Unsupervised Self-prior Embedding Neural Representation for Iterative Sparse-view CT Reconstruction

Xuanyu Tian[1,2], **Lixuan Chen**[3], **Qing Wu**[1], **Chenhe Du**[1], **Jingjing Shi**[4], **Hongjiang Wei**[5], **Yuyao Zhang**[1*]

[1] School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
[2] Lingang Laboratory, Shanghai 200031, China
[3] Electrical and Computer Engineering, University of Michigan, MI 48105, United States
[4] Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China
[5] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200127, China
tianxy@shanghaitech.edu.cn, chenlx@umich.edu, wuqing@shanghaitech.edu.cn, duchenhe@shanghaitech.edu.cn,
walm13@126.com, hongjiang.wei@sjtu.edu.cn, zhangyy8@shanghaitech.edu.cn

Abstract

Emerging unsupervised implicit neural representation (INR) methods, such as NeRP, NeAT, and SCOPE, have shown great potential to address sparse-view computed tomography (SVCT) inverse problems. Although these INR-based methods perform well in relatively dense SVCT reconstructions, they struggle to achieve comparable performance to supervised methods in sparser SVCT scenarios. They are prone to being affected by noise, limiting their applicability in real clinical settings. Additionally, current methods have not fully explored the use of image domain priors for solving SVCT inverse problems. In this work, we demonstrate that imperfect reconstruction results can provide effective image domain priors for INRs to enhance performance. To leverage this, we introduce **S**elf-**p**rior **e**mbedding **ne**ural **r**epresentation (**Spener**), a novel unsupervised method for SVCT reconstruction that integrates iterative reconstruction algorithms. During each iteration, Spener extracts local image prior features from the previous iteration and embeds them to constrain the solution space. Experimental results on multiple CT datasets show that our unsupervised Spener method achieves performance comparable to supervised state-of-the-art (SOTA) methods on in-domain data while outperforming them on out-of-domain datasets. Moreover, Spener significantly improves the performance of INR-based methods in handling SVCT with noisy sinograms. Our code is available at https://github.com/MeijiTian/Spener.

Introduction

X-ray computed tomography (CT) is widely used in clinical diagnosis due to its non-invasive nature and high efficiency (Wang, Yu, and De Man 2008). However, the ionizing radiation of X-ray poses potential risks to biological tissues, while high levels of radiation exposure can increase the lifetime risk of cancer. To mitigate these risks, the medical community generally follows the ALARA (as low as reasonably achievable) principle in clinical practice. Therefore, reducing the radiation dose in the CT acquisition process is a critical and widely researched topic.

The forward acquisition of CT can be formulated as the following linear equation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}, \tag{1}$$

where the $\mathbf{y} \in \mathbb{R}^m$ is the measurement data (*i.e.* sinogram data) with $m = n_v \times n_d$, n_v and n_d is the number of projection views and detectors. $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the forward model (*e.g.* Radon transform), $\mathbf{x} \in \mathbb{R}^n$ is the unknown image intensity of object and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is the system noise. To reduce the radiation of the CT process, one effective solution is to reduce the number of sampling projection views n_v, which is sparse-view CT (SVCT) acquisition. Under the SVCT acquisition setting, the dimension of measurement data $\mathbf{y}$ is much lower than the unknown object $\mathbf{x}$ (*i.e*, $m \ll n$), resulting in the linear equation (Eq. (1)) to be highly ill-posed.

Recently, there has been considerable works on supervised deep learning (DL) for the SVCT inverse problem (Lee et al. 2018; Zhang et al. 2018; Chen et al. 2018; Wu et al. 2021; Ma et al. 2023; Liu et al. 2023). These methods train neural networks to map corrupted images to artifact-free images in end-to-end learning. Leveraging the power of neural networks and data-driven priors, supervised DL methods achieve SOTA performance in SVCT reconstruction. However, these methods face two significant challenges: 1) Learning cost. Developing effective data-driven priors requires collecting large amounts of paired data, which is both challenging and expensive; 2) Out-of-domain (OOD) problem. These methods often suffer from performance degradation when the CT acquisition settings change (*e.g.*, different organs and X-ray acquisition geometry) deviate from the training data. These issues pose substantial obstacles to the practical application of supervised methods in clinical settings.

Implicit Neural Representation (INR) has recently become an innovative unsupervised approach for solving imaging inverse problems without relying on external data. It utilizes multi-layer perception (MLP) to represent the intensity of an object into an implicit function of spatial coordinates. By integrating differentiable physical models, the unknown image intensity can be transformed into the measurement domain. Thus, MLP can learn the implicit function by minimizing the error between the estimated measure-

ments and the acquired data. The inherent learning consistency bias of INR acts as an effective implicit regularization term for solving inverse problems (Rahaman et al. 2019), achieving significant progress in SVCT reconstruction (Sun et al. 2021; Zang et al. 2021; Reed et al. 2021; Zha, Zhang, and Li 2022; Wu et al. 2023; Lin et al. 2023, 2024). However, existing INR-based SVCT methods rely solely on the implicit learning bias of the network as a regularization constraint. When the ill-posedness increases in SVCT reconstruction (*e.g.* sparser view or lower dose in acquisition), they exhibit instability to provide satisfactory performance.

In this work, we demonstrate that the imperfectly reconstructed result of INR can be an effective image domain prior to stabilizing the neural representation. Therefore, we propose a novel unsupervised method, **S**elf-**p**rior **e**mbedding **ne**ural **r**epresentation (**Spener**), which explores the image domain prior generated by INR itself for SVCT reconstruction. To utilize the local image features in the image domain, we incorporate an image encoder to represent the prior image into local image feature vectors providing INR network optimization. Additionally, we employ an iterative reconstruction algorithm, such as plug-and-play half-quadratic splitting (PnP-HQS), to iteratively update the prior image to accelerate the convergence of INR and enhance its performance, particularly in highly ill-posed scenarios.

We conduct extensive experiments to evaluate the performance of our Spener on three public datasets. The results show that Spener achieves comparative performance with the supervised SOTA methods in the in-distribution datasets while outperforming them when the CT acquisition parameters are different from those in the training data. Meanwhile, in noisy acquisition scenarios, Spener significantly outperforms other unsupervised INR methods, highlighting its superiority. Extensive ablation studies further validate the effectiveness and robustness of the proposed method. Our main contributions are summarized as follows:

- We propose Spener, a novel unsupervised robust neural representation, achieving superior performance in solving SVCT reconstruction.

- We demonstrate that imperfect INR reconstruction can serve as an effective image prior and explicit regularization constraint for INR optimization.

- We introduce an image encoder to provide the INR network with local continuous image priors, with experimental results confirming its effectiveness in unsupervised learning.

- We integrate the strengths of iterative algorithms with INR-based methods, improving the model convergence and generalization.

Related Work

Implicit Neural Representation for CT

Consider a 2D CT scanner, the image intensity of scanned object can be formulated as a function of spatial coordinates, defined by

$$f : \mathbf{p} = (x, y) \in \mathbb{R}^2 \longrightarrow \mu(\mathbf{p}) \in \mathbb{R}, \qquad (2)$$

where $\mathbf{p}$ is an arbitrary coordinate and $\mu(\mathbf{p})$ is the corresponding intensity of the object. However, the function f is highly complex, making an analytical solution intractable. Previous works (Shen, Pauly, and Xing 2022; Zha, Zhang, and Li 2022; Wu et al. 2023) have leveraged the consistency bias inherent in INR by employing an MLP network $\mathcal{F}_{\mathbf{\Phi}}$ to approximate the function f. The MLP network can learn the implicit function by incorporating the forward model (*e.g.* Radon transform) to minimize the discrepancy between the predicted and acquired measurement data:

$$\mathbf{\Phi}^* = \arg\min_{\mathbf{\Phi}} \mathcal{L}(\mathbf{A}\mathcal{F}_{\mathbf{\Phi}}, \mathbf{y}), \qquad (3)$$

where $\mathbf{A}$ is the forward model and $\mathbf{y}$ is the acquired measurement data. By adopting the novel coordinate encoding strategy (Müller et al. 2022), these INR-based methods can effectively capture high-frequency signals, providing high-quality reconstructions while maintaining data consistency. However, when the measurement data is undersampled or noisy, these methods tend to generate high-frequency artifacts, resulting in performance degradation.

Plug-and-Play Half-Quadratic Splitting

The classical approach for solving the ill-posed problem in Eq. (1) is by formulating an optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\mathcal{R}(\mathbf{x}), \qquad (4)$$

where the objective function is composed of a data fidelity term $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2/2$ and a regularization term $\lambda\mathcal{R}(\mathbf{x})$. The data fidelity term ensures that the solution conforms to the forward process, maintaining consistency with the measurement data, while the regularization enforces prior knowledge, narrowing the solution space.

Plug-and-play half-quadratic splitting (PnP-HQS) is a widely used iterative framework for solving inverse problems (Zhang et al. 2021). The conventional HQS (Geman and Yang 1995) introduces an auxiliary variable $\mathbf{z}$ to reformulate the original optimization problem Eq. (4) in a constrained manner:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda\mathcal{R}(\mathbf{z}) \quad s.t. \quad \mathbf{z} = \mathbf{x}. \quad (5)$$

HQS solves the objective function by decoupling a data fidelity term and a regularization term, and alternatively solving them in iterative ways:

$$\begin{cases} \mathbf{x}_t = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \frac{\mu}{2}\|\mathbf{x} - \mathbf{z}_{t-1}\|^2 & (6a) \\ \mathbf{z}_t = \arg\min_{\mathbf{z}} \frac{1}{2(\sqrt{\lambda/\mu})^2}\|\mathbf{z} - \mathbf{x}_t\|^2 + \mathcal{R}(\mathbf{z}). & (6b) \end{cases}$$

The PnP-HQS proposes utilizing a general denoising algorithm to provide the solution of Eq (6b), given by

$$\mathbf{z}_t = \mathcal{D}_{\sigma}(\mathbf{x}_t), \qquad (7)$$

where $\mathcal{D}_{\sigma}$ represents the denoising operation with a noise level parameter σ.

In Spener, we introduce an INR-based approach to solve the data fidelity sub-problem. Subsequently, we apply an effective denoiser to regularize the reconstruction obtained from the INR. By iteratively solving these two sub-problems, we achieve a high-fidelity reconstruction.
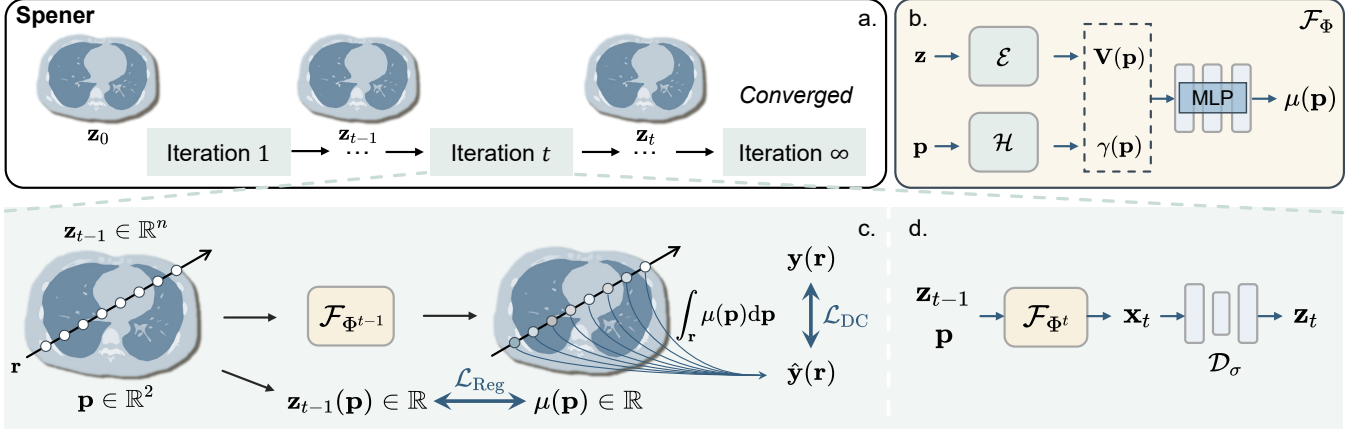
Figure 1: Overview of Spener model, including (a) iterative reconstruction using an image embedding neural network $\mathcal{F}_{\Phi}$, (b) architecture of the image embedding neural network $\mathcal{F}_{\Phi}$, (c) solving the data fidelity subproblem via the image embedding neural network $\mathcal{F}_{\Phi}$, and (d) solving regularization subproblem via a denoiser $\mathcal{D}_{\sigma}$.

Proposed Method

Overview Spener

Figure 1 shows the overall pipeline of Spener. (a) Starting with an initial result $\mathbf{z}_0$, Spener iteratively refines the CT image using the image domain embedding neural network $\mathcal{F}_{\Phi}$. (b) The network $\mathcal{F}_{\Phi}$ leverage the prior image $\mathbf{z}$ and spatial coordinates $\mathbf{p}$ to represent the underlying CT image intensity $\mu(\mathbf{p})$. To further exploit the local image domain prior, $\mathcal{F}_{\Phi}$ incorporates an image encoder $\mathcal{E}$ to extract the local image features. (c) At each iteration t, the INR-based network addresses the data fidelity subproblem. By incorporating the CT physical forward model, the network output $\mu(\mathbf{p})$ is transformed into the measurement data $\hat{\mathbf{y}}(\mathbf{r})$. Finally, the network $\mathcal{F}_{\Phi^{t-1}}$ is optimized by minimizing data consistency loss $\mathcal{L}_{\text{DC}}$ and regularization loss $\mathcal{L}_{\text{Reg}}$. (d) After updating the network, we can generate a new reconstruction of the CT image $\mathbf{x}_t$ from $\mathcal{F}_{\Phi^t}$. The regularization subproblem is then addressed using an effective denoiser $\mathcal{D}_{\sigma}$, which provides a prior image for the next iteration.

Image Prior Embedding Neural Representation

To incorporate the image domain priors, we represent the underlying clean CT images as the function of the spatial coordinates and the prior image:

$$\mathcal{F}: \mathbf{z} \in \mathbb{R}^n, \mathbf{p} \in \mathbb{R}^2 \longrightarrow \mu(\mathbf{p}) \in \mathbb{R}, \quad (8)$$

where $\mathbf{z}$ represents the prior image, $\mathbf{p}$ denotes an arbitrary spatial coordinate and $\mu(\mathbf{p})$ is the image intensity of the desired CT image. The implicit function $\mathcal{F}$ is learned by a neural network $\mathcal{F}_{\Phi}$. Figure 1.(b) shows the architecture of the network $\mathcal{F}_{\Phi}$, which consists of an image encoder $\mathcal{E}$ and a coordinate encoder $\mathcal{H}$. The image encoder $\mathcal{E}$ processes the prior image to extract local neighboring features around the coordinate $\mathbf{p}$, producing the corresponding image local feature $\mathbf{V}(\mathbf{p})$. For the coordinate encoder $\mathcal{H}$, we use hash encoding (Müller et al. 2022) to map the coordinates into a high-dimensional feature vector $\gamma(\mathbf{p})$, facilitating the representation of high-frequency signals. Finally, an MLP net-

work takes the image feature $\mathbf{V}(\mathbf{p})$ and the coordinate feature $\gamma(\mathbf{p})$ as inputs to estimate the image intensity $\mu(\mathbf{p})$.

Network Optimization for Solving Data Fidelity

Figure 1.(c) illustrates the workflow for addressing the data fidelity term in Eq. (6a) using self-prior embedding neural representation. Leveraging the strengths of the INR, the neural network $\mathcal{F}_{\Phi}$ can model a continuous CT image. This allows us to apply the CT forward physical model to transform the neural representation into the measurement domain. For clarity, we denote the acquired sinogram data $\mathbf{y}(\mathbf{r})$ as the integral intensity along an X-ray trajectory $\mathbf{r}$. Given a sinogram data $\mathbf{y}(\mathbf{r})$, we first construct a ray $\mathbf{r}$ and uniformly sample a set of coordinates $\mathbf{p}$ with a fixed interval $\Delta\mathbf{p}$ along the ray. The network $\mathcal{F}_{\Phi}$ takes the prior image $\mathbf{z}_{t-1}$ and the set of coordinates $\mathbf{p}$ as input to estimate the corresponding image intensity $\mu(\mathbf{p})$. Finally, we apply a differential integral operation to obtain the predicted sinogram data $\hat{\mathbf{y}}(\mathbf{r})$ as:

$$\hat{\mathbf{y}}(\mathbf{r}) = \sum_{\mathbf{p} \in \mathbf{r}} \mathcal{F}_{\Phi^{t-1}}(\mathbf{z}_{t-1}, \mathbf{p}) \cdot \Delta\mathbf{p}. \quad (9)$$

Thus, the neural network can achieve data fidelity aligned with measurement data by minimizing the following data consistency loss:

$$\mathcal{L}_{\text{DC}} = \frac{1}{|\mathcal{R}|} \sum_{\mathbf{r} \in \mathcal{R}} |\mathbf{y}(\mathbf{r}) - \hat{\mathbf{y}}(\mathbf{r})|. \quad (10)$$

However, directly minimizing the data consistency loss may lead to overfitting. The latter term in the data fidelity subproblem can be formulated as the regularization loss:

$$\mathcal{L}_{\text{Reg}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathcal{F}_{\Phi^{t-1}}(\mathbf{z}_{t-1}(\mathbf{p}_i), \mathbf{p}_i) - \mathbf{z}_{t-1}(\mathbf{p}_i)\|^2 \quad (11)$$

and the network parameters Φ^t can be optimized by solving the following optimization:

$$\Phi^t = \underset{\Phi^{t-1}}{\arg\min} \, \mathcal{L}_{\text{DC}} + \lambda \cdot \mathcal{L}_{\text{Reg}}, \quad (12)$$

where λ is a hyperparameter that balances the contributions of the two terms in the total loss function. Its impact is analyzed in the following ablation studies.

Prior Image Update for Regularization Solving

Figure 1.(d) illustrates the process for solving the regularization subproblem in Eq. (6b). After the data fidelity subproblem is addressed at iteration t, we can generate a new reconstructed CT image $\mathbf{x}_t$ from updated network $\mathcal{F}_{\Phi^t}$. Specifically, $\mathcal{F}_{\Phi^t}$ takes the prior image from the previous iteration $\mathbf{z}_{t-1}$ and the grid coordinates as input to produce $\mathbf{x}_t$. Then, we leverage an effective denoiser $\mathcal{D}_\sigma$ to address the regularization subproblem. Formally, this process can be expressed as below:

$$\mathbf{z}_t = \mathcal{D}_\sigma(\mathbf{x}_t), \quad \mathbf{x}_t(\mathbf{p}) = \mathcal{F}_{\Phi^t}(\mathbf{z}_{t-1}, \mathbf{p}), \qquad (13)$$

where $\mathbf{z}_t$ is the solution to the regularization subproblem, serving as the new prior image for the next iteration. Numerous studies (Zhang et al. 2017, 2021; Kamilov et al. 2023) have shown that advanced denoisers are crucial for solving the regularization subproblem. In our Spener model, we use the classical BM3D algorithm (Dabov et al. 2007) as the denoiser. The effectiveness of denoiser regularization is further analyzed in ablation studies.

Experiments

Experimental Settings

Datasets   We evaluate the proposed method on three public datasets: AAPM 2016 low-dose CT grand challenge (Mc-Collough et al. 2017), COVID-19 (Shakouri et al. 2021) and CMB-CRC head dataset (Cancer Moonshot Biobank 2022).

The AAPM dataset comprises 5936 full-dose CT images from 10 patients with 1mm slice thickness. In the experiments, we specifically select 1600 slices from 8 patients as the training set, 200 slices from 1 patient as the validation set and 10 slices from 1 patient as the test set. *The training and validation sets are prepared for the two supervised compared methods* (FBPConvNet (Jin et al. 2017) and RegFormer (Xia et al. 2022)), while other compared methods (FBP (Kak and Slaney 2001), CoIL (Sun et al. 2021), SCOPE (Wu et al. 2023)) and proposed Spener can directly recover high-quality image from sparse-view sinogram.

The COVID-19 dataset comprises 3D CT volumes from over 1000 patients with confirmed COVID-19 infections. For our experiments, we chose 10 slices in the axial view as the external test data to evaluate the performance of the compared methods. The CMB-CRC head dataset (Cancer Moonshot Biobank 2022) includes one patient head CT image with 100 slices. For our experiments, we chose 10 slices as external test data to evaluate the performance of compared methods across different organs.

Dataset Simulation   In the experiments, we retrospectively apply Radon transform on the 2D CT slice data. Specifically, we use `torch-radon` library (Ronchetti 2020) in Python with 2D fan-beam geometry to simulate the CT forward acquisition process. *Note that we adopt the same geometry for the AAPM and CMB-CRC head datasets and other different geometry for COVID geometry.* Please refer to the supplementary materials for detailed information on the fan-beam geometry settings.

To mimic the real-world CT acquisition, we introduce a Poisson noise model (Elbakri and Fessler 2002) into the sinogram, defined by $\mathbf{Y}(\mathbf{r}) \sim \mathrm{Poisson}\left\{I_0 \times \mathrm{e}^{-\mathbf{y}(\mathbf{r})} + \epsilon\right\}$, where $\mathbf{Y}(\mathbf{r})$ denotes the transmitted X-ray photon intensity, I_0 is the incident X-ray photon intensity and ϵ denotes the mean of the background events and read-out noise variance. In the experiment, we set $I_0 = 10^6$ as normal dose and $I_0 = 5 \times 10^5$ as low dose acquisition setting.

Implementation Details   In the Spener, we adopt the FBP reconstruction result as the initial prior image. The image encoder includes a two-layer CNN network, with a convolution kernel size of 3 and a feature dim of 48. For the hash encoding used in our model, we set its hyper-parameters as follows: base resolution $N_{\min} = 2$, maximal hash table size $T = 2^{24}$, and resolution growth rate $b = 1.95$. For the iteration process, the training epochs are configured as follows: 1000 epochs for $t = 1$ and 250 for the subsequent iterations. for the regularization, we utilize the BM3D denoiser with $\sigma = 0.01$. Due to space constraints, additional hyper-parameter settings for network training are provided in the supplementary materials.

Methods in Comparison & Metrics   We compare Spener to five representative methods: (1) one analytical reconstruction algorithm (FBP); (2) two supervised DL models (FBP-ConvNet (Jin et al. 2017) and RegFormer (Liu et al. 2023)); and (3) two unsupervised INR-based reconstruction methods (CoIL (Sun et al. 2021) and SCOPE (Wu et al. 2023)). To ensure a fair comparison, we use the codebase released by the authors. All approaches are trained and tested on the same datasets, as detailed in the datasets section.

For evaluating the SVCT reconstruction performance, we utilize two classic metrics, peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM), to assess the reconstruction performance.

Comparison with SOTA Methods

SVCT Reconstructions in Noise-free Scenario   Table 1 shows the quantitative results. On the AAPM dataset (in-domain dataset), our Spener achieves comparable performances with the supervised method RegFormer. Specifically, our Spener produces the best performance in PSNR and the second-best in SSIM. However, on the COVID-19 dataset, which features different CT acquisition geometry, we observe that the two supervised methods (FBPConvNet and RegFormer), trained on the AAPM dataset, suffer from performance drops due to the OOD problem. In contrast, the unsupervised methods demonstrate flexibility and effectiveness, with our Spener achieving the best performance.

Similarly, on the CMB-CRC dataset, the two supervised methods (FBPConvNet and RegFormer) encounter the OOD problem due to data from different organs, resulting in reduced performance. However, the unsupervised methods, particularly our Spener model, achieve better results.

Figure 2 shows the qualitative results, which align with the quantitative comparisons above. On the AAPM
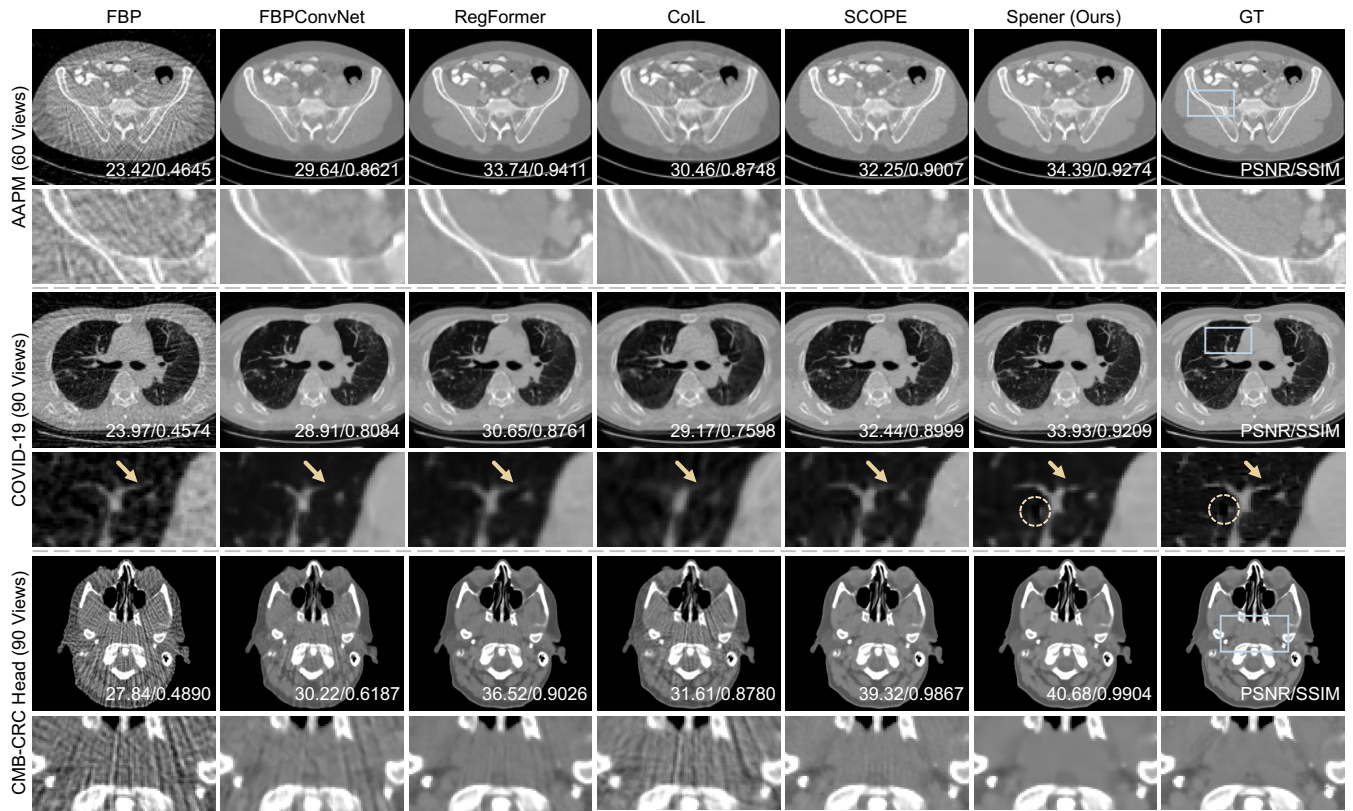
Figure 2: Qualitative results of CT images reconstructed by the compared methods on three datasets. The top two rows show results from the AAPM dataset with 60 views, the middle two rows show results from the COVID-19 dataset with 90 views, and the bottom two rows show results from the CMB-CRC head dataset with 90 views.

Category	Method	AAPM		COVID-19		CMB-CRC Head	
		60 Views	90 Views	60 Views	90 Views	60 Views	90 Views
Analytical	FBP	23.61/0.4786	26.46/0.6050	22.96/0.4578	25.77/0.5574	24.05/0.4116	27.37/0.5086
Supervised	FBPConvNet	29.82/0.8588	31.05/0.8864	28.20/0.7717	29.27/0.8080	27.22/0.5415	29.74/0.6324
	RegFormer	<u>33.78</u>/**0.9399**	<u>34.61</u>/**0.9564**	<u>30.25</u>/0.8436	31.02/0.8810	34.15/0.9264	36.11/0.9102
Unsupervised	CoIL	30.18/0.8604	31.84/0.9032	27.77/0.7156	29.72/0.7819	28.14/0.7867	31.01/0.8763
	SCOPE	32.40/0.8939	34.31/0.9322	30.21/<u>0.8504</u>	<u>32.63</u>/<u>0.8997</u>	<u>36.76</u>/<u>0.9740</u>	<u>38.76</u>/<u>0.9857</u>
	Spener (Ours)	**34.47**/<u>0.9163</u>	**37.16**/<u>0.9430</u>	**31.29/0.8709**	**33.73/0.9181**	**37.56/0.9809**	**40.27/0.9900**

Table 1: Quantitative results of the compared methods on AAPM, COVID-19 and CMB-CRC Head datasets. The best and second performances are highlighted in **bold** and <u>underline</u>, respectively.

dataset, both RegFormer and Spener recover the desired images. However, on the COVID-19 dataset, Spener produces cleaner and more detailed CT reconstructions. Specifically, Spener reconstructs the black 'holes', which are potentially caused by emphysema, more clearly, as indicated by the dashed circle. Additionally, Spener provides a clearer reconstruction of small lung tissues, as indicated by the arrows. On the CMB-CRC dataset, all methods except our Spener exhibit streaking artifacts. Spener consistently presents cleaner and more detailed CT reconstructions across all datasets.

SVCT Reconstructions in Normal and Low Dose Table 2 shows the quantitative results on the AAPM datasets with 60 and 90 input views under both normal and low dose settings. Under the normal dose setting, our Spener achieves the best performance in most cases. Compared to the two supervised methods, Spener also shows notable performance improvements. Under the low dose setting, our Spener demonstrates even more significant improvements. For instance, PSNR respectively improves by 2.84 dB (36.61 vs. 33.77) and 3.25 dB (36.17 vs. 32.92) with 90 input views. These results indicate that Spener is more robust under dose
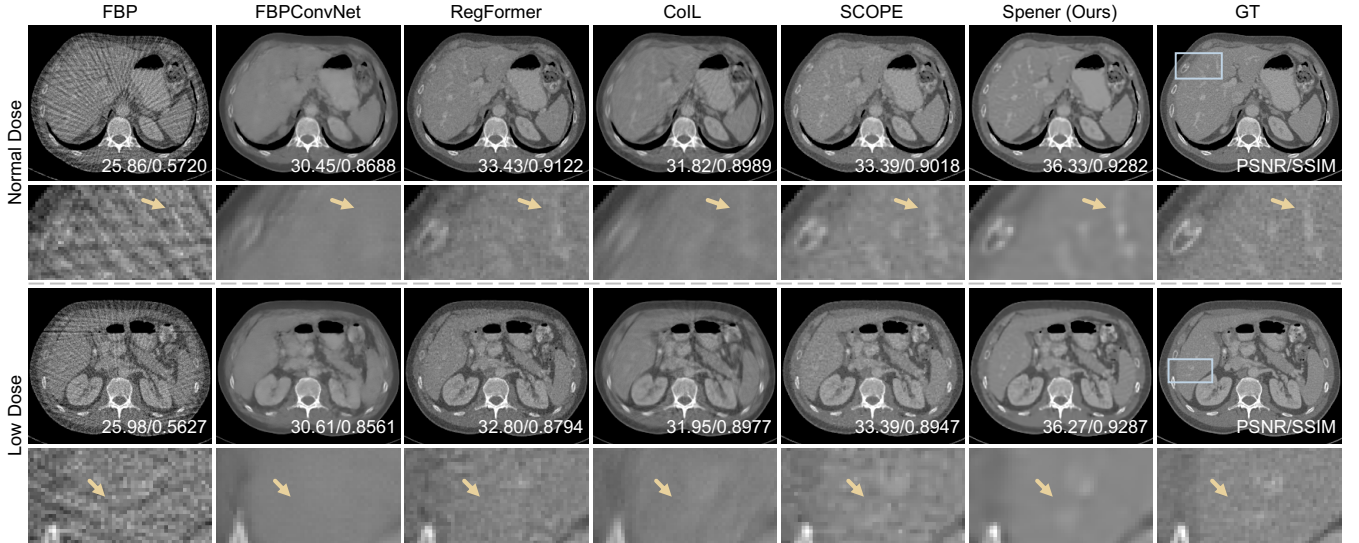
Figure 3: Qualitative results of CT image reconstructed by the compared methods under two dose settings, with both results reconstructed from AAPM dataset with 90 views.

Category	Method	Normal Dose ($I_0 = 10^6$)		Low Dose ($I_0 = 5 \times 10^5$)	
		60 Views	90 Views	60 Views	90 Views
Analytical	FBP	23.25/0.4502	26.01/0.5708	22.92/0.4272	25.59/0.5416
Supervised	FBPConvNet	29.67/0.8527	30.85/0.8787	29.53/0.8473	30.67/0.8721
	RegFormer	<u>33.11</u>/**0.9173**	<u>33.77</u>/<u>0.9256</u>	<u>32.40</u>/<u>0.8865</u>	32.92/0.8901
Unsupervised	CoIL	30.04/0.8555	31.84/0.9001	30.03/0.8557	31.74/0.8966
	SCOPE	32.00/0.8782	33.67/0.9139	31.68/0.8657	<u>33.29</u>/<u>0.9007</u>
	Spener (Ours)	**34.33**/<u>0.9170</u>	**36.61**/**0.9424**	**34.03**/**0.9110**	**36.17**/**0.9349**

Table 2: Quantitative results of compared methods on AAPM dataset under different dose setting. The best performance is highlighted in **bold**, and the second best is <u>underlined</u>.

reduction conditions, which more closely align with real-world CT acquisition scenarios.

Figure 3 shows the qualitative results. As shown in the figure, the current unsupervised INR methods (CoIL and SCOPE) are sensitive to the noise. Specifically, CoIL fails to produce satisfactory results, losing high-frequency details and exhibiting persistent streaking artifacts and blurring. SCOPE reconstructions display some inconsistency points and high-frequency artifacts. However, despite the noise affecting high-frequency details, our proposed Spener successfully recovers these details, as indicated by arrows. Remarkably, Spener reconstructs clean and fine-detail CT images for all dose settings, demonstrating robustness to noise.

Ablation Studies

Effectiveness of Iterative Reconsturction In the Spener framework, the iterative reconstruction process for updating the prior image is crucial for optimal performance. To evaluate the effectiveness of the iteration, we conduct an experiment comparing the performance of Spener with and without iterative prior image updates in SVCT reconstruction. In

the non-iterative scenario, Spener uses the FBP reconstruction as the prior image throughout the entire network optimization process.

Table 3 presents the quantitative results. With iterative updates, Spener achieves significant performance improvements, with approximately a 1 dB increase in PSNR and a 0.014 gain in SSIM. Figure 4 illustrates the qualitative results and performance curves of Spener across different iterations. On the left side of the figure, the progressive reduction of streaking artifacts is evident. The performance curve indicates that quantitative results consistently improve with each iteration, stabilizing around iteration 20.

Effectiveness of Encoder The image encoder in Spener plays a crucial role in extracting effective local image domain priors to enhance network optimization. To explore the effectiveness of the image encoder in Spener, we conducted a comparisons between Spener with and without the image encoder. Specifically, in the absence of an image encoder, the network directly queries the corresponding image intensity from the prior image for prior information during opti-

Strategy	PSNR	SSIM
w/o iteration	36.28±0.69	0.9342±0.0060
w/ iteration	**37.35±0.54**	**0.9484±0.0062**

Table 3: Quantitative results of CT images reconstructed by Spener with and without iteration on the AAPM dataset with 90 views.
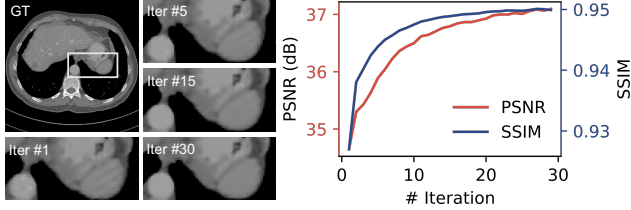

Figure 4: Qualitative and quantitative results of Spener across different iterations on the AAPM dataset with 90 views.

mization. As shown in Table 4, Spener with image encoder achieves significant improvements compared to the version without the encoder.

Settings	Denoiser	PSNR	SSIM
No Noise	×	37.37±0.32	0.9434±0.0050
	√	**37.55±0.63**	**0.9458±0.0056**
Low Dose	×	35.17±0.41	0.9179±0.0065
	√	**36.03±0.61**	**0.9309±0.0051**

Table 5: Quantitative comparisons of the impact of denoiser regularization on the AAPM dataset with 90 view reconstruction under noise-free and low-dose acquisition settings.

Influence of Denoiser Regularization To investigate the impact of denoiser regularization in iterative reconstruction, we evaluated Spener with and without a denoiser on AAPM 90-view reconstructions under no-noise and low-dose conditions. As shown in Table 5, Spener with denoiser outperforms the version without, particularly under low-dose settings, with notable improvements in PSNR and SSIM.

The Choice of Hyperparameter λ To analyze the influence of λ in network optimization, we evaluated the performance of Spener with λ sets as 0, 2.5, and 5 on AAPM 90-view reconstruction under both noise-free and low-dose conditions. Table 6 shows the quantitative results. Overall, the different values of λ do not significantly impact the performance of Spener. In a noise-free setting, Spener with $\lambda = 0$ and $\lambda = 2.5$ achieve a comparable performance. While with a low dose setting, Spener with $\lambda = 2.5$ yields better performance with an approximate 0.1 dB improvement in PSNR.

Figure 5 illustrates the performance curves with different λ settings under low-dose conditions. Spener's performance steadily improves across all λ settings during the initial iterations, with $\lambda = 2.5$ achieving the best results. As iterations

Strategy	PSNR	SSIM
w/o encoder	35.26±0.56	0.9381±0.0068
w/ encoder	**37.35±0.54**	**0.9484±0.0062**

Table 4: Quantitative results of CT images reconstructed by Spener with and without an image encoder on the AAPM dataset with 90 views.

Settings	λ	PSNR	SSIM
No Noise	0.0	37.26±0.61	**0.9456±0.0056**
	2.5	**37.29±0.60**	0.9449±0.0062
	5.0	37.16±0.52	0.9443±0.0054
Low Dose	0.0	35.99±0.44	0.9279±0.0080
	2.5	**36.10±0.49**	0.9321±0.0043
	5.0	35.95±0.53	**0.9322±0.0045**

Table 6: Quantitative comparisons of the influence of λ on AAPM dataset with 90 view reconstruction under noise-free and low-dose acquisition settings.
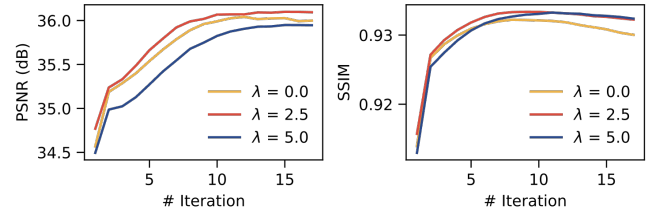

Figure 5: Performance curves of Spener with different λ settings on AAPM dataset with 90 views reconstruction under low-dose acquisition.

continue, a noticeable decline in SSIM occurs when $\lambda = 0$ demonstrates the importance of regularization loss in noisy scenarios. Spener with $\lambda = 2.5$ provides the optimal performance. Therefore, we set $\lambda = 2.5$ for the all experiments.

Conclusion & Discussions

This work introduces Spener, a novel iterative unsupervised SVCT reconstruction method. Unlike SOTA supervised methods, Spener does not require external training data, making it more practical in clinical scenarios with varying acquisition conditions. Compared to SOTA unsupervised INR-based methods, Spener addresses their performance degradation in highly ill-posed SVCT reconstructions, particularly when the measurement data is noisy. Experimental results demonstrate that Spener outperforms current SOTA methods. Despite the success achieved with the current framework in SVCT, there are several areas for improvement: 1) exploring alternative iterative frameworks with INR methods to accelerate model convergence; 2) adopting a more effective image domain encoder to provide robust image priors, enhancing both training speed and performance; 3) extending the framework to address other inverse problems in medical imaging.

Acknowledgments

References

Cancer Moonshot Biobank. 2022. Cancer Moonshot Biobank - Colorectal Cancer Collection (CMB-CRC).

Chen, H.; Zhang, Y.; Chen, Y.; Zhang, J.; Zhang, W.; Sun, H.; Lv, Y.; Liao, P.; Zhou, J.; and Wang, G. 2018. LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE transactions on medical imaging*, 37(6): 1333–1347.

Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8): 2080–2095.

Elbakri, I. A.; and Fessler, J. A. 2002. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE transactions on medical imaging*, 21(2): 89–99.

Geman, D.; and Yang, C. 1995. Nonlinear image recovery with half-quadratic regularization. *IEEE transactions on Image Processing*, 4(7): 932–946.

Jin, K. H.; McCann, M. T.; Froustey, E.; and Unser, M. 2017. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522.

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

Kamilov, U. S.; Bouman, C. A.; Buzzard, G. T.; and Wohlberg, B. 2023. Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1): 85–97.

Lee, H.; Lee, J.; Kim, H.; Cho, B.; and Cho, S. 2018. Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2): 109–119.

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.

Lin, Y.; Yang, J.; Wang, H.; Ding, X.; Zhao, W.; and Li, X. 2024. C^2RV: Cross-Regional and Cross-View Learning for Sparse-View CBCT Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11205–11214.

Liu, J.; Wang, G.; Liu, Z.; Jiang, C.; Pollefeys, M.; and Wang, H. 2023. RegFormer: An Efficient Projection-Aware Transformer Network for Large-Scale Point Cloud Registration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 8417–8426.

Ma, C.; Li, Z.; Zhang, J.; Zhang, Y.; and Shan, H. 2023. FreeSeed: Frequency-band-aware and self-guided network for sparse-view CT reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 250–259. 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.

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.

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.

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.

Ronchetti, M. 2020. TorchRadon: Fast Differentiable Routines for Computed Tomography. arXiv:2009.14788.

Shakouri, S.; Bakhshali, M. A.; Layegh, P.; Kiani, B.; Masoumi, F.; Ataei Nakhaei, S.; and Mostafavi, S. M. 2021. COVID19-CT-dataset: an open-access chest CT image repository of 1000+ patients with confirmed COVID-19 diagnosis. *BMC research notes*, 14(1): 178.

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*.

Sun, Y.; Liu, J.; Xie, M.; Wohlberg, B.; and Kamilov, U. S. 2021. Coil: Coordinate-based internal learning for tomographic imaging. *IEEE Transactions on Computational Imaging*, 7: 1400–1412.

Wang, G.; Yu, H.; and De Man, B. 2008. An outlook on x-ray CT research and development. *Medical physics*, 35(3): 1051–1064.

Wu, Q.; Feng, R.; Wei, H.; Yu, J.; and Zhang, Y. 2023. Self-Supervised Coordinate Projection Network for Sparse-View Computed Tomography. *IEEE Transactions on Computational Imaging*, 9: 517–529.

Wu, W.; Hu, D.; Niu, C.; Yu, H.; Vardhanabhuti, V.; and Wang, G. 2021. DRONE: Dual-domain residual-based optimization network for sparse-view CT reconstruction. *IEEE Transactions on Medical Imaging*, 40(11): 3002–3014.

Xia, W.; Yang, Z.; Zhou, Q.; Lu, Z.; Wang, Z.; and Zhang, Y. 2022. A transformer-based iterative reconstruction model for sparse-view ct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 790–800. Springer.

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 *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 442–452. Springer.

Zhang, K.; Li, Y.; Zuo, W.; Zhang, L.; Van Gool, L.; and Timofte, R. 2021. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10): 6360–6376.

Zhang, K.; Zuo, W.; Gu, S.; and Zhang, L. 2017. Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3929–3938.

Zhang, Z.; Liang, X.; Dong, X.; Xie, Y.; and Cao, G. 2018. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE transactions on medical imaging*, 37(6): 1407–1417.