparse-view CT and interior CT: Introduction to compressed sensing and differentiated backprojection. *Quantitative Imaging in Medicine and Surgery*, 3(3):147–161, 2013.

Donghoon Lee, Sunghoon Choi, and Hee-Joung Kim. High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains. *Medical Physics*, 46(1):104–115, 2019a. doi: 10.1002/mp.13258.

Hoyeon Lee, Jongha Lee, Hyeongseok Kim, Byungchul Cho, and Seungryong Cho. Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2):109–119, 2019b. doi: 10.1109/TRPMS.2018.2867611.

Yingtai Li, Xueming Fu, Han Li, Shang Zhao, Ruiyang Jin, and S. Kevin Zhou. 3DGR-CT: Sparse-view CT reconstruction with a 3D Gaussian representation. *Medical Image Analysis*, 103:103585, 2025.

Yinsheng Li, Ke Li, Chengzhu Zhang, Juan Montoya, and Guang-Hong Chen. Learning to reconstruct computed tomography images directly from sinogram data under a variety of data acquisition conditions. *IEEE Transactions on Medical Imaging*, 38(10):2469–2481, 2019. doi: 10.1109/TMI.2019.2910760.

Yiqun Lin, Jiewen Yang, Hualiang Wang, Xinpeng Ding, Wei Zhao, and Xiaomeng Li. $C^2RV$: Cross-regional and cross-view learning for sparse-view CBCT reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11205–11214, 2024.

Yan Liu, Zhengrong Liang, Jianhua Ma, Hongbing Lu, Ke Wang, Hao Zhang, and William Moore. Total variation-stokes strategy for sparse-view X-ray CT image reconstruction. *IEEE Transactions on Medical Imaging*, 33(3):749–763, 2014. doi: 10.1109/TMI.2013.2295738.

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 *European Conference on Computer Vision*, pp. 405–421, 2020.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):102:1–102:15, 2022.

Shanzhou Niu, Yang Gao, Zhaoying Bian, Jing Huang, Wufan Chen, Gaohang Yu, Zhengrong Liang, and Jianhua Ma. Sparse-view X-ray CT reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 59(12):2997–3017, 2014. doi: 10.1088/0031-9155/59/12/2997.

Daniel Otero Baguer, Johannes Leuschner, and Maximilian Schmidt. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, 36(9): 094004, 2020. doi: 10.1088/1361-6420/aba415.

Jinhui Pan, Hongming Zhang, Wenxiang Wu, Zhiqiang Gao, and Wei Wu. Multi-domain integrative Swin transformer network for sparse-view tomographic reconstruction. *Computers in Biology and Medicine*, 141:105139, 2022.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

Darius Rückert, Yuanhao Wang, Rui Li, Ramzi Idoughi, and Wolfgang Heidrich. NeAT: Neural adaptive tomography. *ACM Transactions on Graphics*, 41(4):55:1–55:13, 2022.

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.

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.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, 2020.

Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S. Kamilov. CoIL: Coordinate-based internal learning for tomographic imaging. *IEEE Transactions on Computational Imaging*, 7:1400–1412, 2021. doi: 10.1109/TCI.2021.3125564.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, 2020.

Xia-Yu Tao, Qi-Si Lin, Zhao Wu, Yong Guan, Yang-Chao Tian, and Gang Liu. Sparse-view phase-contrast and attenuation-based CT reconstruction utilizing model-driven deep learning. *Nuclear Science and Techniques*, 36(4):71, 2025. doi: 10.1007/s41365-025-01642-2.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9446–9454, 2018.

Wim van Aarle, Willem Jan Palenstijn, Jan De Beenhouwer, Thomas Altantzis, Sara Bals, K. Joost Batenburg, and Jan Sijbers. Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics Express*, 24(22):25129–25147, 2016.

Jonas Van der Rauwelaert, Caroline Bossuyt, Stijn E. Verleden, et al. SwinIR-based dual-domain reconstruction for sparse-view computed tomography. *Journal of Nondestructive Evaluation*, 44: 101, 2025. doi: 10.1007/s10921-025-01244-3.

Ge Wang, Jong Chul Ye, Klaus Mueller, and Jeffrey A. Fessler. Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging*, 37(6):1289–1296, 2018.

Jiaying Yang and Shipeng Xie. PE-INeR: Prior-embedded implicit neural representation for sparse-view CBCT reconstruction. *Applied Optics*, 63(35):8907–8916, 2024. doi: 10.1364/AO.534352.

Weihao Yu, Yuanhao Cai, Ruyi Zha, Zhiwen Fan, Chenxin Li, and Yixuan Yuan. $X^2$-Gaussian: 4D radiative gaussian splatting for continuous-time tomographic reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 24728–24738, 2025.

Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: Self-supervised learning-based tomography via sinogram synthesis and prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1960–1970, 2021.

Ruyi Zha, Yanhao Zhang, and Hongdong Li. NAF: Neural attenuation fields for sparse-view CBCT reconstruction. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 442–452, 2022.

Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. $R^2$-Gaussian: Rectifying radiative Gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems*, 2024.

Zhicheng Zhang, Xiaokun Liang, Xiaowei Dong, Yaoqin Xie, and Guohua Cao. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Transactions on Medical Imaging*, 37(6):1407–1417, 2018.

# A Appendix

## A.1 Complete Volumetric Reconstruction Results

The main paper reports the average volumetric reconstruction performance over the five anatomical scenes. Here, we provide the complete per-anatomy results under 6, 8, and 10-view acquisition settings for a more detailed comparison.

Tables 3, 4, and 5 report the 3D PSNR and SSIM for abdomen, chest, foot, head, and jaw. SAGA achieves the highest average PSNR and SSIM under all three view settings. The improvement over 3DGR is most evident in the 6-view case, where severe angular undersampling leaves the reconstruction strongly under-constrained. As the number of projections increases, the performance gap narrows because the additional measurements provide stronger geometric constraints, while SAGA maintains the best average reconstruction accuracy.

The detailed results also reveal anatomy-dependent behavior. SAGA provides particularly clear gains on abdomen, chest, and head, where sparse angular sampling introduces substantial structural ambiguity. On foot and jaw, 3DGR remains competitive in several settings, indicating that the benefit of structure-aware allocation depends on the geometric complexity and sampling characteristics of individual anatomies. Nevertheless, the consistent improvement in the cross-anatomy average shows that SAGA provides a more robust reconstruction strategy across different sparse-view configurations.

Fig. 5, 6, and 7 show the corresponding qualitative comparisons. Under 6-view acquisition, analytical and iterative methods exhibit pronounced streak artifacts and degraded boundaries, while neural-field and Gaussian baselines still show local structural inconsistencies. SAGA preserves more coherent boundaries and internal attenuation structures, with the difference being most visible in the severely undersampled setting. As the number of views increases to 8 and 10, reconstruction quality improves for all methods and the visual differences become less pronounced, consistent with the quantitative results.

## A.2 Complete Novel-View Projection Synthesis Results

The main paper uses novel-view projection synthesis as an auxiliary evaluation of the reconstructed representation. Unlike volumetric PSNR and SSIM, this experiment does not directly measure the accuracy of the recovered 3D attenuation field. Instead, it examines whether a representation optimized from sparse measured views can produce consistent projections at previously unseen angles.

Here, we provide the complete quantitative and qualitative results omitted from the main paper for space reasons. Table 6 reports the full comparison, while Fig. 8–10 show representative unseen-view projections for all five anatomical scenes.

The results follow a similar trend to the volumetric evaluation. Under the 6-view setting, several methods reproduce the observed measurements reasonably well but become less stable when rendered from unobserved directions. Typical failure modes include blurred boundaries, residual streaks, and locally inconsistent projection patterns. These effects become less pronounced as the number of training views increases.

SAGA produces more consistent unseen-view projections across the evaluated view settings. This behavior suggests that the optimized Gaussian field is not restricted to reproducing the measured projection directions and retains useful geometric consistency when queried from new angles. The result is particularly relevant in the ultra-sparse regime, where a low training-view projection error alone does not rule out wrong-depth or view-dependent structures.

We emphasize that novel-view synthesis remains a complementary evaluation. A reconstruction may generate plausible projections at selected unseen angles while still containing volumetric errors. Accordingly, the conclusions of the paper are based primarily on 3D reconstruction quality, with novel-view projection synthesis serving as additional evidence of representation consistency.

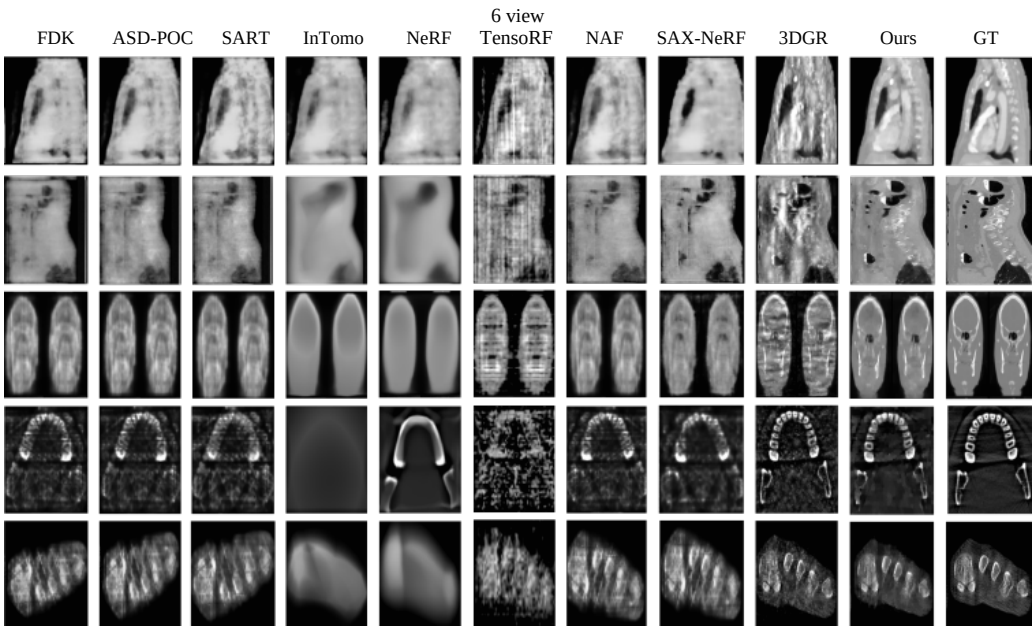


Figure 5: Complete qualitative comparison under the 6-view setting. The rows correspond to chest, abdomen, head, jaw, and foot scenes, while the columns compare analytical, iterative, neural-field, Gaussian-based, and reference reconstructions. The extremely sparse measurements lead to pronounced streaks and structural ambiguity for several baselines, whereas SAGA preserves comparatively coherent boundaries and internal structures.

Table 3: Quantitative comparison of 3D sparse-view CT reconstruction under the 6-view setting. PSNR and SSIM are computed on the reconstructed volume. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 6-view | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Abdomen | | Chest | | Foot | | Head | | Jaw | |
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 21.00 | 0.4600 | 22.15 | 0.6030 | 24.88 | 0.8220 | 25.60 | 0.6760 | 26.57 | 0.6010 |
| ASD-POCS | 21.15 | 0.4640 | 22.29 | 0.6060 | 25.03 | 0.8210 | 25.59 | 0.6800 | 26.60 | 0.6050 |
| SART | 21.00 | 0.4550 | 22.06 | 0.6000 | 24.89 | 0.8220 | 25.45 | 0.6710 | 25.96 | 0.5580 |
| InTomo | 21.86 | 0.5220 | 20.83 | 0.5680 | 22.94 | 0.7220 | 25.09 | 0.6330 | 24.18 | 0.5110 |
| NeRF | 22.27 | 0.5490 | 19.97 | 0.5540 | 23.08 | 0.7500 | 25.22 | 0.6540 | 26.70 | 0.6440 |
| TensoRF | 19.77 | 0.2940 | 19.27 | 0.4030 | 23.38 | 0.6700 | 23.68 | 0.4760 | 23.39 | 0.3040 |
| NAF | 21.35 | 0.4730 | 21.10 | 0.5440 | 25.26 | 0.8270 | 25.78 | 0.6830 | 26.55 | 0.6000 |
| SAX-NeRF | 23.68 | 0.6130 | 23.72 | 0.6580 | 25.34 | 0.8300 | 26.07 | 0.6810 | 26.55 | 0.6040 |
| 3DGR | 23.91 | 0.6330 | 24.42 | 0.6990 | **35.29** | **0.9540** | 29.92 | 0.8210 | **34.72** | **0.8790** |
| Ours | **28.30** | **0.8267** | **31.07** | **0.9000** | 32.89 | 0.9365 | **31.52** | **0.9107** | 31.66 | 0.8256 |

### A.3 Additional Ablation Studies

The main paper studies the contribution of the reconstruction objective, G+P-guided allocation, and the two-stage scheduler. We provide additional experiments below to examine several design choices in greater detail. Unless otherwise specified, all experiments use the chest 6-view setting.

**Allocation Scheduler.** Table 10 examines the effect of stage-aware Gaussian evolution. We further examine the effect of stage-aware Gaussian evolution. Activating all allocation mechanisms from the beginning exposes early densification to unreliable gradients and residual structures. The two-stage schedule instead restricts topology evolution during the initial geometry-formation phase

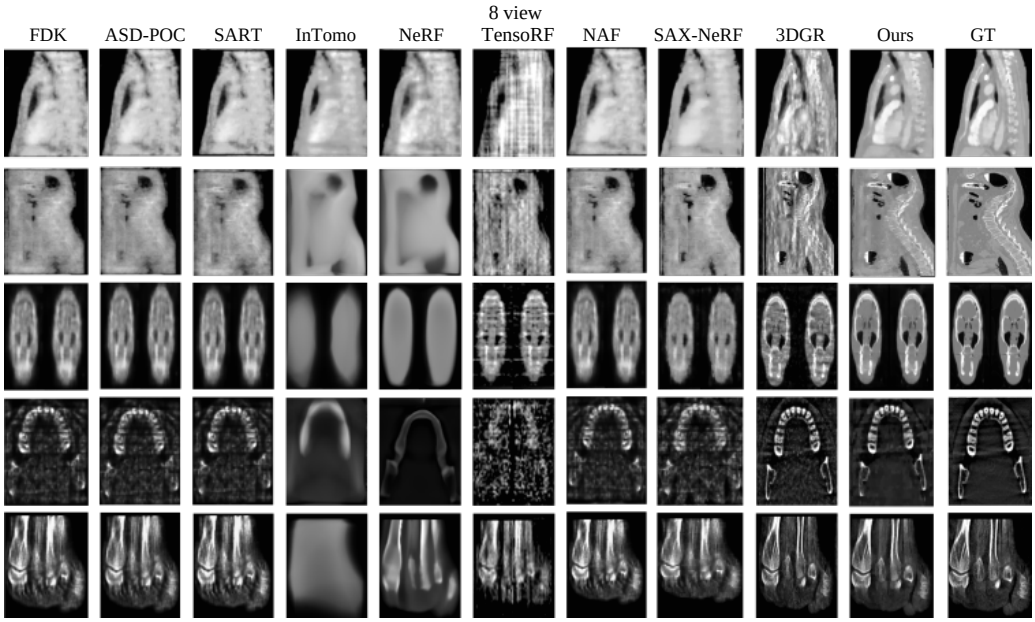


Figure 6: Complete qualitative comparison under the 8-view setting. The rows correspond to chest, abdomen, head, jaw, and foot scenes. With additional projection measurements, most methods recover more stable structures than in the 6-view setting, while SAGA continues to show consistent anatomical boundaries across the evaluated scenes.

Table 4: Quantitative comparison of 3D sparse-view CT reconstruction under the 8-view setting. PSNR and SSIM are computed on the reconstructed volume. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 8-view | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Abdomen | | Chest | | Foot | | Head | | Jaw | |
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 22.54 | 0.5070 | 25.02 | 0.7050 | 25.81 | 0.8440 | 26.34 | 0.6990 | 27.05 | 0.6230 |
| ASD-POCS | 22.41 | 0.5130 | 24.92 | 0.7000 | 26.15 | 0.8490 | 26.91 | 0.7230 | 27.28 | 0.6360 |
| SART | 23.30 | 0.5520 | 24.73 | 0.7020 | 25.27 | 0.8390 | 26.49 | 0.7080 | 27.10 | 0.6230 |
| InTomo | 22.90 | 0.5790 | 23.75 | 0.6670 | 23.55 | 0.7650 | 22.83 | 0.5500 | 24.79 | 0.5290 |
| NeRF | 22.32 | 0.5400 | 22.76 | 0.6210 | 26.62 | 0.8480 | 24.68 | 0.6160 | 28.10 | 0.7100 |
| TensoRF | 21.21 | 0.3680 | 21.56 | 0.4860 | 24.33 | 0.7250 | 25.18 | 0.5510 | 23.09 | 0.2770 |
| NAF | 23.29 | 0.5490 | 24.20 | 0.6850 | 25.60 | 0.8440 | 26.69 | 0.7170 | 27.42 | 0.6420 |
| SAX-NeRF | 24.62 | 0.6500 | 24.84 | 0.6990 | 26.90 | 0.8610 | 27.16 | 0.7400 | 26.66 | 0.5970 |
| 3DGR | 27.94 | 0.7620 | 29.33 | 0.8310 | **37.00** | **0.9680** | 32.87 | 0.8790 | **36.93** | **0.9190** |
| Ours | **29.80** | **0.8644** | **34.76** | **0.9420** | 34.82 | 0.9512 | **34.37** | **0.9484** | 32.95 | 0.8581 |

and enables structure-guided refinement only after the reconstruction has stabilized. This leads to more reliable primitive allocation and reduces premature densification in artifact-dominated regions. Fig. 11 focuses on the timing of Gaussian evolution. When all allocation mechanisms are activated from the beginning, early reconstruction errors can directly influence Gaussian growth. The two-stage strategy instead gives the representation an initial period to stabilize its coarse geometry before structure-guided allocation becomes fully active. The resulting convergence is more stable and reaches a slightly better final solution. This supports the use of the scheduler as a stabilization mechanism rather than as an independent source of reconstruction capacity.

**Loss Composition.** Table 11 isolates the contribution of the individual reconstruction constraints. Projection consistency alone yields poor volumetric reconstruction because the limited measure-

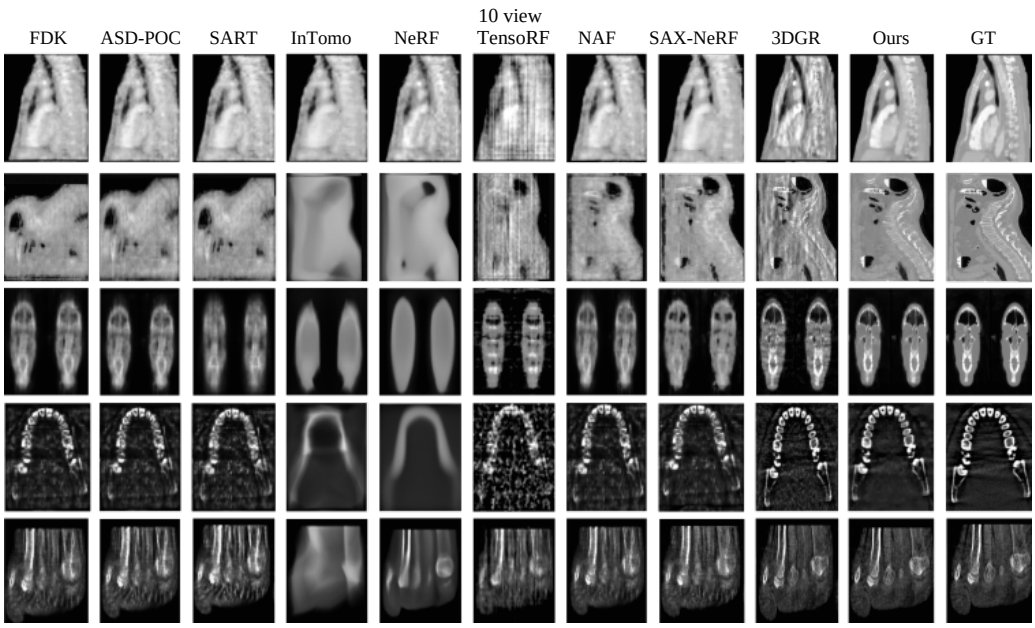


Figure 7: Complete qualitative comparison under the 10-view setting. The rows correspond to chest, abdomen, head, jaw, and foot scenes. The reduced angular ambiguity improves reconstruction quality for all methods, while SAGA maintains stable structural recovery across anatomies.

Table 5: Quantitative comparison of 3D sparse-view CT reconstruction under the 10-view setting. PSNR and SSIM are computed on the reconstructed volume. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 10-view | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Abdomen | | Chest | | Foot | | Head | | Jaw | |
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 23.88 | 0.5810 | 26.32 | 0.7470 | 27.21 | 0.8700 | 27.21 | 0.7340 | 27.77 | 0.6510 |
| ASD-POCS | 24.41 | 0.6060 | 26.36 | 0.7520 | 26.52 | 0.8630 | 27.58 | 0.7480 | 27.99 | 0.6640 |
| SART | 23.81 | 0.5790 | 26.52 | 0.7570 | 25.58 | 0.8480 | 27.22 | 0.7360 | 27.40 | 0.6250 |
| InTomo | 23.09 | 0.5700 | 25.79 | 0.7360 | 23.96 | 0.7650 | 22.82 | 0.5140 | 24.93 | 0.5350 |
| NeRF | 22.14 | 0.5420 | 24.73 | 0.6910 | 28.99 | 0.8810 | 25.06 | 0.6680 | 24.81 | 0.5370 |
| TensoRF | 21.27 | 0.3610 | 21.31 | 0.4750 | 25.92 | 0.7920 | 26.27 | 0.6250 | 24.41 | 0.3710 |
| NAF | 24.16 | 0.5980 | 26.00 | 0.7460 | 27.55 | 0.8770 | 27.36 | 0.7390 | 27.72 | 0.6490 |
| SAX-NeRF | 25.36 | 0.6910 | 25.83 | 0.7350 | 27.90 | 0.8740 | 27.70 | 0.7610 | 27.93 | 0.6650 |
| 3DGR | 28.32 | 0.7760 | 29.87 | 0.8400 | **38.79** | **0.9760** | 33.69 | 0.8950 | **38.24** | **0.9370** |
| Ours | **32.07** | **0.9021** | **35.31** | **0.9484** | 35.95 | 0.9586 | **35.59** | **0.9579** | 34.27 | 0.8860 |

ments leave a large set of projection-compatible solutions. TV regularization provides the largest improvement by stabilizing the global attenuation geometry. The high-frequency and residual terms subsequently improve local structural preservation and suppress inconsistent projection-domain variations. Their combination therefore balances global geometric stability with local detail recovery. Fig. 12 compares the convergence behavior of different loss compositions. Projection-only optimization provides a weak constraint on the 3D solution and converges to a substantially lower-quality reconstruction. Introducing the structural regularization terms progressively raises the attainable reconstruction quality and produces a more stable optimization trajectory. This observation is consistent with the loss ablation reported in the main paper, where the reconstruction objective accounts for the largest performance improvement.

Table 6: Average novel-view projection synthesis results over five anatomical scenes. Each entry reports the average PSNR and SSIM at unseen view angles. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 6-view | | 8-view | | 10-view | |
|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 15.63 | 0.5889 | 13.36 | 0.5606 | 16.60 | 0.6751 |
| ASD-POCS | 13.16 | 0.5737 | 14.62 | 0.6183 | 16.41 | 0.6432 |
| SART | 12.48 | 0.5576 | 13.90 | 0.5875 | 16.13 | 0.6556 |
| InTomo | 20.11 | 0.8253 | 19.54 | 0.8495 | 21.25 | 0.8708 |
| NeRF | 20.86 | 0.7002 | 18.57 | 0.7101 | 20.82 | 0.7214 |
| TensoRF | 22.88 | 0.3952 | 23.65 | 0.4230 | 25.32 | 0.4458 |
| NAF | 12.36 | 0.5433 | 13.90 | 0.6023 | 15.37 | 0.6278 |
| SAX-NeRF | 35.61 | 0.9852 | 35.15 | **0.9968** | 36.46 | 0.9919 |
| 3DGR | 37.34 | 0.9862 | 39.96 | 0.9918 | 42.38 | 0.9933 |
| Ours | **39.93** | **0.9892** | **42.85** | 0.9925 | **44.98** | **0.9940** |

Table 7: Novel-view projection synthesis comparison under the 6-view setting. PSNR and SSIM are computed between rendered projections and reference projections at unseen view angles. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 6-view | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Abdomen | | Chest | | Foot | | Head | | Jaw | |
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 12.73 | 0.1375 | 8.42 | 0.3854 | 34.21 | **0.9996** | 15.07 | 0.6894 | 7.70 | 0.7327 |
| ASD-POCS | 11.32 | 0.1351 | 9.44 | 0.4055 | 22.47 | 0.9169 | 15.16 | 0.6968 | 7.40 | 0.7143 |
| SART | 11.35 | 0.1666 | 12.03 | 0.4648 | 20.37 | 0.9138 | 13.15 | 0.6382 | 5.49 | 0.6045 |
| InTomo | 24.68 | 0.8144 | 7.54 | 0.3454 | 23.67 | 0.9852 | 25.36 | 0.9885 | 19.32 | 0.9929 |
| NeRF | 22.97 | 0.8783 | 9.56 | 0.1589 | 28.18 | 0.8170 | 23.09 | 0.8529 | 20.49 | 0.7938 |
| TensoRF | 17.11 | 0.1619 | 17.35 | 0.2247 | 32.60 | 0.8101 | 28.52 | 0.5523 | 18.82 | 0.2269 |
| NAF | 13.18 | 0.1415 | 8.53 | 0.3660 | 19.73 | 0.9079 | 14.50 | 0.6753 | 5.86 | 0.6260 |
| SAX-NeRF | 38.36 | **0.9993** | 30.64 | 0.9293 | 40.57 | 0.9996 | 38.37 | **0.9992** | 30.12 | **0.9988** |
| 3DGR | 32.40 | 0.9756 | 36.37 | 0.9860 | **47.21** | 0.9951 | 32.77 | 0.9905 | **37.92** | 0.9837 |
| Ours | **39.98** | 0.9888 | **41.75** | **0.9958** | 43.39 | 0.9927 | **40.59** | 0.9954 | 33.92 | 0.9732 |

**TV Regularization Sensitivity** TV regularization provides an important stabilizing constraint when the projection measurements are too sparse to determine a unique volumetric solution. We therefore evaluate the effect of its weight while keeping the remaining HF and RED terms enabled.

Table 12 shows a clear dependence on the TV constraint. Without TV regularization, projection consistency together with the remaining terms is insufficient to recover a stable volume, resulting in substantially lower PSNR and SSIM. Increasing the TV weight progressively improves reconstruction quality, with $\lambda_{\mathrm{tv}} = 0.20$ achieving the best result within the tested range.

This behavior is also reflected in the convergence curves in Fig. 13. Weak or absent TV regularization leads to slower and less stable recovery, whereas the selected setting reaches a higher-quality solution more consistently. These results indicate that, under extreme angular undersampling, TV primarily serves to stabilize the global attenuation structure before finer structural constraints become effective.

**Densification Strategy** The previous experiment concerns the reconstruction objective. We next examine how the Gaussian representation should grow once a sufficiently stable volume has been

Table 8: Novel-view projection synthesis comparison under the 8-view setting. PSNR and SSIM are computed between rendered projections and reference projections at unseen view angles. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 8-view | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Abdomen | | Chest | | Foot | | Head | | Jaw | |
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 12.60 | 0.1800 | 12.06 | 0.4966 | 25.01 | 0.9410 | 12.22 | 0.5982 | 4.89 | 0.5872 |
| ASD-POCS | 13.24 | 0.1898 | 11.38 | 0.4669 | 24.53 | 0.9370 | 15.72 | 0.7437 | 8.26 | 0.7541 |
| SART | 12.41 | 0.1589 | 6.80 | 0.3699 | 25.13 | 0.9385 | 14.43 | 0.6850 | 10.72 | 0.7849 |
| InTomo | 17.75 | 0.8401 | 11.12 | 0.4752 | 24.73 | 0.9911 | 26.69 | 0.9905 | 17.41 | 0.9506 |
| NeRF | 17.60 | 0.8072 | 8.19 | 0.1918 | 24.38 | 0.8565 | 23.45 | 0.8809 | 19.22 | 0.8138 |
| TensoRF | 17.18 | 0.1685 | 19.86 | 0.2829 | 33.49 | 0.8394 | 27.69 | 0.4614 | 20.02 | 0.3625 |
| NAF | 12.34 | 0.2016 | 11.92 | 0.4838 | 24.11 | 0.9341 | 15.92 | 0.7487 | 5.20 | 0.6434 |
| SAX-NeRF | 35.45 | **0.9994** | 33.52 | 0.9862 | 38.21 | **0.9997** | 37.11 | **0.9995** | 31.44 | **0.9991** |
| 3DGR | 33.70 | 0.9884 | 37.45 | 0.9931 | **48.00** | 0.9957 | 39.35 | 0.9931 | **41.31** | 0.9886 |
| Ours | **43.23** | 0.9932 | **46.13** | **0.9969** | 43.89 | 0.9945 | **44.82** | 0.9962 | 36.19 | 0.9817 |

Table 9: Novel-view projection synthesis comparison under the 10-view setting. PSNR and SSIM are computed between rendered projections and reference projections at unseen view angles. Best results are shown in bold with a light-red background, and second-best results are underlined.

| Method | 10-view | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Abdomen | | Chest | | Foot | | Head | | Jaw | |
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| FDK | 12.98 | 0.2150 | 18.82 | 0.7392 | 28.31 | 0.9845 | 15.11 | 0.7232 | 7.77 | 0.7138 |
| ASD-POCS | 14.14 | 0.1635 | 13.49 | 0.5390 | 28.95 | 0.9940 | 14.74 | 0.7164 | 10.75 | 0.8030 |
| SART | 13.81 | 0.2129 | 13.71 | 0.5372 | 28.44 | 0.9960 | 15.80 | 0.7744 | 8.89 | 0.7575 |
| InTomo | 16.40 | 0.7761 | 15.19 | 0.6100 | 24.49 | 0.9821 | 26.74 | 0.9951 | 23.44 | 0.9907 |
| NeRF | 17.71 | 0.8365 | 11.08 | 0.2982 | 29.27 | 0.9151 | 22.87 | 0.8086 | 23.16 | 0.7488 |
| TensoRF | 19.21 | 0.1902 | 21.11 | 0.3032 | 34.54 | 0.7784 | 28.09 | 0.5866 | 23.68 | 0.3704 |
| NAF | 13.90 | 0.1721 | 9.51 | 0.4361 | 25.03 | 0.9377 | 17.18 | 0.7916 | 11.22 | 0.8016 |
| SAX-NeRF | 38.01 | **0.9995** | 31.26 | 0.9614 | 43.06 | **0.9997** | 32.08 | **0.9995** | 37.88 | **0.9995** |
| 3DGR | 36.19 | 0.9895 | 38.15 | 0.9937 | **52.28** | 0.9971 | 42.72 | 0.9945 | **42.57** | 0.9918 |
| Ours | **44.07** | 0.9947 | **47.70** | **0.9973** | 47.97 | 0.9959 | **47.34** | 0.9972 | 37.81 | 0.9851 |

established. This question is important because sparse-view projection gradients do not always indicate reliable 3D structure: a large gradient may arise from a true anatomical boundary, but it may also be caused by an undersampling artifact or depth ambiguity.

As shown in Table 13, allowing the representation to densify is substantially more important than keeping a fixed Gaussian set. Random and gradient-based densification already recover most of this gain, which indicates that the strong reconstruction objective provides the dominant source of stabilization.

The G+P-guided strategy gives a further, smaller improvement over these generic alternatives. Although the numerical margin is modest, it is consistent with the intended role of the allocation field: it does not replace the reconstruction objective, but provides additional evidence for deciding where new Gaussian capacity should be introduced. By combining volumetric gradient information with

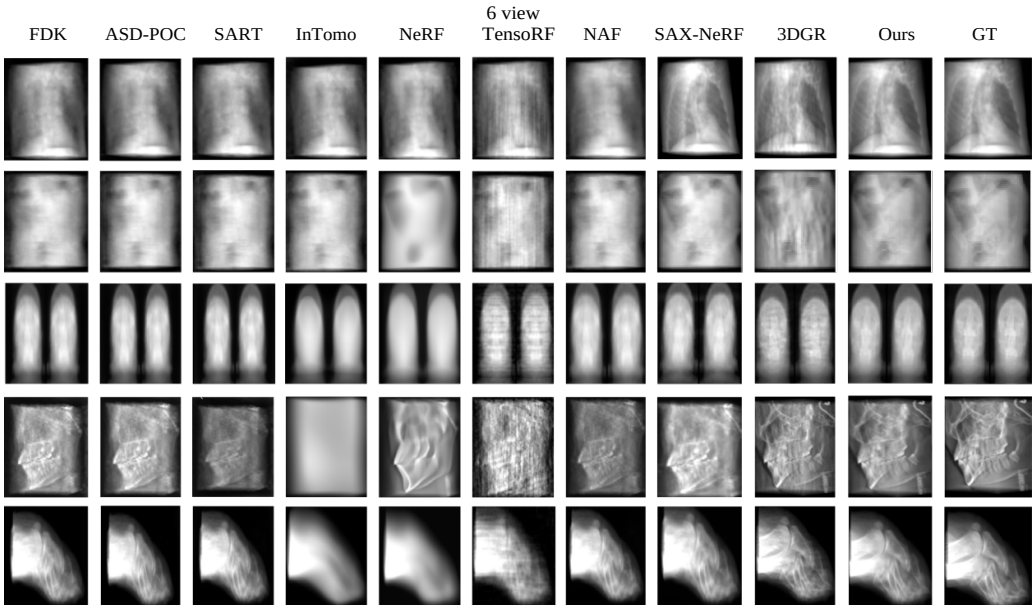


Figure 8: Complete novel-view projection synthesis comparison under the 6-view training setting. The rows show chest, abdomen, head, jaw, and foot examples. SAGA produces comparatively stable projection structures at unseen angles despite the severe angular undersampling during optimization.

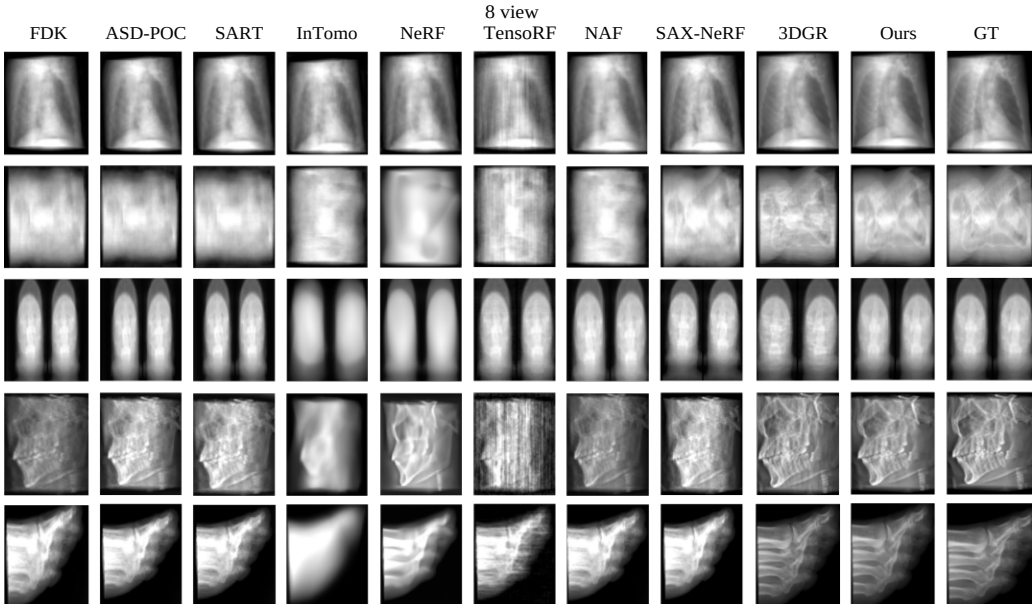


Figure 9: Complete novel-view projection synthesis comparison under the 8-view training setting. Increasing the number of measured views reduces view-dependent artifacts for most methods, while SAGA maintains consistent projection structures across the evaluated anatomies.

residual-backprojection evidence, the allocation rule is less dependent on a single local optimization signal.

**Structure-Cue Source** Having established the effect of structure-guided densification, we further separate the two sources used to construct the allocation field. The gradient cue $G$ describes local structural variation in the reconstructed volume, whereas the projection cue $P$ reflects residual evidence propagated back from the measured views. Their roles are therefore related but not identical.

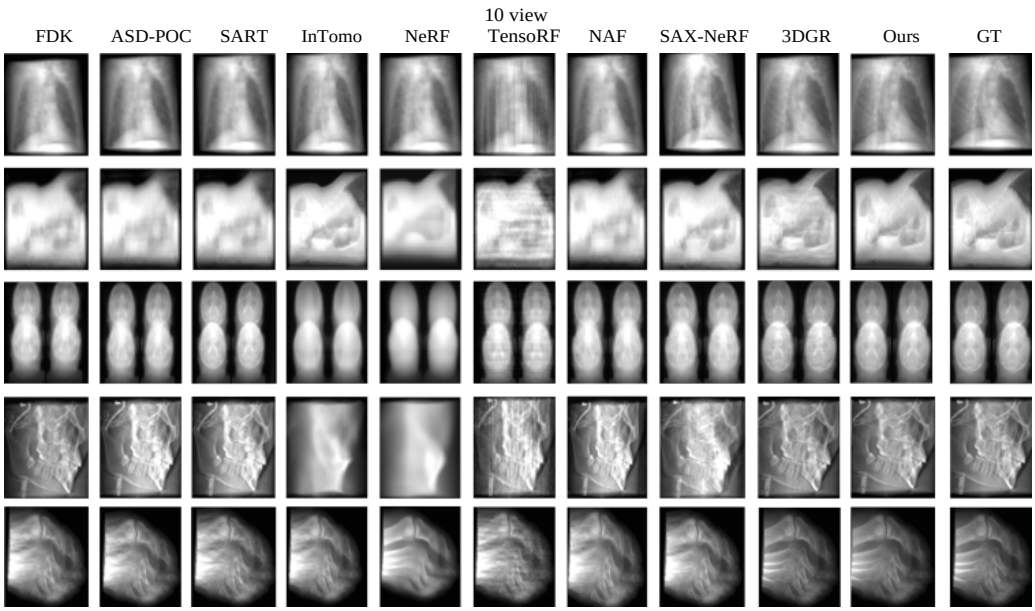


Figure 10: Complete novel-view projection synthesis comparison under the 10-view training setting. With stronger angular constraints, all methods benefit from improved geometric consistency, while SAGA remains stable across different anatomical scenes.

Table 10: Scheduler design (chest, 6-view).

| Scheduler | PSNR | SSIM |
|---|---|---|
| All-on | 30.74 | 0.8947 |
| Two-stage | **31.14** | **0.9008** |

Table 11: Loss composition (chest, 6-view). Terms are added incrementally.

| Loss | PSNR | SSIM |
|---|---|---|
| Projection | 23.45 | 0.6821 |
| +TV | 30.75 | 0.8925 |
| +RED | 30.90 | 0.8954 |
| +HF | 31.04 | 0.8990 |
| +RED+HF | **31.20** | **0.9012** |

Table 14 shows that either cue alone provides useful allocation information. The gradient cue performs slightly better in this experiment, suggesting that local volumetric structure is already a strong indicator once the reconstruction has been stabilized. The projection cue, however, contributes complementary measurement-domain information.

Combining the two cues gives the best PSNR and SSIM. The gain over the single-cue variants is not large, but it is consistent with the purpose of the G+P field: the volumetric cue identifies candidate structural regions, while the residual-backprojection cue provides additional evidence about whether those regions remain insufficiently explained by the measured projections. The fused field therefore offers a more balanced signal for Gaussian allocation than either source alone.

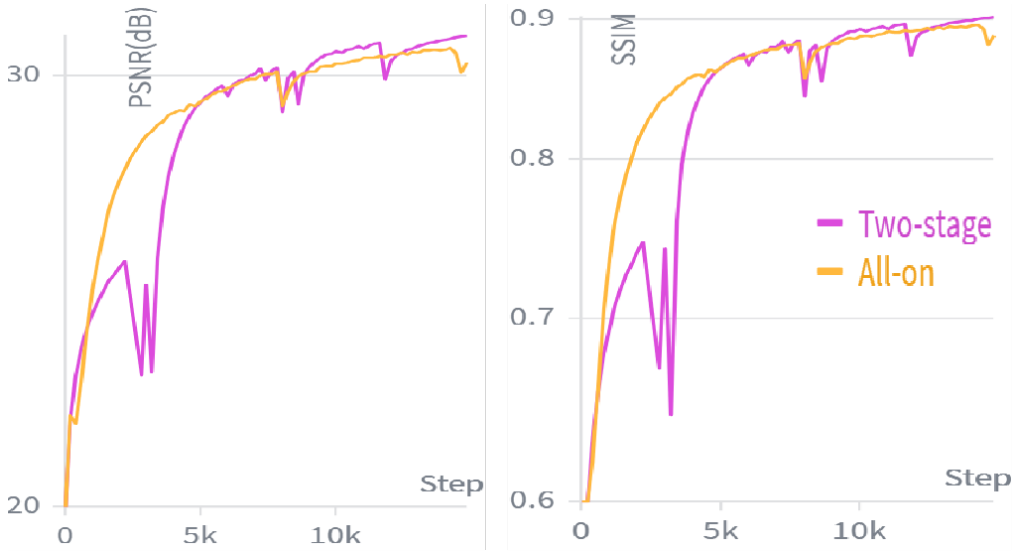


Figure 11: Training convergence of the two-stage scheduler and the all-on schedule on the chest 6-view setting.

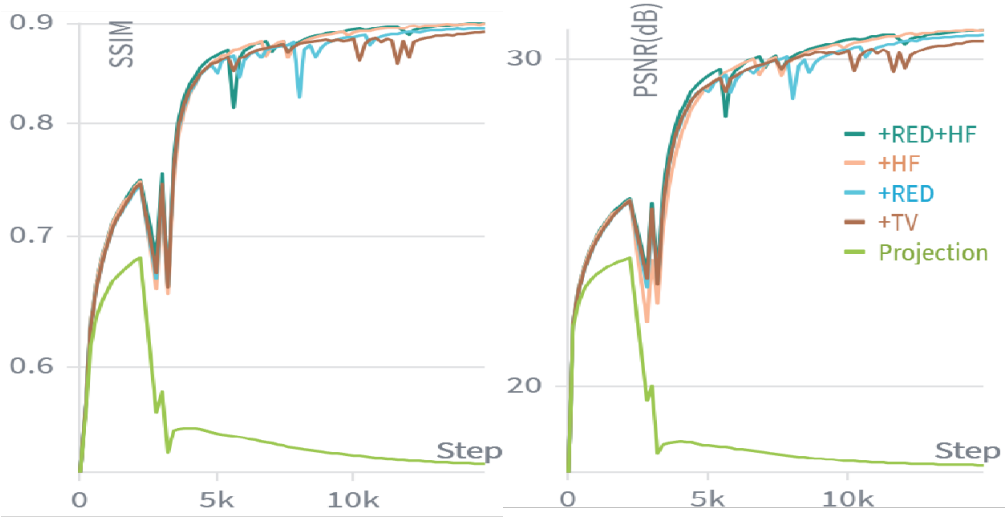


Figure 12: Training convergence under incremental loss composition on the chest 6-view setting.

| TV setting | PSNR↑ | SSIM↑ |
|---|---|---|
| $\lambda_{\text{tv}} = 0.00$ | 23.67 | 0.6875 |
| $\lambda_{\text{tv}} = 0.03$ | 27.14 | 0.8245 |
| $\lambda_{\text{tv}} = 0.10$ | 30.71 | 0.8957 |
| $\lambda_{\text{tv}} = 0.20$ | **31.10** | **0.9002** |

Table 12: Sensitivity to the TV regularization weight on the chest 6-view setting. HF and RED are enabled for all experiments.

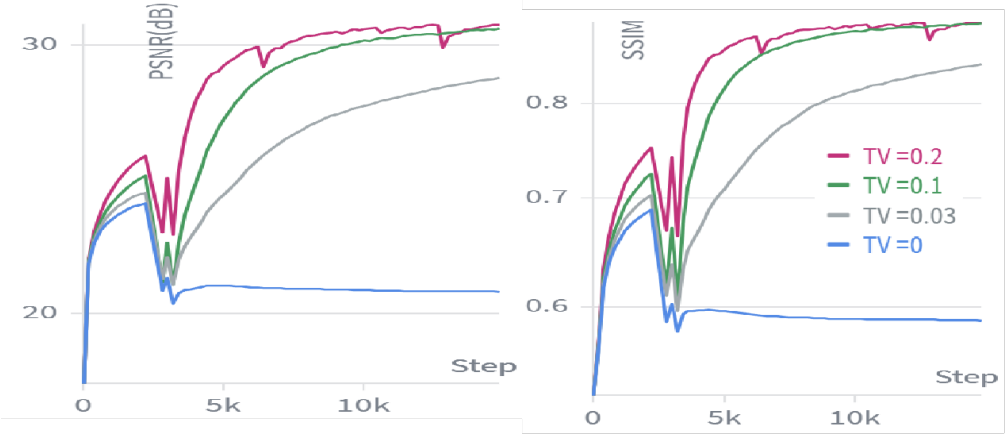


Figure 13: Training convergence with different TV weights on the chest 6-view setting. HF and RED are enabled in all configurations.

| Densification strategy | PSNR↑ | SSIM↑ |
|---|---|---|
| No densification | 28.31 | 0.8434 |
| Random densification | 31.16 | 0.9001 |
| Gradient densification | 31.18 | 0.9004 |
| G+P structure-guided | **31.22** | **0.9016** |

Table 13: Comparison of Gaussian densification strategies on the chest 6-view setting under the same reconstruction objective.

| Structure cue | PSNR↑ | SSIM↑ |
|---|---|---|
| Gradient cue only (G) | 30.45 | 0.8875 |
| Projection cue only (P) | 30.36 | 0.8861 |
| G+P cue | **30.50** | **0.8885** |

Table 14: Ablation of structure-cue sources on the chest 6-view setting.