

Splat-Based Metal Artifact Reduction in Cone-Beam CT via Compact Attenuation Modeling

Supplementary Material

1. Derivation of Backward Propagation

We implement a custom CUDA [1] pipeline that enables automatic differentiation for both the discrete polychromatic forward projection and reconstruction models, both of which are formulated using Gaussian primitives.

1.1. Projection Model

Our polychromatic forward projection is expressed as Equation (1):

$$P(\hat{\mathbf{x}}) = -\log \sum_{k=1}^N \eta_k \exp \left(-\sum_{i=1}^M s_i G_i^P(\hat{\mathbf{x}}) \delta_i \mu_\rho(u_i, k) \right) \quad (1)$$

where

$$s_i = \sqrt{\frac{2\pi|\tilde{\Sigma}_i|}{|\Sigma_i|}},$$

$$G_i^P(\hat{\mathbf{x}}) = \exp \left(-\frac{1}{2}(\hat{\mathbf{x}} - \hat{\mathbf{p}}_i)^T \tilde{\Sigma}_i^{-1}(\hat{\mathbf{x}} - \hat{\mathbf{p}}_i) \right),$$

$$\mu_\rho(u_i, k) = (1 - u_i)^2 \mathbf{b}_s(k) + 2(1 - u_i)u_i \mathbf{b}_m(k) + u_i^2 \mathbf{b}_f(k),$$

$$u_i \in [0, 1],$$

$$\sum_{k=1}^N \eta_k = 1.$$

The baseline method [10] comprises a `Rasterizer_preprocess`, which transforms each Gaussian's position and shape according to the detector pixel coordinates, and a `Rasterizer_render`, which computes the intensity projected onto each detector pixel. While we reuse the original implementation for the `Rasterizer_preprocess`, we design a new `Rasterizer_render` module and manually derive gradients with respect to each intermediate parameter, as

shown in Equations (2)–(5):

$$\left. \frac{d\mathcal{L}}{d\delta_i} \right|_{\hat{\mathbf{x}}} = \frac{d\mathcal{L}}{dP} \cdot \frac{dP}{d\delta_i} = \frac{d\mathcal{L}}{dP} \cdot \left(\frac{C}{A} \right) (s_i G_i^P), \quad (2)$$

$$\left. \frac{d\mathcal{L}}{ds_i} \right|_{\hat{\mathbf{x}}} = \frac{d\mathcal{L}}{dP} \cdot \frac{dP}{ds_i} = \frac{d\mathcal{L}}{dP} \cdot \left(\frac{C}{A} \right) (\delta_i G_i^P), \quad (3)$$

$$\left. \frac{d\mathcal{L}}{dG_i^P} \right|_{\hat{\mathbf{x}}} = \frac{d\mathcal{L}}{dP} \cdot \frac{dP}{dG_i^P} = \frac{d\mathcal{L}}{dP} \cdot \left(\frac{C}{A} \right) (s_i \delta_i), \quad (4)$$

$$\begin{aligned} \left. \frac{d\mathcal{L}}{du_i} \right|_{\hat{\mathbf{x}}} &= \frac{d\mathcal{L}}{dP} \left(\sum_{k=1}^N \frac{dP}{d\mu_\rho(u_i, k)} \cdot \frac{d\mu_\rho(u_i, k)}{du_i} \right) \\ &= \frac{d\mathcal{L}}{dP} \cdot \left(\frac{s_i G_i^P \delta_i}{A} \right) \sum_{k=1}^N B_k \cdot \frac{d\mu_\rho(u_i, k)}{du_i}, \end{aligned} \quad (5)$$

where

$$A = \sum_{k=1}^N B_k,$$

$$B_k = \eta_k \exp \left(-\sum_{i=1}^M s_i G_i^P(\hat{\mathbf{x}}) \delta_i \mu_\rho(u_i, k) \right),$$

$$C = \sum_{k=1}^N (B_k \mu_\rho(u_i, k)),$$

$$\begin{aligned} \frac{d\mu_\rho(u_i, k)}{du_i} &= 2 \left[(1 - u_i)(\mathbf{b}_m(k) - \mathbf{b}_s(k)) \right. \\ &\quad \left. + u_i(\mathbf{b}_f(k) - \mathbf{b}_m(k)) \right]. \end{aligned}$$

All these equations are then implemented in the CUDA backward function.

1.2. Reconstruction Model

Our reconstruction is expressed as Equation (6):

$$\mu(\mathbf{x}, k) = \sum_{i=1}^M G_i^R(\mathbf{x}) \delta_i \mu_\rho(u_i, k), \quad (6)$$

where

$$G_i^R(\mathbf{x}) = \exp \left(-\frac{1}{2}(\mathbf{x} - \mathbf{p}_i)^T \Sigma_i^{-1}(\mathbf{x} - \mathbf{p}_i) \right).$$

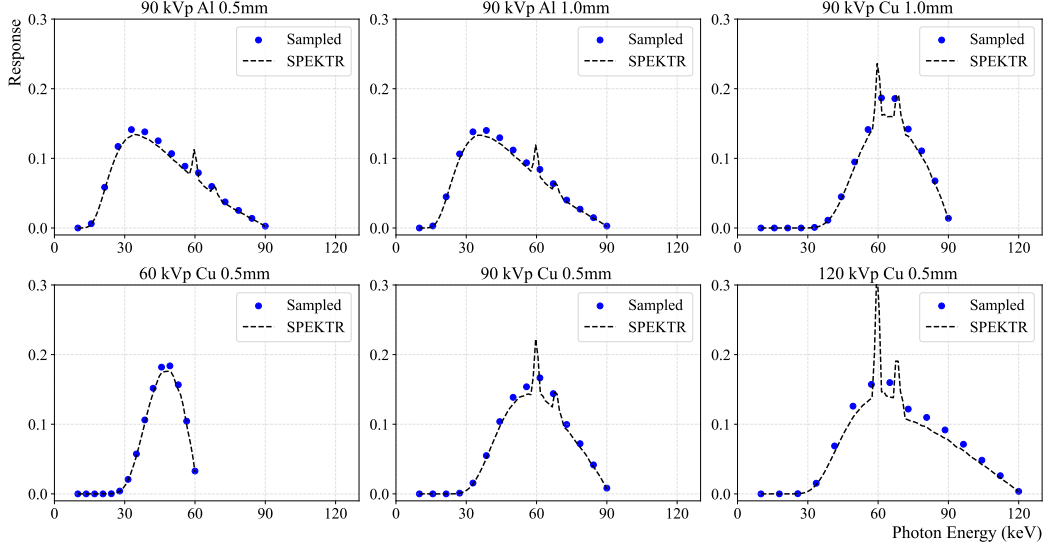


Figure 1. Energy spectra generated using the SPEKTR simulator for various CT system configurations. Each curve represents the normalized photon distribution corresponding to a specific combination of tube voltage and filtration settings. The blue dots denote the 15 uniformly sampled energy components used by our method.

The baseline method [10] comprises a `Voxelizer_preprocess`, which transforms each Gaussian’s position and shape according to the voxel grid coordinates, and a `Voxelizer_render`, which computes the intensity accumulated within each voxel. While we reuse the original implementation for the `Voxelizer_preprocess`, we design a new `Voxelizer_render` module and manually derive gradients with respect to each intermediate parameter, as shown in Equations (7)–(9):

$$\begin{aligned} \left. \frac{d\mathcal{L}}{d\delta_i} \right|_{\mathbf{x}} &= \sum_{k=1}^N \frac{d\mathcal{L}}{d\mu(k)} \cdot \frac{d\mu(k)}{d\delta_i} \\ &= \sum_{k=1}^N \frac{d\mathcal{L}}{d\mu(k)} \cdot G_i^R \mu_\rho(u_i, k), \end{aligned} \quad (7)$$

$$\begin{aligned} \left. \frac{d\mathcal{L}}{dG_i^R} \right|_{\mathbf{x}} &= \sum_{k=1}^N \frac{d\mathcal{L}}{d\mu(k)} \cdot \frac{d\mu(k)}{dG_i^R} \\ &= \sum_{k=1}^N \frac{d\mathcal{L}}{d\mu(k)} \cdot \delta_i \mu_\rho(u_i, k), \end{aligned} \quad (8)$$

$$\begin{aligned} \left. \frac{d\mathcal{L}}{du_i} \right|_{\mathbf{x}} &= \sum_{k=1}^N \frac{d\mathcal{L}}{d\mu(k)} \cdot \frac{d\mu(k)}{d\mu_\rho(u_i, k)} \cdot \frac{d\mu_\rho(u_i, k)}{du_i} \\ &= \sum_{k=1}^N \frac{d\mathcal{L}}{d\mu(k)} \cdot G_i^R \delta_i \cdot \frac{d\mu_\rho(u_i, k)}{du_i}. \end{aligned} \quad (9)$$

All these equations are then implemented in the CUDA backward function.

2. X-ray Spectrum

Our method relies on the X-ray spectrum response $\eta(E)$, which depends on the CT system configuration used during data acquisition—specifically, the tube voltage, filter material, and filter thickness. Following prior work [9], we employ the SPEKTR simulator [5] to generate polychromatic X-ray spectra at 1 keV resolution. The simulated spectra are subsequently resampled into 15 uniformly spaced energy components, matching the number of spectral bins used in our model. These interpolated spectral samples are then incorporated into our polychromatic forward projection formulation. Figure 1 visualizes the energy-dependent spectra for the CT settings considered in our experiments.

3. Scanning and Reconstruction Configuration

Tables 1–2 summarize the scanning and reconstruction configurations used for both the synthetic and real datasets. To assess the robustness and generality of our method, we evaluate performance across a wide range of acquisition settings, including variations in tube voltage, filtration, geometry, and object scale. These configurations closely follow the parameters of practical CBCT systems and ensure diverse conditions for validating our polychromatic forward-projection and reconstruction framework.

4. Fairness of Comparison

Polyner [9] and Park et al. [4] are joint reconstruction and metal artifact reduction methods originally developed for **2D fan-beam** geometry. To enable a fair comparison with our **3D cone-beam** method, we extend both im-

Table 1. Scanning and reconstruction parameters for the synthetic datasets. The left three columns correspond to the main evaluation scenes, and the four rightmost columns denote the additional settings used in the supplementary experiments, including variations in tube voltage and filtration. All other geometric parameters are kept consistent across the datasets.

Parameters	Synthetic Dataset						
	Lung	Teeth	Broccoli	Lung (supple)	Teeth (supple)	Pancreas (supple)	Pepper (supple)
Source voltage (kV)	90	←	←	60/90/120	←	90	←
Filter (mm)	Cu(1.0)	Al(0.5)	←	Cu(0.5)	←	Cu(1.0)	←
Metal material	Fe	Ti	Al	Fe	Ti	←	Fe
Volume resolution	256×256×256	←	←	←	←	←	←
Volume size (mm)	50.0×50.0×50.0	←	←	←	←	←	←
Angle range (°)	[0, 360)	←	←	←	←	←	←
Number of angles	720	←	←	←	←	←	←
Detector resolution	512×512	←	←	←	←	←	←
Detector size (mm)	145.485	←	←	←	←	←	←
Source-to-object distance (mm)	200.962	←	←	←	←	←	←
Source-to-detector distance (mm)	501.309	←	←	←	←	←	←

Table 2. Scanning and reconstruction parameters for the real datasets. Each object is scanned using high-resolution CBCT settings, with variations in object scale, and detector geometry inherent to the acquisition system. These diverse acquisition conditions enable a comprehensive evaluation of metal artifact behavior in real-world scenarios.

Parameters	Real Dataset					
	Walnut	Avocado	Chicken	Garlic	Enoki Mushroom	Blueberry
Source voltage (kV)	90	←	←	←	←	←
Source current (μA)	166	←	300	←	←	←
Filter (mm)	Al(1.0)	←	←	←	←	←
Metal material	Fe	←	←	←	←	←
Volume resolution	256×256×256	←	←	←	←	←
Volume size (mm)	40.0×40.0×40.0	60.0×60.0×60.0	120.0×120.0×120.0	80.0×80.0×80.0	←	←
Angle range (°)	[0, 360)	←	←	←	←	←
Number of angles	720	←	←	←	←	←
Detector resolution	512×512	←	←	←	←	←
Detector size (mm)	104.773	145.485	229.902	104.773	145.485	←
Source-to-object distance (mm)	200.962	←	334.942	←	←	←
Source-to-detector distance (mm)	501.309	←	←	←	←	←

Table 3. Quantitative comparison on additional synthetic scenes

Method	Pancreas (Ti)		Pepper (Fe)	
	PSNR3D	SSIM3D	PSNR3D	SSIM3D
FDK	23.99	0.974	11.92	0.894
LIMAR	26.76	0.985	16.33	0.952
NMAR	23.44	0.972	16.65	0.962
ACDNet	15.41	0.776	10.77	0.879
DICDNet	15.29	0.770	10.75	0.878
OSNet	15.20	0.766	10.63	0.875
Polyner	28.17	0.992	22.30	0.995
Park et al.	15.77	0.695	12.69	0.863
Ours	33.73	0.998	25.13	0.997

plementations to the cone-beam setting. LIMAR [2] and NMAR [3] are inherently applicable to 3D reconstruction and are therefore directly comparable. For learning-based methods (ACDNet [7], DICDNet [6], and OSCNet [8]), we use publicly available pre-trained models and apply them slice-wise to FDK-reconstructed volumes, following their original design.

5. Robustness to Tube Voltage Variation

To evaluate the stability of our method under different X-ray source energy conditions, we generated synthetic projection

images at 60, 90, and 120 kVp and assessed the corresponding reconstruction quality. The Lung and Teeth scenes were simulated using a 0.5 mm Cu filter to match realistic CBCT acquisition settings. As shown in Figure 2, our method maintains consistent reconstruction performance across different tube voltages, demonstrating strong robustness to variations in the underlying X-ray spectrum.

6. Additional Discussion

We compute 3D PSNR and SSIM to evaluate reconstruction quality on the synthetic dataset. For PSNR, the ground truth (GT) volume is defined using the effective energy described in the main paper, and—consistent with prior metal artifact reduction studies [4, 9]—the metric is computed over all regions excluding the metal mask. For real datasets, no reliable GT is available; therefore, we report only qualitative comparisons rather than quantitative scores.

7. Additional Synthetic Results

We conducted additional experiments on two newly introduced synthetic scenes: a pancreas scene with titanium (Ti)

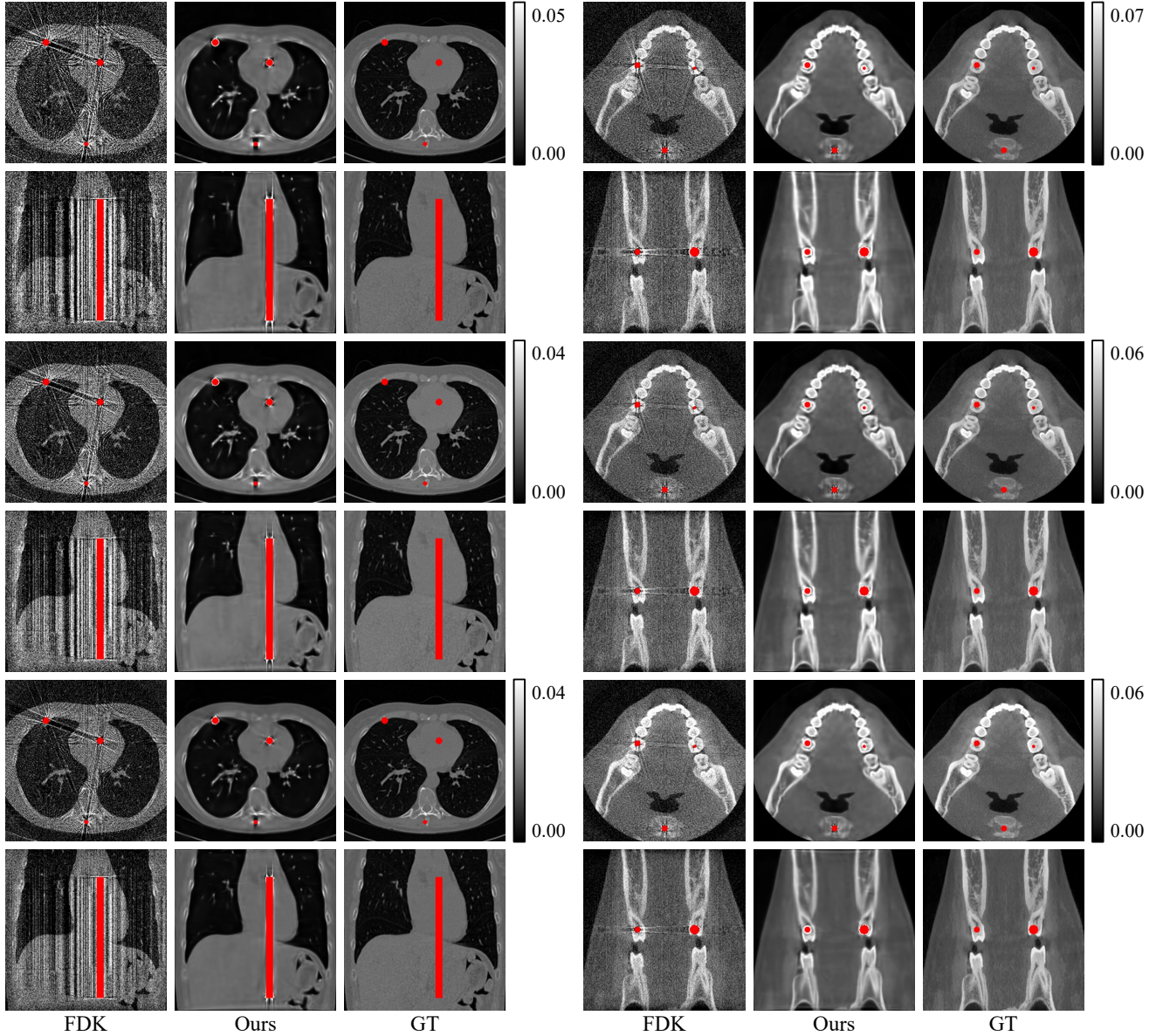


Figure 2. Evaluation of reconstruction robustness under varying tube voltages (60, 90, and 120 kVp). Despite substantial changes in the underlying X-ray spectrum, our method consistently produces high-quality reconstructions. Rows 1–2 show results at 60 kVp, rows 3–4 at 90 kVp, and rows 5–6 at 120 kVp.

and a pepper scene with iron (Fe). Quantitative results, summarized in Table 3, show consistent improvements over all baseline methods. Qualitative results are shown in Figure 4.

8. Additional Real Results

We provide additional orthogonal slices for the real-data results in Figure 3.

References

- [1] Design Guide. Cuda c programming guide. *NVIDIA*, July, 29:31, 2013. 1
- [2] Willi A Kalender, Robert Hebel, and Johannes Ebersberger. Reduction of ct artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987. 3
- [3] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß. Normalized metal artifact reduction (nmar) in computed tomography. *Medical physics*, 37(10):5482–5493, 2010. 3
- [4] Hyoungh Suk Park, Jin Keun Seo, and Kiwan Jeon. Implicit

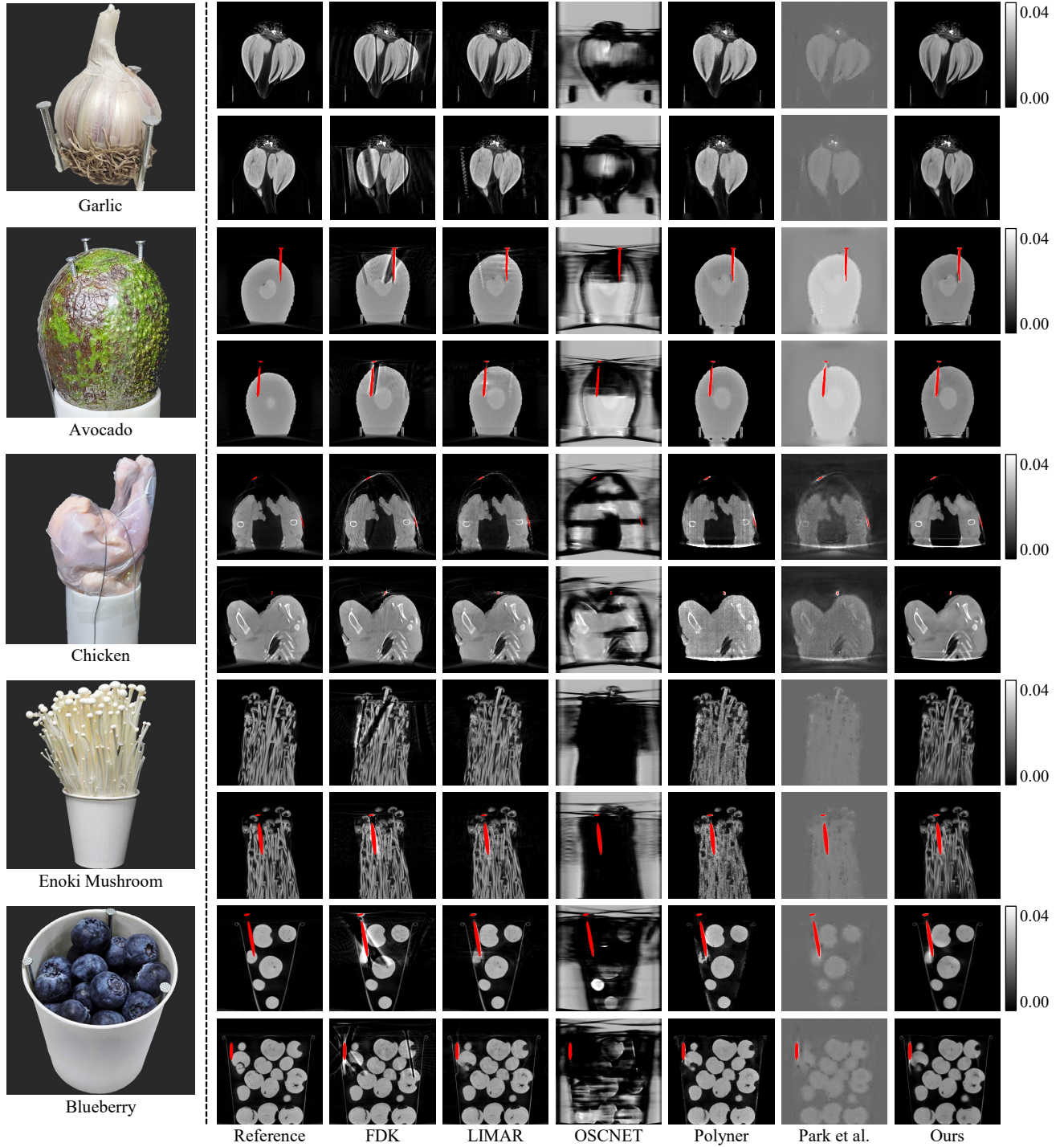


Figure 3. Qualitative comparison on additional slices for real scenes.

neural representation-based method for metal-induced beam hardening artifact reduction in x-ray ct imaging. *Medical Physics*. **2**, **3**

- [5] J Punnoose, J Xu, A Sisniega, W Zbijewski, and JH Siewersden. Technical note: spektr 3.0—a computational tool for x-ray spectrum. *Medical Physics*, 43(8):4711–4717, 2016. **2**

- [6] Hong Wang, Yuexiang Li, Nanjun He, Kai Ma, Deyu Meng, and Yefeng Zheng. Dicednet: Deep interpretable convolutional dictionary network for metal artifact reduction in ct images. *IEEE Transactions on Medical Imaging*, 41(4):869–880, 2022. **3**

- [7] Hong Wang, Yuexiang Li, Deyu Meng, and Yefeng Zheng.

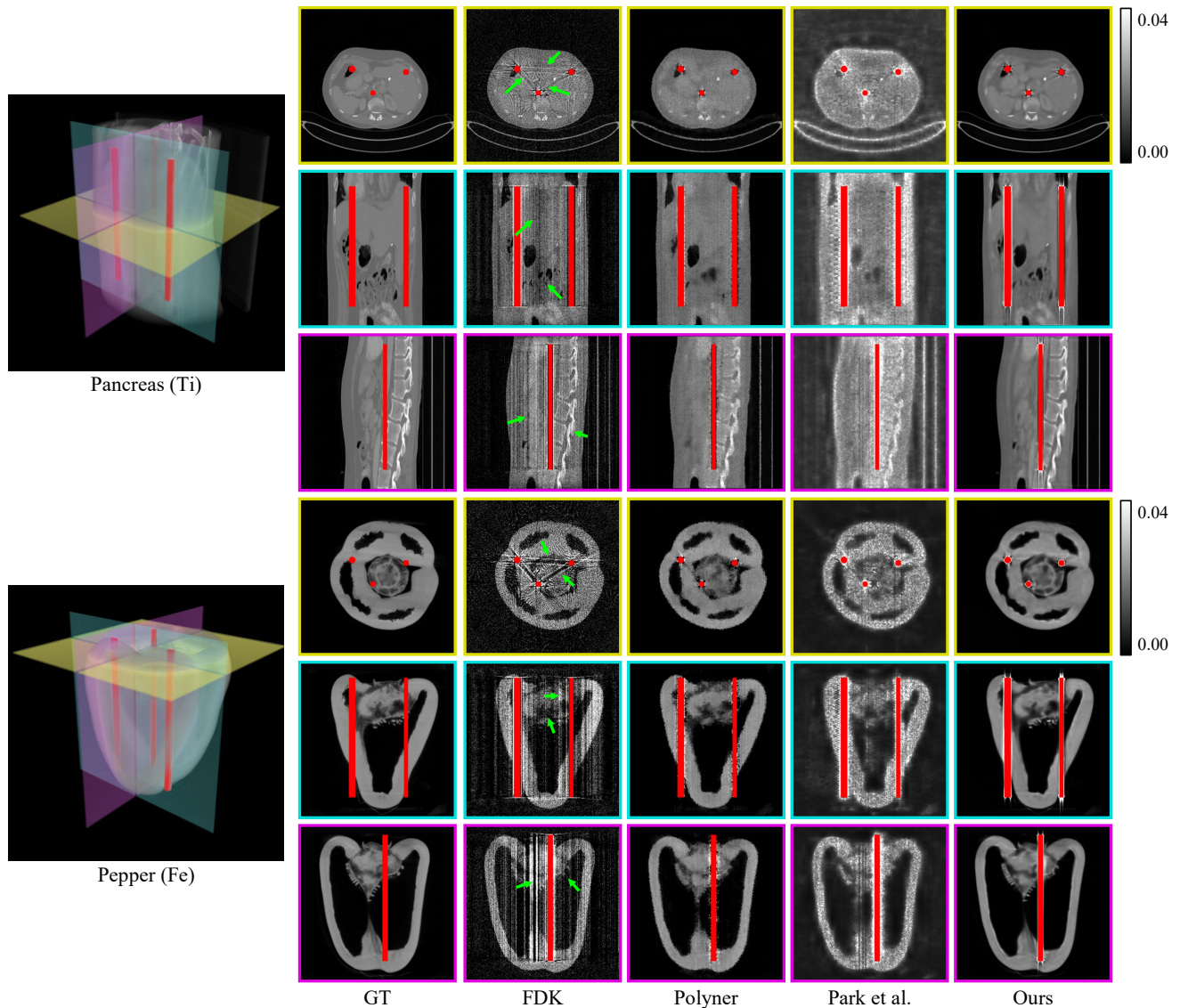


Figure 4. Qualitative comparison on additional synthetic scenes.

- Adaptive convolutional dictionary network for ct metal artifact reduction. In *The 31st International Joint Conference on Artificial Intelligence*. IEEE, 2022. 3
- [8] Hong Wang, Qi Xie, Dong Zeng, Jianhua Ma, Deyu Meng, and Yefeng Zheng. Oscnet: Orientation-shared convolutional network for ct metal artifact learning. *IEEE Transactions on Medical Imaging*, 43(1):489–502, 2024. 3
- [9] Qing Wu, Lixuan Chen, Ce Wang, Hongjiang Wei, S Kevin Zhou, Jingyi Yu, and Yuyao Zhang. Unsupervised polychromatic neural representation for CT metal artifact reduction. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 2, 3
- [10] Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. R²-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*,

2024. 1, 2