PRISM-NAF: PHYSICALLY REALISTIC INTENSITY-DOMAIN SPECTRAL MODELING OF NEURAL ATTENUATION FIELDS FOR SPARSE-VIEW CBCT RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural attenuation fields (NAF) provide a data-efficient approach to sparse-view cone-beam CT (CBCT) reconstruction, yet most methods retain a simplified detector model: each pixel is represented by a single center ray, while spectral and low-frequency detector effects are modeled with weakly constrained spatial degrees of freedom. Under sparse views, these mismatches can be absorbed into the reconstructed attenuation field, yielding visually plausible but systematically distorted anatomy. We propose Physically Realistic Intensity-Domain Spectral Modeling of NAF (PRISM-NAF), a detector-aware neural attenuation field that structures these degrees of freedom at the measurement level. Coherent Aperture Aggregation samples multiple physically realizable paths over each area-based detector pixel; Material-Coupled Response replaces a free voxel-wise beam-hardening spread with a compact, globally shared monotone function of attenuation; and a coarse non-negative detector-plane intensity field provides a low-bandwidth correction at the stage where additive photon flux enters the measurement. We interpret the latter as an effective nuisance correction rather than a calibrated scatter estimate. Experiments on two real anthropomorphic phantom datasets spanning three anatomies and multiple sparse-view settings show that PRISM-NAF improves perceptual reconstruction quality and vascular-structure fidelity over state-of-the-art methods, while reducing structured artifacts and preserving fine anatomical boundaries.

1 INTRODUCTION

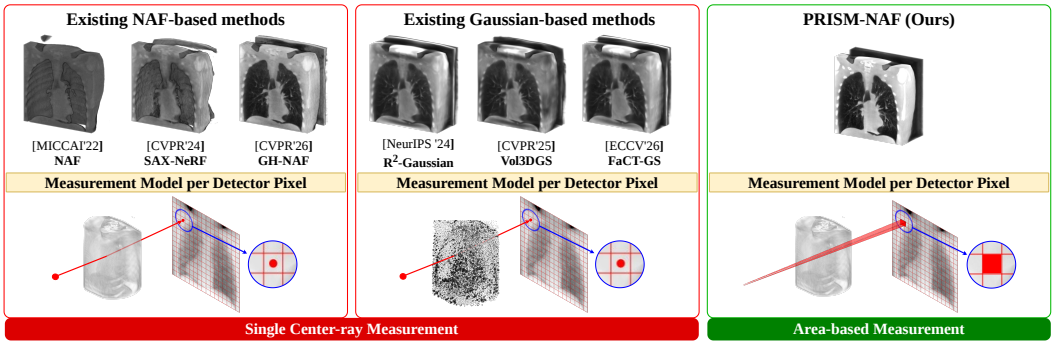


Figure 1: **Different representations, the same point-pixel approximation.** Representative NAF and Gaussian-based methods form one detector sample per pixel, so the area pixel is represented by a single point sample. PRISM-NAF instead models the pixel area using multiple geometry-consistent paths and aggregates their transmitted intensities into one measurement.

Cone-beam computed tomography (CBCT) reconstructs three-dimensional attenuation distributions from X-ray projections acquired around an object (Feldkamp et al., 1984). High-quality reconstruction, however, typically requires many projection views, increasing acquisition time and cumulative

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.

radiation exposure (Bian et al., 2010; Jia et al., 2010). Sparse-view CBCT therefore aims to recover high-quality volumes from substantially fewer projections. Deep learning-based methods can reduce streak artifacts and noise, but generally require representative paired training data and may generalize poorly across scanner geometries or acquisition protocols (Jin et al., 2017; Adler & Öktem, 2018). Neural Attenuation Fields (NAF) provide a scan-specific, self-supervised alternative by directly optimizing a continuous attenuation field from the projections of a single scan (Zha et al., 2022). Subsequent neural-field and Gaussian-based methods have improved scene representation, structural recovery, and reconstruction efficiency (Cai et al., 2024; Oh et al., 2026; Zha et al., 2024; Talegaonkar et al., 2025; Pieta et al., 2026).

Despite these advances, most existing methods retain a common detector approximation: each detector pixel is represented by a single center-associated ray or projected footprint (Zha et al., 2022; 2024). Thus, while scene representations have evolved, the detector pixel itself is still largely treated as a point-like measurement (Fig. 1). In practice, a detector pixel corresponds to an area over which multiple source-to-detector paths contribute to a single measurement (Liu et al., 2019). These paths can traverse different tissues and material boundaries, particularly near anatomical edges and fine structures (Glover & Pelc, 1980; Joseph & Spital, 1981). Representing the entire pixel with a single center ray can therefore create a systematic mismatch between the modeled and observed projection measurements (Liu et al., 2019). In sparse-view reconstruction, where only a limited number of projections constrain the attenuation field, such mismatch can be absorbed into the reconstructed volume, leading to boundary distortion or structured artifacts even when projection agreement remains high (Zha et al., 2024).

To address this limitation, we propose PRISM-NAF, which models each detector pixel as an area-based measurement rather than a point-like one. Its core component, Coherent Aperture Aggregation (CAA), samples multiple geometry-consistent sub-rays across the pixel area, evaluates the transmitted intensity along each path, and aggregates these contributions to form the final pixel measurement. CAA therefore improves reconstruction not simply by using more rays, but by better matching how the detector measurement is formed, reducing the need for the attenuation field to compensate for an inaccurate center-ray model. We further introduce Material-Coupled Response (MCR) to model path-dependent effective spectral variation using a compact attenuation-coupled response, and Detector-Plane Scatter (DPS) to capture broad additive detector-domain mismatch with a low-bandwidth correction. Together, these components introduce modeling flexibility at the corresponding stages of measurement formation, rather than allowing all mismatch to be absorbed into the three-dimensional attenuation field. Experiments on two real anthropomorphic phantom datasets spanning three anatomies and multiple sparse-view settings show that PRISM-NAF improves perceptual reconstruction quality and vascular-structure fidelity over the evaluated neural-field and Gaussian-based methods. Component-wise analyses further confirm the benefit of area-based detector modeling and the complementary effects of MCR and DPS. Our main contributions are summarized as follows:

- **Area-based detector modeling.** We replace the conventional approximation of a single center ray per detector pixel with CAA, which represents each detector pixel using multiple geometry-consistent paths across the pixel area and aggregates their transmitted intensities.
- **Structured measurement-side modeling.** We combine path-dependent attenuation-coupled spectral modeling with a low-bandwidth detector-plane correction, reducing the tendency of the attenuation field to absorb forward-model mismatch.
- **Real sparse-view CBCT validation.** We demonstrate improved perceptual quality and vascular-structure fidelity on two real anthropomorphic phantoms across three anatomies and multiple sparse-view settings.

2 Related Work

Sparse-view neural and Gaussian reconstruction. Coordinate-based tomography optimizes implicit fields directly from measured projections (Sun et al., 2021; Shen et al., 2024; Zang et al., 2021). NAF introduced a hash-encoded attenuation field for sparse-view CBCT (Zha et al., 2022); SAX-NeRF adds structure-aware ray processing (Cai et al., 2024), and GH-NAF adapts hash-grid levels under projection discrepancy (Oh et al., 2026). Gaussian approaches instead use explicit primitives and differentiable X-ray rasterization: R^2-Gaussian rectifies projected Gaussian kernels, Vol3DGS

integrates volumetric Gaussian transmittance, and FaCT-GS accelerates Gaussian ray-integral rendering (Zha et al., 2024; Talegaonkar et al., 2025; Pieta et al., 2026). These works primarily improve the *volume representation or ray renderer*. PRISM-NAF complements these advances by explicitly modeling the integration of transmitted intensities over the area of each detector pixel.

Support and physics-aware forward models. Detector support can produce nonlinear partial-volume and edge-gradient inconsistencies (Glover & Pelc, 1980; Joseph & Spital, 1981), while model-based CT has separately incorporated polychromatic attenuation, detector/focal-spot blur, and scatter (Brooks & Di Chiro, 1976; Herman, 1979; De Man et al., 2001; Tilley et al., 2016; Siewerdsen & Jaffray, 2001). Neural approaches likewise model energy- or material-dependent attenuation (Wu et al., 2023; Park et al., 2025; Birklein et al., 2025); GH-NAF uses the same closed-form effective spectral correction as PRISM-NAF but predicts a spatially free spread channel (Oh et al., 2026). We retain that spectral form but constrain its latent spread through MCR and evaluate it separately on the aperture paths before intensity aggregation. DPS then supplies only a low-bandwidth additive detector-plane variable. Anti-aliased fields such as Mip-NeRF and Zip-NeRF integrate spatial footprints to control representation aliasing (Barron et al., 2021; 2023); CAA instead approximates the nonlinear *detector measurement* by integrating final transmitted intensities over the physical pixel support.

3 PRELIMINARIES

3.1 NEURAL ATTENUATION FIELDS

We represent the attenuation volume with a multiresolution hash encoder $\mathcal{H}$ followed by a lightweight MLP F_θ (Müller et al., 2022):

$$\big(\mu_\theta(\mathbf{x}),\, \beta_\theta^2(\mathbf{x})\big) = \mathcal{F}_\theta(\mathcal{H}(\mathbf{x}))\,, \qquad \mathbf{x} \in \mathbb{R}^3, \tag{1}$$

where $\mu_\theta(\mathbf{x}) \geq 0$ is the linear attenuation coefficient and $\beta_\theta^2(\mathbf{x}) > 0$ is a local reliability variable used in the heteroscedastic projection loss, following Kendall & Gal (2017) and its use in neural CBCT reconstruction by Oh et al. (2026). For notational clarity, all log-projection quantities below are expressed in optical-depth units and the incident intensity is normalized to one.

For a source-to-detector ray $\mathbf{r}(s) = \mathbf{o} + s\widehat{\mathbf{d}}$ parameterized by arc length s, the conventional monochromatic log projection is

$$P_{\mathrm{mono}}(\mathbf{r}) = \int \mu_\theta(\mathbf{r}(s))\; ds \approx \sum_j \mu_\theta(\mathbf{x}_j)\, \Delta_j, \tag{2}$$

where Δ_j is the physical length of interval j. Existing neural-field and Gaussian-based CBCT renderers typically approximate each detector pixel by evaluating only the ray through its center.

3.2 AREA DETECTOR-PIXEL MEASUREMENT

Under the point-source and uniform intra-pixel sensitivity approximations used in this work, let $A_v(\mathbf{u})$ denote the sensing area of detector pixel $\mathbf{u}$ in projection view v. If $P_v(\boldsymbol{\xi})$ is the effective primary log projection along the path terminating at $\boldsymbol{\xi} \in A_v(\mathbf{u})$, the pixel forms

$$P_{\mathrm{pix},v}(\mathbf{u}) = -\log\left[\frac{1}{|A_v(\mathbf{u})|}\int_{A_v(\mathbf{u})} \exp(-P_v(\boldsymbol{\xi}))\; d\boldsymbol{\xi}\right]. \tag{3}$$

Thus, a detector pixel averages transmitted intensities over its area and applies the logarithm only afterward. Eq. 3 is distinct from both a center-ray point approximation and an aperture average of log projections. The present model addresses the geometric pixel footprint; focal-spot blur and detector point-spread effects are outside its scope.

4 METHOD

PRISM-NAF augments a neural attenuation field with a detector-aware measurement model. As illustrated in Fig. 2, CAA models detector-pixel support using geometry-consistent sub-rays, while

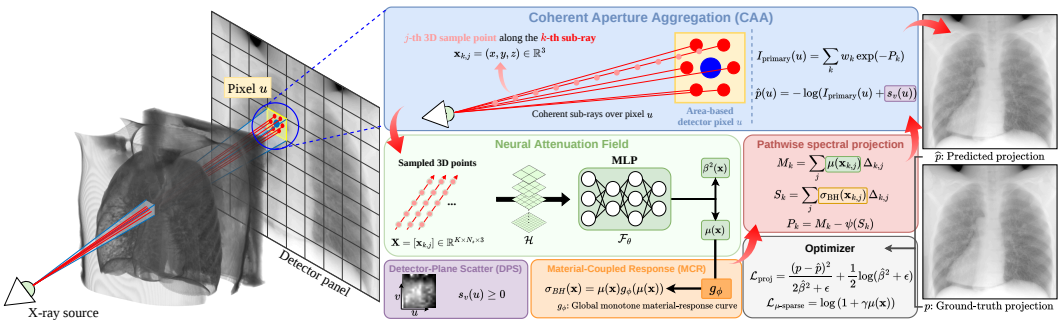


Figure 2: **Overview of PRISM-NAF.** For each area-based detector pixel, CAA evaluates multiple geometry-consistent sub-rays through the NAF and aggregates their transmitted intensities over the pixel area. MCR models the pathwise effective spectral response, while DPS corrects low-bandwidth detector-plane mismatch, together forming a detector-aware area measurement model.

MCR parameterizes the pathwise effective spectral response through a compact attenuation-coupled function shared across the volume. We additionally use DPS as a low-bandwidth per-view detector-plane correction before the final logarithm. CAA and MCR define the shared measurement operator, whereas DPS provides auxiliary flexibility for broad view-specific mismatch.

4.1 Coherent Aperture Aggregation

Geometry-consistent sub-rays. Given detector pixel $\mathbf{u}$ with center $\mathbf{e}(\mathbf{u})$, half-width vectors $\mathbf{h}_1, \mathbf{h}_2$, and X-ray source $\mathbf{o}$, we define the detector endpoint and ray direction for aperture coordinates $(a_k, b_k) \in [-1, 1]^2$ as follows:

$$\mathbf{e}_{\mathbf{u},k} = \mathbf{e}(\mathbf{u}) + a_k \mathbf{h}_1 + b_k \mathbf{h}_2, \qquad \widehat{\mathbf{d}}_{\mathbf{u},k} = \frac{\mathbf{e}_{\mathbf{u},k} - \mathbf{o}}{\|\mathbf{e}_{\mathbf{u},k} - \mathbf{o}\|_2}. \tag{4}$$

Each detector endpoint $\mathbf{e}_{\mathbf{u},k}$ lies within the pixel aperture and defines a straight path from the source $\mathbf{o}$. Here, *coherent* refers to geometric consistency along each source-to-detector path, not wave coherence.

Footprint-aware field evaluation. Following Zip-NeRF (Barron et al., 2023), hash features are weighted by the footprint standard deviation supplied for each query sample. For a network query containing J samples, the encoding is

$$\overline{\mathcal{H}} = \frac{1}{J} \sum_{q=1}^{J} \left(\bigoplus_{\ell=1}^{L} a_\ell(\rho_q) \mathcal{H}_\ell(\mathbf{x}_q) \right), \qquad a_\ell(\rho) = \operatorname{erf}\left(\frac{1}{\sqrt{8\rho^2 f_\ell^2}} \right), \qquad f_\ell = \frac{N_\ell - 1}{2b}, \tag{5}$$

where ρ_q is the footprint scale in physical units, N_ℓ is the grid resolution, and b is the spatial half-extent. Individual point queries use $J = 1$. This weighting attenuates fine-grid features relative to the query footprint. The filtered encoding replaces $\mathcal{H}$ in Eq. 1, and MCR derives the spectral spread from the predicted attenuation. Footprint dependence is omitted below for brevity; detector-level aggregation remains defined over transmitted intensities in Eq. 8.

Pathwise effective spectral projection. For each sub-ray, attenuation and effective beam-hardening spread obtained from the filtered field queries are integrated independently along the corresponding path. Using the attenuation-coupled spread σ_{BH} defined in Sec. 4.2, we compute

$$M_{v,\mathbf{u},k} = \sum_j \mu(\mathbf{x}_{v,\mathbf{u},k,j}) \Delta_{v,\mathbf{u},k,j}, \qquad S_{v,\mathbf{u},k} = \sum_j \sigma_{\mathrm{BH}}(\mathbf{x}_{v,\mathbf{u},k,j}) \Delta_{v,\mathbf{u},k,j}, \tag{6}$$

where $\Delta_{v,\mathbf{u},k,j}$ denotes the physical integration interval. Following the symmetric spectral model of Park et al. (2025), the effective log projection of each path is

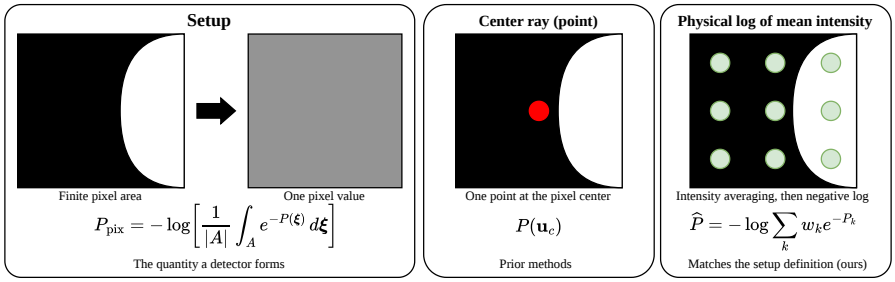


Figure 3: **Detector-pixel measurement.** A physical detector pixel forms one measurement by integrating transmitted intensities over its spatial support before the logarithm is applied. Conventional methods approximate this measurement using a single center ray, whereas CAA evaluates multiple geometry-consistent paths over the detector pixel.

$$P_{v,\mathbf{u},k} = M_{v,\mathbf{u},k} - \psi(S_{v,\mathbf{u},k})\,, \qquad \psi(S) = \log\left(\frac{\sinh S}{S}\right), \quad \psi(0) = 0. \tag{7}$$

Because the spectral response is evaluated independently for each path, different locations within a detector pixel can exhibit different attenuation and effective spectral responses before contributing to the same measured pixel value.

Area intensity aggregation. Each pathwise log projection is converted to transmitted primary intensity. The area detector measurement is then approximated by quadrature over the aperture paths:

$$I_{\text{primary},v}(\mathbf{u}) = \sum_{k=1}^{K} w_k \exp(-P_{v,\mathbf{u},k})\,, \qquad w_k \geq 0, \quad \sum_k w_k = 1. \tag{8}$$

Aperture samples are arranged in a centroid-preserving low-discrepancy pattern and assigned equal quadrature weights $w_k = 1/K$.

The corresponding detector-pixel log measurement is

$$\widehat{P}_{\text{primary},v}(\mathbf{u}) = -\log I_{\text{primary},v}(\mathbf{u}). \tag{9}$$

This ordering follows the area measurement process illustrated in Fig. 3: pathwise transmitted intensities are first integrated over the detector support, and the logarithm is applied only after the pixel intensity has been formed.

4.2 Material-Coupled Response

MCR constrains the effective beam-hardening response to vary with local attenuation rather than treating it as an independent spatial field. For the pathwise projection $P = M - \psi(S)$, the response to a relative change in spectral spread is

$$\frac{\partial P}{\partial \log S} = -S\psi'(S) \approx -\frac{S^2}{3}, \qquad S \to 0, \tag{10}$$

whereas $\partial P/\partial \log M = M$. Thus, in the small-spread regime, the spectral term is more weakly constrained by projection supervision than the primary attenuation term. We therefore parameterize the local spectral spread as

$$\sigma_{\text{BH}}(\mathbf{x}) = \mu(\mathbf{x})\, g_\phi(\mu(\mathbf{x}))\,, \tag{11}$$

where g_ϕ is a globally shared, non-negative one-dimensional response. We use a non-decreasing parameterization as a mild ordering prior over the effective attenuation range, rather than as a universal

material law. The key restriction is that the spectral response is coupled to attenuation and shared across the volume, substantially reducing the spatial degrees of freedom available to the spectral term.

4.3 Detector-Plane Scatter-Inspired Correction

Broad additive mismatch, including contributions associated with scatter, enters the measurement at the detector plane rather than along the primary ray. We therefore introduce DPS as a non-negative, low-bandwidth detector-plane intensity field $s_v(\mathbf{u})$ for each projection view. The field is represented at a coarse detector-plane resolution and interpolated to the measurement resolution before being added to the primary intensity:

$$I_{\mathrm{det},v}(\mathbf{u}) = I_{\mathrm{primary},v}(\mathbf{u}) + s_v(\mathbf{u}), \qquad \widehat{p}_v(\mathbf{u}) = -\log I_{\mathrm{det},v}(\mathbf{u}). \tag{12}$$

The low-bandwidth parameterization restricts DPS to broad detector-domain variation and limits its capacity to represent fine anatomical structure. Because projection supervision does not uniquely separate physical scatter from other low-frequency measurement discrepancies, we interpret s_v as a scatter-inspired detector-plane correction rather than as a calibrated scatter estimate.

Unlike CAA and MCR, s_v is defined independently for each observed projection and is not intended to extrapolate to unseen acquisition angles. Its role is to prevent broad view-specific detector mismatch from being absorbed directly into the reconstructed attenuation fields.

4.4 Complete Forward Operator

Combining Eqs. 7, 8, and 12, the complete measurement operator is

$$\widehat{p}_v(\mathbf{u}) = -\log\left[\sum_{k=1}^{K} w_k \exp(-M_{v,\mathbf{u},k} + \psi(S_{v,\mathbf{u},k})) + s_v(\mathbf{u})\right]. \tag{13}$$

The operator makes the measurement order explicit: pathwise attenuation and effective spectral response, area aggregation of primary intensities, additive low-bandwidth detector-plane correction, and the final logarithm.

Joint non-identifiability. The projection measurements constrain the total intensity in Eq. 13, but not a unique decomposition into attenuation, effective spectral response, and detector-plane correction. The structured parameterizations of g_ϕ and s_v restrict their degrees of freedom but do not guarantee unique physical recovery. We therefore treat μ, g_ϕ, and s_v as jointly optimized effective variables rather than individually calibrated physical quantities.

4.5 Training Objective

The neural field additionally predicts a local uncertainty variable $\beta^2(\mathbf{x})$ following prior work (Kendall & Gal, 2017; Oh et al., 2026). We aggregate this quantity along each geometry-consistent path and across the detector pixel using a weighted sum, obtaining the detector-level uncertainty $\widehat{\beta}_v^2(\mathbf{u})$.

Given the measured log projection $p_v(\mathbf{u})$, the reconstruction is optimized using

$$\begin{aligned}\mathcal{L} = \mathbb{E}_{v,\mathbf{u}}\left[\frac{\big(p_v(\mathbf{u}) - \widehat{p}_v(\mathbf{u})\big)^2}{2(\widehat{\beta}_v^2(\mathbf{u}) + \epsilon)} + \frac{1}{2}\log\Big(\widehat{\beta}_v^2(\mathbf{u}) + \epsilon\Big)\right] \\ + \lambda_\mu \mathbb{E}_{v,\mathbf{u},k,j}[\log(1 + \gamma\,\mu(\mathbf{x}_{v,\mathbf{u},k,j}))]\,.\end{aligned} \tag{14}$$

The first term is a heteroscedastic projection loss, while the second provides a sparsity prior on the reconstructed attenuation field. The attenuation and reliability fields, the shared material response, and the detector-plane corrections are optimized jointly from projection supervision without requiring a ground-truth attenuation volume.

Table 1: Quantitative comparison of perceptual reconstruction quality across sparse-view CBCT methods, averaged over the evaluated view settings (50/60/70/80 views for Lungman chest and 50/70 views for pediatric chest and head). Best in **bold**, second-best underlined.

	LPIPS ↓			MANIQA ↑		
Method	Lungman chest	Pediatric chest	Pediatric head	Lungman chest	Pediatric chest	Pediatric head
NAF [MICCAI'22 Oral]	0.680	0.538	0.404	0.363	0.318	0.405
SAX-NeRF [CVPR'24]	0.646	0.550	0.416	0.302	0.379	0.454
R^2-Gaussian [NeurIPS'24]	0.603	0.554	0.391	0.290	0.356	0.478
Vol3DGS [CVPR'25 Highlight]	0.620	0.567	0.385	0.288	0.371	0.479
GH-NAF [CVPR'26]	0.576	**0.496**	0.405	0.236	0.326	0.490
FaCT-GS [ECCV'26]	0.606	0.569	0.420	0.311	0.357	0.491
PRISM-NAF (Ours)	**0.527**	0.497	**0.375**	**0.372**	**0.461**	**0.498**

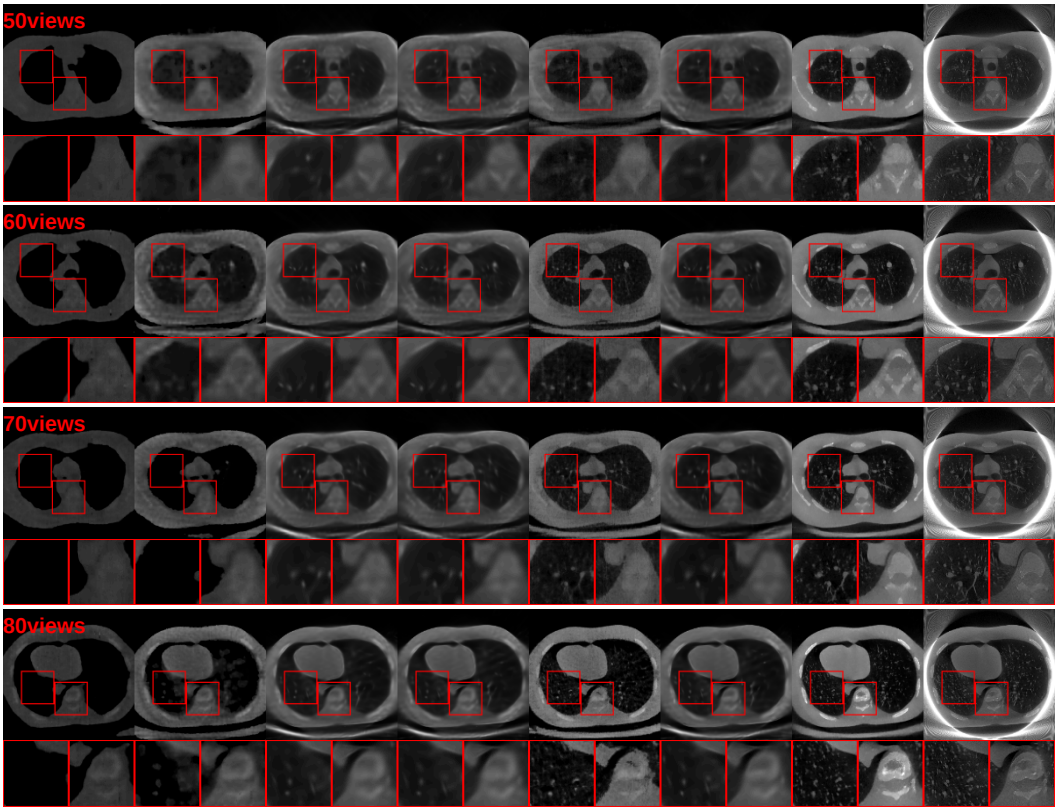


Figure 4: Qualitative comparison on the Lungman chest phantom across sparse-view settings.

5 Experiments

5.1 Experimental Settings

Datasets. We evaluate PRISM-NAF on two real anthropomorphic CBCT phantom datasets, each acquired with 720 projections: a Lungman chest phantom and a pediatric phantom scanned at the chest and head. Because ground-truth attenuation volumes are unavailable, we use the corresponding 720-view FDK reconstructions (Feldkamp et al., 1984) as dense-view structural references. We evaluate 50/60/70/80-view settings for the Lungman chest and 50/70-view settings for the pediatric chest and head. Detailed descriptions of the datasets, acquisition geometry, and preprocessing are provided in Appendix A.

Baselines. We compare with NAF (Zha et al., 2022), SAX-NeRF (Cai et al., 2024), R^2-Gaussian (Zha et al., 2024), Vol3DGS (Talegaonkar et al., 2025), FaCT-GS (Pieta et al., 2026), and GH-NAF (Oh et al., 2026), using their official implementations. All methods share the same

Table 2: Quantitative comparison of vascular-structure preservation across sparse-view CBCT methods, averaged over all evaluated view settings. Best in **bold**, second-best underlined.

	Vesselness corr. ↑		Vessel Dice ↑		Centerline ratio → 1		Vessel-count ratio → 1	
Method	Lungman chest	Pediatric chest	Lungman chest	Pediatric chest	Lungman chest	Pediatric chest	Lungman chest	Pediatric chest
NAF [MICCAI'22 Oral]	0.321	0.901	0.260	0.833	0.11	0.75	0.11	0.83
SAX-NeRF [CVPR'24]	0.574	0.934	0.462	0.867	0.24	0.85	0.22	0.89
R^2-Gaussian [NeurIPS'24]	0.766	0.833	0.613	0.760	0.18	0.52	0.18	0.56
Vol3DGS [CVPR'25 Highlight]	0.764	0.847	0.610	0.771	0.18	0.58	0.19	0.64
GH-NAF [CVPR'26]	0.846	0.924	0.692	0.858	0.40	**0.95**	0.54	1.03
FaCT-GS [ECCV'26]	0.764	0.834	0.612	0.759	0.18	0.53	0.19	0.59
PRISM-NAF (Ours)	**0.924**	**0.964**	**0.786**	**0.907**	**0.60**	0.94	**0.77**	**0.99**

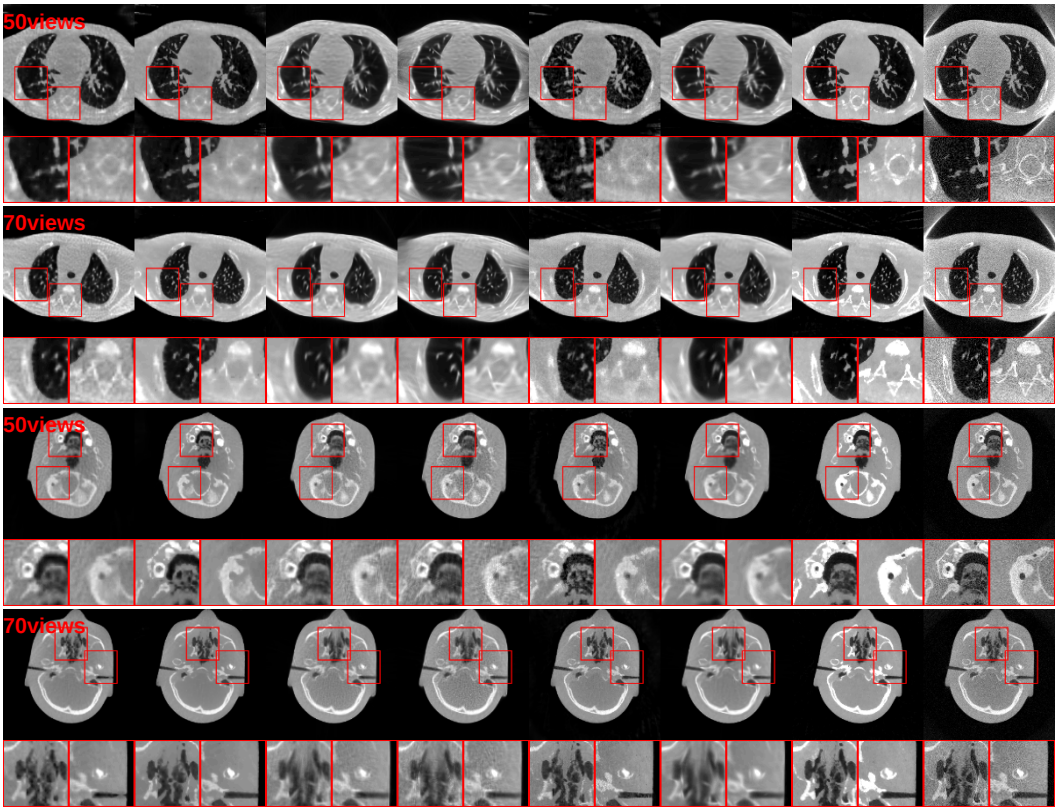


Figure 5: Qualitative comparison on the pediatric chest/head phantom across sparse-view settings.

sparse-view subsets, preprocessing, reconstruction grid, and display window. The 720-view FDK reconstruction serves only as a reference and is not ranked as a sparse-view baseline. Method-specific hyperparameters follow the settings recommended in the corresponding papers and official implementations.

Metrics. Perceptual quality is evaluated using MANIQA (Yang et al., 2022) and LPIPS (Zhang et al., 2018). MANIQA provides a reference-free assessment, while LPIPS is computed after global affine intensity alignment to the 720-view FDK reference, which may retain beam-hardening and scatter artifacts. Vascular-structure fidelity is assessed using Frangi vesselness (Frangi et al., 1998), with vesselness correlation, Dice overlap, centerline-length ratio, and vessel-count ratio measured within fixed lung regions. All methods are evaluated using identical slices, display windows, regions, and metric protocols.

5.2 Reconstruction Results on Real Anthropomorphic Phantoms

Lungman chest phantom. Fig. 4 and Tables 1 and 2 show that PRISM-NAF consistently improves reconstruction quality across sparse-view settings. Prior neural-field methods retain structured nonuniformity or texture-like artifacts, whereas Gaussian-based methods tend to suppress

Table 3: Cumulative ablation on the Lungman chest phantom at 50 views, with each proposed component added sequentially.

Variant	CAA	MCR	DPS	LPIPS ↓	MANIQA ↑	Vessel Dice ↑
Center-ray renderer	✗	✗	✗	0.556	0.341	0.731
+ area aggregation	✓	✗	✗	0.545	0.339	0.736
+ material-coupled response	✓	✓	✗	**0.535**	0.360	**0.752**
+ detector-plane correction	✓	✓	✓	0.544	**0.365**	0.737

noise at the expense of fine pulmonary detail. In contrast, PRISM-NAF produces more spatially coherent lung regions while preserving thin vessels and soft-tissue boundaries.

The quantitative results are consistent with these observations. PRISM-NAF provides the strongest overall perceptual quality on the Lungman phantom and consistently improves vesselness agreement, overlap, and recovered vascular extent. Importantly, these gains are observed jointly: artifact suppression is accompanied by improved preservation of fine vascular structure rather than by indiscriminate spatial smoothing.

Pediatric phantom. Fig. 5 shows that the improvements observed on the Lungman phantom are consistently reproduced on the independently acquired pediatric dataset. In the chest, PRISM-NAF reduces spatially structured reconstruction artifacts and intensity nonuniformity while preserving pulmonary branches and air–tissue interfaces.

In the head, PRISM-NAF yields sharper and more coherent bone–soft-tissue boundaries and better preserves fine anatomical structures than the competing methods. These qualitative improvements are consistent with the perceptual evaluation across both anatomies, indicating that the proposed measurement model remains effective across distinct anatomical regions and acquisition settings.

5.3 Ablation Study

Component-wise ablation. Table 3 and Fig. 6 show that the three components affect complementary aspects of reconstruction quality. CAA provides the first structural improvement over the center-ray formulation, while MCR gives the strongest agreement with the dense-view structural reference together with a clear gain in reference-free perceptual quality. Adding DPS further improves the reference-free perceptual score and visibly reduces broad intensity nonuniformity, although it does not yield the best FDK-referenced metrics. This behavior is consistent with its role as a view-specific low-frequency nuisance correction rather than a universally transferable component of the shared measurement model. Further ablation studies and supplementary evaluations are provided in Appendix B.

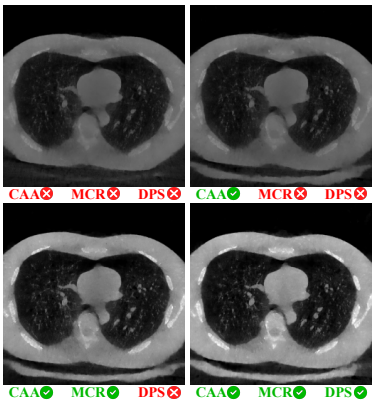


Figure 6: Qualitative cumulative ablation of CAA, MCR, and DPS.

6 Conclusion And Limitation

In this work, we introduced PRISM-NAF, a neural attenuation fields framework that explicitly models area measurement formation in flat-panel detectors. The framework integrates primary intensity over each detector pixel, models beam-hardening effects through a shared attenuation-dependent response, and incorporates an additive detector-plane correction before the logarithmic transform. Experiments on two real anthropomorphic phantom datasets demonstrate the strongest overall perceptual quality and vascular-structure fidelity among the evaluated sparse-view methods. These findings highlight the importance of jointly considering attenuation-field representation and measurement-model fidelity in neural CBCT reconstruction. The evaluation remains limited in anatomical diversity and lacks voxel-wise ground-truth attenuation, restricting conclusions about clinical generalization and direct quantitative validation of the individual physical components. Future work will extend evaluation to more diverse, clinically representative data and controlled phantom experiments with known attenuation properties.

AI USAGE DISCLOSURE

In this work, we used generative AI tools only to edit the manuscript for readability, including correcting typographical errors and refining wording for a formal academic style. We did not use generative AI tools to develop theoretical or conceptual frameworks, formulate mathematical claims or proofs, propose or refine hypotheses, design research methodology or experiments, implement methods, clean or reformat datasets, conduct qualitative or thematic data analysis, or interpret results. Tasks involving translation were not applicable to this work. We reviewed all AI-assisted edits and verified that they did not alter the scientific content or claims of the paper. We take responsibility for the final content of this work, including any text produced or edited with the aid of generative AI.

REPRODUCIBILITY STATEMENT

All datasets, sparse-view subsets, projection geometry, reconstruction grids, and evaluation regions are shared across methods and are described in Sec. 5 and Appendix A. Baselines use official public implementations with their recommended settings. Our code, configuration files, and evaluation scripts will be released upon publication.

ETHICS STATEMENT

All experiments were conducted using anthropomorphic phantoms, without human participants or patient data.

REFERENCES

Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018. doi: 10.1109/TMI.2018.2799231.

Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5835–5844. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00580.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-NeRF: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19640–19648. IEEE, 2023.

Junguo Bian, Jeffrey H. Siewerdsen, Xiao Han, Emil Y. Sidky, Jerry L. Prince, Charles A. Pelizzari, and Xiaochuan Pan. Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Physics in Medicine & Biology*, 55(22):6575–6599, 2010. doi: 10.1088/0031-9155/55/22/001.

Lukas Birklein, Elmar Schömer, Ulrich Schwanecke, and Ralf Schulze. Polychromatic neural CBCT reconstruction through density-attenuation modeling. *Physics in Medicine & Biology*, 70(24): 245021, 2025.

Rodney A Brooks and Giovanni Di Chiro. Beam hardening in x-ray reconstructive tomography. *Physics in medicine & biology*, 21(3):390–398, 1976.

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

Bruno De Man, Johan Nuyts, Patrick Dupont, Guy Marchal, and Paul Suetens. An iterative maximum-likelihood polychromatic algorithm for ct. *IEEE transactions on medical imaging*, 20(10):999–1008, 2001.

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.

Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 130–137. Springer, 1998.

G. H. Glover and N. J. Pelc. Nonlinear partial volume artifacts in x-ray computed tomography. *Medical physics*, 7(3):238–248, 1980.

Gabor T Herman. Correction for beam hardening in computed tomography. *Physics in Medicine & Biology*, 24(1):81–106, 1979.

Xun Jia, Yifei Lou, Ruijiang Li, William Y. Song, and Steve B. Jiang. GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation. *Medical Physics*, 37(4):1757–1760, 2010. doi: 10.1118/1.3371691.

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. doi: 10.1109/TIP.2017.2713099.

Peter M Joseph and RD Spital. The exponential edge-gradient effect in x-ray computed tomography. *Physics in Medicine & Biology*, 26(3):473–487, 1981.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

Xin Liu, Buxin Chen, Zheng Zhang, Dan Xia, Emil Y Sidky, and Xiaochuan Pan. Optimization-based reconstruction for correcting non-linear partial volume artifacts in ct. In *Proc. SPIE*, volume 10948, pp. 109482Q, 2019.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022. doi: 10.1145/3528223.3530127.

Seong Je Oh, Ju Hwan Lee, Chae Yeon Lim, Donghwan Lee, Myung Jin Chung, and Kyungsu Kim. GH-NAF: Grid-Adaptive Hash-Level-Attended Neural Attenuation Fields for Discrepancy-Aware CBCT. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 41762–41772, 2026.

Hyoung Suk Park, Jin Keun Seo, and Kiwan Jeon. Implicit neural representation-based method for metal-induced beam hardening artifact reduction in x-ray CT imaging. *Medical Physics*, 52(4): 2201–2211, 2025.

Pawel Tomasz Pieta, Rasmus Juul Pedersen, Sina Borgi, Jakob Sauer Jørgensen, Jens Wenzel Andreasen, and Vedrana Andersen Dahl. Fact-gs: Fast and scalable ct reconstruction with gaussian splatting. In *European Conference on Computer Vision*, pp. 225–245. Springer, 2026.

Liyue Shen, John Pauly, and Lei Xing. Nerp: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):770–782, 2024. doi: 10.1109/TNNLS.2022.3177134.

Jeffrey H. Siewerdsen and David A. Jaffray. Cone-beam computed tomography with a flat-panel imager: magnitude and effects of x-ray scatter. *Medical physics*, 28(2):220–231, 2001.

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.

Chinmay Talegaonkar, Yash Belhe, Ravi Ramamoorthi, and Nicholas Antipa. Volumetrically consistent 3d gaussian rasterization. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10953–10963. IEEE, 2025.

Steven Tilley, Jeffrey H Siewerdsen, and J Webster Stayman. Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise. *Physics in Medicine & Biology*, 61(1):296–319, 2016.

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. *Advances in Neural Information Processing Systems*, 36:69605–69624, 2023. doi: 10.52202/075280-3048.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. MANIQA: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1190–1199. IEEE, 2022.

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*, pp. 1940–1950. IEEE, 2021.

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*, pp. 442–452. Springer, 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*, volume 37, pp. 44907–44934, 2024. doi: 10.52202/079017-1427.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595. IEEE, 2018.

APPENDIX

A EXPERIMENTAL DETAILS

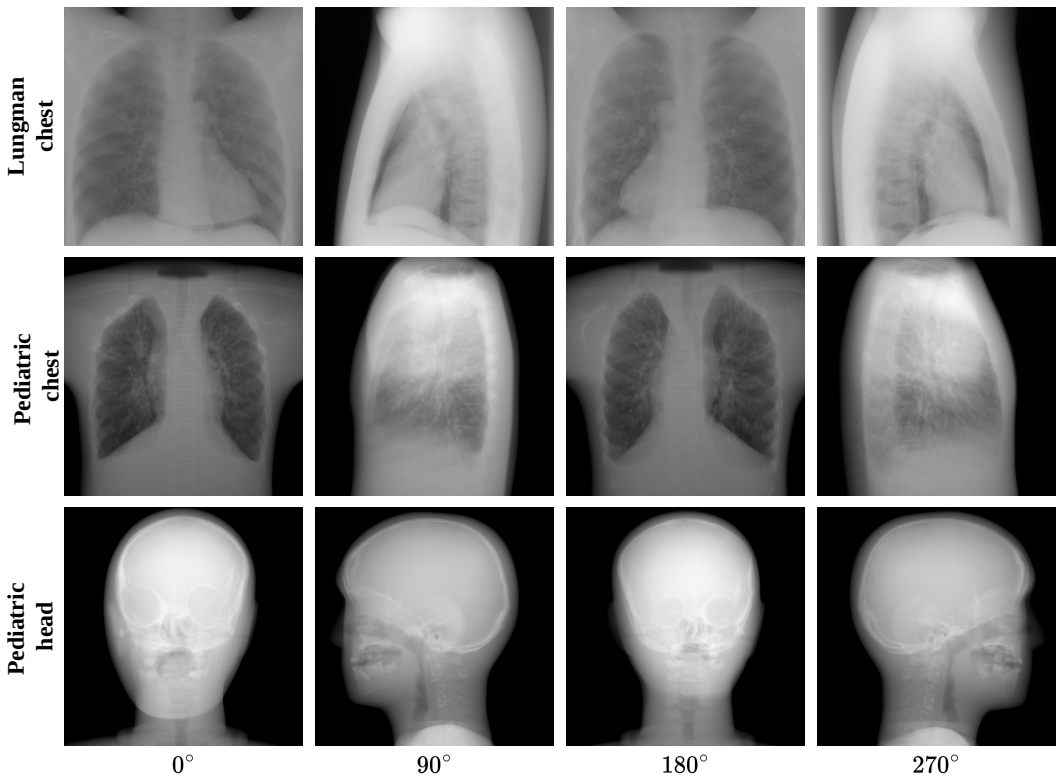
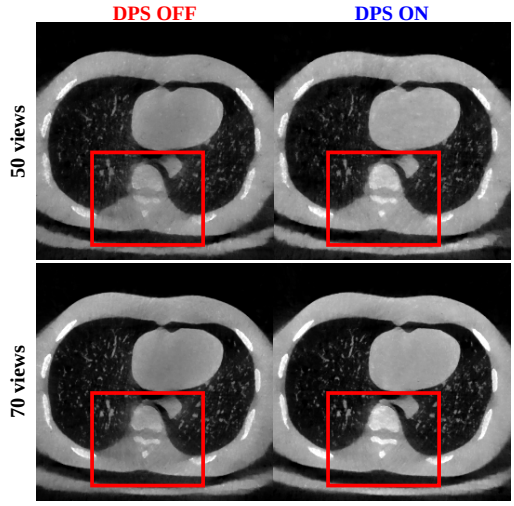


Figure 7: Representative projection images from the CBCT datasets.

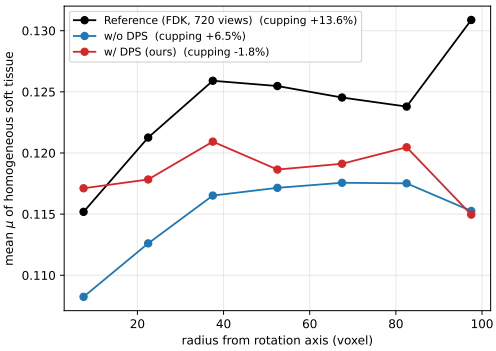
A.1 Dataset and Acquisition Details

This section provides detailed descriptions of the real CBCT datasets used in our experiments. To evaluate PRISM-NAF under realistic acquisition conditions, we acquired CBCT scans of two anthropomorphic phantoms using two distinct CBCT systems with different detector geometries and acquisition configurations. Representative projection samples are shown in Fig. 7.

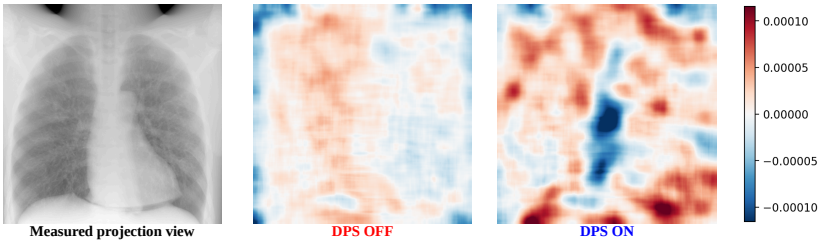
Lungman chest phantom. The Lungman chest phantom was acquired using a mobile CBCT system. The scan consists of 720 projections acquired over a full 360° rotation at a uniform angular interval of 0.5°. The native detector resolution is 1516×1516 pixels with a detector pixel pitch of $0.28\,\mathrm{mm}$. The source-to-object distance (SOD) is $620\,\mathrm{mm}$ and the source-to-detector distance (SDD) is $920\,\mathrm{mm}$. In the source-centered coordinate system, the detector center is offset by $-1.4\,\mathrm{mm}$ and $-56.66\,\mathrm{mm}$ along the detector x- and y-axes, respectively.



(a) Reconstructions



(b) Cupping profile



(c) Low-frequency projection residual

Figure 8: Effect of the detector-plane correction on the Lungman chest phantom. **(a)** Axial reconstructions without and with DPS at 50 and 70 views. **(b)** Radial attenuation profiles measured in a homogeneous soft-tissue region for the 720-view FDK reference and reconstructions without and with DPS. **(c)** Representative measured projection and the corresponding low-frequency projection residuals without and with DPS.

In practical mobile CBCT acquisition, the imaging geometry can exhibit non-negligible uncertainty and calibration drift, including small but systematic deviations in both in-plane and out-of-plane projection angles. To remain consistent with the geometry used by the clinical reconstruction pipeline, we adopt the calibrated system configuration provided by the manufacturer's reconstruction software as the geometry of our forward model. This configuration includes the refined in-plane and out-of-plane projection angles obtained through the system's internal geometric calibration procedure. For computational efficiency, all projection images are downsampled by a factor of two, from 1516×1516 to 758×758 pixels, before reconstruction.

Pediatric phantom. The pediatric phantom was acquired using a cone-beam X-ray system in which the X-ray source and detector remain fixed while the phantom is rotated. We separately scan the chest and head regions of the phantom, yielding two anatomical datasets. Each scan consists of 720 projections acquired over a full 360° phantom rotation at a uniform angular interval of 0.5°. The native detector resolution is 2090×2090 pixels with a detector pixel pitch of $0.21\,\mathrm{mm}$. The SOD is $605\,\mathrm{mm}$ and the SDD is $863\,\mathrm{mm}$.

Because the source and detector remain mechanically fixed throughout acquisition and the angular sampling is generated by rotating the phantom, the projection geometry is specified directly by the prescribed phantom rotation angles, without additional view-dependent geometric correction. For computational efficiency, all projection images are downsampled by a factor of two, from 2090×2090 to 1045×1045 pixels, before reconstruction. The same acquisition geometry and preprocessing procedure are used for both the pediatric chest and head scans.

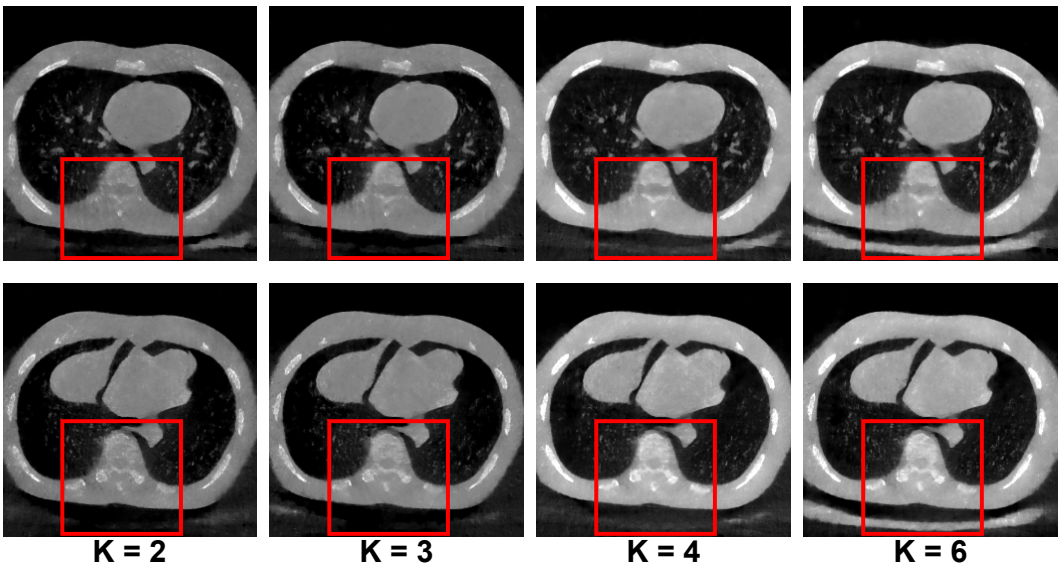


Figure 9: Qualitative comparison of reconstructions obtained with different numbers of coherent sub-rays. Increasing K improves local uniformity and boundary consistency.

Reconstruction volume. For all datasets and all compared methods, we reconstruct an attenuation volume of size $256 \times 256 \times 256$. Within each dataset, all methods use the same reconstruction grid and acquisition geometry to ensure a consistent comparison.

A.2 Implementation Details

Neural attenuation fields. We use a multiresolution hash encoder with 12 levels and 4 features per level, with resolutions increasing geometrically from 16 to 2048 and a hash-table size of 2^{24} per level. The encoded spatial bound is set from the acquisition geometry. The decoder consists of three 128-wide MLP layers, a 128-dimensional bottleneck, and output heads for attenuation $\mu(\mathbf{x})$ and heteroscedastic variance $\beta^2(\mathbf{x})$, both constrained to be non-negative. Hash features are prefiltered according to the physical query footprint to suppress frequencies finer than the corresponding detector footprint.

Coherent Aperture Aggregation. Each detector pixel is represented by $K = 6$ coherent sub-rays sampled over the rectangular pixel aperture using a centroid-preserving low-discrepancy pattern. Each detector offset is fixed along its source-to-detector path, and the offsets are jittered during training while preserving the pixel centroid. Pathwise transmitted intensities are aggregated in the intensity domain using log-sum-exp. All experiments assume a point X-ray source because measured focal-spot sizes are unavailable.

Material and detector-plane parameterization. The material-coupled response $g_\phi(\mu)$ is represented by a globally shared monotone function with 12 sigmoid knots and 25 trainable parameters, evaluated on $u(\mu) = 1 - \exp(-\mu/\mu_{\text{ref}})$ with $\mu_{\text{ref}} = 0.3$. The detector-plane correction is parameterized independently for each projection view by a 16×16 non-negative field, stored in the log domain and bilinearly interpolated to the detector resolution.

Optimization. Each sub-ray is integrated using 320 samples with stratified perturbation during training. All models are optimized for 1000 epochs with 1024 rays per step using Adam and gradient-norm clipping at 1.0. The learning rates are 10^{-2} for the hash embeddings, 5×10^{-4} for the MLP, 5×10^{-3} for the material response, and 10^{-2} for the detector-plane fields. The hash and MLP learning rates are halved every 200 epochs, while the material-response and detector-plane learning rates remain fixed. The attenuation sparsity weight is linearly increased to 5×10^{-5}.

Table 4: Comparison with GH-NAF using its original configuration and PRISM-NAF's encoder and training configuration on the Lungman chest phantom.

			50 views			70 views		
Method	$\mathcal{H}$	$\mathcal{F}_\theta$	LPIPS↓	MANIQA↑	Dice↑	LPIPS↓	MANIQA↑	Dice↑
GH-NAF	Own	Own	0.588	0.189	0.632	0.567	0.244	0.727
GH-NAF	Ours	Own	0.594	0.212	0.628	0.568	0.244	0.668
GH-NAF	Ours	Ours	0.561	0.301	0.718	0.554	0.252	0.756
PRISM-NAF (Ours)	Ours	Ours	**0.544**	**0.365**	**0.737**	**0.522**	**0.354**	**0.803**

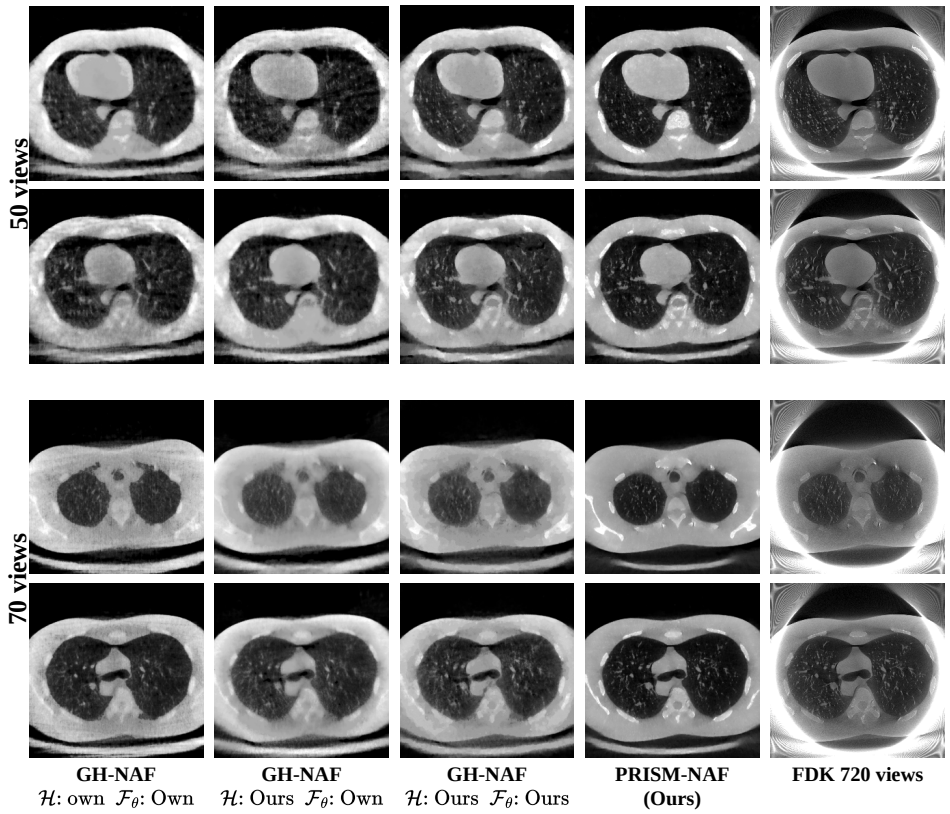


Figure 10: Qualitative comparison with GH-NAF using its original configuration and variants incorporating PRISM-NAF's hash encoder and network architecture on the Lungman chest phantom. All images use the same display window.

A.3 Rationale for Material-Coupled Response

Relation to existing correction models. Park et al. (Park et al., 2025) jointly predict attenuation and a beam-hardening factor, while GH-NAF (Oh et al., 2026) uses an auxiliary discrepancy field. Their nonlinear correction can be expressed in the common form

$$P = M - \psi(S), \qquad \psi(S) = \log\left(\frac{\sinh S}{S}\right), \tag{15}$$

where M is the attenuation line integral and S is the integrated correction field, with any global correction scale absorbed into its definition. This formulation specifies how the two fields contribute to a projection, but does not impose a shared functional relationship between their local values.

Why couple the response to attenuation? Under sparse-view supervision, compensating adjustments to attenuation and the auxiliary field can produce similar projection predictions, leaving their separation weakly constrained. MCR restricts this freedom through

$$\sigma_{\mathrm{BH}}(\mathbf{x}) = \mu(\mathbf{x}) g_\phi(\mu(\mathbf{x}))\,, \tag{16}$$

where g_ϕ is a globally shared, non-negative, non-decreasing function. The multiplicative factor makes the spectral spread vanish as attenuation approaches zero. The shared function assigns the same response to locations with equal attenuation, preventing the spectral field from varying independently at those locations. Its non-decreasing parameterization further imposes an ordering prior on the relative spectral response.

MCR therefore retains the established nonlinear rendering form while constraining its latent parameterization. Attenuation serves as a proxy for material-dependent behavior; the shared monotone relationship is an effective regularizing assumption, not a universal material law or a guarantee of unique physical recovery.

B Ablation Study

B.1 Complementarity of beam-hardening and scatter modeling

Effect of the detector-plane correction. Fig. 8 demonstrates the effect of DPS on broad intensity nonuniformity that remains after the pathwise spectral correction. Without DPS, the reconstructed soft-tissue region exhibits a clear radial intensity bias at both 50 and 70 views. Introducing DPS substantially suppresses this broad shading while preserving the local anatomical structures, producing visibly more uniform soft-tissue attenuation (Fig. 8a).

The radial analysis in Fig. 8b confirms this observation quantitatively. With CAA and MCR alone, the homogeneous soft-tissue region exhibits $+6.5\,\%$ cupping, whereas adding DPS reduces the radial bias to $-1.8\,\%$, yielding an approximately flat attenuation profile. Notably, the 720-view FDK reference itself exhibits $+13.6\,\%$ cupping because it is reconstructed from the same polychromatic and scatter-contaminated measurements. Therefore, the flattening produced by DPS is not expected to improve every FDK-referenced metric, despite correcting the within-material nonuniformity in the physically expected direction.

Fig. 8c further shows that DPS actively changes the low-frequency structure of the detector-domain mismatch rather than behaving as an inactive auxiliary variable. The residual becomes spatially redistributed after introducing the coarse detector-plane field. Its low-frequency RMS does not decrease, indicating that DPS should not be interpreted as directly recovering or subtracting the true physical scatter distribution. Taken together, the reconstruction-space and projection-domain diagnostics show that DPS provides an effective low-bandwidth detector-plane correction that substantially improves soft-tissue uniformity, while its learned field remains an effective nuisance variable rather than a calibrated scatter estimate.

B.2 Effect of the Number of Coherent Sub-rays

We examine the sensitivity of CAA to the number of coherent sub-rays $K \in \{2, 3, 4, 6\}$ while keeping the network architecture, optimization schedule, and all other rendering components fixed.

As shown in Fig. 9, increasing K progressively improves the spatial coverage of the detector-pixel area. With fewer sub-rays, localized intensity leakage and boundary nonuniformity remain visible in the highlighted soft-tissue region. These artifacts are gradually reduced as K increases, and $K = 6$ produces the cleanest and most spatially coherent reconstruction among the evaluated settings. This trend indicates that denser aperture sampling provides a more stable approximation of the area measurement. We therefore use $K = 6$ for CAA in all main experiments.

B.3 Comparison with GH-NAF

GH-NAF (Oh et al., 2026) also incorporates beam-hardening effects into its projection model, making it a relevant baseline for assessing the contribution of our measurement formulation. To determine whether the performance advantage of PRISM-NAF can be attributed to differences in the hash encoding or network architecture, GH-NAF is evaluated under progressively matched representations. The first variant adopts our hash encoder, decoder width, and training configuration; the second additionally incorporates our footprint-aware feature filtering. Both variants retain GH-NAF's original rendering formulation, output parameterization, and loss functions.

As shown in Table 4, matching the encoder and decoder does not eliminate the performance gap. Although adopting the complete representation improves GH-NAF, PRISM-NAF maintains lower LPIPS and higher MANIQA and vessel Dice at both evaluated view counts. The qualitative comparisons in Fig. 10 support these findings. The persistent advantage under matched representations indicates that the improvements of PRISM-NAF cannot be attributed solely to its hash encoding or network architecture, supporting the contribution of its detector-aware measurement model.

B.4 Held-Out Projection Consistency

The reference-based metrics in the main paper are computed against a 720-view FDK reconstruction. This reconstruction may contain beam-hardening cupping, noise, and intensity-scaling bias, which can affect the evaluation independently of the accuracy of the recovered attenuation field. To assess consistency with measured data directly, projection-domain evaluation is performed on 50 held-out views from the 720-view Lungman chest scan. These views are excluded from all training subsets, providing a common evaluation set for the 50-, 60-, 70-, and 80-view reconstructions. Each trained model predicts the held-out projections using its native forward operator, the calibrated acquisition geometry, and the same preprocessing as in training. RMSE is computed over detector pixels with positive measured line integrals, without additional spatial masking, post-hoc rescaling, or parameter fitting. For PRISM-NAF, predictions are generated using the shared CAA+MCR forward model; DPS is omitted because its parameters are defined only for observed training views.

Table 5: Held-out projection consistency on the Lungman chest phantom (RMSE in units of 10^{-3}).

Method	50 views	60 views	70 views	80 views
NAF (Zha et al., 2022)	4.94 ± 2.31	5.18 ± 2.45	4.41 ± 1.92	4.34 ± 1.81
SAX-NeRF (Cai et al., 2024)	1.64 ± 0.81	1.51 ± 0.74	3.76 ± 1.40	2.18 ± 0.92
GH-NAF (Oh et al., 2026)	1.20 ± 0.48	1.41 ± 0.53	1.21 ± 0.43	1.19 ± 0.59
Vol3DGS (Talegaonkar et al., 2025)	1.14 ± 0.56	1.11 ± 0.54	1.12 ± 0.54	1.11 ± 0.53
FaCT-GS (Pieta et al., 2026)	1.04 ± 0.63	1.01 ± 0.59	1.00 ± 0.56	1.01 ± 0.54
R^2-Gaussian (Zha et al., 2024)	$\mathbf{1.02} \pm 0.60$	$\mathbf{0.99} \pm 0.58$	$\mathbf{1.00} \pm 0.60$	$\mathbf{0.98} \pm 0.59$
PRISM-NAF (Ours)	1.29 ± 0.71	1.22 ± 0.65	1.22 ± 0.74	1.02 ± 0.53

Global projection consistency. Table 5 shows that R^2-Gaussian and FaCT-GS achieve lower global projection RMSE than PRISM-NAF across the evaluated settings. The error of PRISM-NAF decreases from 1.29×10^{-3} at 50 views to 1.02×10^{-3} at 80 views, approaching the corresponding R^2-Gaussian result of 0.98×10^{-3}. However, lower global projection RMSE does not necessarily imply greater structural fidelity in the reconstructed volume. This metric aggregates residuals across detector pixels without distinguishing their anatomical significance, so localized discrepancies associated with fine structures may have a limited contribution to the overall score.

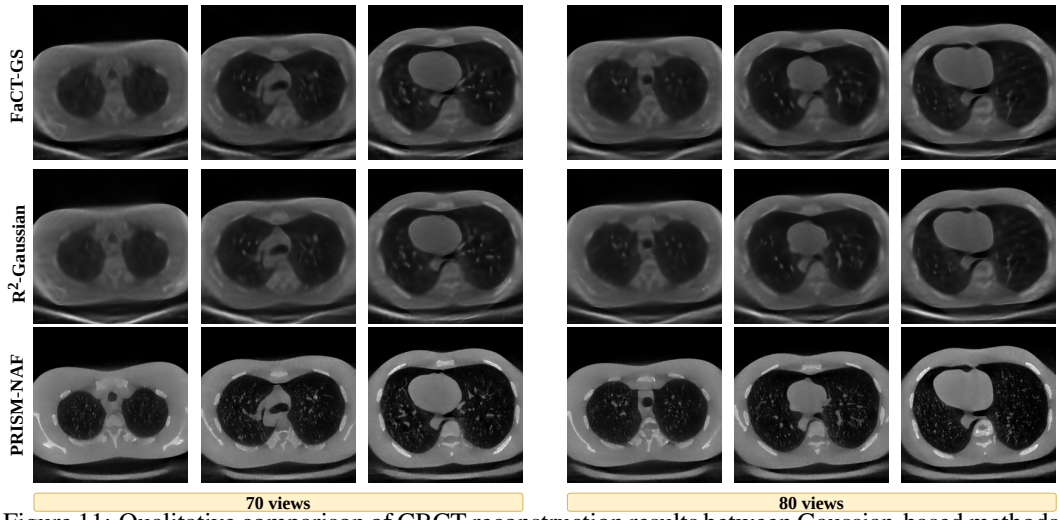


Figure 11: Qualitative comparison of CBCT reconstruction results between Gaussian-based methods and PRISM-NAF on the Lungman chest phantom at 70 (left) and 80 (right) views.

Relation to structural preservation. Fig. 11 illustrates this distinction through a direct comparison of the reconstructed CT images. The Gaussian-based reconstructions exhibit a smoother pulmonary appearance, accompanied by reduced visibility of thin vessels and less distinct tissue boundaries. PRISM-NAF retains these fine structures more clearly despite its slightly higher global projection RMSE. This observation is consistent with the vascular-structure evaluation in the main paper and suggests that improved average projection agreement can coexist with reduced preservation of fine anatomical detail. Held-out projection RMSE and image-domain structural assessment therefore capture complementary aspects of reconstruction quality; the former alone is insufficient to determine which reconstruction better preserves pulmonary anatomy.