Unsupervised Multi-Parameter Inverse Solving for Reducing Ring Artifacts in 3D X-Ray CBCT

Qing Wu[1], Hongjiang Wei[2], Jingyi Yu[1], Yuyao Zhang[1,3,*]

[1]School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
[2]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200127, China
[3]State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
wuqing@shanghaitech.edu.cn, hongjiang.wei@sjtu.edu.cn, yujingyi@shanghaitech.edu.cn, zhangyy8@shanghaitech.edu.cn

Abstract

Ring artifacts are prevalent in 3D cone-beam computed tomography (CBCT) due to non-ideal responses of X-ray detectors, substantially affecting image quality and diagnostic reliability. Existing state-of-the-art (SOTA) ring artifact reduction (RAR) methods rely on supervised learning with large-scale paired CT datasets. While effective in-domain, supervised methods tend to struggle to fully capture the physical characteristics of ring artifacts, leading to pronounced performance drops in complex real-world acquisitions. Moreover, their scalability to 3D CBCT is limited by high memory demands. In this work, we propose Riner, a new unsupervised RAR method. Based on a theoretical analysis of ring artifact formation, we reformulate RAR as a multi-parameter inverse problem, where the non-ideal responses of X-ray detectors are parameterized as solvable physical variables. Using a new differentiable forward model, Riner can jointly learn the implicit neural representation of artifact-free images and estimate the physical parameters directly from CT measurements, without external training data. Additionally, Riner is memory-friendly due to its ray-based optimization, enhancing its usability in large-scale 3D CBCT. Experiments on both simulated and real-world datasets show Riner outperforms existing SOTA supervised methods.

Code — https://github.com/iwuqing/Riner

1 Introduction

3D cone-beam computed tomography (CBCT) enables detailed visualization of internal structures, providing critical information for applications ranging from medical diagnosis and biological research to materials science (Scarfe and Farman 2008; Venkatesh and Elluru 2017). CBCT acquisition involves a series of 2D projections captured by a 2D X-ray detector array at varying angles, as illustrated in Fig. 1. However, due to inherent physical hardware limitations (Boas, Fleischmann et al. 2012), individual detectors within the 2D array often struggle to maintain ideal, consistent signal responses, especially in μm-scale imaging using micro-CT scanners (Yousuf and Asaduzzaman 2010; Rashid, Lee, and Hasan 2012). These non-ideal responses lead to nonlinear physical distortions in CT measurement data, suffering from severe ring artifacts when traditional linear reconstruction algorithms like FDK (Feldkamp, Davis, and Kress 1984), are applied. Such ring artifacts can markedly degrade the quality and reliability of resulting CT images.

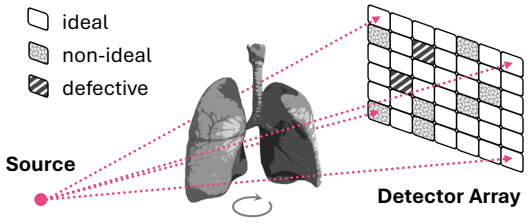

Figure 1: Illustration of 3D CBCT acquisition. An X-ray source emits cone-shaped beams that pass through objects and are received by a 2D detector array. Due to hardware limitations (Boas, Fleischmann et al. 2012), the X-ray detectors can be ideal (accurate response), non-ideal (fluctuating response), or defective (no response).

Current supervised deep learning (DL) methods (Chang et al. 2020; Trapp et al. 2022; Wang et al. 2023, 2022) represent the state-of-the-art (SOTA) in the CT reconstructions. These methods train neural networks, such as UNet (Ronneberger, Fischer, and Brox 2015) and ViT (Dosovitskiy et al. 2020), on large-scale paired datasets to map artifact-corrupted images to artifact-free outputs. While promising, they often face two key limitations: *1) High data collection costs and limited generalization*. Collecting paired CT images is often prohibitively expensive, especially for medical and biological applications. As a result, most models are trained on simulated datasets generated from clean CT images (Fang, Li, and Chen 2020; Trapp et al. 2022). However, simulation errors (*e.g.*, differences in image resolution, detector characteristics, and CT geometry) can substantially degrade performance on unseen real-world data; *2) Scalability issues in 3D CBCT.* Supervised DL models typically are difficult to scale to 3D due to high memory requirements. Adapting 2D models by processing slice-by-slice often causes discontinuities along the Z-axis, compromising 3D volumetric consistency. These limitations confine the use of supervised DL methods to constrained research settings, limiting their broader applicability.

*Corresponding author.

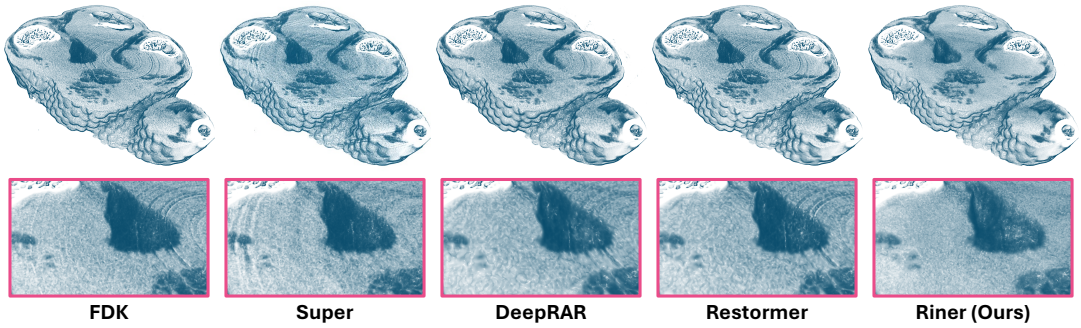

Figure 2: We propose Riner, an unsupervised method for reducing ring artifacts in 3D X-ray CBCT imaging. On a real-world *Chicken foot* sample (see Appendix) with image dimension of 512×512×80 and an ultra-high resolution of 60×60×60 μm^3, acquired by a commercial Bruker SKYSCAN 1276 micro-CT scanner, our Riner effectively removes ring artifacts and reconstructs high-quality CT volumes, outperforming SOTA model-based algorithms (FDK (Feldkamp, Davis, and Kress 1984) and Super (Vo, Atwood, and Drakopoulos 2018)) and SOTA supervised deep learning models (DeepRAR (Trapp et al. 2022) and Restormer (Zamir et al. 2022))

In this paper, we propose Riner, a new unsupervised CT ring artifact reduction (RAR) method. Unlike end-to-end supervised learning, Riner reformulates RAR as a multi-parameter inverse problem grounded in X-ray CT physics. Theoretically, we analyze the physical origins of ring artifacts and identify two key nonlinear physical effects: inconsistent responses and invalid measurements. Motivated by the theoretical results, we introduce solvable physical parameters to characterize these two nonlinear effects. By leveraging a new differentiable forward model, Riner jointly learns the implicit neural representation (INR) of high-quality images and estimates the physical parameters directly from raw CT measurements. Estimating these parameters effectively corrects the two physical effects, improving image quality. While the inherent spectral bias of INR (Rahaman et al. 2019) helps regularize the ill-posed inverse problem, enabling high-quality CT reconstructions. Moreover, Riner uses a ray-based optimization scheme, making it memory-efficient and thus enhancing its usability in large-scale 3D CBCT settings.

We evaluate the proposed Riner on five datasets, including three simulated and two real-world cases, under both 2D fan-beam (FB) CT and 3D CBCT protocols. Experimental results show that our unsupervised method consistently outperforms existing SOTA techniques across all settings. Extensive ablation studies further validate the effectiveness of each key component in Riner. *To the best of our knowledge, Riner is the first unsupervised RAR method to outperform supervised approaches.*

2 Related Work

Advances for Ring Artifact Reduction Ring artifact reduction (RAR) remains a long-standing and challenging problem in X-ray CT imaging (Chen et al. 2024). Traditional hardware-based correction methods (Zhu et al. 2013) can deliver satisfactory performance, but they typically require specialized hardware designs and significantly increase system costs. To overcome these limitations, many studies have explored software-based solutions. Early model-based approaches (Rivers 1998; Münch et al. 2009; Vo, Atwood, and Drakopoulos 2018) rely on handcrafted filters to correct CT measurements, but their performance is often limited due to insufficient priors. More recently, supervised deep learning (DL) methods (Chang et al. 2020; Fang, Li, and Chen 2020; Trapp et al. 2022; Wang et al. 2023, 2022; Zamir et al. 2022; Chen et al. 2022; Guo et al. 2024b,a) have achieved SOTA results by learning data-driven priors from large-scale paired datasets. However, these end-to-end models often struggle to fully capture the underlying physical properties of ring artifacts, leading to poor generalization on out-of-domain (OOD) data. In contrast, we reformulate RAR as a multi-parameter inverse problem within a physics-driven framework to improve robustness and generalization.

Multi-Parameter Inverse Problem Multi-parameter inverse problems (MPIP) involve estimating multiple unknown parameters of a complex system from measurement data. By incorporating domain-specific physical priors, MPIP-based approaches have shown strong potential in a range of applications, including physics (Dalla Riva, Luzzini, and Musolino 2022; Cristofol et al. 2011; Zhu et al. 2023), partial differential equations (Chen and Liu 1984), and remote sensing (Kostsov 2015). Motivated by the success of this framework, we conduct a theoretical analysis of the physical mechanisms underlying CT ring artifact formation and reformulate the RAR problem as a multi-parameter inverse problem. This leads to Riner, a new unsupervised approach that improves CT reconstruction quality without relying on external training data.

Implicit Neural Representation for CT Imaging Implicit neural representation (INR) is an unsupervised unified framework for solving visual inverse problems (Essakine et al. 2024), where a multi-layer perceptron (MLP) is optimized to represent continuous signals of interest. Neural radiance fields (Mildenhall et al. 2021) primarily integrate volume rendering techniques into INR, achieving breakthrough progress in novel view synthesis. By incorporating a differentiable Radon transform to simulate the CT acquisition

process, INR has been successfully extended to various CT tasks, such as sparse-view reconstruction (Shen, Pauly, and Xing 2022; Wu et al. 2023b; Zha, Zhang, and Li 2022; Rückert et al. 2022; Du et al. 2024; Zang et al. 2021), dynamic CT (Reed et al. 2021; Zhang et al. 2023; Birklein et al. 2023), and metal artifact reduction (Wu et al. 2023a, 2024; Lee, Ahn, and Baek 2024). However, existing INR-based methods do not address the challenging RAR problem, as they typically assume ideal X-ray detectors and thus ignore the nonlinear physical distortions caused by hardware imperfections. In contrast, our method introduces a new physical formulation that explicitly models non-ideal detector behavior by estimating additional parameters, enabling effective RAR reconstructions.

3 Proposed Method

3.1 Theoretical Analysis for CT Ring Artifacts

Lambert-Beer's law (Lambert 1760; Beer 1852) formulates the physical process of X-ray CT acquisition as below:

$$I(\theta, s) = \alpha_s \cdot I_0 \cdot \mathrm{e}^{-\int_{L(\theta,s)} \mu(\mathbf{x}) \mathrm{d}\mathbf{x}}, \quad \forall L(\theta, s) \in \mathbf{\Pi}, \quad (1)$$

where $\mathbf{\Pi}$ denotes the set of all X-rays, $L(\theta, s)$ is an X-ray received by an X-ray detector s at a projection view θ, and $I(\theta, s)$ is the number of received photons. I_0 is the number of photons emitted by an X-ray source. $\alpha_s \geq 0$ is the response factor of the detector s, measuring its detection ability. $\mu(\mathbf{x})$ represents the attenuation coefficient of observed objects at the position $\mathbf{x}$.

By assuming ideal X-ray detectors (*i.e.*, $\alpha_s = 1, \forall s$), the CT measurement data is given by:

$$\rho(\theta, s) \triangleq -\ln \frac{I(\theta, s)}{\alpha \cdot I_0} = \int_{L(\theta,s)} \mu(\mathbf{x}) \mathrm{d}\mathbf{x}. \quad (2)$$

Then, one can reconstruct high-quality CT images μ from the ideal measurement data ρ using linear algorithms, such as FDK (Feldkamp, Davis, and Kress 1984).

However, in practice, the responses of detectors are often non-ideal (*i.e.*, $\alpha_s \neq 1, \forall s$) and difficult to maintain consistently (*i.e.*, $\alpha_{s_1} \neq \alpha_{s_2}, \forall s_1, s_2 \in S$, where S denotes the set of the detectors) due to hardware limitations (Boas, Fleischmann et al. 2012). The measurement can be defined by:

$$\begin{aligned} \widetilde{\rho}(\theta, s) &\triangleq -\ln \frac{I(\theta, s)}{I_0} \\ &= \begin{cases} -\ln \alpha_s + \int_{L(\theta,s)} \mu(\mathbf{x}) \mathrm{d}\mathbf{x}, & \text{if } \alpha_s > 0 \\ \texttt{NaN}, & \text{if } \alpha_s = 0. \end{cases} \end{aligned} \quad (3)$$

From Eq. 3, we observe two nonlinear physical effects:

- **Inconsistent Responses (IR)**, where non-ideal fluctuating (*i.e.*, $\alpha_s > 0$ and $\alpha_s \neq 1$), inconsistent (*i.e.*, $\alpha_{s_1} \neq \alpha_{s_2}, \forall s_1, s_2 \in S$) detectors introduce an extra nonlinear term $-\ln \alpha_s$;
- **Invalid Measurements (IM)**, where the partial measurements by defective (*i.e.*, $\alpha_s = 0$) detectors are invalid.

These two physical distortions make the inverse problem of reconstructing the unknown CT images μ from the real measurements $\widetilde{\rho}$ inherently nonlinear. Therefore, traditional linear algorithms, such as FDK (Feldkamp, Davis, and Kress 1984), are prone to severe ring artifacts in CT results.

3.2 Problem Formulation

Existing supervised DL methods (Trapp et al. 2022; Zamir et al. 2022) mostly treat RAR as a post-processing denoising task. They first construct paired datasets $\{(\mu, \hat{\mu})_i\}_{i=1}^N$ by simulating artifact-corrupted images $\hat{\mu}$ from clean CT scans μ, and train networks to learn a mapping $\mu = \phi(\hat{\mu})$. While effective in-domain, supervised learning lacks physical modeling of artifact formation, leading to poor generalization on OOD data. Moreover, their high memory demands limit applicability to 3D CBCT.

In contrast, we formulate RAR as a multi-parameter inverse problem grounded in CT physics. As highlighted in our theoretical analysis (Sec. 3.1), the IR and IM effects are fundamental causes of CT ring artifacts. To address them, we introduce two vectors: $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_s, \ldots] \in \mathbb{R}^{|S|}$ to model detector response factors, and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_s, \ldots] \in \mathbb{R}^{|S|}$ to eliminate invalid measurements. Joint estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ enables effective correction of both IR and IM, resulting in artifact-free reconstructions. Thus, our goal is to recover the image μ and estimate $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ from measurement data $\widetilde{\rho}$ in an *unsupervised* way. Formally, we aim to solve the following optimization problem:

$$\begin{aligned} \mu^*,\ \boldsymbol{\alpha}^*,\ \boldsymbol{\beta}^* = \underset{\mu,\ \boldsymbol{\alpha},\ \boldsymbol{\beta}}{\arg\min} \quad & \mathcal{L}(\widetilde{\rho},\ \widehat{\rho}) \\ \text{subject to} \quad & \widehat{\rho} = \boldsymbol{A}(\mu,\ \boldsymbol{\alpha},\ \boldsymbol{\beta}) \\ & \boldsymbol{\alpha} \succeq \mathbf{0}, \quad \boldsymbol{\beta} \in \{0, 1\}^{|S|}, \end{aligned} \quad (4)$$

where $\mu^*,\ \boldsymbol{\alpha}^*,\ \boldsymbol{\beta}^*$ denote the underlying optimal solutions, $\widehat{\rho}$ represents estimated measurement, $\mathcal{L}$ is the loss function defined in Eq. (7), and $\boldsymbol{A}$ is a differentiable physical forward model that characterizes the CT acquisition process under the RAR setting, as detailed in Sec. 3.3.

Parametrizing CT Images via INR Due to the introduction of the physical parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, this inverse problem (Eq. 4) is highly ill-posed, where multiple feasible solutions exist. Our Riner introduces INR to reconstruct the underlying CT image μ. Specifically, we represent the CT image as a continuous function of spatial coordinates parametrized by an MLP network as follows:

$$f_{\mathbf{\Phi}} : \mathbf{x} = (x, y, z) \in \mathbb{R}^3 \rightarrow \mu(\mathbf{x}) \in \mathbb{R}, \quad (5)$$

where $\mathbf{\Phi}$ denote the learnable weights of the MLP network. Numerous studies (Mildenhall et al. 2021; Tancik et al. 2020; Rahaman et al. 2019; Sitzmann et al. 2020) have demonstrated that neural networks, when used to approximate images, typically capture low-frequency global structures first and then gradually recover high-frequency details. We aim to leverage this inherent learning prior to constrain the solution space of the ill-posed inverse problem (Eq. 4), resulting in excellent CT reconstructions.

3.3 Differential Physical Forward Model

Following the problem formulation in Eq. (4) and Eq. (5), we optimize the MLP network $f_{\mathbf{\Phi}}$ to map spatial coordinates $\mathbf{x}$ to CT intensities $\mu(\mathbf{x})$. To achieve this, we introduce a new *differential* physical forward model, which converts

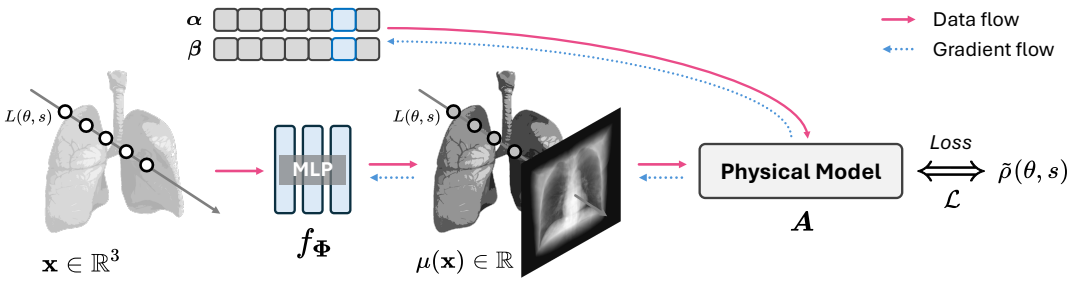

Figure 3: Overview of the proposed Riner method. Given raw measurements $\tilde{\rho}(\theta, s)$, an MLP network $f_{\boldsymbol{\Phi}}$ receives multiple spatial coordinates $\mathbf{x}$ along an X-ray path $L(\theta, s)$ as input and predicts the corresponding CT intensities $\mu(\mathbf{x}) = f_{\boldsymbol{\Phi}}(\mathbf{x})$. These predicted intensities, $\mu(\mathbf{x}), \forall \mathbf{x} \in L(\theta, s)$, the response factor α_s, and the mask β_s are then used to generate estimated measurements $\widehat{\rho}(\theta, s)$ via a differentiable physical model $\boldsymbol{A}$ (Eq. 6). Finally, the MLP network $f_{\boldsymbol{\Phi}}$ and parameters α_s, β_s jointly are optimized by minimizing the loss function $\mathcal{L}$ (Eq. 7) *without* using any external data.

the MLP-predicted intensities $\mu(\mathbf{x}) = f_{\boldsymbol{\Phi}}(\mathbf{x})$ and the variables (α_s, β_s) into estimated measurements $\widehat{\rho}(\theta, s)$, while allowing for gradient back-propagation. Building on the theoretical results in Eq. (3), our forward model $\boldsymbol{A}$ is defined as follows:

$$\boldsymbol{A}: \quad \widehat{\rho}(\theta, s) = \Big[-\ln \alpha_s + \sum_{\mathbf{x} \in L(\theta, s)} \mu(\mathbf{x}) \cdot \Delta \mathbf{x} \Big] \cdot \beta_s, \quad \text{with} \quad \alpha_s = \max(\alpha_s^0,\, \epsilon), \quad \beta_s = \sigma(\beta_s^0), \tag{6}$$

where α_s^0 is initialized to 1, and $\epsilon = 1 \times 10^{-8}$ is a small constant. These ensure the non-negativity of the response factor α_s and the validity of the logarithmic computation. $\Delta \mathbf{x} = \|\mathbf{x}_a - \mathbf{x}_b\|_2$ represents the Euclidean distance between adjacent coordinates $\mathbf{x}_a, \mathbf{x}_b \in L(\theta, s)$ along the X-ray path. While the variable β_s acts as a binary mask, where $\beta_s = 0$ for defective detectors and 1 otherwise. It is generated using the Sigmoid function with an initial value of $\beta_s^0 = 1$.

Superior to the conventional linear integral model (Eq. 2) (Shen, Pauly, and Xing 2022; Zha, Zhang, and Li 2022; Wu et al. 2023b; Zang et al. 2021; Lin et al. 2023), our model effectively addresses the CT RAR problem by introducing additional physical parameters for more accurate acquisition modeling. Specifically, it includes **1) Paramerization of the detector responses**, where our model explicitly estimates the response factor α_s to alleviate the IR effect; **2) Removal of invalid measurements**, where we learn the mask β_s to block the gradient back-propagation from invalid signals measured by defective X-ray detectors, thereby eliminating the IM effect.

3.4 Model Optimization

Fig. 3 shows the ray-based optimization pipeline for our Riner model. We first sample a set of spatial coordinates $\mathbf{x}$ along any X-ray $L(\theta, s)$ at a pre-defined and fixed interval $\Delta \mathbf{x}$. Then, we feed these coordinates into the MLP network $f_{\boldsymbol{\Phi}}$ to predict the corresponding CT intensities $\mu(\mathbf{x})$. Furthermore, we conduct our differentiable forward model $\boldsymbol{A}$ (Eq. 6) to convert these MLP-predicted intensities $\mu(\mathbf{x}) = f_{\boldsymbol{\Phi}}(\mathbf{x}), \forall \mathbf{x} \in L(\theta, s)$, the corresponding learnable response factor α_s, and the mask β_s into estimated measurement $\widehat{\rho}(\theta, s)$. Finally, the MLP network $f_{\boldsymbol{\Phi}}$ and the parameters (α_s, β_s) are jointly optimized using gradient descent back-propagation algorithms, such as Adam (Kingma 2014), to minimize the loss function $\mathcal{L}$, which is defined by:

$$\mathcal{L} = \underbrace{\sum_{L(\theta, s) \in \bar{\boldsymbol{\Pi}}} \big\| \widehat{\rho}(\theta, s) - \tilde{\rho}(\theta, s) \cdot \beta_s \big\|_1}_{\text{Data Consistency}} + \lambda \cdot \underbrace{\sum_{s \in \bar{S}} -\beta_s^2}_{\text{Negative } \ell_2}, \tag{7}$$

where $\bar{\boldsymbol{\Pi}} \subseteq \boldsymbol{\Pi}$ is a random subset of X-rays, and $\bar{S} \subseteq S$ is the corresponding detector subset. The data consistency minimizes the discrepancy between the predicted and actual measurements, while the negative ℓ_2 regularizes the masks β_s, preventing all of them from converging to zero (*i.e.*, identifying all detectors as defective), thereby avoiding optimization degradation. $\lambda = 0.01$ is a hyperparameter that controls the contribution of the regularization.

Fortunately, *this ray-based optimization is memory-friendly*, as each iteration only involves forward and backward passes over a small set of sampled voxels, rather than the entire 3D volume as in supervised methods (Chang et al. 2020; Trapp et al. 2022; Zamir et al. 2022). This enables our approach to scale seamlessly to CBCT scans. After model optimization, the MLP network $f_{\boldsymbol{\Phi}^*}$ and the vectors $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ represent the desired CT image, detector responses, and defective detectors, respectively. We finally reconstruct the high-quality CT image μ^* by feeding all coordinates $\mathbf{x}$ into the well-optimized network $f_{\boldsymbol{\Phi}^*}$.

4 Experiments

4.1 Experimental Settings

Datasets We perform experiments on three simulated datasets and two real-world datasets as follows: **1) Simulated 2D FBCT DeepLesion** (Yan et al. 2018), consisting of 5,261 2D slices of a 256×256 size from multiple 3D volumes. We split these 2D slices into three parts: 5,061 slices for the training set, 100 slices for the validation set, and 100 slices for the test set. *Notably, our unsupervised Riner does not access the training and validation sets used for training supervised baselines*; **2) Simulated 2D FBCT LIDC** (Armato III et al. 2011), consisting of 100 2D slices of a 256×256 size from multiple 3D volumes for an additional

Category	Method	DeepLesion		LIDC		AAPM	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Model-based	FBP/FDK	13.81±3.01	0.505±0.129	12.29±2.86	0.452±0.153	15.26±1.24	0.695±0.062
	Norm	24.37±1.95	0.774±0.067	24.61±2.31	0.786±0.049	26.86±0.50	0.865±0.024
	WaveFFT	20.70±2.48	0.663±0.096	19.44±2.22	0.622±0.118	23.83±1.17	0.777±0.051
	Super	29.11±3.43	0.815±0.084	31.24±3.14	0.855±0.051	30.06±3.64	0.915±0.026
Supervised	HyUNet	33.73±1.88	0.893±0.035	32.46±1.54	0.854±0.051	31.49±1.58	0.871±0.041
	DeepRAR	34.61±2.13	0.893±0.043	30.34±3.54	0.802±0.085	28.35±1.61	0.821±0.053
	NAFNet	36.51±1.91	0.937±0.030	36.07±1.91	0.917±0.051	33.89±1.69	0.907±0.034
	Restormer	<u>37.31±1.92</u>	<u>0.947±0.028</u>	<u>36.69±1.90</u>	<u>0.925±0.051</u>	<u>33.94±1.50</u>	<u>0.907±0.031</u>
	MambaIRv2	36.77±1.88	0.939±0.029	35.83±1.87	0.912±0.050	33.23±2.03	0.901±0.037
Unsupervised	SinoRAR	32.70±1.77	0.881±0.034	32.63±1.21	0.879±0.035	N/A	N/A
	Riner (Ours)	**39.02±2.18**	**0.967±0.019**	**39.11±2.17**	**0.962±0.030**	**36.53±1.28**	**0.949±0.020**

Table 1: Quantitative results of our Riner and compared methods on three simulated datasets, including 2D FBCT DeepLesion (Yan et al. 2018), 2D FBCT LIDC (Armato III et al. 2011), and 3D CBCT AAPM (McCollough et al. 2017). The best and second performances are highlighted in **blod** and <u>underline</u>, respectively.

test set; **3) Simulated 3D CBCT AAPM** (McCollough et al. 2017), consisting of 10 3D volumes, each with sizes of 256×256×100, used to evaluate our Riner model and baselines within 3D CBCT geometry; **4) Real-world 2D FBCT Walnut**, consisting of a 2D slice with sizes of 512×512 and a physical resolution of 60×60 μm^2, acquired by a commercial Bruker SKYSCAN 1276 micro-CT scanner using 2D FBCT geometry; **5) Real-world 3D CBCT Chicken foot**, consisting of a 3D volume with sizes of 512×512×80 and a physical resolution of 60×60×60 μm^3, acquired by a commercial Bruker SKYSCAN 1276 micro-CT scanner using 3D CBCT geometry. *We provide details on data preprocessing in the Appendix.*

Baselines and Metrics We compare Riner against 10 cutting-edge methods, including 1) 4 model-based algorithms (FBP/FDK (Kak and Slaney 2001; Feldkamp, Davis, and Kress 1984), Norm (Rivers 1998), WaveFFT (Münch et al. 2009), and Super (Vo, Atwood, and Drakopoulos 2018)); 2) 5 supervised DL models (HyUNet (Chang et al. 2020), DeepRAR (Trapp et al. 2022), NAFNet (Chen et al. 2022), Restormer (Zamir et al. 2022), and MambaIRv2 (Guo et al. 2024b,a)); 3) 1 unsupervised DL model (SinoRAR (Shi et al. 2024)). *The five supervised methods are trained using the training and validation sets of the DeepLesion dataset (Yan et al. 2018)*. Moreover, we use PSNR and SSIM as quantitative metrics. *More implementation details about the baselines can be found in the Appendix.*

Implementation Details We implement our Riner using PyTorch (Paszke et al. 2019). The network $f_{\mathbf{\Phi}}$ consists of a hash encoding module (Müller et al. 2022) followed by two fully connected (FC) layers, with a ReLU activation applied after the first FC layer. The hash encoding accelerates optimization and enhances the reconstruction of high-frequency details. We adopt the hyperparameters suggested in the original paper (Müller et al. 2022): $L = 10$, $T = 2^{10}$, $F = 8$, $N_{\text{min}} = 2$, and $b = 2$. During each step, we randomly sample a subset of 80 X-rays (*i.e.*, $|\bar{\mathbf{\Pi}}| = 80$) from 2 detectors (*i.e.*, $|\bar{S}| = 2$) across 40 projection views. We use the Adam optimizer (Kingma 2014) with default settings. The learning rate is set to 10^{-3} and decays by a factor of 0.5 every 1,000 iterations over a total of 4,000 iterations. The optimization is performed on a single NVIDIA RTX 4090 GPU, requiring only around 3 GB of memory. *All hyperparameters of Riner are tuned using 10 samples from the DeepLesion dataset (Yan et al. 2018) and remain unchanged in all other samples.*

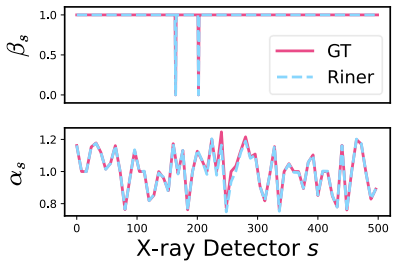

Figure 4: Parameter estimations of our Riner on a sample (#9) of DeepLesion dataset (Yan et al. 2018).

4.2 Main Results

Comparisons on Simulated Datasets Table 1 shows the quantitative results. Riner consistently achieves the best performance across all three simulated datasets, surpassing the best-performing supervised baseline, Restormer, by +1.71 dB, +2.42 dB, and +2.59 dB in PSNR, respectively. Among model-based methods, Super achieves the best numerical results due to its modeling of multiple types of ring artifacts. Among the supervised baselines, NAFNet, Restormer, and MambaIRv2 perform well on 2D datasets due to their advanced architectures, but degrade notably on the 3D data. In contrast, our Riner maintains superior and stable RAR performance across both 2D and 3D datasets, owing to its physics-informed formulation.

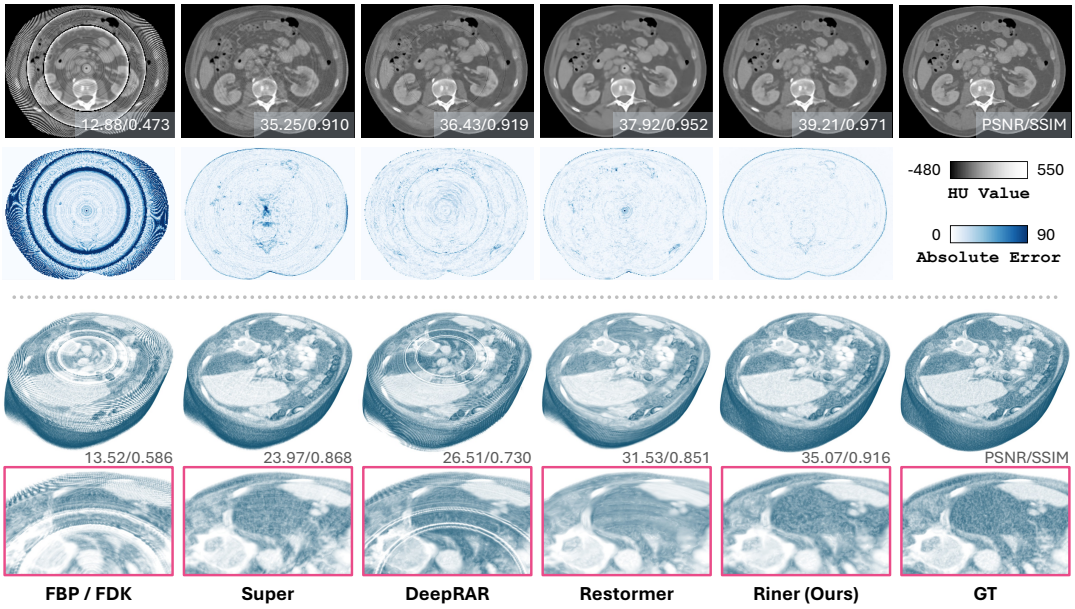

Figure 5: Qualitative results of our Riner and compared methods on two representative samples of three simulated datasets (2D FBCT DeepLesion (Yan et al. 2018) and 3D CBCT AAPM (McCollough et al. 2017)).

The core idea of our Riner for addressing the CT RAR problem is to explicitly estimate the detector physical parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Thus, accurately estimating these parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is critical for enabling reliable CT reconstructions. In Fig. 4, we visualize the parameters estimated by our Riner alongside GTs on a sample of the DeepLesion dataset. These results confirm that Riner can accurately estimate the physical parameters, enabling reliable CT results.

Fig. 5 demonstrates qualitative comparisons. Visually, our Riner consistently reconstructs clear, high-fidelity CT images with fine structures, closely matching the GTs and significantly outperforming all baselines.

Comparisons on Real-World Datasets Micro-CT imaging is widely used in fields such as biomedical research and materials science (Ritman 2011). However, ring artifacts are common in Micro-CT due to its μm-scale ultra-high physical resolution (Yousuf and Asaduzzaman 2010; Rashid, Lee, and Hasan 2012). Here, we validate the effectiveness of our Riner on real-world Micro-CT imaging. We scan a *Walnut* sample and a *Chicken foot* sample using a commercial Bruker SKYSCAN 1276 micro-CT scanner under 2D FBCT and 3D CBCT protocols (see Appendix). We only report qualitative results since artifact-free GT images are difficult to obtain. As demonstrated in Fig. 2 and Fig. 6, our Riner effectively removes ring artifacts while preserving image details, significantly outperforming baselines.

4.3 Ablation Studies

Influence of Physical Forward Model The proposed forward model $\boldsymbol{A}$ introduce the learnable variables $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ to correct the IM and IR effects, respectively. Here, we investigate their effectiveness. Fig. 7 and Table 2 show the qualitative and quantitative results. There are four cases:

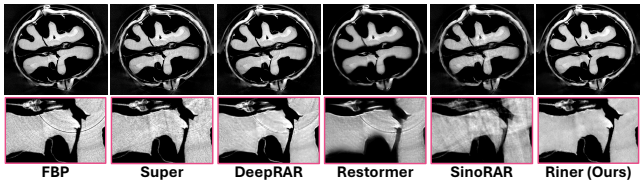

Figure 6: Qualitative results of our Riner and compared methods on a real-world *Walnut* sample (see Appendix) with sizes of 512×512 and an ultra-high resolution of 60×60 μm^2, acquired by a commercial Bruker SKYSCAN 1276 micro-CT scanner.

- w/o $\boldsymbol{\beta}, \boldsymbol{\alpha}$ (Integral Model): *i.e.*, degrading our physical model to the traditional integral model in Eq. (2). We observe that although the integral model reduces some ring artifacts using the continuous prior of INR, it still contains significant artifacts due to the IM and IR effects.
- w/o $\boldsymbol{\beta}$: When removing the mask $\boldsymbol{\beta}$, our Riner fails to handle the IM effect. The image contains black ring artifacts of the IM effect, indicated by **the pink arrow** in Fig. 7. While the artifacts of the IR effect are effectively reduced by estimating the response $\boldsymbol{\alpha}$.
- w/o $\boldsymbol{\alpha}$: This case is opposite to case 2. The IM effect artifacts are removed, but the IR effect artifacts remain.
- Full Model: Our Riner method can effectively eliminate both types of ring artifacts.

Generally, the results of ablating the physical model align perfectly with the expectations of our theoretical modeling for the IR and IM effects, showing its effectiveness.

Influence of Negative ℓ_2 Regularization Riner explicitly learns the mask $\boldsymbol{\beta}$ to remove signals from defective detec-

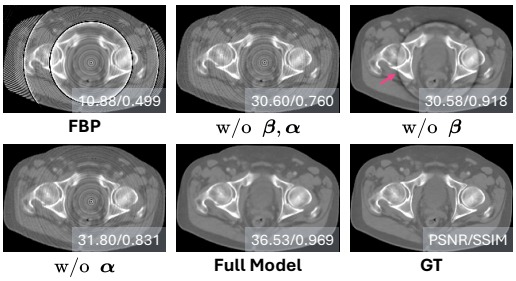

Figure 7: Qualitative results of our Riner ablating the physical forward model (Eq. 6) on a representative sample (#6) of DeepLesion dataset (Yan et al. 2018).

Physical Model		PSNR	SSIM
Mask β	Factor α		
✗	✗	35.34±2.26	0.890±0.050
✗	✓	31.05±9.16	0.894±0.130
✓	✗	35.72±1.93	0.910±0.033
✓	✓	**38.98±1.41**	**0.974±0.005**

Table 2: Quantitative results of our Riner ablating the physical forward model (Eq. 6) on 10 samples of DeepLesion dataset (Yan et al. 2018).

tors, thereby addressing the IM effect. To avoid the trivial solution of $\boldsymbol{\beta} = \mathbf{0}$, we incorporate a negative ℓ_2 regularization term into the loss $\mathcal{L}$ (Eq. 7). We evaluate its effect by testing different weights $\lambda \in \{0, 0.01, 1\}$. Fig. 8 and Table 3 show both qualitative and quantitative results. Without regularization (*i.e.*, $\lambda = 0$), the optimization quickly collapses to $\boldsymbol{\beta} = \mathbf{0}$, failing to reconstruct detailed CT images. In contrast, with a strong regularization of $\lambda = 1$, the mask $\boldsymbol{\beta}$ converges to $\mathbf{1}$, meaning no defective detectors are identified. Therefore, the black ring artifacts caused by IM effects remain. The best performance is achieved at $\lambda = 0.01$, which achieves a balanced regularization. *Based on this ablation study, we set* $\lambda = 0.01$ *as the default in all experiments.*

Influence of Network Architecture The network $f_{\boldsymbol{\Phi}}$ consists of hash encoding (Müller et al. 2022) and a 2-layer MLP. Here, we study the influence of network architecture on RAR reconstruction. Specifically, we implement an alternative using Fourier encoding (Tancik et al. 2020) followed by a 4-layer MLP, while keeping all other settings identical for a fair comparison. Fig. 9 shows the reconstructed CT images. These results demonstrate that both networks achieve excellent performance, with the hash encoding yielding a slight improvement of +0.25 dB in PSNR (40.15 vs. 39.90). This suggests that the superior RAR performance of our Riner model is mainly attributed to the proposed physical modeling that effectively captures the nonlinear IM and IR effects, rather than the network architecture itself.

5 Conclusion and Limitation

In this paper, we present Riner, a novel approach to address the RAR problem in 3D X-ray CBCT. Unlike existing end-to-end RAR methods, our approach formulates RAR as a multi-parameter inverse problem, targeting the physical origins of ring artifact formation. Using a differentiable physical model, Riner jointly learns the neural representation of the artifact-free CT image and estimates the detector response parameters directly from raw measurements.

While effective, the efficiency of our method can be further improved. As an unsupervised approach, Riner requires case-specific optimization, taking about 30 seconds for a 2D slice of 256×256 and 10 minutes for a 3D volume of 256×256×100 on an NVIDIA RTX 4090 GPU. Advanced techniques (*e.g.*, 3D GS) may improve its efficiency.

Acknowledgment This work was supported by the National Natural Science Foundation of China (Nos. 62571328 and W2431046), the Core Facility Platform of Electronics, SIST, ShanghaiTech University, the MoE Key Lab of Intelligent Perception and Human-Machine Collaboration (ShanghaiTech University), and the Shanghai Frontiers Science Center of Human-centered Artificial Intelligence.

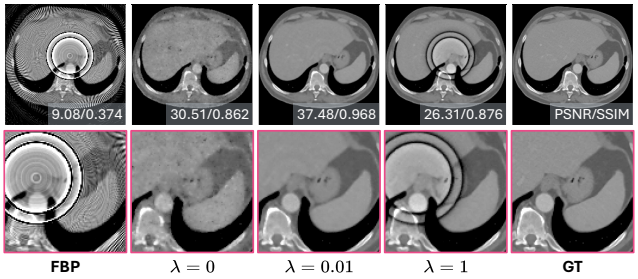

Figure 8: Qualitative results of our Riner ablating the negative ℓ_2 term (Eq. 7) on a representative sample (#3) of DeepLesion dataset (Yan et al. 2018).

Weight	PSNR	SSIM
$\lambda = 0$	31.63±1.77	0.881±0.034
$\lambda = 0.01$	**38.98±1.41**	**0.974±0.005**
$\lambda = 1$	31.55±8.97	0.896±0.134

Table 3: Quantitative results of our Riner ablating the negative ℓ_2 term (Eq. 7) on 10 samples of DeepLesion dataset (Yan et al. 2018).

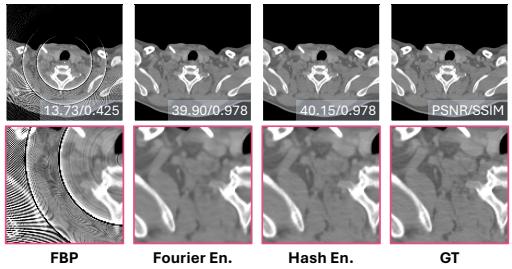

Figure 9: Qualitative results of our Riner ablating the network $f_{\boldsymbol{\Phi}}$ on a representative sample (#4) of DeepLesion dataset (Yan et al. 2018).

References

Armato III, S. G.; McLennan, G.; Bidaut, L.; McNitt-Gray, M. F.; Meyer, C. R.; Reeves, A. P.; Zhao, B.; Aberle, D. R.; Henschke, C. I.; Hoffman, E. A.; et al. 2011. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, 38(2): 915–931.

Beer. 1852. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen der Physik*, 162(5): 78–88.

Birklein, L.; Schömer, E.; Brylka, R.; Schwanecke, U.; and Schulze, R. K. W. 2023. Neural Deformable Cone Beam CT. Eurographics.

Boas, F. E.; Fleischmann, D.; et al. 2012. CT artifacts: causes and reduction techniques. *Imaging Med*, 4(2): 229–240.

Chang, S.; Chen, X.; Duan, J.; and Mou, X. 2020. A CNN-based hybrid ring artifact reduction algorithm for CT images. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(2): 253–260.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022. Simple baselines for image restoration. In *European conference on computer vision*, 17–33. Springer.

Chen, Y.; and Liu, J. 1984. An iterative numerical algorithm for solving multi-parameter inverse problems of evolutional partial differential equations. *Journal of Computational Physics*, 53(3): 429–442.

Chen, Z.; Xi, X.; Han, Y.; Tan, S.; Wang, C.; Sun, Y.; Li, L.; and Yan, B. 2024. Research status and prospect of ring artifact correction method for X-ray CT. In *International Conference on Optoelectronic Information and Functional Materials (OIFM 2024)*, volume 13183, 245–252. SPIE.

Cristofol, M.; Garnier, J.; Hamel, F.; and Roques, L. 2011. Uniqueness from pointwise observations in a multi-parameter inverse problem. *Communications on Pure and Applied Analysis*, 11(1): 173–188.

Dalla Riva, M.; Luzzini, P.; and Musolino, P. 2022. Multi-parameter analysis of the obstacle scattering problem. *Inverse Problems*, 38(5): 055004.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*.

Du, C.; Lin, X.; Wu, Q.; Tian, X.; Su, Y.; Luo, Z.; Wei, H.; Zhou, S. K.; Yu, J.; and Zhang, Y. 2024. DPER: Diffusion Prior Driven Neural Representation for Limited Angle and Sparse View CT Reconstruction. *arXiv preprint arXiv:2404.17890*.

Essakine, A.; Cheng, Y.; Cheng, C.-W.; Zhang, L.; Deng, Z.; Zhu, L.; Schönlieb, C.-B.; and Aviles-Rivero, A. I. 2024. Where Do We Stand with Implicit Neural Representations? A Technical and Performance Survey. *arXiv preprint arXiv:2411.03688*.

Fang, W.; Li, L.; and Chen, Z. 2020. Removing ring artefacts for photon-counting detectors using neural networks in different domains. *IEEE access*, 8: 42447–42457.

Feldkamp, L. A.; Davis, L. C.; and Kress, J. W. 1984. Practical cone-beam algorithm. *Josa a*, 1(6): 612–619.

Guo, H.; Guo, Y.; Zha, Y.; Zhang, Y.; Li, W.; Dai, T.; Xia, S.-T.; and Li, Y. 2024a. MambaIRv2: Attentive State Space Restoration. *arXiv preprint arXiv:2411.15269*.

Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2024b. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, 222–241. Springer.

Kak, A. C.; and Slaney, M. 2001. *Principles of computerized tomographic imaging*. SIAM.

Kingma, D. P. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kostsov, V. 2015. General approach to the formulation and solution of the multi-parameter inverse problems of atmospheric remote sensing with measurements and constraints of different types. *International Journal of Remote Sensing*, 36(11): 2963–2994.

Lambert, J.-H. 1760. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Sumptibus viduae Eberhardi Klett, typis Christophori Petri Detleffsen.

Lee, J.; Ahn, J.; and Baek, J. 2024. Neural attenuation fields for metal artifact reduction in dental CT. In *Medical Imaging 2024: Physics of Medical Imaging*, volume 12925, 173–179. SPIE.

Lin, Y.; Luo, Z.; Zhao, W.; and Li, X. 2023. Learning deep intensity field for extremely sparse-view CBCT reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 13–23. Springer.

McCollough, C. H.; Bartley, A. C.; Carter, R. E.; Chen, B.; Drees, T. A.; Edwards, P.; Holmes III, D. R.; Huang, A. E.; Khan, F.; Leng, S.; et al. 2017. Low-dose CT for the detection and classification of metastatic liver lesions: results of the 2016 low dose CT grand challenge. *Medical physics*, 44(10): e339–e352.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15.

Münch, B.; Trtik, P.; Marone, F.; and Stampanoni, M. 2009. Stripe and ring artifact removal with combined wavelet—Fourier filtering. *Optics express*, 17(10): 8567–8591.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; and Courville, A. 2019. On the spectral bias of neural networks. In *International conference on machine learning*, 5301–5310. PMLR.

Rashid, S.; Lee, S. Y.; and Hasan, M. K. 2012. An improved method for the removal of ring artifacts in high resolution CT imaging. *EURASIP Journal on Advances in Signal Processing*, 2012: 1–18.

Reed, A. W.; Kim, H.; Anirudh, R.; Mohan, K. A.; Champley, K.; Kang, J.; and Jayasuriya, S. 2021. Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2258–2268.

Ritman, E. L. 2011. Current status of developments and applications of micro-CT. *Annual review of biomedical engineering*, 13(1): 531–552.

Rivers, M. 1998. Tutorial introduction to X-ray computed microtomography data processing. *University of Chicago*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Rückert, D.; Wang, Y.; Li, R.; Idoughi, R.; and Heidrich, W. 2022. Neat: Neural adaptive tomography. *ACM Transactions on Graphics (TOG)*, 41(4): 1–13.

Scarfe, W. C.; and Farman, A. G. 2008. What is cone-beam CT and how does it work? *Dental Clinics of North America*, 52(4): 707–730.

Shen, L.; Pauly, J.; and Xing, L. 2022. 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.

Shi, L.; Liu, C.; Jiang, X.; Liu, Y.; Yang, P.; Guo, S.; and Zhao, X. 2024. Ring Artifacts Removal Based on Implicit Neural Representation of Sinogram Data. *arXiv preprint arXiv:2409.15731*.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33: 7462–7473.

Tancik, M.; Srinivasan, P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J.; and Ng, R. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33: 7537–7547.

Trapp, P.; Amato, C.; Sawall, S.; Kachelrieß, M.; and Vöth, T. 2022. DeepRAR: a CNN-based approach for CT and CBCT ring artifact reduction. In *Medical Imaging 2022: Physics of Medical Imaging*, volume 12031, 337–344. SPIE.

Venkatesh, E.; and Elluru, S. V. 2017. Cone beam computed tomography: basics and applications in dentistry. *Journal of istanbul University faculty of Dentistry*, 51(3 Suppl 1): 102–121.

Vo, N. T.; Atwood, R. C.; and Drakopoulos, M. 2018. Superior techniques for eliminating ring artifacts in X-ray microtomography. *Optics express*, 26(22): 28396–28412.

Wang, C.; Shang, K.; Zhang, H.; Li, Q.; and Zhou, S. K. 2022. DuDoTrans: dual-domain transformer for sparse-view CT reconstruction. In *International Workshop on Machine Learning for Medical Image Reconstruction*, 84–94. Springer.

Wang, H.; Li, Y.; Zhang, H.; Meng, D.; and Zheng, Y. 2023. InDuDoNet+: A deep unfolding dual domain network for metal artifact reduction in CT images. *Medical Image Analysis*, 85: 102729.

Wu, Q.; Chen, L.; Wang, C.; Wei, H.; Zhou, S. K.; Yu, J.; and Zhang, Y. 2023a. Unsupervised polychromatic neural representation for ct metal artifact reduction. *Advances in Neural Information Processing Systems*, 36: 69605–69624.

Wu, Q.; Feng, R.; Wei, H.; Yu, J.; and Zhang, Y. 2023b. Self-supervised coordinate projection network for sparse-view computed tomography. *IEEE Transactions on Computational Imaging*, 9: 517–529.

Wu, Q.; Guo, X.; Chen, L.; Liu, Y.; He, D.; Wang, X.; Chen, X.; Zhang, Y.; Zhou, S. K.; Yu, J.; et al. 2024. Solving energy-independent density for ct metal artifact reduction via neural representation. *arXiv preprint arXiv:2405.07047*.

Yan, K.; Wang, X.; Lu, L.; Zhang, L.; Harrison, A. P.; Bagheri, M.; and Summers, R. M. 2018. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9261–9270.

Yousuf, M.; and Asaduzzaman, M. 2010. An efficient ring artifact reduction method based on projection data for micro-CT images. *Journal of Scientific Research*, 2(1): 37–45.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zang, G.; Idoughi, R.; Li, R.; Wonka, P.; and Heidrich, W. 2021. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1960–1970.

Zha, R.; Zhang, Y.; and Li, H. 2022. NAF: neural attenuation fields for sparse-view CBCT reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 442–452. Springer.

Zhang, Y.; Shao, H.-C.; Pan, T.; and Mengke, T. 2023. Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR). *Physics in Medicine & Biology*, 68(4): 045005.

Zhu, L.; Cui, Y.; Du, Y.; Wei, D.; and Deng, Y. 2023. Research on the multi-parameter identification of thermal-acoustic-solid coupling problem in an inhomogeneous medium. *Journal of Mathematical Analysis and Applications*, 521(2): 126920.

Zhu, Y.; Zhao, M.; Li, H.; and Zhang, P. 2013. Micro-CT artifacts reduction based on detector random shifting and fast data inpainting. *Medical physics*, 40(3): 031114.