Learned Full-sampling Reconstruction from Incomplete Data

Weilin Cheng, Yu Wang, Hongwei Li, and Yuping Duan*

***Abstract*—Sparse-view and limited-angle Computed Tomography (CT) are very challenging problems in real applications. Due to the high ill-posedness, both analytical and iterative reconstruction methods may present distortions and artifacts for such incomplete data problems. In this work, we propose a novel reconstruction model to jointly reconstruct a high-quality image and its corresponding high-resolution projection data. The model is built up by deploying regularization on both CT image and projection data, as well as by introducing a novel full-sampling condition to fuse information from both domains. Inspired by the success of deep learning methods in imaging, we utilize the convolutional neural networks to embed and learn both the interrelationship between raw data and reconstructed images and prior information such as regularization, which is implemented in an end-to-end training process. Numerical results demonstrate that the proposed approach outperforms both variational and popular learning-based reconstruction methods for the sparse-view and limited-angle CT problems.**

***Index Terms*—Tomography, sparse-view, limited-angle, Radon inpainting, full-sampling condition, deep learning**

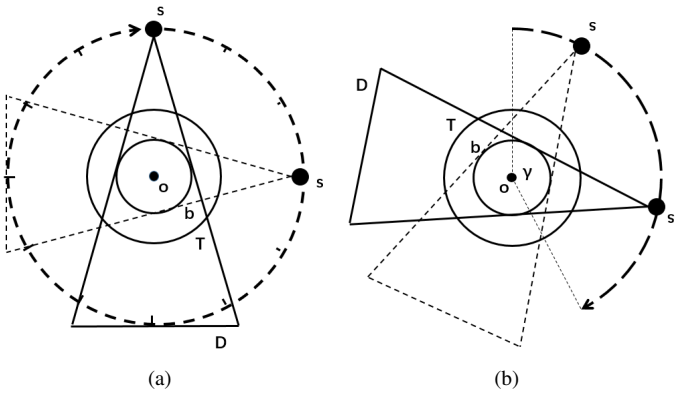


Fig. 1. Sketch map of fan-beam scanning geometry for CT reconstruction: (a) Sparse-view CT Reconstruction; (b) Limited-angle CT Reconstruction. In the figures, s denotes the X-ray source, T denotes the turntable, b denotes the field-of-view, D denotes the detector array, o denotes the rotation center of the X-ray source, and γ denotes the rotation angle, which is less than 180° plus the fan-angle for limited-angle CT.

I. Introduction

X-RAY computed tomography (CT) plays an important role in the diagnostic of cancer and radiotherapy. One major concern of CT is the radiation dose imposed on patients. Sparse-view CT is a popular technique used to reduce the radiation dose by decreasing the number of projections. On the other hand, the projection data of the mobile C-arm CT [1] and dental CT [2] are normally obtained within an angular range less than π, which are known as the limited-angle reconstruction problem. Image reconstruction from sparse-view and limited-angle measurements is typical examples of ill-posed inverse problems. Fig. 1 depicts the fan-beam scanning geometry of sparse-view CT and limited-angle CT, where the projection data is acquired with large scanning angular intervals in (a) and within a limited angular range in (b). Due to the insufficient projections, the conventional reconstruction methods such as the filtered back-projection (FBP) [3], algebraic reconstruction technique (ART) [4] and simultaneous algebraic reconstruction technique (SART) [5] may produce severe streaking artifacts and blurring artifacts.

Asterisk indicates corresponding author.

W. Cheng is with Center for Applied Mathematics, Tianjin University, Tianjin 300072, China.

Y. Wang is with Alibaba Group, Hangzhou, China.

H. Li is with School of Mathematical Sciences and Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China.

*Y. Duan is with Center for Applied Mathematics, Tianjin University, Tianjin, China 300072. E-mail: yuping.duan@tju.edu.cn.

Such inverse problems can be formulated as to estimate an image $u \in X$ from the measurement $g \in Y$

$$g = Au + n, \tag{1}$$

where the reconstruction space X and data space Y are typically Hilbert spaces, $A : X \to Y$ is a linear operator representing the forward system matrix (also named projection matrix) consisting of elements defined according to intersection lengths between the beamlets and the pixels of the image u, and $n \in Y$ is the random noise generated during the imaging process. Due to the lack of scanning views, the linear operator A has a large kernel space and the linear system (1) is sensitive to small perturbations. Numerous regularization based algorithms have been proposed for CT reconstruction including total variation (TV) [6], [7], [8], nonlocal TV [9], tight frame [10], dictionary learning [11], etc. Compared with analytical reconstruction algorithms, the TV-based model can significantly improve the reconstruction quality. However, there are still some drawbacks. First, it assumes that the CT images are piecewise constant, which may lead to unrealistic reconstruction, especially for the limited-angle problem. Second, the iterative algorithms are usually time-consuming and suffer from tricky parameter tuning.

Thanks to the development of deep convolutional neural networks (CNN) in a broad range of computer vision tasks, deep learning techniques have been actively used in the medical imaging community [12], [13], [14]. We follow the review work [15] to categorize the deep CNN models for CT

reconstruction into two categories: post-processing and raw-to-image. Post-processing is the kind of methods to enhance the initial low-quality reconstructed image, e.g., FBP reconstructed image, to its high-quality counterpart. Chen *et al.* [16] designed a symmetrical convolutional and deconvolutional neural network with shortcut connections to approximate the mapping between the degraded images and clean images. Jin *et al.* [17] used the U-net as a post-processing step after the reconstruction of the FBP to mitigate noises and artifacts. Zhang *et al.* [18] fed the FBP reconstruction to a DenseNet and deconvolution based network for sparse-view CT reconstruction. Because the post-processing methods rely on initial reconstruction, plenty of information contained in the raw data is ignored, which may lose effect for noisy or incomplete data.

For raw-to-image methods, the reconstructed images are directly estimated from the raw data in the Radon domain. The pioneering work of Yang *et al.* [19] reformulated an ADMM algorithm for compressive sensing MR imaging into a deep network whose parameters are learned end-to-end in the training phase. Since then, the unrolling technique has become popular to devise various neural network architectures for better CT reconstructions in terms of both quality and efficiency. Adler and Öktem [20] developed a partially learned gradient descent scheme for the solution of ill-posed inverse problems such as (1). Kang *et al.* [21] constructed a deep CNN network in the wavelet domain, which trained the wavelet coefficients from the CT images after applying the contourlet transform. Chen *et al.* [22] unfolded the field of experts (FoE) regularized CT reconstruction model into a deep learning network, all parameters of which can be learned from the training process. Adler and Öktem [23] proposed a fully learned algorithm based on the primal-dual formulation of (1) for CT reconstruction. Gupta *et al.* [24] presented a novel image reconstruction method by replacing the projector in a projected gradient descent method with a CNN. Meinhardt *et al.* [25] used the plug-and-play strategy to replace the regularization term with a denoising neural network. Sun *et al.* [26] proposed an online plug-and-play algorithm based on the proximal gradient method with convergence analysis for image reconstruction. Liu *et al.* [27] integrated a deep regularization term into the variational model to learn the data consistency from the observed data. Kang *et al.* [28] developed a deep convolutional framelet denoising for low-dose CT via wavelet residual network. The aforementioned methods are designed for the general CT reconstruction problem. When dealing with incomplete data, especially limited-angle CT, they may suffer from artifacts such as boundary distortion, edge blurring, and intensity biases.

Recently, the deep learning-based CT reconstruction techniques have been developed for the incomplete data reconstruction problem. Kelly and Matthews [29] utilized the neural network as a quasi-projection operator with a least-square problem in an iterative way, where the CNN is trained to encode high-level information of images being imaged. Han and Ye [30] proposed new multi-resolution deep learning schemes based on the frame condition to overcome the limitation of U-net for sparse-view CT. Pelt *et al.* [31] developed a mixed-scale dense convolutional neural network architecture to improve tomography reconstruction from limited data. Würfl *et al.* [32] mapped the Feldkamp-Davis-Kress (FDK) algorithm to the neural networks by introducing a novel cone-beam back-projection layer for limited-angle problems. Dong, Li and Shen [33] proposed a joint spatial-Radon domain reconstruction (JSR) model for sparse-view CT imaging and was recently reformulated into the feed-forward deep network for incomplete data problems [34]. Wang *et al.* [35] proposed an ADMM-based deep reconstruction algorithm for limited-angle reconstruction by replacing one part of the ADMM method for denoising with a deep CNN. Bubba *et al.* [36] developed a hybrid reconstruction framework that fuses model-based sparse regularization with data-driven deep learning for limited-angle computed tomography. These learning-based methods can achieve better reconstruction results than conventional reconstruction methods for incomplete data.

II. Related works and our contributions

A. Variational regularization and PD-net

Following the Bayesian approach [37], the solution of the inverse problem (1), i.e., the posterior density, can be obtained based on the knowledge of the measurements. The joint probability of observing both u and g is defined as

$$P(u,g) = P(g|u)P(u) = P(u|g)P(g).$$

By re-arranging the above equation, we obtain the well-known Bayesian formula, i.e.,

$$\underbrace{P(u|g)}_{\text{a-posteriori}} = \frac{\overbrace{P(g|u)}^{\text{likelihood}}\ \overbrace{P(u)}^{\text{a-priori}}}{\underbrace{P(g)}_{\text{normalization}}}. \tag{2}$$

Thus, a reconstructed image providing the measurement can be obtained by calculating the point estimates of the posterior density such as the *maximum a posteriori (MAP)* estimation. By taking the negative logarithm, we arrive at the following minimization problem

$$u = \arg\min_u \ \mathcal{D}(Au, g) + \mathcal{R}(u), \tag{3}$$

where $\mathcal{D}(Au,g) := -\log\big(P(g|u)\big)$ is the data fidelity term and $\mathcal{R}(u) := -\log(P(u)$ is the regularization term. The task of solving (3) mainly includes: 1) how to define the data fidelity to describe the interrelationship between g and u; and 2) how to model the regularization according to the prior information of u.

Adler and Öktem [23] applied the Chambolle-Pock primal-dual algorithm [38] for the general minimization problem (3) as follows

$$\begin{cases} p^{k+1} = (\mathcal{I} + \tau\partial\mathcal{D}^*)^{-1}(p^k, Au^{k+1}, g), \\ u^{k+1} = (\mathcal{I} + \sigma\partial\mathcal{R})^{-1}(u^k, A^*p^k), \end{cases} \tag{4}$$

where $\mathcal{I}$ denotes the identity operator, $\mathcal{D}^*$ denotes the conjugate of $\mathcal{D}$, A^* represents the backward projection, p is the dual variable of u and σ, τ are two positive parameters. Instead

of choosing handcrafted data fidelity and regularization, the resolvent operators are learnt by the neural networks such as

$$\begin{cases} p^{k+1} = \Gamma_{\theta^p}(p^k, Au^{k+1}, g), \\ u^{k+1} = \Lambda_{\theta^u}(u^k, A^*p^k), \end{cases}$$

where Γ_{θ^p} and Λ_{θ^u} are neural networks parameterized by θ^p and θ^u, respectively. Given a set of samples $(g_1, u_1^*), \ldots, (g_N, u_L^*)$, by minimizing the following empirical loss

$$\mathcal{L}_{PD}(\theta) = \frac{1}{L}\sum_{i=1}^{L} \|u_i(\theta) - u_i^*\|^2, \tag{5}$$

the PD-net has a significant performance boost compared to the handcrafted reconstruction models and the post-processing methods.

B. JSR model and JSR-net

The joint spatial-Radon domain reconstruction (JSR) model was proposed in [33] to suppress the artifacts induced by noise and incomplete data, which is formulated as

$$\min_{u,f} \ \mathcal{F}(u, f) + \mathcal{R}(u, f) \tag{6}$$

with

$$\mathcal{F}(u, f) = \frac{1}{2}\|R_{\Gamma^c}(f-g)\|_2^2 + \frac{\alpha}{2}\|R_{\Gamma}(Au-f)\|_2^2 + \frac{\beta}{2}\|R_{\Gamma^c}(Au-g)\|_2^2,$$

and

$$\mathcal{R}(u, f) = \lambda_1\|W_1 u\|_{1,2} + \lambda_2\|W_2 f\|_{1,2}.$$

Note that u and f in (6) denote the underlying CT image and the restored projection data, respectively. The notation R_Γ is a restriction operator with Γ indicating the missing data region and Γ^c indicating the complement of Γ. Besides, W_1, W_2 are tight frame transforms, and α, β, λ_1, λ_2 are positive parameters. The minimization problem (6) is solved by an alternative optimization algorithm, two sub-minimization problems of which can be efficiently solved by the split Bregman iteration. The JSR-model improves the quality of CT reconstruction by enforcing the data consistency in the image domain and the Radon domain simultaneously. However, two main drawbacks limit its practical usage: 1) As reported in [33], the computational cost of the JSR model is three times that of the TV model; 2) The reconstruction quality of the JSR model for the limited-angle problem is still not satisfactory as shown by the numerical experiments in [34].

Recently, Zhang *et al.* [34] extended the JSR model (6) into the JSR-net for sparse-view and limited-angle CT reconstruction, which is a feed-forward network by unrolling the iterative scheme of the JSR model. More specifically, the JSR-net used the CNNs to approximate not only the inverse operators but also the thresholding operators. In the training process, the following structure-semantic-ℓ_2 loss is used

$$\mathcal{L}_{JSR}(\theta) = \sum_{i=1}^{L}\Bigg(\underbrace{\|u_i(\theta) - u_i^*\|^2}_{\mathcal{L}_{SD}} + \alpha_1 \underbrace{\|R_{\Gamma^c}A(u_i(\theta) - u_i^*)\|^2}_{\mathcal{L}_{RD}} + \alpha_2 \underbrace{\big(1 - \mathrm{SSIM}(u_i(\theta), u_i^*)\big)}_{\mathcal{L}_{SSIM}} + \alpha_3 \underbrace{\|\mathrm{sem}(u_i(\theta)) - \mathrm{sem}(u_i^*)\|^2}_{\mathcal{L}_{sem}}\Bigg), \tag{7}$$

where $\mathcal{L}_{SD}$ and $\mathcal{L}_{RD}$ are the ℓ_2 loss in the spatial domain and Radon domain, respectively. $\mathcal{L}_{SSIM}$ is the error summation of the SSIM between the reconstructed image and the ground truth over mini batches. $\mathcal{L}_{sem}$ is the ℓ_2 norm of the difference between the level sets of the reconstructed image and ground truth. As reported in [34], the JSR-net achieves better reconstruction quality than the JSR model by utilizing multi-domain image features.

C. Our contributions

In this paper, we propose a novel, end-to-end trainable approach for sparse-view and limited-angle CT reconstruction by utilizing prior information in both the image domain and the Radon domain. We extend our previous work [39] in terms of both model development and evaluation. The current model is derived by minimizing the data fidelity and two regularization terms on the reconstructed image and its high-resolution projection data, respectively. Besides, we introduce a data consistency term as the constraint to the minimization problem by employing the full-sampling condition. Both the penalty method and the primal-dual method are applied in designing the alternating minimization algorithm, where all resolvent operators are replaced by the convolutional neural networks. To fully exploit the advantages of our full-sampling reconstruction network (FSR-net), a joint mean square error of the reconstructed image and projection data is introduced as the loss function for training the network. We evaluate the performance of the proposed model on both synthetic ellipse dataset and AAPM human phantom dataset, which shows improved image quality especially on the preservation of tissue textures.

The differences between our FSR-net and the two relevant works, i.e., PD-net and JSR-net, can be clarified as follows. Compared with PD-net, we solve a different minimization problem, where both the image and the corresponding full projection data are unknown variables. Although both methods use the unrolled iterative learning scheme, the employed algorithms are different such that the PD-net is the reconstruction subproblem in our algorithm. The Radon domain inpainting was introduced in [33] in a variational formulation and extended to JSR-net in [34]. The main difference between JSR-net and our model lies in the data fidelity, where ours is designed using the prior knowledge on the relationship between the CT image and the full-sampling projection data while JSR-net depends on the regularization term on acquired projection data. Relying on the stronger data consistency and a better choice to encode the priors existing between the image domain and the Radon domain, our model can significantly improve the reconstruction quality.

III. Learned Full-sampling Reconstruction

A. The full-sampling reconstruction

For sparse-view and limited-angle reconstruction problems, either only a few number of projections are acquired by equally sampling the scanning angular range or only the projections in a limited scanning angular range are acquired. Both cases result in insufficient sampling such that the information

contained in g is far from being enough. In such conditions, the system matrix A has much fewer rows than columns, which means that the linear inverse problem (1) has infinitely many solutions. Thus, it is crucial to select the optimal solution from the solution set.

We develop a new model to simultaneously reconstruct the CT image and its corresponding higher resolution projection data. Instead of solely relying on the regularization of acquired projection data, we introduce the following full-sampling condition. Assume $A_F : X \to Z$ to be the full-sampling system matrix defined on the full angular range, i.e., 2π. Observing that the high-resolution projection data f and the reconstructed image u should satisfy the following full-sampling condition

$$A_F u = f, \tag{8}$$

we consider the following constrained minimization problem by jointly reconstructing the spatial and Radon domain data

$$\begin{aligned} &\min_{u,f} \ \mathcal{D}(Au, g) + \mathcal{R}(u) + \mathcal{F}(f) \\ &\text{s.t.,} \ \ A_F u = f, \end{aligned} \tag{9}$$

where $\mathcal{R}(u)$ and $\mathcal{F}(f)$ represent the regularization term for the reconstructed image u and projection data f, respectively. The merits of the full-sampling condition (8) are twofold:

- Because the full-sampling system matrix A_F has much more rows than A, the full-sampling condition can efficiently reduce the size of the solution set such that the optimal solution of the reconstruction problem (9) becomes easier to track;
- The full-sampling condition can bind the projection data and the image tighter. When involved in the convolutional neural networks, it helps to squeeze and encode as much as possible the priors regarding the relationship between images and their projection data in the training process, i.e., the mapping from the projection data to their reconstructions could be better modeled and learned.

The constrained minimization problem (9) can be further reformulated into an unconstrained minimization problem using the penalty method as follows

$$\min_{u,f} \ \mathcal{D}(Au, g) + \mathcal{R}(u) + \frac{1}{2\mu}\|A_F u - f\|^2 + \mathcal{F}(f), \tag{10}$$

where μ is a positive regularization parameter. The functional $\mathcal{R}(\cdot)$ and $\mathcal{F}(\cdot)$ can be chosen according to prior knowledge on u and f such as TV regularization and tight frame, and the functional $\mathcal{D}(\cdot,\cdot)$ can be explicitly defined based on noise distribution in the observed data, e.g., L^2 norm for Gaussian noise. In our work, we use the CNNs to learn not only the prior information of u and f, but also the interrelationship between g and u.

B. Learned alternating direction algorithm

Because all terms in (10) contain the variable u, we introduce a new variable $\tilde{u}$ and rewrite it as the following minimization problem

$$\min_{u,f,\tilde{u}} \mathcal{D}(A\tilde{u}, g) + \mathcal{R}(\tilde{u}) + \frac{1}{2r}\|\tilde{u} - u\|^2 + \frac{1}{2\mu}\|A_F u - f\|^2 + \mathcal{F}(f), \tag{11}$$

where r is a positive parameter. The main advantage is that we decouple the minimization problem (10) into the reconstruction step and consistency step such that the alternating direction method can be used to solve the multi-variable optimization problem (11). Starting from an initial guess u^0, we compute a sequence of iterates

$$\tilde{u}^1, (u^1, f^1), \tilde{u}^2, (u^2, f^2), \ldots, \tilde{u}^k, (u^k, f^k),$$

where $k \in \{1, 2 \ldots, I\}$ denotes the kth iteration and I denotes the total number of the iteration. The variables $\tilde{u}^k$ and (u^k, f^k) can be estimated from the following minimization problems, respectively

$$\begin{cases} \text{Reconstruction:} \ \min\limits_{\tilde{u}} \mathcal{D}(A\tilde{u}, g) + \mathcal{R}(\tilde{u}) + \frac{1}{2r}\|\tilde{u} - u^{k-1}\|^2, \\ \text{Consistency:} \ \min\limits_{u,f} \mathcal{F}(f) + \frac{1}{2\mu}\|A_F u - f\|^2 + \frac{1}{2r}\|\tilde{u}^k - u\|^2. \end{cases}$$

In the reconstruction step, we need to solve a minimization problem similar to the regularization model (3). Thus, we apply the learned primal-dual algorithm as follows

$$\begin{cases} p^{k+1} = \arg\min\limits_{p} \ \mathcal{D}^*(p, g) - \langle A\tilde{u}^k, p\rangle + \frac{1}{2\tau}\|p - p^k\|_2^2, \\ \tilde{u}^{k+1} = \arg\min\limits_{\tilde{u}} \ \mathcal{R}(\tilde{u}) + \langle A\tilde{u}, p^{k+1}\rangle + \frac{1}{2r}\|\tilde{u} - u^k\|_2^2, \end{cases}$$

with τ being a positive parameter. The consistency step is to reconstruct both the image data and full Radon domain data by fitting with the image generated in the reconstruction step, which is also a minimization problem involving multiple variables and can be efficiently solved by the alternating minimization algorithm as follows

$$\begin{cases} f^{k+1} = \arg\min\limits_{f} \ \mathcal{F}(f) + \frac{1}{2\mu}\|A_F u^k - f\|^2 + \frac{1}{2\sigma}\|f - f^k\|_2^2, \\ u^{k+1} = \arg\min\limits_{u} \ \frac{1}{2\mu}\|A_F u - f^{k+1}\|^2 + \frac{1}{2r}\|u - \tilde{u}^{k+1}\|_2^2, \end{cases}$$

with σ being a positive parameter. Consequently, we use the neural networks to learn the solutions to the variables in each iteration, which gives

$$\begin{cases} p^{k+1} = (\mathcal{I} + \tau\partial\mathcal{D}^*)^{-1}(p^k, \tau A\tilde{u}, g), \\ \tilde{u}^{k+1} = (\mathcal{I} + r\partial\mathcal{R})^{-1}(u^k, rA^* p^{k+1}), \\ f^{k+1} = \left((\sigma + \mu)\mathcal{I} + \sigma\mu\partial\mathcal{F}\right)^{-1}(\mu f^k, \sigma A_F u^k), \\ u^{k+1} = (\mathcal{I} + rA_F^* A_F)^{-1}(rA_F^* f^{k+1} + \mu\tilde{u}^{k+1}), \end{cases} \tag{12}$$

with A^* and A_F^* representing the backward projections.

In our implementations, the following techniques are utilized for better efficiency:

- Following the PD-net, we extend the function spaces of the variables to allow the algorithm to memorize between iterations, e.g.,

$$u = [u^{(1)}, u^{(2)}, \ldots, u^{(N_u)}].$$

- Instead of using artificial variables τ, r, σ, and μ, we allow the networks to learn how to combine the operators and the variables. In this way, the resolvent operators can vary with iterations, which can help to further improve the reconstruction quality.
- Although the subproblem concerning u^{k+1} in the consistency step can be solved by conventional reconstruction algorithms, we use the learned resolvent operator to save computations.

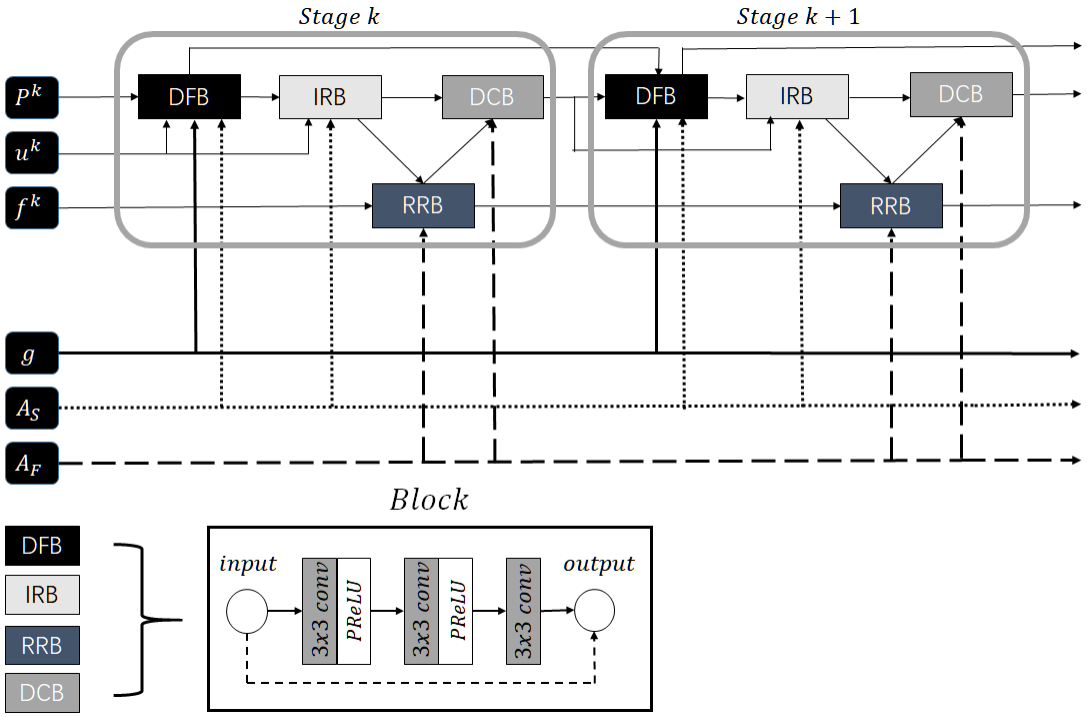


Fig. 2. The data flow graph of the FSR-net to solve the tomography problem. According to Algorithm 1, there are four variables in each stage which are computed sequentially corresponding to DFB (data fidelity block), IRB (image domain regularization block), RRB (radon domain regularization block), and DCB (data consistency block). As shown, each block has the same architecture with three conv-layers.

- We assume the constraint $\tilde{u} = u$ holds unconditionally during the iterations process. Therefore, f^{k+1} is calculated based on $A_F\tilde{u}^{k+1}$ rather than $A_F u^k$ as $\tilde{u}^{k+1}$ has been already updated in the previous step.

Now, we are ready to summarize the unrolled alternating direction algorithm for solving the full-sampling reconstruction problem (11) as Algorithm 1. Henceforth, we call it the *Learned Full-Sampling Reconstruction Network* (FSR-net).

Algorithm 1 Learned full-sampling reconstruction algorithm

Initialize u^0, f^0, p^0
for $k = 0, \ldots, I$, **do**
  $p^{k+1} \leftarrow \Gamma_{\theta^p}(p^k, Au^k, g)$;
  $\tilde{u}^{k+1} \leftarrow \Lambda_{\theta^{\tilde{u}}}(u^k, A^* p^{k+1})$;
  $f^{k+1} \leftarrow \Pi_{\theta^f}(f^k, A_F\tilde{u}^{k+1})$;
  $u^{k+1} \leftarrow \Xi_{\theta^u}(\tilde{u}^{k+1}, A_F^* f^{k+1})$;
return u^I, f^I

IV. Implementation

Our FSR-net is generated based on the full-sampling reconstruction model (11), which is implemented in Python using Operator Discretization Library (ODL), the ASTRA Toolbox and Tensorflow 1.8.0 on a Tesla P100 GPU. Tensorflow is an end-to-end open-source platform for machine learning, ASTRA toolbox is a MATLAB and Python toolbox of high-performance GPU primitives for 2D and 3D tomography, and the ODL is a Python library for fast prototyping focusing on inverse problems.

A. Network architecture

The unrolling strategy is a discriminative learning method by unrolling an iterative optimization algorithm into a hierarchical architecture. Fig. 2 depicts the network structures of the kth and $(k+1)$th stages. In each stage, there are four blocks are corresponding to the four variables, i.e., the data fidelity block (DFB) w.r.t. p, image domain regularization block (IRB) w.r.t. $\tilde{u}$, Radon domain regularization block (RRB) w.r.t. f and data consistency block (DCB) w.r.t. u. The networks take three inputs: the incomplete projection data g, the system matrix A, and the full-sampling system matrix A_F. We also illustrate the detailed structures of each block, which involves a 3-layer network. The total depth of the network depends on the number of the stages contained in the network, which is chosen to balance the receptive fields and the total number of parameters. For all the four variables in Algorithm 1, the learned resolvent operators have the same form as

$$\mathrm{Id} + \mathrm{W}_{w_3,b_3} \circ \mathrm{PR}_{c_2} \circ \mathrm{W}_{w_2,b_2} \circ \mathrm{PR}_{c_1} \circ \mathrm{W}_{w_1,b_1},$$

where Id is the identity operator to make the network similar to a residual network [40]. The merit of the residual structure is that the network is easy to be trained as each update just need to learn a small offset from the identity operator. The operator W_{w_j,b_j} is the convolution operator including both weights w_j and biases b_j, $j = 1, 2, 3$, i.e.,

$$\mathrm{W}_{w_j,b_j} : X^n \to X^m,$$

the k-th component of which is given by

$$\left[\mathrm{W}_{w_j,b_j}\left([u^{(1)},...,u^{(n)}]\right)\right]^{(k)} = b_j^{(k)} + \sum_{i=1}^{n} w_j^{(i,k)} * u^{(i)}$$

with $b_j \in X^m$, $w_j \in X^{n\times m}$, and m, n denoting the dimensions. We adopt the parametric rectified linear units (PReLU) as the activation function, i.e.,

$$\mathrm{PR}_{c_j}(u) = \begin{cases} u, & \text{for } u \geq 0, \\ -c_j u, & \text{for } u < 0, \end{cases} \tag{13}$$

where c_j is automatically calculated by network feedback, and each channel uses a different c_j.

It is well-known that too few channels will weaken the expression ability of the network, while too many channels will increase the difficulties in training. We set the initial channels of the variables in each stage as $N_u = N_{\tilde{u}} = 6$ and $N_p = N_f = 7$. We use the convolution kernel of size 3×3 and set the numbers of channels in each stage as $9 \to 32 \to 32 \to 7$ for p, $7 \to 32 \to 32 \to 6$ for $\tilde{u}$ and u, and $8 \to 32 \to 32 \to 7$ for f, where the differences in numbers are due to the inputs.

Moreover, we choose the Xavier initialization scheme for the convolution parameters and the zero initialization for all biases. The convolution stride is set as 1 and the padding strategy is chosen as 'SAME' in the network.

B. The full-sampling system matrix

Suppose the dimension of the system matrix A be $M \times N^2$ for the circular fan-beam CT such that the image array is of the size $N \times N$. The number of ray integrations M can be defined as

$$M = N_{\text{views}} \times N_{\text{bins}}, \tag{14}$$

where N_{views} denotes the number of views and N_{bins} denotes the number of bins on the detector. To be specific, the scanning arc, either 2π for sparse-view CT or γ for limited-angle CT, is divided into N_{views} equally spaced angular intervals, while the detector D is equally divided into N_{bins} units. In the following, we determine our full-sampling system matrix according to the pioneering work [41] by deploying either its invertibility or numerical stability property.

- Invertible Full-Sampling (IFS) system matrix

To guarantee the invertibility of the system matrix, A_F should have more rows than columns in case there are linearly dependent rows, i.e.,

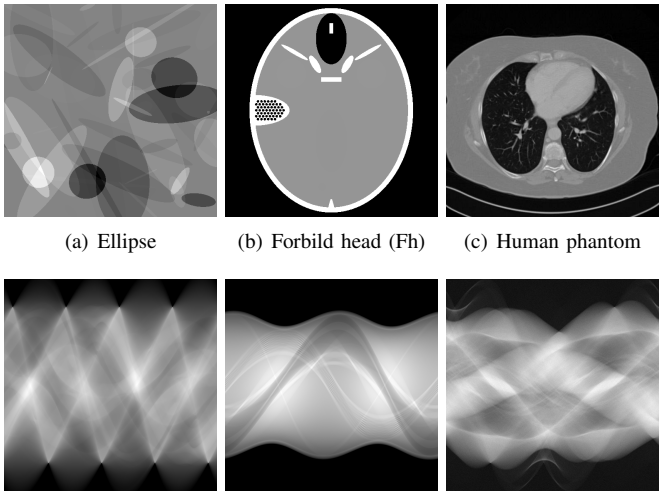
$$N_{\text{views}} \times N_{\text{bins}} \geq N^2,$$

where N_{views} is used to equally divide the 2π arc.

- Stable Full-Sampling (SFS) system matrix

Alternatively, we can define the full-sampling system matrix to guarantee the numerical stability by requiring the condition number of A_F as small as possible, which can be realized by using $2N$ samples in both the view and bin directions, i.e.,

$$N_{\text{views}} = N_{\text{bins}} = 2N.$$



(a) Ellipse (b) Forbild head (Fh) (c) Human phantom

(d) Ellipse sinogram (e) Fh sinogram (f) Human sinogram

Fig. 3. Examples of data which are used for training and validation. Left: randomly generated ellipses. Middle: validation data generated by the modified Forbild head phantom. Right: validation data generated from the human phantoms.

It should be noted that instead of generating and storing the full-sampling matrix A_F, we compute the matrix-vector products involving A_F and A_F^* on the fly in our code implementations, to avoid possible huge memory consumption. More specifically, we use the function 'FanFlatGeometry' with 'astra_cuda' in the ODL to construct an operator, which acts as the forward projection operator to obtain the Radon domain data with GPU implementation. Besides, we use the function 'as_tensorflow_layer' in ODL to ensure the operator can act on tensors and generate tensors in our neural network. The backward projection is implemented in the same way.

C. Network loss and optimization

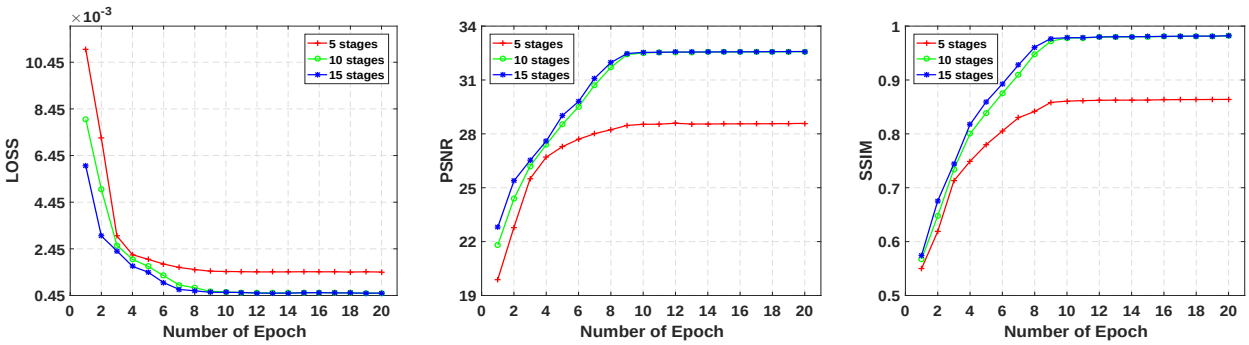
Let $\Theta = \{\theta^p, \theta^{\tilde{u}}, \theta^f, \theta^u\}$ be the parameters in the network and L be the number of training samples. We use the following joint mean squared error (JMSE) as the loss function for our FSR-net

$$\mathcal{L}(\Theta) = \frac{1}{2L}\sum_{i=1}^{L}\left(\left\|u_i(\Theta) - u_i^*\right\|_2^2 + \alpha\left\|f_i(\Theta) - A_F u_i^*\right\|_2^2\right), \tag{15}$$

where $u_i(\Theta)$ and $f_i(\Theta)$ denote the i-th reconstructed image and Radon domain data, u^* denotes the reference image, and α is the trade-off parameter. In the numerical experiments, we fix $\alpha = 1$ for all applications. The gradient of our loss function is computed by Tensorflow automatically and applied to the network by the back propagation.

For a fair comparison, most experimental parameters are set as the same as the PD-net in [23]. We adopt the adaptive moment estimation (Adam) [42] to optimize the learning rate by setting the parameter $\beta_2 = 0.99$ and other parameters to their default values. The learning rate schedule is set according to cosine annealing [43] to improve training stability such that

$$\eta^t = \frac{\eta^0}{2}\left(1 + \cos\left(\pi\frac{t}{t_{max}}\right)\right)$$



(a) LOSS (b) PSNR (c) SSIM

Fig. 4. The values of loss function, PSNR and SSIM with respect to different numbers of stages and numbers of epochs in our network, where red line is obtained using 5 stages, green line is obtained by 10 stages, and the blue line is obtained by 15 stages. All curves are evaluated on the Forbild head phantom with noiseless 60 views.

with the initial learning rate being $\eta^0 = 10^{-3}$ for all experiments. To further improve the training stability, the global gradient norm clipping is performed by limiting the gradient norm to 1. Besides, the batch size is set to 1 for both the ellipse data and human phantom data.

V. Numerical Results

In this section, we evaluate our FSR-net on both sparse-view and limited-angle reconstruction problems using both analytical phantom and human phantom datasets.

A. Datasets and settings

The examples of the two datasets used in our numerical experiments are exhibited in Fig. 3. More details about the datasets are described below.

1) Ellipse phantom: The training dataset includes 2000 randomly generated ellipse images of size 512×512 as shown in Fig. 3(a). We use the fan-beam geometry to generate the projection data and choose the Forbild head phantom [44] as the testing image, where the display window is set as [0,1] for the Forbild head phantom as shown in Fig. 3(b). In order to meet the requirement of being the IFS and SFS system matrices, we set $N_{\text{views}} = N_{\text{bins}} = 512$ and $N_{\text{views}} = N_{\text{bins}} = 1024$, respectively.

2) Human phantom: We use the clinical data "The 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge" [45], which contains 10 full-dose scans of the ACR CT accreditation phantom. We use 9 data as the training profile and leave 1 data for the evaluation, resulting in 2164 images of size 512×512 for training and 214 images for testing. The two-dimensional fan-beam geometry is adopted with geometrical parameters displayed in TABLE I. For limited-angle reconstruction, we set the scanning angular interval as 1 degree.

TABLE I
Reconstruction parameters of human phantom.

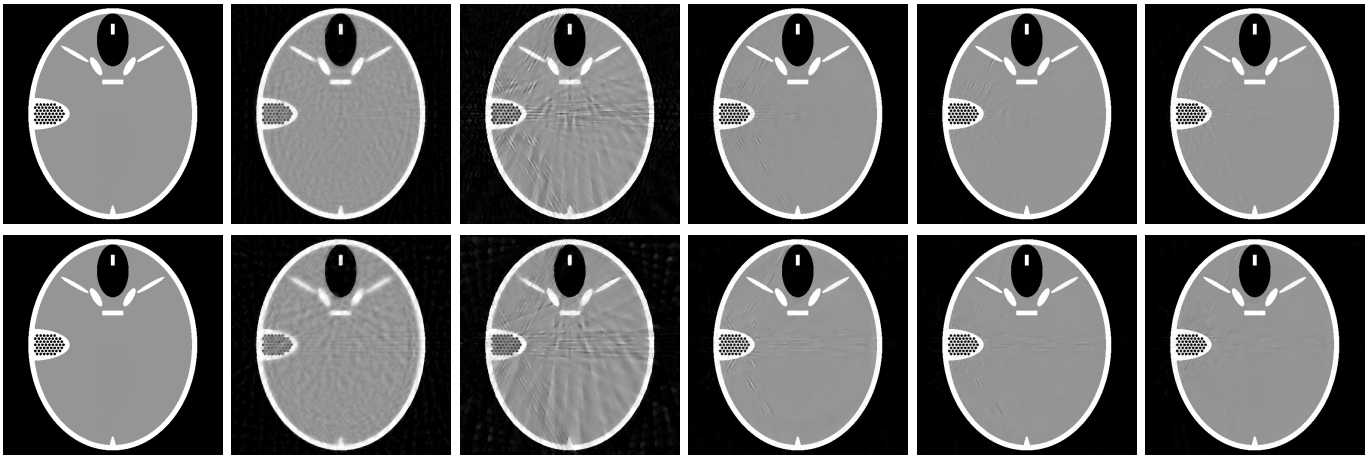
Parameters	Value
Scanning type	Fan-beam scan
Reconstructed image size	512×512
Number of detector units	736
X-ray source to axis distance	500 mm
Axis to detector distance	500 mm

B. Comparison algorithms

We compare our FSR-net with several state-of-the-art algorithms including both the variational method and learning based methods, i.e.,

- TV model: *the TV regularized reconstruction model* in [7]. We tuned the balance parameter $\lambda \in [0.03, 0.09]$, the step size for the primal value $\tau \in [0.5, 0.9]$ and the step size for the dual value $\sigma \in [0.2, 0.5]$ for different experiments.
- FBP-Unet: *the FBP-Unet reconstruction* in [17][1]. It is a method combining the FBP reconstruction with the Unet as the postprocessing to improve image quality. We use the Xavier to initialize the network parameters. And the loss function is the mean squared loss of the reconstructed image and the ground truth.
- JSR-net: *the Joint Spatial-Radon domain reconstruction* in [34][2]. It is a feed-forward network by unrolling the iterative scheme of the JSR model. We use the Xavier initialization and define the loss function as $\mathcal{L} = \mathcal{L}_{SD} + \alpha_1 \mathcal{L}_{RD} + \alpha_2 \mathcal{L}_{sem}$ with $\alpha_1 = 0.01$ and $\alpha_2 = 100$ for all experiments. Note that $\mathcal{L}_{sem}$ is not used in JSR-net to guarantee a fair comparison.
- PD-net: *the Leaned Primal-Dual network* in [23][3]. The network is a deep unrolled neural network with 10 stages. The number of initialization channels for primal values and dual values is set to 5. The Xavier initialization and the mean squared loss of the reconstructed image and the ground truth are used in all experiments.

[1]https://github.com/panakino/FBPConvNet
[2]http://bicmr.pku.edu.cn/ dongbin/Publications.html
[3]https://github.com/adler-j/learned_primal_dual



(a) Clean (b) TV (c) FBP-Unet (d) JSR-net (e) PD-net (f) FSR-net

Fig. 5. From row one to row two are reconstruction comparison on the Forbild head phantom with 60 views and 45 views, where the results of the FSR-net is obtained using the SFS system matrix.

C. Parameter behavior

In the first place, we test the impact of the stage number in our learnable algorithm by using the number of stages $k = 5, 10, 15$ to construct our FSR-net on the ellipse dataset. Considering the sparse-view problem with $N_{\text{views}} = 60$, a total of 20 epochs are used in the training stage and the Forbild head phantom image is used as the testing image. The values of the loss function, PSNR, and SSIM are tracked and plotted in Fig. 4, which show that three models with 5, 10 and 15 stages all converge to steady states within 10 epochs in the training process. More importantly, the network's learning capacity is improved as the stage number increases from $k = 5$ to $k = 10$, while the improvement of both PSNR and SSIM is negligible by further increasing the stage number to $k = 15$. Therefore, we fix the stage number as $k = 10$ and the number of the epoch as 10 in the following experiments.

Because each stage in our model involves four 3-layer networks, our model has a total of 120 convolution layers, which gives 4.9×10^5 parameters. The parameters of the JSR-net and PD-net are about 2.4×10^5 in their original papers. For a fair comparison, we double the number of stages in the JSR-net and PD-net in our evaluation. TABLE II summarizes the number of parameters contained in the learning-based algorithms, where the FBP-Unet has the largest amount of parameters.

TABLE II
COMPARISON OF MODEL PARAMETERS IN LEARNING BASED METHODS.

Learning based models	Number of parameters
FBP-Unet	10^7
JSR-net	4.9×10^5
PD-net	4.9×10^5
FSR-net	4.9×10^5

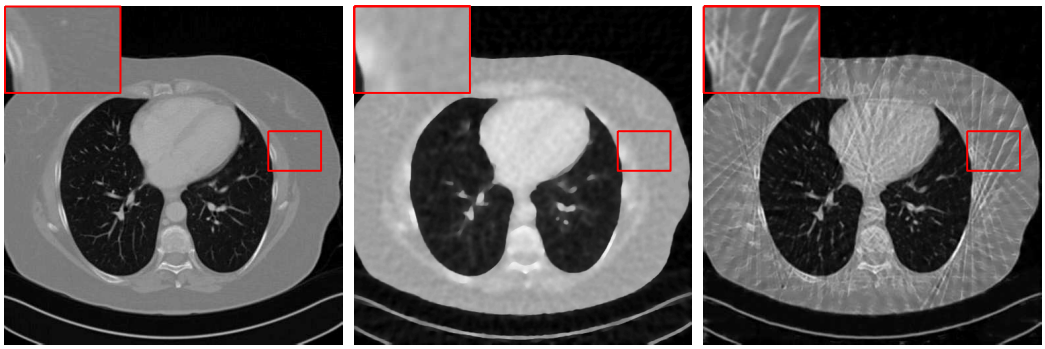
D. Sparse-view CT reconstruction

In this subsection, we evaluate the performance of our FSR-net as well as the comparison methods on sparse-view CT reconstruction problems.

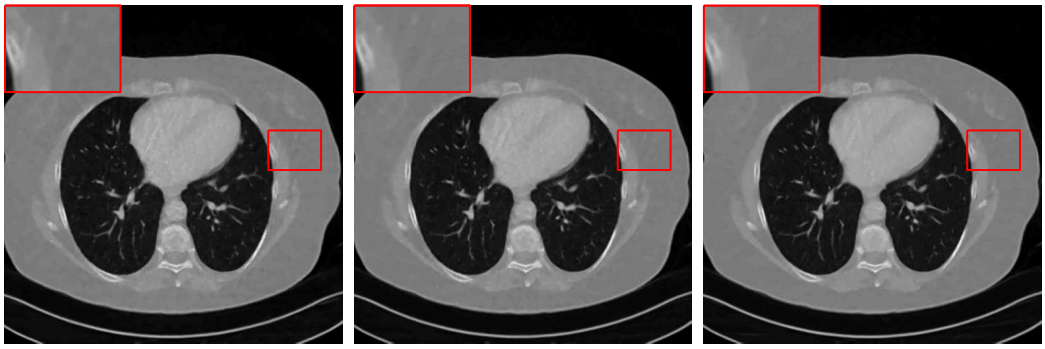
TABLE III
COMPARISON OF RECONSTRUCTION METHODS ON THE FORBILD HEAD PHANTOM IN TERMS OF PSNR, SSIM AND RUNTIME (MS).

A_F	N_{views}	Methods	PSNR	SSIM	time
IFS	60	FBP	11.6544	0.4958	**402**
		TV	23.8021	0.8167	52683
		FBP-Unet	22.8821	0.7216	883
		JSR-net	30.2901	0.9562	986
		PD-net	31.0141	0.9799	922
		FSR-net	**32.5733**	**0.9817**	1160
	45	FBP	11.5491	0.3642	**240**
		TV	23.2326	0.7587	51255
		FBP-Unet	22.3061	0.7206	843
		JSR-net	29.7891	0.9539	953
		PD-net	30.2281	0.9755	846
		FSR-net	**31.1191**	**0.9761**	1107
SFS	60	FBP	11.6549	0.4984	**418**
		TV	25.2022	0.8504	54713
		FBP-Unet	24.1667	0.7247	1053
		JSR-net	34.2271	0.9578	1153
		PD-net	36.0857	0.9818	1122
		FSR-net	**37.9404**	**0.9819**	1453
	45	FBP	11.5876	0.4431	**265**
		TV	23.5966	0.7717	53000
		FBP-Unet	23.0098	0.7244	1051
		JSR-net	32.6525	0.9564	1135
		PD-net	34.0341	0.9778	1029
		FSR-net	**35.1258**	**0.9781**	1407

1) Results on ellipses: Both $N_{\text{views}} = 60$ and $N_{\text{views}} = 45$ are considered as representatives of the sparse-view problem. We tabulate the PSNR and SSIM in TABLE III, which clearly shows that our FSR-net produces the best reconstruction results with both IFS and SFS system matrices. Obviously, the advantage of our FSR-net is more prevalent when the SFS condition is satisfied, which gives a 1.8 dB and 1.0 dB higher PSNR than the PD-net on 60 views and 45 views projection data, respectively. From the perspective of computational efficiency, it is clear that the FBP method is the fastest and



(a) Clean image (b) TV (c) FBP-Unet



(d) JSR-net (e) PD-net (f) FSR-net

Fig. 6. Sparse-view reconstruction experiment on a human phantom data with 60 projections, where the reconstruction of FSR-net is obtained by the SFS system matrix.

our model performs slightly slower than other learning-based methods due to the computations on full-sampling system matrices. We display the reconstruction results with 60 views and 45 views projection data in Fig. 5. As shown, obvious artifacts can be observed in the reconstructed images of the TV model and FBP-Unet, while the JSR-net, PD-net, and our FSR-net can preserve geometric structures and keep the homogeneity in the smooth regions. More importantly, we can find out that fewer artifacts are presented in the reconstructed images by our method, which are consistent with the quantitative results in TABLE III.

2) Results on human phantom data: We further evaluate the performance of sparse-view reconstruction on the AAPM human phantom dataset. Both 60 views and 45 views are used for the validation. The PSNR, SSIM and CPU time of all comparison methods on the noiseless and noisy (5% white Gaussian noise) Radon domain data are displayed in TABLE IV, where the IFSR-net and SFSR-net denote the reconstruction using the IFS and SFS system matrix, respectively. Note that because the number of bins is fixed as $N_{\text{bins}} = 736$, we approximate the SFS system matrix by increasing the number of projections to $N_{\text{views}} = 1425$ such that the dimension of A_{F} is around $4N^2$. Similar to the experiment on the ellipse dataset, all the deep unrolling algorithms offer significant gain in terms of PSNR and SSIM over the TV model and post-processing method FBP-Unet. Among the three unrolled learning-based methods, the FSR-net produces the best PSNR values while the PD-net achieves the second place. Besides, we also have the following two observations based on TABLE IV:

TABLE IV
COMPARISON OF SPARSE-VIEW RECONSTRUCTION METHODS FOR THE HUMAN PHANTOM IN TERMS OF PSNR, SSIM AND RUNTIME(MS).

Noise	N_{views}	Methods	PSNR	SSIM	time
0	60	FBP	14.2361	0.6414	**400**
		TV	29.4829	0.9082	53512
		FBP-Unet	27.3711	0.8458	1053
		JSR-net	35.4462	0.9647	1152
		PD-net	35.9844	0.9693	1065
		IFSR-net	36.6895	0.9733	1296
		SFSR-net	**37.2069**	**0.9752**	1486
	45	FBP	14.1802	0.6066	**242**
		TV	28.0876	0.8827	52120
		FBP-Unet	25.3767	0.8171	1021
		JSR-net	33.4384	0.9542	1106
		PD-net	34.6781	0.9639	1023
		IFSR-net	35.3797	0.9672	1235
		SFSR-net	**36.1099**	**0.9706**	1396
5%	60	FBP	14.0922	0.4989	**400**
		TV	27.0301	0.8548	53523
		FBP-Unet	26.0313	0.8261	1046
		JSR-net	29.8641	0.9217	1156
		PD-net	30.0261	0.9237	1076
		IFSR-net	30.4563	0.9321	1253
		SFSR-net	**30.4681**	**0.9331**	1486
	45	FBP	13.9732	0.4389	**242**
		TV	25.7201	0.8183	52126
		FBP-Unet	24.7362	0.7983	1023
		JSR-net	28.5931	0.9048	1103
		PD-net	29.1028	0.9135	1001
		IFSR-net	29.6814	0.9253	1205
		SFSR-net	**29.7915**	**0.9263**	1396

TABLE V
COMPARISON OF LIMITED-ANGLE RECONSTRUCTION METHODS FOR THE HUMAN PHANTOM IN TERMS OF PSNR, SSIM AND RUNTIME(MS).

Noise	N_{views}	Methods	PSNR	SSIM	time
0	150°	FBP	13.6239	0.5128	**775**
		TV	30.4607	0.9278	56523
		FBP-Unet	24.5191	0.9192	1046
		JSR-net	32.8648	0.9638	1256
		PD-net	34.3603	0.9723	1116
		IFSR-net	34.8423	0.9733	1453
		SFSR-net	**35.6669**	**0.9764**	1586
	120°	FBP	13.4822	0.4689	**602**
		TV	27.5898	0.9196	60026
		FBP-Unet	22.3886	0.8846	1023
		JSR-net	28.2803	0.9376	1273
		PD-net	29.5207	0.9452	1131
		IFSR-net	30.0843	0.9454	1415
		SFSR-net	**30.8151**	**0.9499**	1556
	90°	FBP	13.0982	0.4662	**422**
		TV	23.4676	0.8681	54926
		FBP-Unet	20.0601	0.8221	1013
		JSR-net	23.9463	0.8794	1243
		PD-net	24.2297	0.8837	1091
		IFSR-net	25.4361	0.9033	1385
		SFSR-net	**25.8551**	**0.9084**	1526
5%	150°	FBP	13.5911	0.4854	**776**
		TV	25.8815	0.8091	56532
		FBP-Unet	23.8923	0.8703	1033
		JSR-net	28.7462	0.9059	1285
		PD-net	30.3766	0.9301	1113
		IFSR-net	30.8763	0.9303	1429
		SFSR-net	**30.9411**	**0.9324**	1582
	120°	FBP	13.4418	0.4008	**602**
		TV	23.5852	0.7891	60032
		FBP-Unet	21.1637	0.8072	1023
		JSR-net	25.3458	0.8555	1278
		PD-net	27.1539	0.9037	1132
		IFSR-net	27.7241	0.9079	1417
		SFSR-net	**27.7293**	**0.9103**	1551
	90°	FBP	13.0314	0.3881	**430**
		TV	19.9501	0.6918	54917
		FBP-Unet	18.5181	0.7481	1015
		JSR-net	20.2241	0.7971	1241
		PD-net	22.6047	0.8612	1099
		IFSR-net	23.9399	0.8744	1382
		SFSR-net	**24.2494**	**0.8761**	1525

- The comparison between the results on the noiseless and noisy projection data shows that the superiority of our FSR-net is more prevalent with the noiseless projection data.
- The comparison between the results of IFSR-net and SFSR-net illustrates that the reconstruction quality can be improved by enlarging the dimension of the full-sampling system matrix.

Fig. 6 displays the original image and the reconstruction results of TV, FBP-Unet, JSR-net, PD-net, and SFSR-net on noiseless 60 views projection data, which provides the visual inspection on the advantageous of our FSR-net. Both the reconstruction results of the TV model and FBP-Unet present artifacts and the TV model produces better quality than the FBP-Unet, which illustrates that the post-processing methods are more sensitive to the deterioration in the projection data. As shown, three deep unrolling models work quite well on the 60 views projection data. By examining the zoomed-in regions, it can be found out that fine structures and features are better preserved by our FSR-net.

E. Limited-angle CT reconstructions

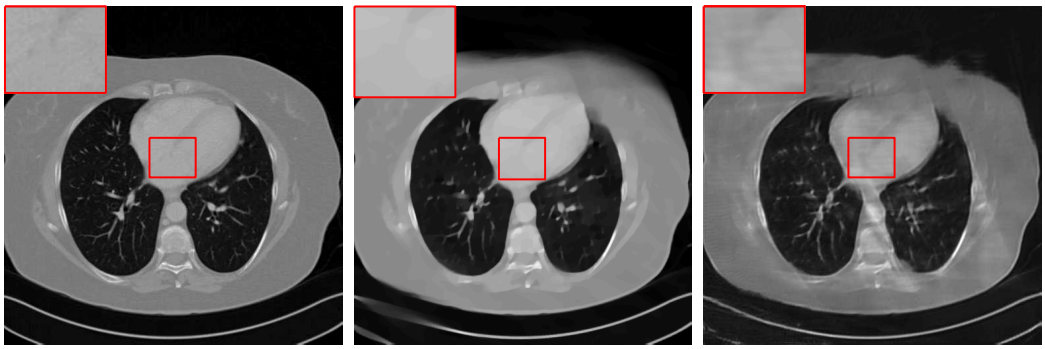
In this subsection, we conduct a series of experiments on the limited-angle problem using the AAPM human phantom dataset. The experiments are set up as follows: the observed data g is generated with the scanning angular range $[0, 150°]$, $[0, 120°]$ and $[0, 90°]$ without noise or with 5% white Gaussian noise.

We tabulate both the PSNR and SSIM of our model and the comparison methods in TABLE V. As can be seen, the values of both PSNR and SSIM obtained by all competing methods drop when either the scanning angular range shrinks from 150° to 90° or the noises are introduced into the projection data. The results of limited-angle reconstruction concur with the sparse-view reconstruction. For different combinations of the scanning angular range (i.e., 150°, 120° and 90°) and the noise level (i.e., 0% or 5%), our FSR-net consistently outperforms the other reconstruction methods. The advantage of our model is more prevalent when the SFR system matrix is used in which case about 1.5 dB improvement is achieved in PSNR over the PD-net for noiseless reconstruction problems.

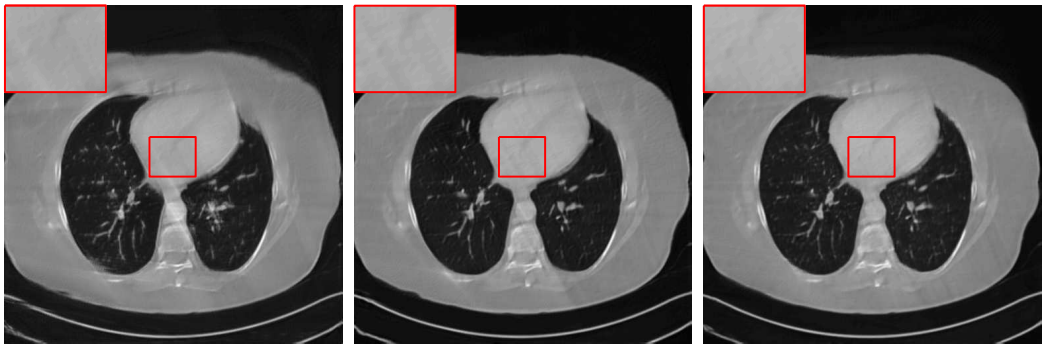
The original image and the reconstructed images from different methodologies with scanning angular range of 120° are displayed in Fig. 7. It can be seen that both the TV model and FBP-Unet suffer from severe streak artifacts and distortions. The comparison between FBP-Unet and other deep unrolling methods shows that the iterative algorithms (i.e., JSR-net, PD-net, and SFSR-net) outperform the post-processing methods as no obvious distortions are presented in the reconstructed images of the iterative methods. By further comparing the reconstruction results of the JSR-net, PD-net, and our FSR-net, it can be found out that our model can better preserve image details than the other two algorithms, e.g., no obvious steak artifacts existing in the zoomed-in region. Both the quantitative and qualitative results confirm that projection data can be filled up in the missing angular range due to the use of the full-sampling system matrix.

VI. DISCUSSION

Both JSR-net and FSR-net aim to enhance the resolution of the projection data in the Radon domain for CT reconstruction from incomplete data. The results in TABLE III, IV, and V show that the reconstruction results of JSR-net are no better than PD-net, while our FSR-net always produces better reconstruction than PD-net. We believe that the reasons are threefold: 1) From model-wise, we explore stronger prior knowledge on the high-resolution projection data such that the relationship between the image and projection data is explicitly defined through the full-sampling system matrix; 2) In algorithm design, we allow the network to learn how to combine the variables while JSR-net uses the fixed forms for some variables; 3) The loss function used in our model also differs from the JSR-net. We minimize the ℓ^2 norm of



(a) Clean image (b) TV (c) FBP-Unet



(d) JSR-net (e) PD-net (f) FSR-net

Fig. 7. Limited-angle reconstruction experiment of a human phantom with 120° scanning angular range, where the reconstruction of FSR-net is obtained by the SFS system matrix.

image difference and projection data difference, while the loss function of JSR-net is derived solely based on the image function.

Since our algorithm simultaneously reconstructs the image and the high-resolution projection data, we compare the FBP result of the observed data g and the reconstructed projection data f in TABLE VI. As shown, the values of PSNR obtained by f are consistently higher than ones obtained by g, which demonstrates that our model works well for both sparse-view and limited-angle problems. We notice that the advantage of the reconstructed projection data decreases as the observed data becomes very insufficient such as the limited-angle problem with angular range $90°$. For severely degenerated projection data, e.g., scanning angular range of $90°$, we may need stronger data fidelity to further improve the quality of the projection data.

TABLE VI
COMPARISON BETWEEN THE FBP OF THE GIVEN AND RECONSTRUCTED PROJECTION DATA, WHERE FBP-W-F DENOTES THE FBP OF f.

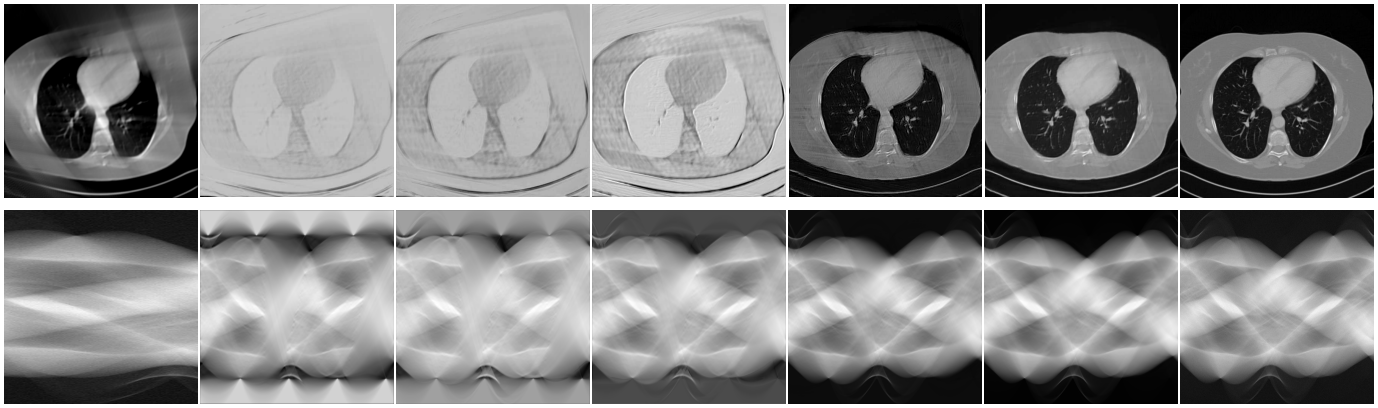
Settings / Methods	sparse-view		limited-angle		
	60 view	45 view	150°	120°	90°
FBP	14.0922	13.9732	13.5911	13.4418	13.0314
FBP-w-f	19.8017	19.0803	20.4825	18.5719	15.9486

As the number of bins is given as $N_{\text{bins}} = 736$ for the AAMP phantom dataset, we approximate the SFS system matrix by increasing the number of projections such that the dimension of A_F is guaranteed to be the same as the SFS system matrix. To evaluate the accuracy of such an approximation, we compare the reconstruction results with the ones obtained by the SFS system matrix with $N_{\text{bins}} = N_{\text{views}} = 2N$. As shown in TABLE VII, higher PSNR values are always achieved by the full-sampling system matrix with $N_{\text{bins}} = 2N$, which means that the number of bins is also an important factor for controlling the reconstruction quality.

TABLE VII
COMPARISON ON THE AAPM DATASET BETWEEN THE SFSR-NET AND SFS SYSTEM MATRICES.

Settings / $N_{\text{views}} \times N_{\text{bins}}$	sparse-view		limited-angle		
	60 view	45 view	150°	120°	90°
1425×736	37.2069	36.1099	30.9411	27.7293	24.2494
1024×1024	37.3099	36.5921	31.2032	27.7353	24.2539

Last but not least, we visualize the iterates to see how the iterative algorithm works. The limited-angle experiment with scanning angular range $120°$ is used as an example, for which both the reconstructed image u and the inpainted projection data f at the selected stages are displayed in Fig. 8. As the iteration number increases, we observe that the artifacts contained in the reconstructed image get less and less, and the information in projection data becomes richer and richer. It indicates that the joint reconstruction of the image u and its projection data f can promote each other, which coincides



(a) FBP and g (b) 3rd stage (c) 5th stage (d) 7th stage (e) 9th stage (f) 10th stage (g) Ground truth

Fig. 8. The FBP reconstruction, observed projection data, reconstructed image and projection data together with the ground truth image and the full-sampling projection data of the limited-angle reconstruction experiment with 120° angular scanning range, where windows are selected to cover most of the range of the values.

with our expectations.

VII. Conclusion and future works

We proposed a novel iterative reconstruction model by fitting the reconstructed image with its corresponding measurements in the Radon domain through the full-sampling condition. The deep unrolling algorithm was developed by replacing the proximal operators with learned operators, which can reconstruct the CT images directly from the raw projection data. The applications on the sparse-view and limited-angle CT reconstruction problems demonstrate that our learned FSR-net achieves state-of-the-art results.

In this work, we mainly focus on the fan-beam CT reconstruction. The basic idea, however, can be generally applied to other system matrix classes with appropriate parameter setting according to [46], [47]. Although our FSR-net outperforms other reconstruction schemes for limited-angle reconstruction, for the very challenging case with 90° scanning angular range, however, the results of our method are still not close to diagnostic quality. One of the reasons is that we use the images to generate the ground truth full-sampling projection data, which may limit the potential of our FSR-net in learning the full-sampling Radon domain data. Indeed, it is more reasonable to use the ground truth of the full-sampling Radon data in the loss function to enable the network to learn better priors. Also, our FSR-net is trained based on the joint mean squared norm (15) for a fair comparison, which may also lead to over-smoothing in the results. More advanced loss functions such as the structure-semantic-ℓ^2 loss [34], perceptual loss [48] might help to further improve the performance of our FSR-net. As observed, the FSR-net outperforms other algorithms on noisy projection data, but the advantage has declined compared to the noiseless case. Therefore, we plan to incorporate denoising networks into our FSR-net as future work to investigate additional mechanisms for further improvement.

Acknowledgment

The authors would like to thank the authors of [17], [23], [34] for sharing the CT reconstruction codes and the referees for providing us numerous valuable suggestions to revise this paper. The work was supported by National Natural Science Foundation of China (NSFC 11701418), Major Science and Technology Project of Tianjin 18ZXRHSY00160 and Recruitment Program of Global Young Expert. The third author Hongwei Li was supported by National Natural Science Foundation of China (NSFC) (61971292) and key research project of the Academy for Multidisciplinary Studies, Capital Normal University. And he is also grateful to Beijing Higher Institution Engineering Research Center of Testing and Imaging as well as Beijing Advanced Innovation Center for Imaging Technology for funding his research.

References

[1] S. Schafer, S. Nithiananthan, D. Mirota, A. Uneri, J. Stayman, W. Zbijewski, C. Schmidgunst, G. Kleinszig, A. Khanna, and J. Siewerdsen, "Mobile c-arm cone-beam CT for guidance of spine surgery: Image quality, radiation dose, and integration with interventional guidance," *Medical Physics*, vol. 38, no. 8, pp. 4563–4574, 2011.

[2] M. Cho, H. Kim, H. Youn, and S. Kim, "A feasibility study of digital tomosynthesis for volumetric dental imaging," *Journal of Instrumentation*, vol. 7, no. 03, p. P03007, 2012.

[3] L. A. Feldkamp, L. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, 1984.

[4] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography," *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.

[5] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm," *Ultrasonic Imaging*, vol. 6, no. 1, pp. 81–94, 1984.

[6] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[7] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle–pock algorithm," *Physics in Medicine & Biology*, vol. 57, no. 10, p. 3065, 2012.

[8] M. G. McGaffin and J. A. Fessler, "Alternating dual updates algorithm for x-ray CT reconstruction on the gpu," *IEEE Transactions on Computational Imaging*, vol. 1, no. 3, pp. 186–199, 2015.

[9] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *Journal of Scientific Computing*, vol. 42, no. 2, pp. 185–197, 2010.

[10] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," *Physics in Medicine & Biology*, vol. 56, no. 13, p. 3787, 2011.

[11] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose x-ray CT reconstruction via dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1682–1697, 2012.

[12] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.

[13] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 269–276, 2019.

[14] S. Ravishankar, J. C. Ye, and J. A. Fessler, "Image reconstruction: From sparsity to data-adaptive methods and machine learning," *Proceedings of the IEEE*, vol. 108, no. 1, pp. 86–109, 2019.

[15] H.-M. Zhang and B. Dong, "A review on deep learning in medical image reconstruction," *Journal of the Operations Research Society of China*, pp. 1–30, 2020.

[16] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[17] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[18] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A sparse-view CT reconstruction method based on combination of densenet and deconvolution," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1407–1417, 2018.

[19] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 10–18.

[20] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, p. 124007, 2017.

[21] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction," *Medical Physics*, vol. 44, no. 10, pp. e360–e375, 2017.

[22] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "Learn: Learned experts assessment-based reconstruction network for sparse-data CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1333–1347, 2018.

[23] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.

[24] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. Mccann, and M. Unser, "CNN-based projected gradient descent for consistent image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1440–1453, 2018.

[25] T. Meinhardt, M. Moller, C. Hazirbas, and D. Cremers, "Learning proximal operators: Using denoising networks for regularizing inverse imaging problems," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1781–1790.

[26] Y. Sun, B. Wohlberg, and U. S. Kamilov, "An online plug-and-play algorithm for regularized image reconstruction," vol. 5, no. 3, pp. 395–408, 2019.

[27] J. Liu, T. Kuang, and X. Zhang, "Image reconstruction by splitting deep learning regularization from iterative inversion," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 224–231.

[28] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1358–1369, 2018.

[29] B. Kelly, T. P. Matthews, and M. A. Anastasio, "Deep learning-guided image reconstruction from incomplete data," *ArXiv Preprint ArXiv:1709.00584*, 2017.

[30] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.

[31] D. M. Pelt, K. J. Batenburg, and J. A. Sethian, "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks," *Journal of Imaging*, vol. 4, no. 11, p. 128, 2018.

[32] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1454–1463, 2018.

[33] B. Dong, J. Li, and Z. Shen, "X-ray CT image reconstruction via wavelet frame based regularization and radon domain inpainting," *Journal of Scientific Computing*, vol. 54, no. 2-3, pp. 333–349, 2013.

[34] H. Zhang, B. Dong, and B. Liu, "JSR-Net: A deep network for joint spatial-radon domain CT reconstruction from incomplete data," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3657–3661, 2019.

[35] J. Wang, L. Zeng, C. Wang, and Y. Guo, "ADMM-based deep reconstruction for limited-angle CT," *Physics in Medicine & Biology*, vol. 64, no. 11, p. 115011, 2019.

[36] T. A. Bubba, G. Kutyniok, M. Lassas, M. Maerz, W. Samek, S. Siltanen, and V. Srinivasan, "Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography," *Inverse Problems*, vol. 35, no. 6, p. 064002, 2019.

[37] J. P. Kaipio and E. Somersalo, "Statistical and computational inverse problems," *Technometrics*, vol. 48, no. 1, pp. 146–146, 2005.

[38] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[39] W. Cheng, Y. Wang, Y. Chi, X. Xie, and Y. Duan, "Learned full-sampling reconstruction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 375–384.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[41] J. S. Jorgensen, E. Y. Sidky, and X. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 460–473, 2013.

[42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.

[43] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *International Conference on Learning Representations*, 2017.

[44] Z. Yu, F. Noo, F. Dennerlein, A. Wunderlich, G. Lauritsch, and J. Hornegger, "Simulation tools for two-dimensional experiments in x-ray computed tomography using the forbild head phantom," *Physics in Medicine & Biology*, vol. 57, no. 13, p. N237, 2012.

[45] C. H. McCollough, A. C. Bartley, R. E. Carter, B. Chen, T. A. Drees, P. Edwards, D. R. Holmes III, A. E. Huang, F. Khan, S. Leng *et al.*, "Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 low dose CT grand challenge," *Medical Physics*, vol. 44, no. 10, pp. e339–e352, 2017.

[46] R. Huesman, "The effects of a finite number of projection angles and finite lateral sampling of projections on the propagation of statistical errors in transverse section reconstruction," *Physics in Medicine & Biology*, vol. 22, no. 3, p. 511, 1977.

[47] F. Natterer and F. Wübbeling, *Mathematical methods in image reconstruction*. SIAM, 2001.

[48] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.