A total variation prior unrolling approach for computed tomography reconstruction

Pengcheng Zhang, Shuhui Ren, Yi Liu, Zhiguo Gui, Hong Shangguan, Yanling Wang, Huazhong Shu, Yang Chen

▶ **To cite this version:**

Pengcheng Zhang, Shuhui Ren, Yi Liu, Zhiguo Gui, Hong Shangguan, et al.. A total variation prior unrolling approach for computed tomography reconstruction. Medical Physics : The international journal of medical physics research and practice, 2023, ⟨10.1002/mp.16307⟩. ⟨hal-04012755⟩

HAL Id: hal-04012755

https://hal.science/hal-04012755v1

Submitted on 24 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

A Total Variation Prior Unrolling Approach for Computed Tomography Reconstruction

Pengcheng Zhang,[1] Shuhui Ren,[1] Yi Liu,[1] Zhiguo Gui,[1, a)] Hong Shangguan,[2] Yanling Wang,[3] Huazhong Shu,[4] and Yang Chen[4,5,6,7, a)]

[1] *State Key Laboratory of Dynamic Testing Technology, North University of China, Taiyuan 030051, China*

[2] *School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China*

[3] *School of Information, Shanxi University of Finance and Economics, Taiyuan 030006, China*

[4] *Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China*

[5] *Centre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), F-3500 Rennes, France*

[6] *Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China*

[7] *Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China*

Abstract

Background: With the rapid development of deep learning technology, deep neural networks can effectively enhance the performance of computed tomography (CT) reconstructions. One kind of commonly used method to construct CT reconstruction networks is to unroll the conventional iterative reconstruction (IR) methods to convolutional neural networks (CNNs). However, most unrolling methods primarily unroll the fidelity term of IR methods to CNNs, without unrolling the prior terms. The prior terms are always directly replaced by neural networks.

Purpose: In conventional IR methods, the prior terms play a vital role in improving the visual quality of reconstructed images. Unrolling the hand-crafted prior terms to CNNs may provide a more specialized unrolling approach to further improve the performance of CT reconstruction. In this work, a primal-dual network (PD-Net) was proposed by unrolling both the data fidelity term and the total variation (TV) prior term, which effectively preserves the image edges and textures in the reconstructed images.

Methods: By further deriving the Chambolle–Pock (CP) algorithm instance for CT reconstruction, we discovered that the TV prior updates the reconstructed images with its divergences in each iteration of the solution process. Based on this discovery, CNNs were applied to yield the divergences of the feature maps for the reconstructed image generated in each iteration. Additionally, a loss function was applied to the predicted divergences of the reconstructed image to guarantee that the CNNs' results were the divergences of the corresponding feature maps in the iteration. In this manner, the proposed CNNs seem to play the same roles in the PD-Net as the TV prior in the iterative reconstruction methods. Thus, the TV prior in the CP algorithm instance can be directly unrolled to CNNs.

Results: The datasets from the Low-Dose CT Image and Projection Data and the Piglet dataset were employed to assess the effectiveness of our proposed PD-Net. Compared with conventional CT reconstruction methods, our proposed method effectively preserves the structural and textural information in reference to ground truth.

Conclusions: The experimental results show that our proposed PD-Net framework is feasible for the implementation of CT reconstruction tasks. Owing to the promising results yielded by our proposed neural network, this study is intended to inspire further development of unrolling approaches by enabling the direct unrolling of hand-crafted prior terms to CNNs.

ACCEPTED MANUSCRIPT / CLEAN COPY

Keywords: deep learning, computed tomography, unrolling approach, convolutional neural network, total variation, Chambolle–Pock algorithm

1. INTRODUCTION

A major challenge in X-ray computed tomography (CT) science is solving the inverse problem of CT reconstruction, determining the visual quality of the reconstructed images. Many state-of-the-art methods have been proposed to solve this inverse problem and they can be divided into three categories: analytical methods, iterative reconstruction (IR) methods and artificial neural network methods. The filtered back-projection (FBP) algorithm[1] is a classical analytical method that is widely used in real clinical applications owing to its high computational efficiency. However, reconstructed images may degrade dramatically with a small number of inputs or heavily polluted projections.[2]

In IR methods, the cost function of the inverse problem generally involves data fidelity term and prior terms. The data fidelity term is employed to model the physical processes underlying the problem. The prior term captures prior domain knowledge (e.g. low-rank, sparsity, smoothness, and non-local self-similarity). Various elaborately designed prior terms have been investigated to model the statistical characteristics of sinogram data or reconstructed images, including total variation (TV),[3-5] nonlocal patches,[6] dictionary learning,[7] sparsifying transform learning,[8, 9] etc. Frequently, their prior terms are non-smooth, which remarkably increases the difficulty of solving the inverse problem. Based on the inverse problem's prior terms, the prevailing approaches include adaptive steepest descent projection onto convex set algorithm,[3] Chambolle–Pock (CP) algorithm,[4] alternating direction method of multipliers (ADMM),[10] iterative shrinkage/thresholding algorithm (ISTA),[11] and others. Compared with other analytical methods, IR methods yield more promising results for CT reconstruction.

Recently, with the rapid development of deep-learning technologies, deep neural networks have further improved CT reconstruction performance.[12] Thus, major efforts have been dedicated to reconstructing promising images from sinogram data, such as via unrolling approaches,[13-19] model based approaches,[20] domain transform approaches,[21] and dual-domain approaches.[22-26] The unrolling approach constructs CT reconstruction networks by unrolling IR methods. Despite the tremendous success achieved by the extant iterative algorithms used in signal processing, unrolling approaches achieve better performance than the original IR methods.[27] By unrolling a simple gradient descent scheme, Chen et al.[14] constructed a deep reconstruction network based on learned experts' assessments for sparse-data CT reconstruction. This reconstruction network was further improved by utilizing a graph convolution to extract the features of the low-dimensional manifold for low-dose CT reconstruction.[15] Using a similar unrolling strategy, Liu et al.[16] reduced the computational and memory complexity of the deep reconstruction network using stochastic approximations of the data-consistency layers. Adler and Öktem[13] constructed the Learned Primal-Dual reconstruction network which unfolded the CP algorithm to absorb its convergence guarantee.[4] The ADMM approach was also unrolled into deep reconstruction networks, owing to its high computational efficiency.[17, 18] Xiang et al.[19] cast a fast ISTA into a deep reconstruction network in which the shrinkage/thresholding operator was replaced by a neural network. These state-of-the-art unrolling approaches exhibit good generalizability under different scanning conditions. Various unrolling approaches [13-19, 28-30] mainly unroll the fidelity term to convolutional neural networks (CNNs), without unrolling the prior terms. In unrolling approaches, the prior term is always replaced by a CNN. In conventional IR methods, the hand-crafted prior terms play an important role in solution efficiency and determine the image quality of CT reconstruction. Unrolling the hand-crafted prior term to a CNN may provide

ACCEPTED MANUSCRIPT / CLEAN COPY

a more tailor-made unrolling approach to improve the CT image quality.

Owing to the potential to perform edge preservation in reconstructed images, the TV prior has received considerable attentions in the field of CT reconstruction.[3, 4, 31] In this work, to further improve IR methods based on the TV prior, we proposed a primal-dual network (PD-Net), which unrolls both the data fidelity and TV prior terms of the cost function to the CNNs. Owing to the convergence guarantee, the CP algorithm effectively solves the inverse problem based on the TV prior.[4] In this work, we provided a detailed derivation of the CP algorithm instance for CT reconstruction. After analyzing this derived instance, we discovered that the inclusion of the TV prior in the CP algorithm instance is to update the reconstructed images with its divergences in each iteration of the solution process. Based on this discovery, CNNs were applied to yield the divergences of the feature maps for the reconstructed image generated in each iteration. Additionally, a loss function was applied to the predicted divergences of the reconstruct image to ensure that the yielded results of the proposed CNNs are indeed the divergences of the corresponding feature maps in the iteration. In this manner, the proposed CNNs are founded to provide the same roles in the PD-Net as the TV prior in the IR methods. Thus, the TV prior in the CP algorithm instance can be directly unrolled to CNNs. Our PD-Net is constructed by unrolling both the data fidelity term and the TV prior in the CP algorithm instance. By doing so, the PD-Net effectively preserves the structural and textural information in the reconstruction results. The main contribution of our proposed PD-Net is a direct improvement to unrolling approaches by unrolling the TV prior in the IR methods to the CNNs.

2. METHODS

2.A. CP Algorithm Based on TV Prior

By incorporating a data-fidelity term and a penalty or prior term in the cost function, model-based iterative reconstruction methods can effectively restore the desired images from sinogram data. The TV prior plays a vital role in CT reconstruction owing to its robust performance in preserving image edges and textures.[3,4] Based on the TV prior, the cost function of the inverse problem for CT reconstruction is defined as:

$$\min_{u}\left\{\frac{1}{2}\|Au-g\|_2^2+\lambda\left\|(|\partial u|)\right\|_1\right\}, \tag{1}$$

where A is the system matrix, g are the sinogram data, and u is the reconstructed image obtained by solving this minimization problem. The first term is the data fidelity term, and the second is the TV prior. The regularization parameter λ controls the balance of these two terms. The TV prior is the ℓ_1-norm of the gradient-magnitude image, which is calculated by the operator ∂ (in Appendix).

Many kinds of optimization algorithms[18] have been proposed to solve the minimization problem (1) and to obtain the desired images. The CP algorithm[32] effectively achieves a robust primal-dual scheme solution owing to its convergence guarantee. For a given primal minimization,

$$\min_{x}\{F(Kx)+G(x)\}, \tag{2}$$

the dual maximization is expressed as:

$$\max_{y}\{-F^*(y)-G^*(-K^Ty)\}, \tag{3}$$

ACCEPTED MANUSCRIPT / CLEAN COPY

where x and y are the vectors in spaces X and Y, K is a linear transform from X to Y, the T superscript denotes matrix transposition, and F and G are two convex functions whose convex conjugations are F^* and G^*, respectively. The conjugate of a convex function can be computed using the Legendre transform.[4] In the solution process, the CP algorithm iteratively computes the primal minimization and dual maximization and achieves convergence when the duality gap between them is zero.[4] Given convex functions F and G, the primal minimization and the dual maximization can be analytically solved with the proximal mappings $prox_\sigma[F^*]$ and $prox_\tau[G]$, as discussed in Ref. [4]. For the convex function H, the proximal mapping is calculated by the following minimization:

$$prox_\sigma[H](z) = \arg\min_{z'} \left\{ H(z') + \frac{\|z - z'\|_2^2}{2\sigma} \right\}. \tag{4}$$

The derivation of the proximal mappings $prox_\sigma[F^*]$ and $prox_\tau[G]$ is a key step for the CP algorithm to solve the minimization problem (1).

Algorithm. 1. Pseudocode for N steps of the ℓ_2^2-TV CP algorithm. The constant L is the ℓ_2-norm of the matrix K, σ and τ are non-negative CP algorithm parameters. The vectors p_i and q_i are two temporary variables. $div = -\partial^T$.

$L = \|(A, \partial)\|_2$; $\tau \leftarrow 1/L$; $\sigma \leftarrow 1/L$; $\theta \leftarrow 1$; $i \leftarrow 0$;

initialize u_0, p_0 and q_0 to zero values

$\bar{u}_0 \leftarrow u_0$

repeat

$p_{i+1} \leftarrow \left(p_i + \sigma(A\bar{u}_i - g) \right) / (1 + \sigma)$

$q_{i+1} \leftarrow \lambda(q_i + \sigma\partial\bar{u}_i) / \max\left(\lambda \mathbf{1}_1, |q_i + \sigma\partial\bar{u}_i|\right)$

$u_{i+1} \leftarrow u_i - \tau A^T p_{i+1} + \tau div(q_{i+1})$

$\bar{u}_{i+1} \leftarrow u_{i+1} + \theta(u_{i+1} - u_i)$

$i \leftarrow i + 1$

until $i \geq N$

As derived in Ref. [4] for the minimization problem (1) based on the TV prior, the convex functions F and G in the CP algorithm are:

$$F(y, z) = \frac{1}{2}\|y - g\|_2^2 + \lambda \|(|z|)\|_1, \tag{5}$$

$$G(x) = 0, \tag{6}$$

where $x = u$, $y = Au$, and $z = \partial u$. Terms $\frac{1}{2}\|y - g\|_2^2$ and $\lambda\|(|z|)\|_1$ in Equation (5) are both convex functions. Thus, the function $F(y, z)$ is convex because the sum of two convex functions is also convex. The

ACCEPTED MANUSCRIPT / CLEAN COPY

linear transforms K are A and ∂ for the two terms of $F(y,z)$, respectively. The proximal mappings $prox_{\sigma}[F^{*}]$ and $prox_{\tau}[G]$ can be analytically derived as:[4]

$$prox_{\sigma}[F^{*}](y,z)=\left(\frac{y-\sigma g}{1+\sigma},\frac{\lambda z}{\max(\lambda \mathbf{1}_{I},|z|)}\right), \tag{7}$$

$$prox_{\tau}[G](x)=x, \tag{8}$$

where $\mathbf{1}_{I}$ is an image with all pixels set to 1. The corresponding CP pseudocode is shown in Alg. 1. As derived from Equation (7), the role played by the TV prior plays in the CP algorithm is to update vector u_i with vector $\tau div(q_{i+1})$.

Algorithm. 2. New pseudocode for N steps of the ℓ_2^2-TV CP algorithm.

1: $L=\left\|(A,\partial)\right\|_2$; $\tau\leftarrow 1/L$; $\sigma\leftarrow 1/L$; $\theta\leftarrow 1$; $i\leftarrow 0$;

2: initialize u_0, p_0 and d_0 to zero values

3: $\overline{u}_0\leftarrow u_0$

4: **repeat**

5: $p_{i+1}\leftarrow\sum_{k=0}^{i}\frac{\sigma}{(1+\sigma)^{i-k+1}}(A\overline{u}_k-g)$

6: $d_{i+1}\leftarrow\sum_{k=0}^{i}\frac{\lambda^{i-k+1}\sigma}{\prod_{m=k}^{i}\xi_m}\partial\cdot(\overline{u}_k)$

7: $u_{i+1}\leftarrow u_i-\tau A^{T}p_{i+1}+\tau d_{i+1}$

8: $\overline{u}_{i+1}\leftarrow u_{i+1}+\theta(u_{i+1}-u_i)$

9: $i\leftarrow i+1$

10: **until** $i\geq N$

2.B. PD-Net

The learned reconstruction scheme based on the CP algorithm was well-investigated for CT reconstruction in Ref. [13]. The Learned Primal-Dual reconstruction[13] replaces the proximal operators in the CP algorithm with neural networks whose parameters are learned from big data. However, this learned reconstruction model only focuses on the unfoldment of the data fidelity term in cost function (1) to CNNs without considering the prior term. It is well-known that good CT reconstruction performance can be achieved by incorporating an elaborately designed prior into the cost function of the inverse problem.[14] In this work, to improve the performance of the learned reconstruction scheme, our PD-Net is constructed by unrolling both the data-fidelity term and the TV prior in minimization problem (1) to the CNNs.

As shown in Alg. 1, there are three proximal operators for the vectors p_i (in the projection domain), q_i (in the image domain), and u_i (in the image domain), respectively. Because p_0=0, the formula for the vector p_{i+1} in Line 5 of Alg. 1 can be rewritten as:

ACCEPTED MANUSCRIPT / CLEAN COPY

$$p_{i+1}=\sum_{k=0}^{i}\frac{\sigma}{(1+\sigma)^{i-k+1}}(A\bar{u}_k-g)\cdot \tag{9}$$

Thus, the vector p_{i+1} represents the sum of the difference between the forward projection of image $\bar{u}_i$ and sinogram data g. As $q_0=0$, the proximal operator for the vector q_{i+1} in Line 6 of Alg. 1 can be reformulated as:

$$q_{i+1}=\sum_{k=0}^{i}\frac{\lambda^{i-k+1}\sigma}{\prod_{m=k}^{i}\xi_m}\partial\bar{u}_k\,, \tag{10}$$

where $\xi_i=\max(\lambda\mathbf{1}_I,|q_i+\sigma\partial\bar{u}_i|)$. Obviously, vector q_{i+1} represents the sum of the gradients of images $\bar{u}_i$. The corresponding formula $div(q_{i+1})$ in Alg. 1 can be redefined as:

$$d_{i+1}=div(q_{i+1})=\sum_{k=0}^{i}\frac{\lambda^{i-k+1}\sigma}{\prod_{m=k}^{i}\xi_m}\partial\cdot(\bar{u}_k)\,, \tag{11}$$

where the operator $\partial\cdot=-\partial^T\partial$ (in Appendix) represents the calculation of the divergences for an image. Thus, d_{i+1} is the sum of the weighted divergences of images $\bar{u}_i$. Given parameters ξ_i, Alg. 1 can be rewritten as Alg. 2 without changing the final reconstruction results. As mentioned, the role played by TV prior in the CP algorithm instance is to update vector u_i with vector $div(q_{i+1})$ (i.e., vector d_{i+1}). Vector $\bar{u}_{i+1}$ is calculated

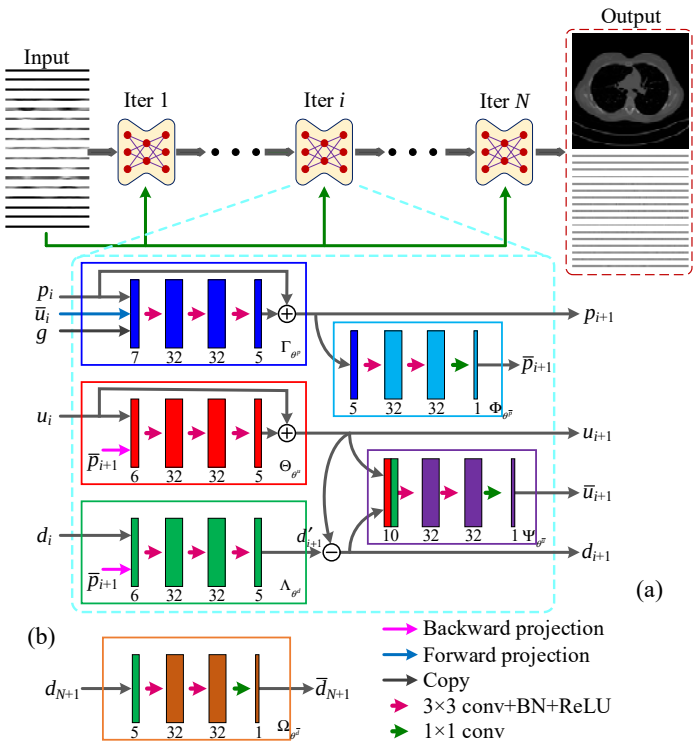

FIG. 1. Network architecture of (a) our proposed PD-Net and (b) the CNN $\Omega_{\theta^{\bar{n}}}$.

ACCEPTED MANUSCRIPT / CLEAN COPY

by $u_{i+1}+\theta(u_{i+1}-u_i)$. Thus, the inclusion of the TV prior in the cost function (1) is to update the vector u_{i+1} with its weighted divergences d_{i+1} in each iteration of the solution process. In this work, six kinds of CNNs were proposed to replace the proximal operators of the vectors p_i, d_i, and u_i, unrolling both the data fidelity term and the TV prior in minimization problem (1) to the CNNs.

The proposed PD-Net is an unrolling approach. In this work, a fixed number $N=10$ of iterations were unrolled to form the PD-Net for simplicity and better utility. For the deep-learning solution, multi-channel convolutions are employed to improve the feature extraction capabilities of neural networks. In each convolution layer, the effective information of a variable is usually represented by multi-channel feature maps. However, for the iteration algorithm, only a single vector of each variable is passed to the next iteration, such as p_{i+1}, d_{i+1}, and u_{i+1} in Alg. 2. Thus, for conventional unrolling approaches, multi-channel feature maps must be reduced to a single feature map at the end of each iteration. This feature map reduction inevitably decreases the feature extraction capability of reconstruction networks. To address this drawback, we used a strategy similar to the Learned Primal-Dual[13] and passed all feature maps to the next iteration. In this work, the number of feature maps that persist between iterations were set to $N_p=5$, $N_d=5$, and $N_u=5$ for p_i, d_i and u_i, respectively.

Algorithm. 3. Pseudocode for the PD-Net.

1: initialize u_0, p_0 and d_0 to zero values

2: **repeat**

3: $p_{i+1} \leftarrow \Gamma_{\theta^p}(p_i, A\overline{u}_i, g)$; $\overline{p}_{i+1} \leftarrow \Phi_{\theta^{\overline{p}}}(p_i)$; // The operator in the line 5 of Alg. 2.

4: $u_{i+1} \leftarrow \Theta_{\theta^u}(u_i, A^T\overline{p}_{i+1})$; // The operator $u_i - A^T\overline{p}_{i+1}$ in the line 7 of Alg. 2.

5: $d_{i+1} \leftarrow u_{i+1} - \Lambda_{\theta^d}(d_i, A^T\overline{p}_{i+1})$; // The operator in the line 6 of Alg. 2.

6: $\overline{u}_{i+1} \leftarrow \Psi_{\theta^{\overline{u}}}(u_{i+1}, d_{i+1})$; // The operators in the lines 7 and 8 of Alg. 2.

7: **until** $i \geq N$

8: **output** $\overline{u}_{N+1}$; $\overline{d}_{N+1} \leftarrow \Omega_{\theta^{\overline{d}}}(d_{i+1})$

The architecture of the proposed PD-Net is shown in Fig. 1 and the corresponding pseudocode is displayed in Alg. 3. The CNN Γ_{θ^p} (the blue block in Fig. 1(a)) is employed to replace the operator in Line 5 of Alg. 2. The inputs of this CNN include sinogram data g, feature maps $\overline{u}_i$, and feature maps p_i. The latter two kinds of feature maps are generated from the previous iteration. The output of this CNN is p_{i+1}, which is further

ACCEPTED MANUSCRIPT / CLEAN COPY

reduced to vector $\overline{p}_{i+1}$ by CNN $\Phi_{\theta^{\overline{p}}}$ (the light blue block in Fig. 1(a)). The operator $u_i - \tau A^T p_{i+1}$, in Line 7 of Alg. 2, is replaced by CNN Θ_{θ^u} (the red block in Fig. 1(a)). Through CNN Θ_{θ^u}, feature maps u_{i+1} is generated by updating feature maps u_i with vector $A^T \overline{p}_{i+1}$ generated in the current iteration. The vectors u_{i+1} represent the feature maps of the reconstructed image $\overline{u}_{i+1}$ in the current iteration. Combined with vector $A^T \overline{p}_{i+1}$, feature maps d_i are applied to restore feature maps d'_{i+1} in CNN Λ_{θ^d} (the green block in Fig. 1(a)). Because feature maps d_i are the divergences of feature maps u_i, the restored feature maps d'_{i+1} are also the feature maps of the reconstructed image $\overline{u}_{i+1}$ in the current iteration. Compared with CNN Θ_{θ^u}, CNN Λ_{θ^d} utilizes the extra feature maps d_i to yield feature maps d'_{i+1}. The outputs d'_{i+1} of CNN Λ_{θ^d} contain the more divergence information of the reconstructed image $\overline{u}_{i+1}$ than those of CNN Θ_{θ^u}. Thus, the differences d_{i+1} of feature maps d'_{i+1} and u_{i+1} can be employed to represent the divergences of feature maps u_{i+1} for the reconstructed image $\overline{u}_{i+1}$ in the current iteration. Moreover, a loss function is applied to the predicted divergences of the reconstructed image (Fig. 1), to ensure that feature maps d_{i+1} are the divergences of feature maps u_{i+1}. The vector d_{i+1} in Line 6 of Alg. 2 can be calculated by subtracting feature maps d'_{i+1} from feature maps u_{i+1}. The operator for the sum of the terms $u_i - \tau A^T p_{i+1}$ and τd_{i+1} in Line 7 of Alg. 2 and the operator in Line 8 are implemented by CNN $\Psi_{\theta^{\overline{u}}}$ (the purple block in Fig. 1(a)), whose output $\overline{u}_{i+1}$ is the reconstructed image of the current iteration. The outputs of the PD-Net are the reconstructed image $\overline{u}_{N+1}$ and its divergence $\overline{d}_{N+1}$ which is generated by CNN $\Omega_{\theta^{\overline{d}}}$ (Fig. 1(b)) with feature maps d_{N+1} after the N th iteration.

In this work, there are 3 convolutions per CNN. The pixel size and number of channels of each convolution are shown in Fig. 1. All convolutions are followed by batch normalization (BN) and a rectified linear unit (ReLU), apart from the last ones in CNNs $\Phi_{\theta^{\overline{p}}}$, $\Psi_{\theta^{\overline{u}}}$, and $\Omega_{\theta^{\overline{d}}}$. The parameters of these six kinds of CNNs are θ^p, $\theta^{\overline{p}}$, θ^d, $\theta^{\overline{d}}$, θ^u, and $\theta^{\overline{u}}$, and their parameters contain parameters ξ_i from Alg. 2. All these parameters are learned from big data. In this work, a ray-driven method and a pixel-driven method[33] are employed to perform the forward projection $A\overline{u}_i$ and the back-projection $A^T \overline{p}_{i+1}$, respectively. There are two outputs $\overline{u}_{N+1}$ and $\overline{q}_{N+1}$ for the PD-Net. During training, the ℓ_2-norm is utilized as the loss to minimize the differences between predicted images $\overline{u}_{N+1}$ and actual images $\hat{u}$, and the differences between the predicted divergences $\overline{d}_{N+1}$ and the divergences $\hat{d}$ calculated from actual images $\hat{u}$. Therefore, the total loss of the PD-Net can be defined as:

ACCEPTED MANUSCRIPT / CLEAN COPY

$$L_{total} = \left\| \bar{u}_{N+1} - \hat{u} \right\|_2^2 + \gamma \left\| \bar{d}_{N+1} - \hat{d} \right\|_2^2, \tag{12}$$

where γ is the weight parameter. In this work, this parameter is set to 0.5.

As displayed in Fig. 1, six kinds of CNNs employed are closely combined in the PD-Net, which is trained in an end-to-end manner. In this work, after analytically deriving the CP algorithm instance, we discovered that the TV prior contributes mainly to updating the reconstructed images by its divergences in each iteration of the solution process. Based on this discovery, two CNNs are proposed to yield the divergences of feature maps u_i for the reconstructed image in each iteration. Additionally, to guarantee that the yielded results of these two CNNs are the same as the divergences of feature maps u_{i+1} during the iteration, a loss function is applied to the predicted divergences of the reconstructed image. Hence, the two proposed CNNs can play the same role in the PD-Net as the TV prior in the IR methods. Thus, the TV prior in the CP algorithm instance can be unrolled to CNNs. In this work, by unrolling the TV prior to the CNNs, our PD-Net effectively preserves the structural and textural information in the reconstruction results.

2.C. Datasets

The Low-Dose CT Image and Projection Data was employed to assess the effectiveness of our proposed PD-Net. The provided projection data were acquired from a helical trajectory but are not suitable for fan-beam CT reconstruction. In this work, we utilized a fan-beam geometry to simulated projection data in 2D CT scanning settings through the forward projection of the provided CT images. Chest CT images from 20 patients were selected, and we randomly chose 3,357 images from 18 patients for training and 100 images from the remaining two patients for testing. We performed the forward projections for the chosen images by equally sampling along the rotation angle direction. The linear detector has 768 bins with physical lengths of 1 mm. The distance is 1068.0 mm from the X-ray source to the detector arrays, and 595.0 mm from the X-ray source to the center of rotation. The datasets comprise simulated projection data and the corresponding chosen CT images. These simulated projection data are employed to reconstruct the 512×512-pixel CT images. The physical height and width of each pixel are both 0.5859 mm. The simulated projection data are the noise-free data $\mathbf{y}$, without considering the noise introduced during acquisition. To model a realistic clinical environment, Poisson noise and electronic noise[5] were added to the noise-free data $\mathbf{y}$ by

$$\hat{\mathbf{y}} = \ln \frac{I_0}{Poisson\left(I_0 \exp(-\mathbf{y})\right) + Normal(0, \varepsilon^2)}, \tag{13}$$

where I_0 is the number of X-ray photons employed to penetrate the object, and ε^2 denotes the variance used to simulate the electronic noise from the equipment measurement error. For the normal dose dataset, the projection data were generated under a 360-view scanning condition. The noise was added to the projection data with the parameters $I_0 = 10^6$ and $\varepsilon^2 = 10$.[15] Under the 360-view scanning condition, we also simulated the low-dose datasets of different dose levels, namely 15%, 10% and 5%. To add noise of the different levels to the datasets, parameter I_0 was set to 1.5×10^5, 10^5, and 5×10^4, respectively.[15] The sparse dataset was simulated under the 120-view scanning condition. The noise was added to the projection data with $I_0 = 10^6$.

ACCEPTED MANUSCRIPT / CLEAN COPY

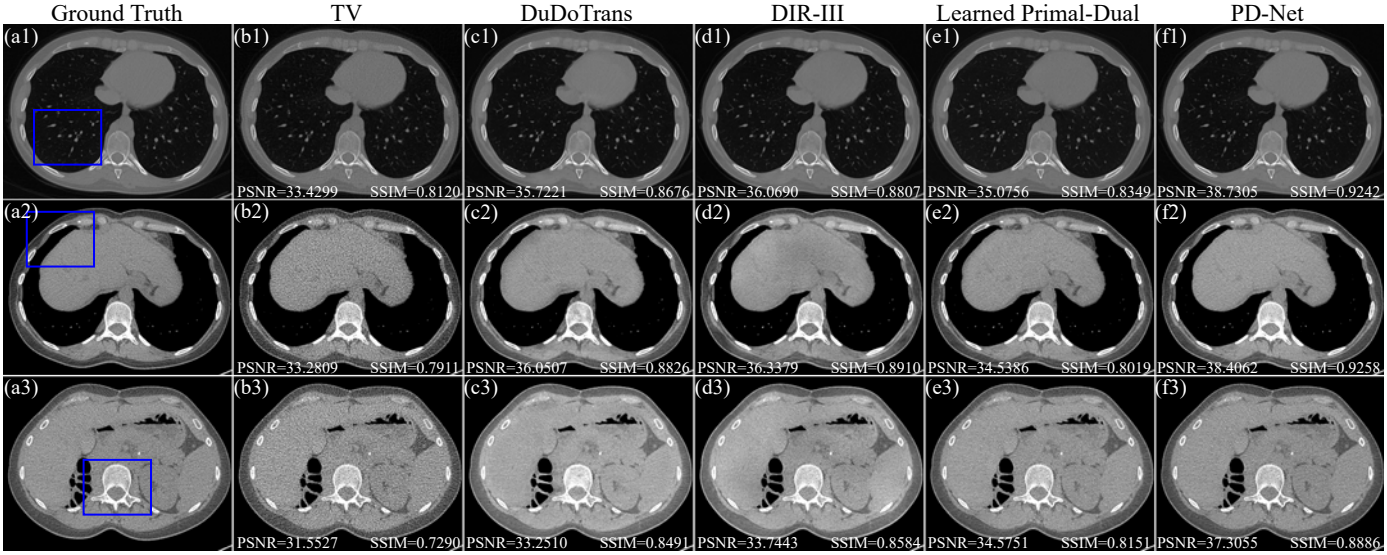

FIG. 2. CT reconstruction results on normal dose data simulated from the Low-Dose CT Image and Projection Data. The display windows are [-1050, 1950] HU, [-250, 450] HU, and [-320, 480] HU for images in the three rows, respectively.

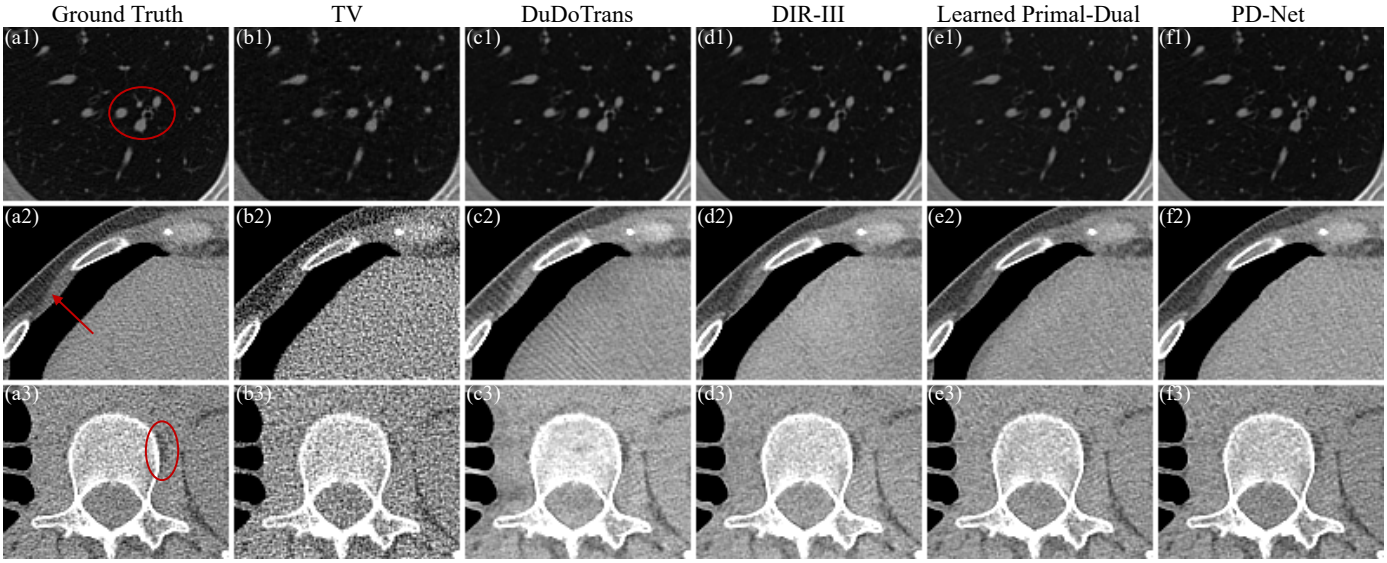

FIG. 3. Zoomed parts from Fig. 2.

To further investigate the ability of the proposed PD-Net, we also performed experiments on the Piglet dataset,[34] which includes four different sub-datasets. We tested PD-Net on the sub-dataset of the normal dose CT, which contains 565 images for training and 190 images for testing. Using the same scanning conditions for the normal dose dataset simulated from the Low-Dose CT Image and Projection Data, we also generated a normal dose dataset from the Piglet dataset.

2.D. Implementation Details

The PD-Net model was implemented on the PyTorch deep learning toolbox.[35] To learn the parameters of the PD-Net model from the dataset, the loss function (12) was optimized end-to-end using the Adam optimizer.[36] The initial learning rate was set to 10^{-4} and was decreased by 20% after each epoch. The experiments were performed on a workstation computer platform with an Intel® Core™ i7-9700K CPU, 32 GB RAM, and an NVIDIA RTX 3090 GPU.

ACCEPTED MANUSCRIPT / CLEAN COPY

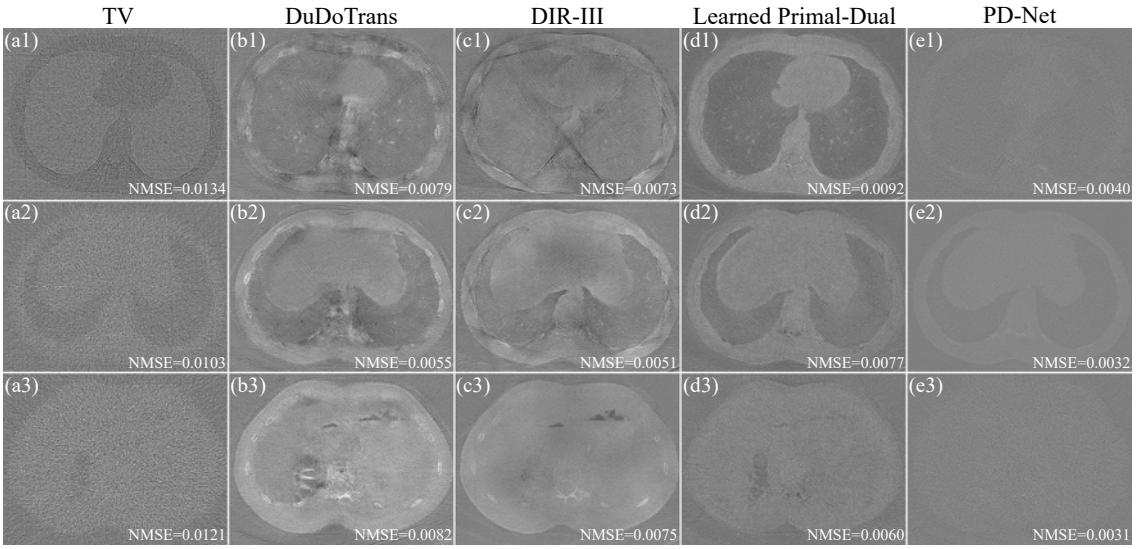

FIG. 4. Residual images of normal dose data reconstructions for the different methods in Fig. 2. The display window is [−10, 10] HU.

TABLE I. Quantitative measures (average scores ± standard deviations) of the reconstructions on normal dose data simulated from the Low-Dose CT Image and Projection Data.

Metrics	TV	DuDoTrans	DIR-III	Learned Primal-Dual	PD-Net
PSNR	32.5454±1.0149	34.9937±0.7886	35.1899±1.0018	34.5478±0.7318	**37.7627±1.0817**
SSIM	0.7801±0.0356	0.8575±0.0169	0.8693±0.0178	0.8230±0.0175	**0.9071±0.0182**
NMSE	0.0126±0.0015	0.0074±0.0018	0.0071±0.0020	0.0085±0.0027	**0.0039±0.0008**

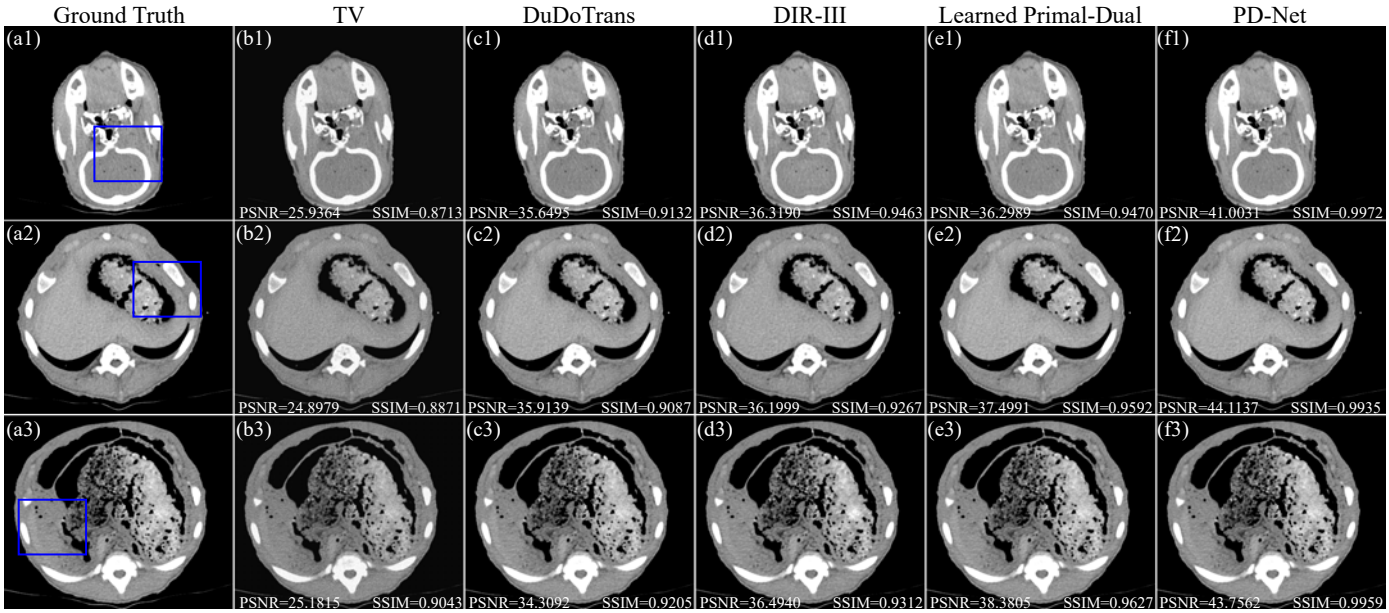

FIG. 5. CT reconstruction results on normal dose data simulated from the Piglet dataset. The display windows are [-1050, 1950] HU for images in the three rows.

Our proposed PD-Net was compared to several widely used reconstruction algorithms, including the FBP algorithm (Ram-Lak filter), TV algorithm,[3] DuDoTrans,[37] DIR-III,[38] and Learned Primal-Dual.[13] The
260 Tomographic Iterative GPU-based Reconstruction (TIGRE) toolbox[39] was employed to execute the TV algorithm. The training settings of DuDoTrans, DIR-III, and Learned Primal-Dual were based on Refs. [37], [38] and [13], respectively. The normalized mean square error (NMSE), peak signal-to-noise ratio (PSNR), and

ACCEPTED MANUSCRIPT / CLEAN COPY

structural similarity index measure (SSIM) [40] were employed to quantitatively evaluate the performance of our proposed PD-Net and the comparison models.

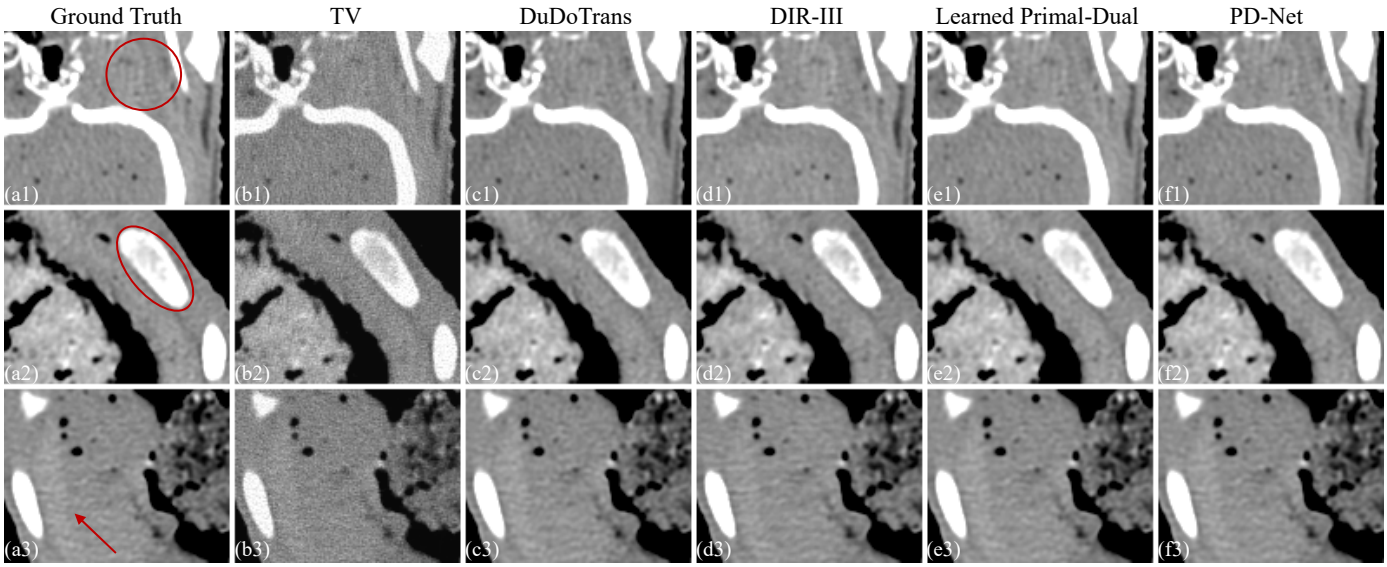

FIG. 6. Zoomed parts from Fig. 5.

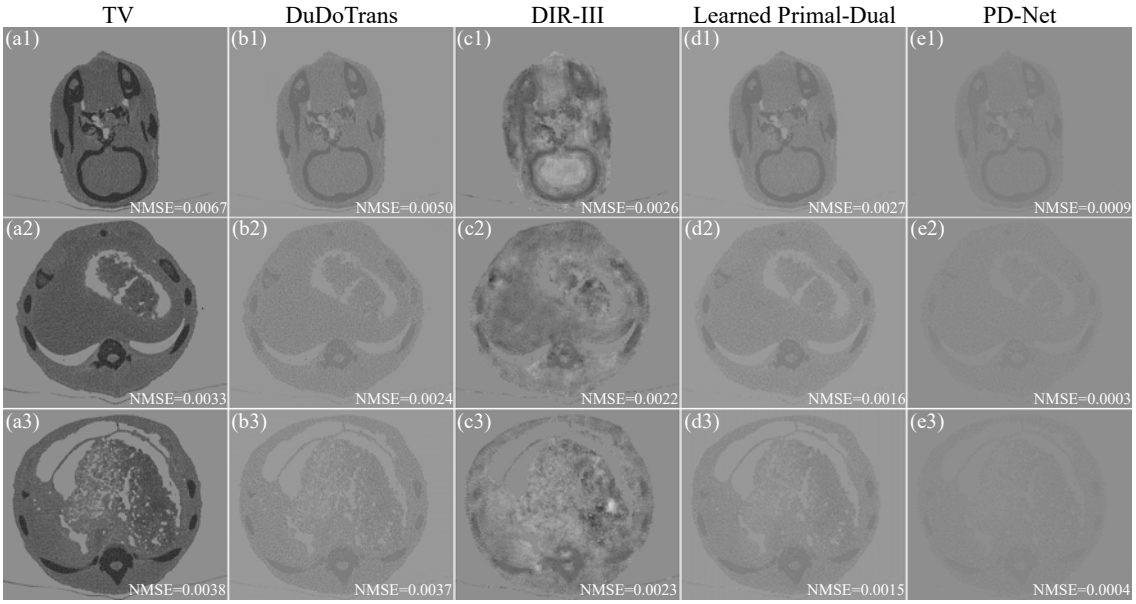

FIG. 7. Residual images of normal dose data reconstructions for the different methods in Fig. 5. The display window is [−10, 10] HU.

TABLE II. Quantitative measures of the reconstructions on normal dose data simulated from the Piglet dataset.

Metrics	TV	DuDoTrans	DIR-III	Learned Primal-Dual	PD-Net
PSNR	26.0673±1.3502	35.3099±1.2407	35.9514±1.2205	38.3275±1.2774	**44.8953±1.6320**
SSIM	0.8736±0.0327	0.9041±0.0239	0.9131±0.0187	0.9276±0.0190	**0.9952±0.0020**
NMSE	0.0285±0.0060	0.0057±0.0094	0.0030±0.0012	0.0019±0.0014	**0.0004±0.0002**

3. RESULTS

3.A. Validation with Normal Dose Data

To demonstrate the effectiveness of our proposed PD-Net, we performed a qualitative comparison on the normal dose data simulated from the Low-Dose CT Image and Projection Data. Figure 2 presents the

ACCEPTED MANUSCRIPT / CLEAN COPY

reconstructed images of the proposed PD-Net and competing methods. Three reconstruction cases from the testing dataset are shown in the three rows. For a better evaluation of the visual quality of the reconstructed images in Fig. 2, the images over three regions of interest (ROIs) marked by the blue rectangles are zoomed-in and depicted in Fig. 3. The corresponding residual images of the reconstruction results are shown in Fig. 4. The TV algorithm's reconstructed images were shown to be contaminated by the piecewise constant effect (Figs. 2(b) and 3(b)), and the images reconstructed by the DuDoTrans algorithm suffered from small streak artifacts (Figs. 2(c) and 3(c)). As shown in Figs. 2(d) and 3(d), the DIR-III algorithm works well with noise suppression and artifact removal but blurs some local structural features. The Learned Primal-Dual algorithm improved the visual quality of the reconstructed images (Figs. 2(e) and 3(e)). However, some tiny structures were lost in the reconstruction results. In Figs. 2 and 3, the proposed PD-Net produced promising images in terms of texture preservation (Fig. 3(f)). Upon visual inspection, our proposed PD-Net achieves better reconstruction accuracy than those of the competing algorithms (Fig. 4). Table I lists the average scores and standard deviations of PSNR, SSIM, and NMSE for all testing phantom results. Our proposed PD-Net achieved the best numerical values among all competing methods.

To further assess the ability of the proposed PD-Net, we performed a CT reconstruction using normal dose data simulated from the Piglet dataset. The reconstructed images of the different methods are shown in Fig. 5. Their performances are visually differentiated in Fig. 5 and over the three ROIs in Fig. 6. The corresponding residual images of the reconstruction results are displayed in Fig. 7. All reconstruction methods efficiently reconstruct tomography images from the projection data. Compared to the competing algorithms, PD-Net suppresses artifacts and maintains structural features better (Fig. 6), yielding the most similar reconstruction results to the ground truth (Fig. 7). According to human perceptual evaluations, our proposed PD-Net exhibits better visual quality with respect to authentic structural and textural information (Fig. 6). Consistent with the visual results, the proposed PD-Net achieved the highest PSNR and SSIM, and the lowest NMSE (Table II). It can be seen that our proposed PD-Net indisputably improved these three metrics.

3.B. Validation with Low Dose Data

To evaluate the performance of our proposed PD-Net for different dose levels, a dataset simulated with 10% doses was utilized for training and testing. The reconstructed images are shown in Figs. 8-10. Because the radiation dose was severely reduced, the reconstruction results of the TV algorithm were seriously degraded with the loss of textural and structural information (Figs. 8(b) and 9(b)). In Figs. 8 and 9, the DuDoTrans, DIR-III and the Learned Primal-Dual algorithms efficiently suppressed the noise in the reconstructed images. However, the DuDoTrans results were contaminated by strong artifacts (Figs. 9(c1)-(c3)). The DIR-III algorithm blurred some crucial structural information (Figs. 9(d1)-(d3)). Additionally, the Learned Primal-Dual's results suffered from the loss of texture details (Figs. 9(e1)-(e3)). Our proposed algorithm exhibits better visual quality with respect to fine structural details, as shown in Figs. 9 and 10. For the low-dose data, the statistical quantitative results of the different reconstruction algorithms are shown in Table III. It is clear that our proposed PD-Net obtains better numerical values than all competing methods.

3.C. Validation with Sparse Data

To further demonstrate the effectiveness of our proposed PD-Net, we also performed a CT reconstruction using sparse data. The reconstructed images obtained under the scanning condition of 120 views are demonstrated in Fig. 11, and a quantitative analysis of ROIs is provided in Fig. 12. The difference images of the reconstruction results are displayed in Fig. 13. As observed in Figs. 11(b) and 12(b), there are severe streak artifacts in the images reconstructed by the TV algorithm. The DuDoTrans and DIR-III algorithms greatly suppressed the streak artifacts in the reconstruction results (Figs. 11 and 12). However, there are still some tiny

ACCEPTED MANUSCRIPT / CLEAN COPY

streak artifacts in the DuDoTrans results (Figs. 11(c1)-(c3)). Figures 12(d1)-(d3) clearly show that the reconstruction results of the DIR-III algorithm suffered from losing diagnostic information. The Learned Primal-Dual algorithm effectively recovered the desired images from the projection data, but they suffered from the loss of structural information and spatial resolution (Figs. 12(e1)-(e3)). Compared with the Learned Primal-Dual algorithm, our proposed PD-Net preserved more anatomical details, thereby yielding images of higher visual quality (Figs. 12(f1)-(f3)). Table IV lists the quantitative measures of the reconstructed images on sparse data. As shown in Table IV, the PD-Net scored the highest PSNR and SSIM, and the lowest NMSE.

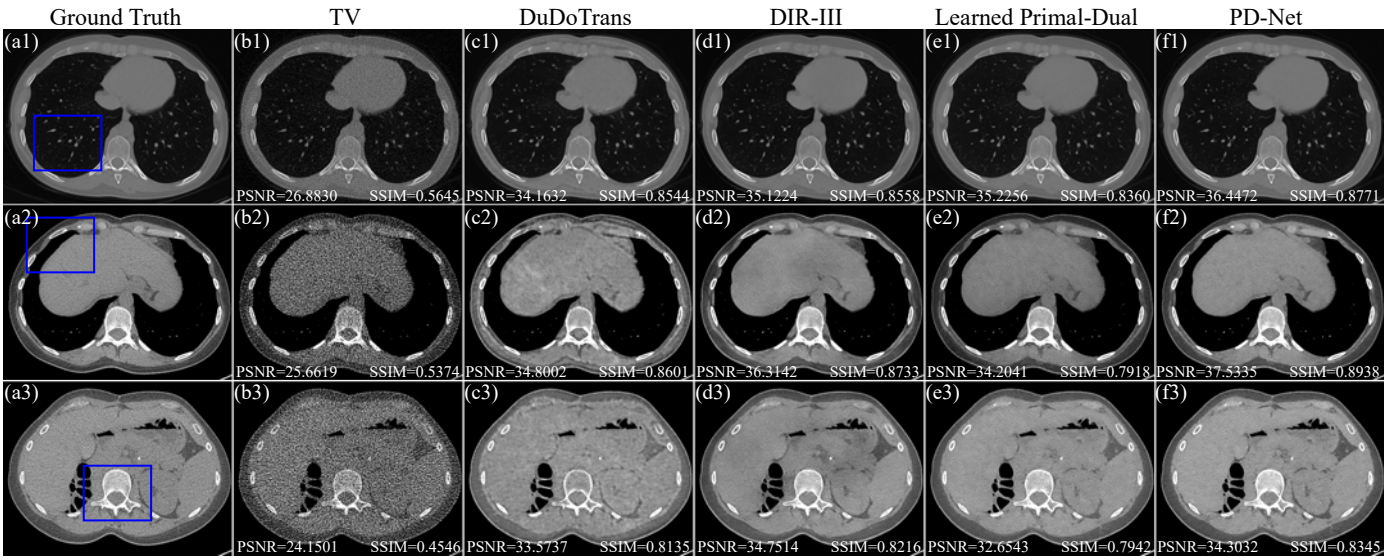

FIG. 8. CT reconstruction results on low-dose data. The display windows are [-1050, 1950] HU, [-250, 450] HU, and [-320, 480] HU for images in the three rows, respectively.

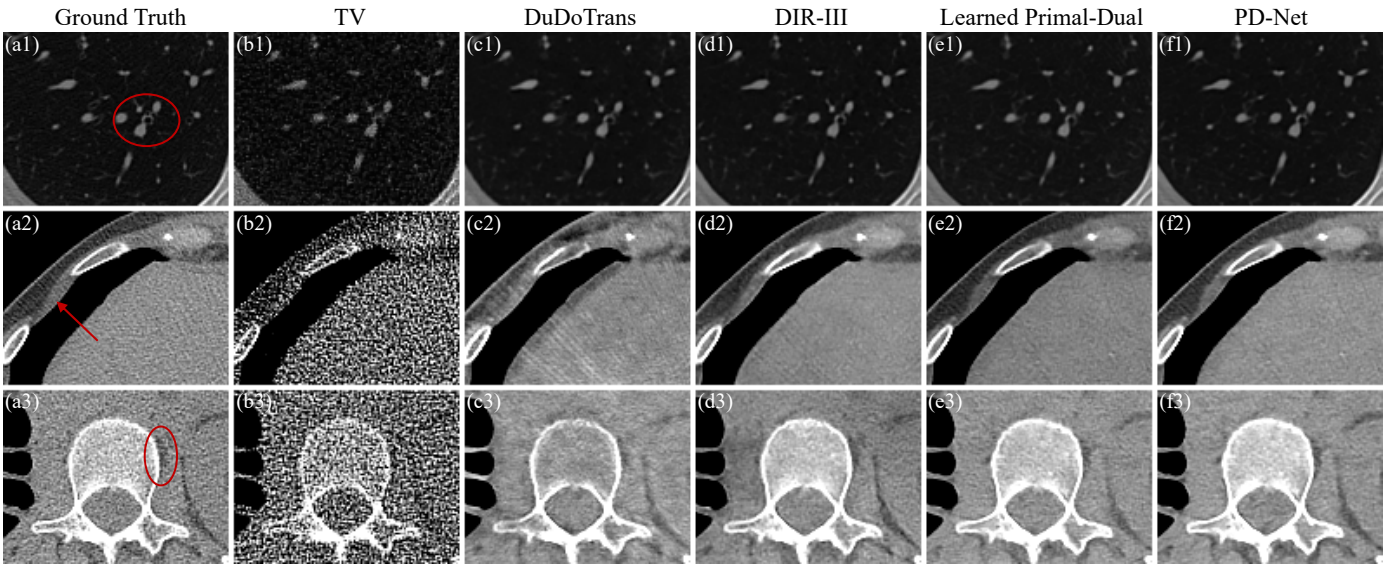

FIG. 9. Zoomed parts from Fig. 8.

3.D. Network Design Evaluation

In this section, we present an efficacy assessment of the combination chosen for the proposed PD-Net framework. The proposed PD-Net framework consists of six kinds of CNNs, as discussed in Section 2.B. However, we can also construct the Learned Primal-Dual by only utilizing CNNs Γ_{θ^p} and Θ_{θ^u} (Sup. 1 in the

ACCEPTED MANUSCRIPT / CLEAN COPY

supplementary file). We removed the BN and ReLU operators after each convolution and added a parametric ReLU (PReLU) after the first two convolutions. A feature map of the outputs of CNN Γ_{θ^p} was selected as output $\overline{p}_{i+1}$, and that of the outputs of CNN Θ_{θ^u} was used as output $\overline{u}_{i+1}$. Based on the Learned Primal-Dual, the Primal-Dual-ReLU was obtained by removing all PReLU operators and adding the BN and ReLU operators after each convolution (Sup. 2). To further improve the Primal-Dual-ReLU, the extra CNNs $\Phi_{\theta^{\overline{p}}}$ and $\Psi_{\theta^{\overline{u}}}$

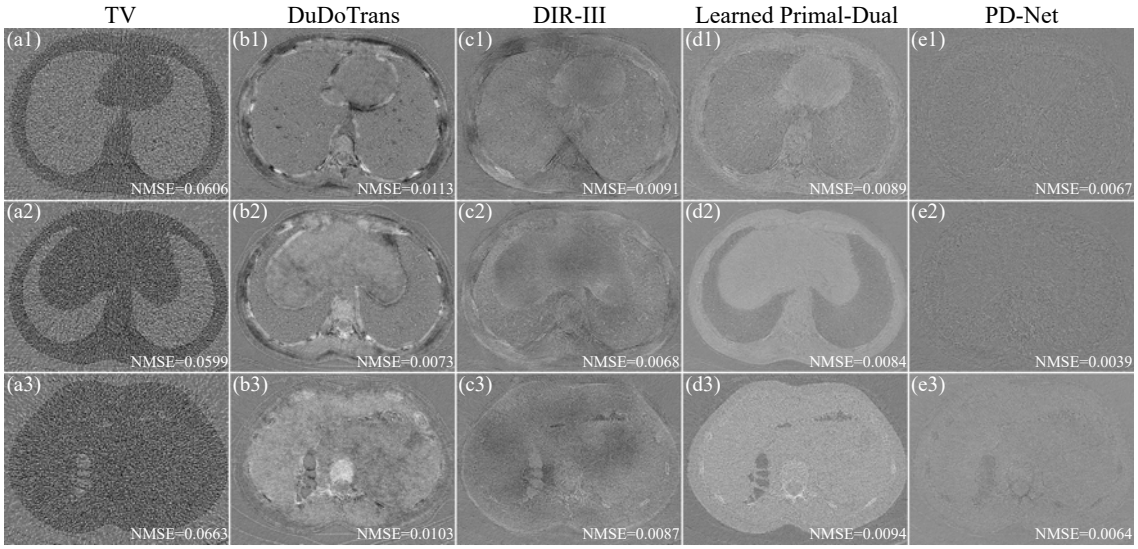

FIG. 10. Residual images of low-dose data reconstructions for the different methods in Fig. 8. The display window is [−10, 10] HU.

TABLE III. Quantitative measures of the reconstructions on low-dose data.

Metrics	TV	DuDoTrans	DIR-III	Learned Primal-Dual	PD-Net
PSNR	25.2906±1.4185	33.4349±1.0586	34.5303±0.9938	33.8402±1.0074	**35.4082±1.1821**
SSIM	0.5299±0.0484	0.8293±0.0227	0.8393±0.0249	0.7951±0.0285	**0.8590±1.1821**
NMSE	0.0673±0.1000	0.0105±0.0024	0.0082±0.0020	0.0097±0.0031	**0.0067±0.0016**

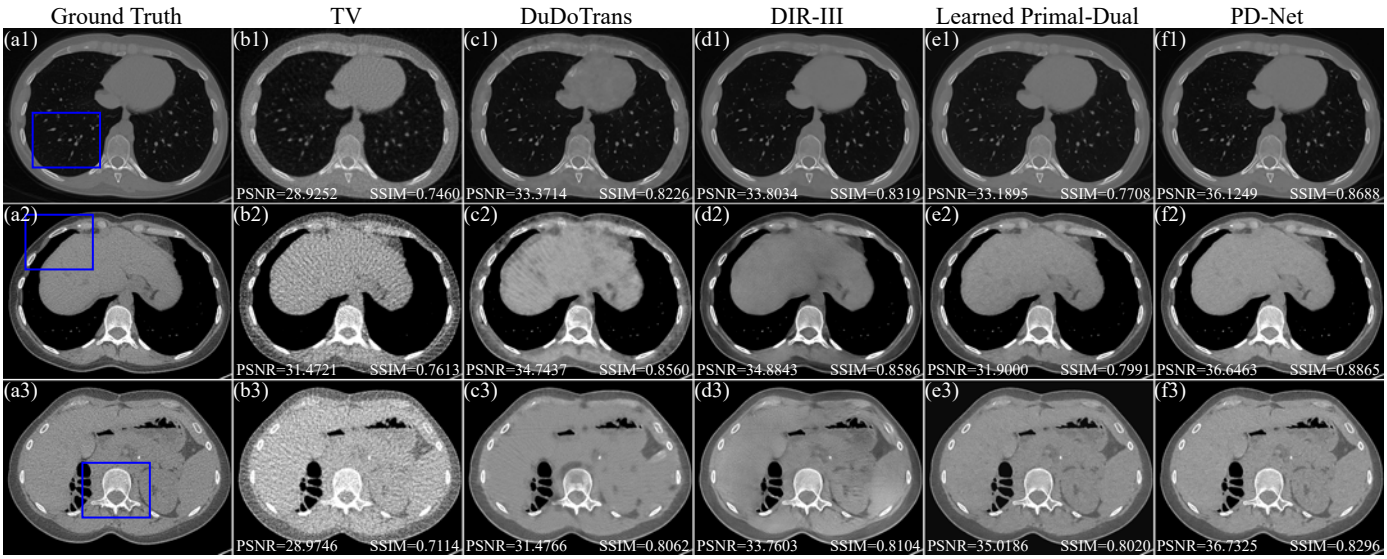

FIG. 11. CT reconstruction results on sparse data. The display windows are [-1050, 1950] HU, [-250, 450] HU, and [-320, 480] HU for images in the three rows, respectively.

ACCEPTED MANUSCRIPT / CLEAN COPY

were employed to construct the network architecture of the Primal-Dual-Fusion (Sup. 3). In these two kinds of CNNs, outputs $\overline{p}_{i+1}$ and $\overline{u}_{i+1}$ were obtained by extracting the effective information from the feature maps p_{i+1} and u_{i+1}, respectively. Our proposed PD-Net further improved the Primal-Dual-Fusion by unrolling the TV prior in the cost function via CNNs Λ_{θ^d} and $\Omega_{\theta^{\bar{d}}}$ (Fig. 1).

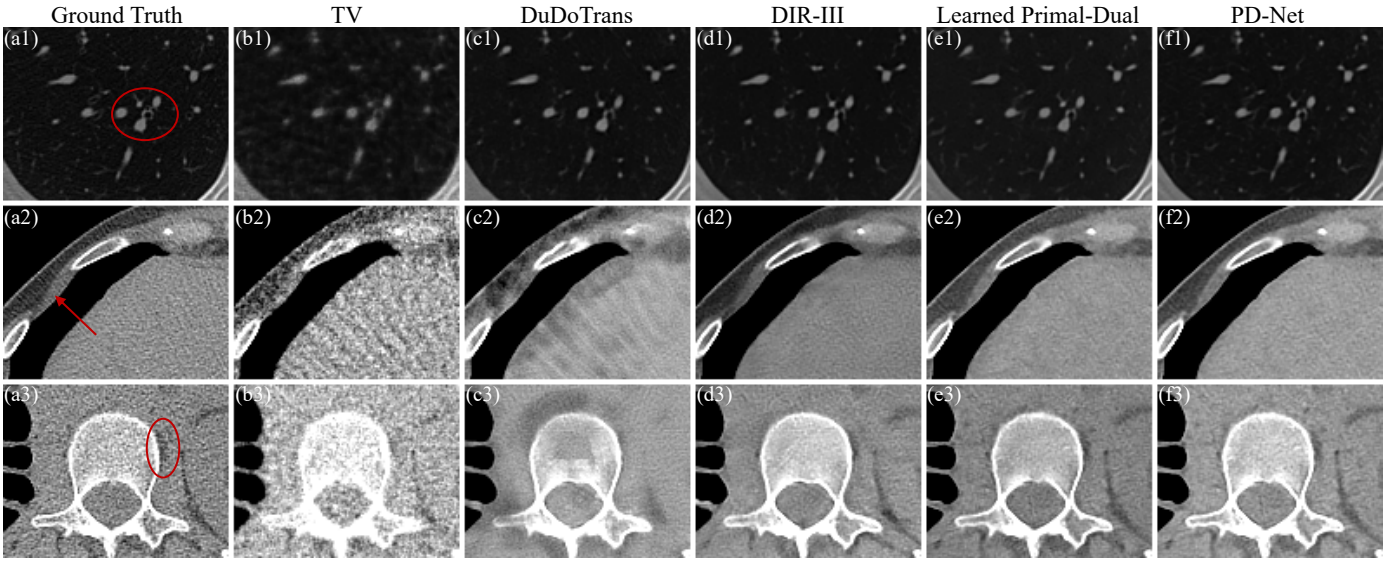

FIG. 12. Zoomed parts from Fig. 11.

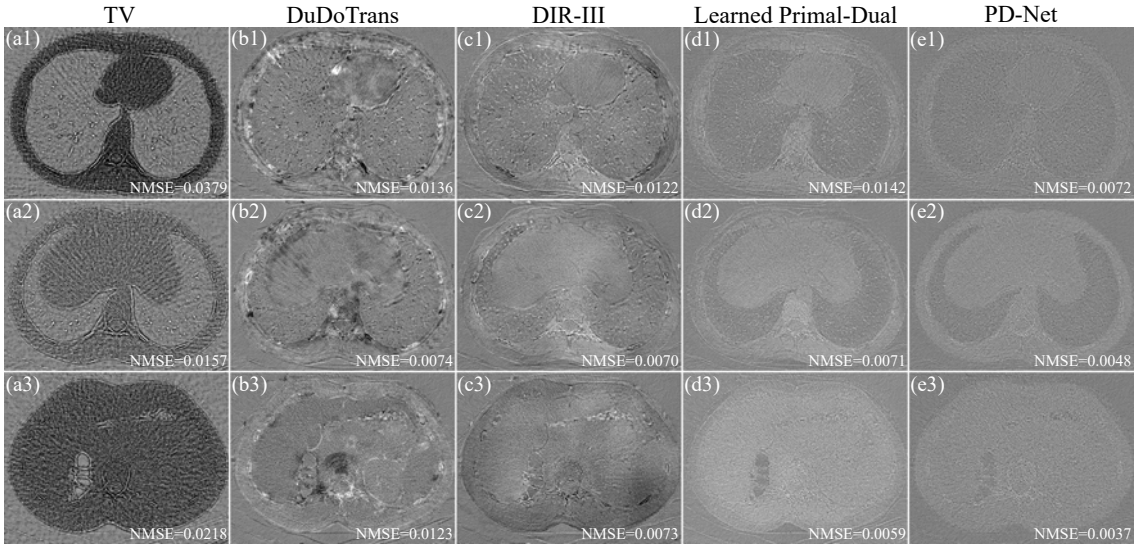

FIG. 13. Residual images of sparse data reconstructions for the different methods in Fig. 11. The display window is [−10, 10] HU.

TABLE IV. Quantitative measures of the reconstructions on sparse data.

Metrics	TV	DuDoTrans	DIR-III	Learned Primal-Dual	PD-Net
PSNR	28.8094±1.4618	33.1754±0.9225	33.4723±0.7394	34.0128±0.8184	**35.4554±1.0040**
SSIM	0.7241±0.0287	0.8114±0.0244	0.8201±0.0230	0.7937±0.0202	**0.8498±0.0249**
NMSE	0.0316±0.0136	0.0113±0.0032	0.0104±0.0025	0.0092±0.0024	**0.0066±0.0015**

The images in the first row of Fig. 14 present the reconstruction results of the different network designs on normal dose data. The images over the ROIs of the reconstruction results are zoomed-in and depicted in the second row of Fig. 14. The Learned Primal-Dual suffered from the loss of subtle structural information (Figs.

ACCEPTED MANUSCRIPT / CLEAN COPY

14(b1) and (b2)). The Primal-Dual-ReLU yielded reconstruction results with better subtle details than the Learned Primal-Dual (Figs. 14(c1) and (c2)). Compared with the Primal-Dual-ReLU, the Primal-Dual-Fusion further improved the reconstructed results by preserving more structural features (Figs. 14(d1) and (d2)). However, the Primal-Dual-Fusion also blurred some local textures. As shown in Fig. 14, the proposed PD-Net exhibited better visual quality and texture preservation than the other network designs. To quantitatively evaluate the statistical properties of images reconstructed by different network designs, we calculated the average scores and standard deviations of the PSNR, SSIM, and NMSE results for all testing phantom results. As shown in Table V, our proposed PD-Net scored the highest PSNR and SSIM, and the lowest NMSE. These promising reconstruction performances indicate the efficacy of the combination chosen for the proposed PD-Net framework.

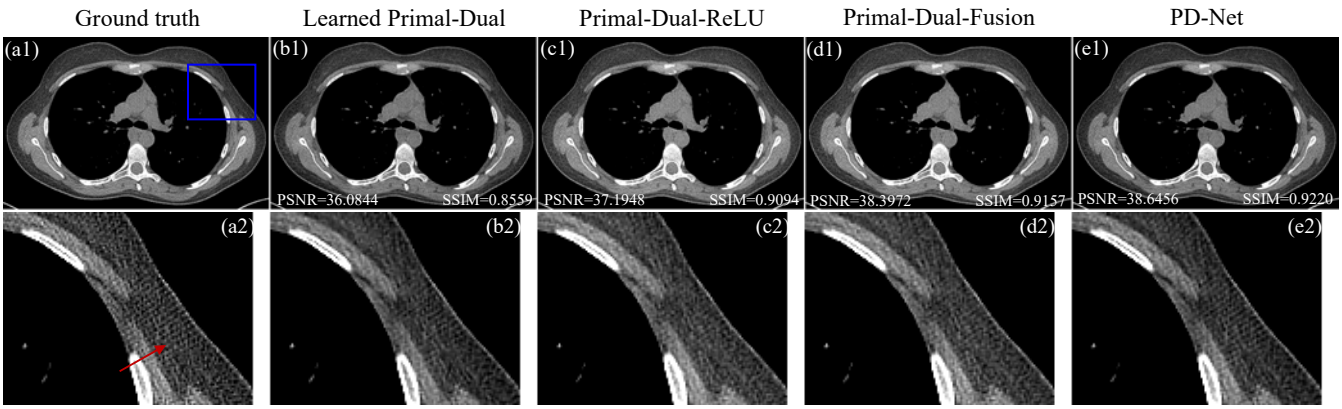

FIG. 14. CT reconstruction results on normal dose data for different network architectures. The display window is [-250, 450] HU.

TABLE V. Quantitative measures of the reconstructions for different network architectures.

Metrics	Learned Primal-Dual	Primal-Dual-ReLU	Primal-Dual-Fusion	PD-Net
PSNR	34.5478±0.7318	36.8346±1.0672	37.2678±1.0136	**37.7627±1.0817**
SSIM	0.8230±0.0175	0.8976±0.0158	0.9010±0.0177	**0.9071±0.0182**
NMSE	0.0085±0.0027	0.0048±0.0011	0.0041±0.0008	**0.0039±0.0008**

3.E. Robustness Analysis

In clinical settings, scanning doses may vary based on the task at hand. Training multiple networks for different dose levels makes it clinically infeasible to apply most learning based algorithms.[41-43] To verify the robustness of our proposed PD-Net, we utilized a model trained with 10% dose data to test the results at different dose levels: 15% and 5%. The chest reconstruction results are displayed in Fig. 15. The images in the first, third and fifth rows reflect the reconstructed images of the 5%, 10% and 15% dose cases, respectively. The images over ROIs are displayed in the second, fourth and sixth rows of Fig 15. It can be seen that the noise in the reconstructed images for FBP and TV algorithms dramatically increased alongside the reduction of scanning dose. As shown in Figs. 15(d1)-(d6), our proposed PD-Net effectively reduced noise and yielded images with more accurate anatomical details. The noise in the images reconstructed by our proposed PD-Net were reduced gradually with the increase of the radiation dose. The statistical quantitative results of the reconstructed images are shown in Table VI. The experimental results show that our proposed PD-Net can successfully reconstruct images from projection data with slightly different distributions.

Additionally, a blind reader study was conducted as a visual assessment. Two experienced radiologists (J. Li with 17 years of experience, W. Du with 9 years of experience) were invited to independently score 10 groups of randomly selected CT images. Each group included images obtained by the different reconstruction algorithms mentioned in this paper and the ground-truth versions. The radiologists were blind to which image was from which reconstruction algorithm. All images were evaluated on noise reduction, sharpness, diagnostic acceptance

ACCEPTED MANUSCRIPT / CLEAN COPY

and overall quality using a five-point scale (1 = “Unacceptable”, 2 = “Substandard”, 3 = “Acceptable”, 4 = “Above average”, and 5 = “Excellent”). The subjective criterion of diagnostic acceptance is applied to evaluate whether the image quality of the reconstruction results are acceptable for clinical diagnosis. The quality scores were reported as the average scores and standard deviations means. The quality scores of reconstructed images are shown in table VII. It can be seen that our proposed PD-Net achieved the highest scores. Here, the readers only examined the diagnozability in the reconstructed images and did not actually perform any detection/diagnosis on the reconstruction results.

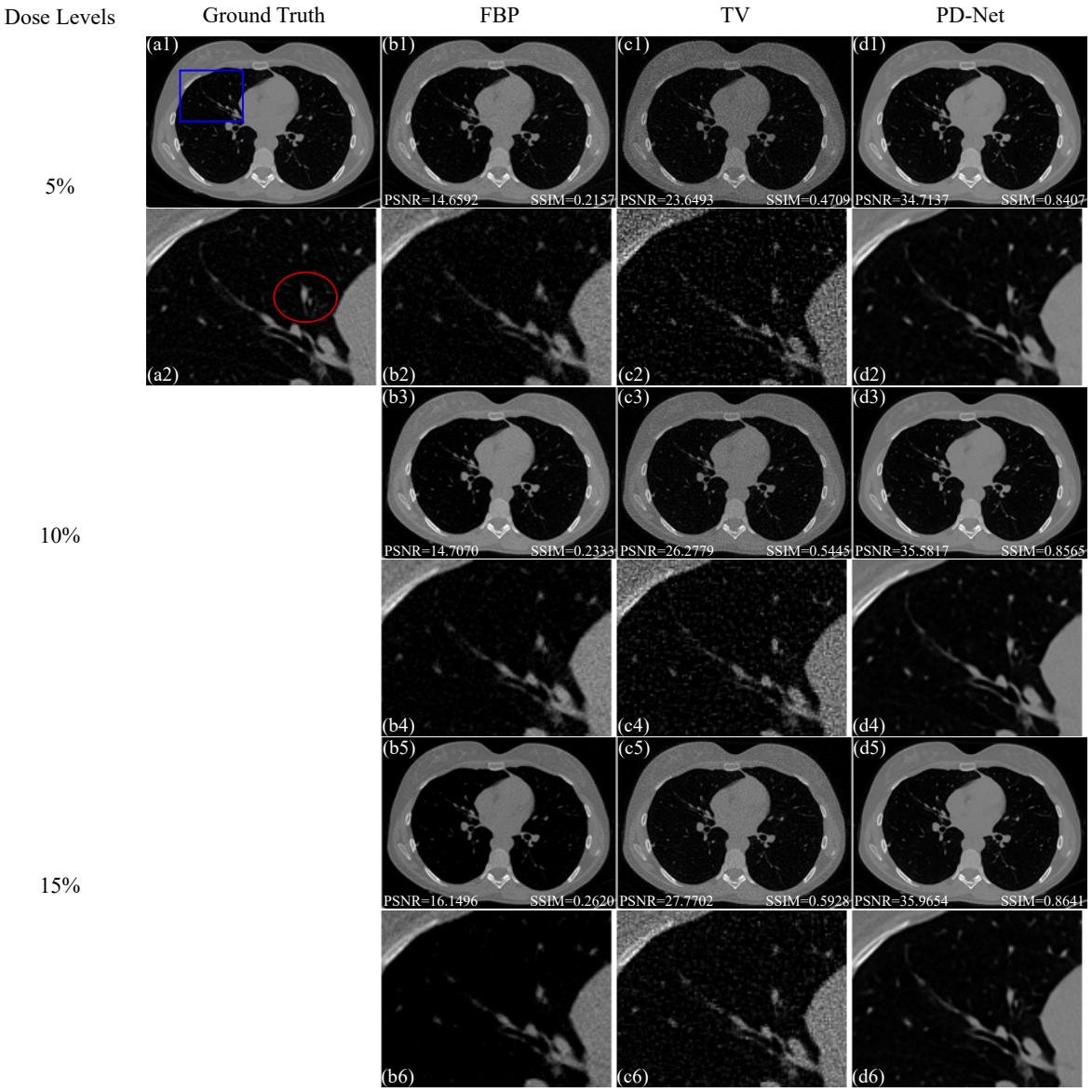

FIG. 15. CT reconstruction results on normal dose data. The display window is [-250, 450] HU.

TABLE VI. Quantitative measures of the reconstruction results for robustness analysis.

Metrics	5%	10%	15%
PSNR	34.3973±1.2255	35.4082±1.1821	35.7857±1.1969
SSIM	0.8425±0.0264	0.8590±1.1821	0.8662±0.0248
NMSE	0.0084±0.0020	0.0067±0.0016	0.0061±0.0013

3.F. Parameter Experiments

ACCEPTED MANUSCRIPT / CLEAN COPY

The number of feature maps for p_i, d_i, and u_i that persist between iterations directly affect the visual quality of the reconstructed images. In this work, we present the quantitative measures of the reconstructions with respect to the different feature map numbers. For these reconstruction results, the boxplots of the three selected metrics (i.e., PNSR, SSIM, and NMSE) are displayed in Fig. 16. The proposed PD-Net method achieves the poorest image quality metrics, when all feature map numbers are set to 1. However, these metrics are improved with an increase in the number feature maps. The improvement is limited, when the number is larger than 5. Thus, the feature maps passed between iterations were set to 5 in this work.

TABLE VII. Visual assessment scores by two radiologist readers.

Algorithms	Ground truth	TV	DuDoTrans	DIR-III	Learned Primal-Dual	PD-Net
Noise Reduction	-	2.90±0.8307	3.95±0.7703	4.00±0.7532	4.15±0.7263	**4.55±0.5895**
Sharpness	-	3.25±0.6225	3.65±0.6427	3.70±0.6852	3.60±0.6633	**4.05±0.5895**
Diagnostic Acceptability	-	3.55±0.4975	3.85±0.6825	3.90±0.6457	4.00±0.7746	**4.50±05916**
Overall Quality	4.60±0.4899	3.15±0.5723	4.10±0.5308	4.15±0.4729	4.20±0.4000	**4.55±0.4975**

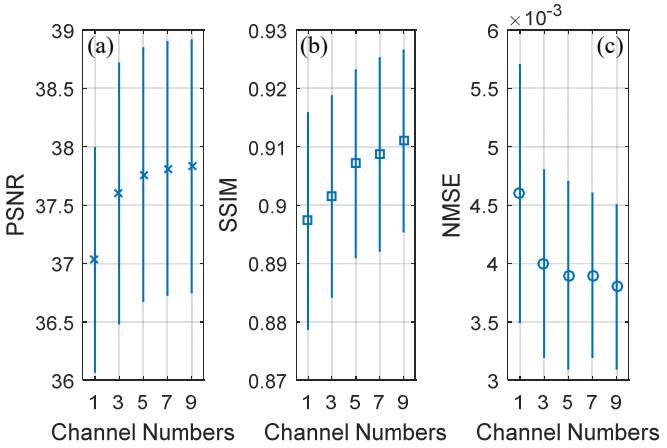

FIG. 16. Comparison of indices for the proposed method of different feature map numbers. Comparison of boxplots of (a) PSNR, (b) SSIM and (c) NMSE for reconstruction results.

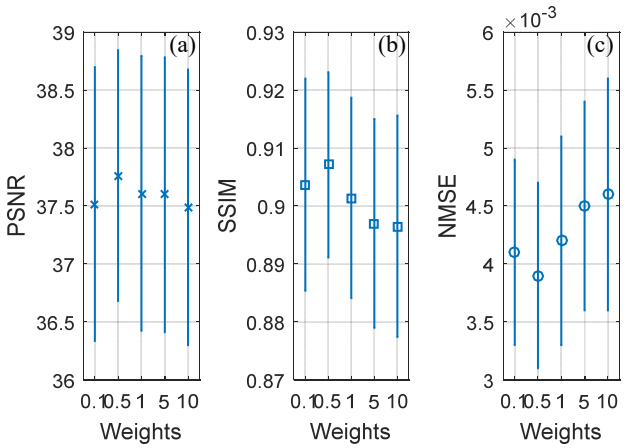

FIG. 17. Comparison of indices for the proposed method with different weights. Comparison of boxplots of (a) PSNR, (b) SSIM and (c) NMSE for reconstruction results.

TABLE VIII. Time costs of the different methods for training and testing on normal dose data.

	TV	DuDoTrans	DIR-III	Learned Primal-Dual	PD-Net
Train	-	46 h	37 h	18 h	60 h
Test	1963 s	0.2681 s	0.1647 s	0.1443 s	0.3715 s

The hyper-parameter λ controls the tradeoff between the two per-pixel ℓ_2 losses in Equation (12). To find the optimal value of this hyper-parameter, we performed parameter experiments using normal dose data. Fig. 17 displays the boxplots of PNSR, SSIM, and NMSE for the proposed PD-Net with weights of 0.1, 0.5, 1, 5 and 10. It can be seen that the proposed PD-Net method achieved the best image quality metrics, when λ equals 0.5. Thus, we used this setting in this work.

4. DISCUSSION

When including the TV prior in the cost function of an inverse problem, the IR methods yield promising reconstructed images with critical feature preservation that is useful for diagnosis. In this work, to further improve upon these methods, we developed a PD-Net unrolling approach in which both the data-fidelity term

ACCEPTED MANUSCRIPT / CLEAN COPY

and the TV prior of the cost function were unrolled to the CNNs. By deriving the CP algorithm instance for CT reconstruction, we discovered that the main contribution of including the TV prior in IR methods is to update the reconstructed images by its weighted divergences in each iteration of the solution process. Based on this discovery, CNNs were proposed to yield the divergences of feature maps for the reconstructed image in each iteration. Additionally, we applied a loss function to the predicted divergences, ensuring that the yielded results of the proposed CNNs were indeed the divergences of feature maps for the reconstructed image generated in the iteration. In this manner, the proposed CNNs play the same roles in the PD-Net as the TV prior in the IR methods. Thus, the TV prior in the CP algorithm instance can be unrolled directly to the CNNs in the PD-Net. The experimental results (Figs. 2-13) show that our proposed PD-Net framework achieves robust and superior performances in texture preservation.

Compared to the Learned primal-dual model,[13] our proposed PD-Net improved the unrolling approaches by unrolling the TV prior in the IR methods to the CNNs. Moreover, the proposed CNNs in our PD-Net performed the same roles as the TV prior in the IR methods. The proposed PD-Net produced high-quality CT images with better anatomical feature preservations (Figs. 2-13) and great improvements in image quality metrics is shown in Tables I-IV. The experimental results clearly demonstrate that unfolding the TV prior in the IR methods to the CNNs further improves the feature extraction capability of the unrolling approaches. ISTA-Net [28] is an unrolling approach, which is constructed by appropriately unrolling the iterative algorithm. To improve reconstruction performance, ISTA-Net adopts a CNN to replace the hand-crafted sparsity regularizer. Without giving the analytical formula of sparsity transform, ISTA-Net cannot fully unroll the prior term. Compared with ISTA-Net, the proposed PD-Net was constructed by unrolling both the data-fidelity term and the prior term. The PD-Net affords a tailor-made unrolling approach with the unfoldment of the hand-crafted TV prior.

Conventional IR methods adopt an iteration technique to solve the inverse problem. In each iteration, the TV prior contributes to updating the reconstructed images with its weighted divergences.[44] However, the iteration technique may blur image edges, especially those of inconspicuous edges. On one hand, the noise in the images may lead to incorrect estimations of image divergences in each iterate. On the other hand, the inappropriate selection of divergence weights may also reduce the visual quality of the reconstruction results. To overcome these drawbacks, a loss function was applied to the predicted divergences to avoid the incorrect estimation of the image divergences in the PD-Net. Moreover, divergence weights were learned from big data, reducing the blur of reconstructed images.

For conventional unrolling approaches, IR methods directly unroll to neural networks for CT reconstruction. In unrolling approaches,[13-19] the number of feature maps that persist between the iterations is usually set to 1. However, the reduction of persistent feature maps between the iterations may decrease the feature extraction capability of CT reconstruction networks. As observed in Fig. 10, the image quality metrics were the poorest when the feature map number was set to 1. Thus, to improve the visual quality of the reconstructed images, more feature maps were employed to transmit effective feature information between the iterations. As discussed in Ref. [45], the increase in the persistent feature maps may also improve the convergence rate of the reconstruction networks in the unrolling approaches.

The runtimes of the different reconstruction algorithms on normal dose data are shown in Table VIII. It can be seen that learning-based reconstruction algorithms take many hours for training. However, the testing time for a single image was dramatically reduced. When the model is trained, the three learning based approaches mentioned in this work will take far less execution time than is consumed by the TV algorithm, which facilitates clinical work flows. Owing to the calculations of CNN convolutions for image divergences, the computational time of our PD-Net was slightly longer than that of the Learned Primal-Dual algorithm. More effort is needed to reduce the computational time of the PD-Net to meet the requirements of clinical practice.

ACCEPTED MANUSCRIPT / CLEAN COPY

In this work, the projection data simulated from real CT images were employed for the experiments. Based on these simulated projection data, our proposed PD-Net yielded promising results with high image quality and quantitative metrics. However, the simulated projection data differed from the clinical CT data. In future research, we plan to further validate and assess the performance of our proposed PD-Net method on clinical data. The robustness analysis showed that the proposed PD-Net is capable of reconstructing images from projection data with slightly different distributions. However, the PD-Net may not work correctly with the increase of the data distribution changes at test time. Future work will also focus on the robustness analysis for the larger data distribution changes.

5. CONCLUSIONS

In this work, we have developed a PD-Net method for the unrolling approaches applied to CT reconstructions. To enhance the feature extraction capability of the unrolling approaches, the TV prior of the IR methods was unrolled to the CNNs in the PD-Net. The experimental results demonstrate that our proposed PD-Net framework is feasible for the implementation of CT reconstruction tasks. Compared with conventional primal-dual networks,[13] our proposed PD-Net method effectively preserves the structural and textural information in reference to the ground truth. To further improve the unrolling approaches, future work will focus on the unfolding of the other sophisticated prior terms that are widely used in IR methods.

ACKNOWLEDGMENTS

The authors would like to thank the institution, the Mayo Clinic, for providing the data (Low-Dose CT Image and Projection Data) used in this study. This work was supported in part by the State Key Project of Research and Development Plan under Grant 2022YFC2401600, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and development Programs in Jiangsu Province of China under Grants BE2021703 and BE2022768, in part by the Natural Science Foundation of Shanxi Province under Grant 201901D211246 and 202103021223308, and in part by the Research Project Supported by Shanxi Scholarship Council of China under Grant 2021-111.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

APPENDIX: THE FORMS OF THE SPATIAL GRADIENT AND THE SPATIAL DIVERGENCE

The explicit forms of the spatial gradient ∂ and the spatial divergence $\partial\cdot$ in two dimensions are defined herein as follows. We use $x\in I$ to refer to an $M\times N$ image and $x_{i,j}$ to represent the (i,j)th pixel of x. To specify the linear transform ∂, we introduce two gradient images $\partial_c x\in I$ and $\partial_r x\in I$:

$$\partial_c x_{i,j}=\begin{cases} x_{i+1,j}-x_{i,j} & i<M \\ -x_{i,j} & i=M \end{cases}, \tag{A.1}$$

$$\partial_r x_{i,j}=\begin{cases} x_{i,j+1}-x_{i,j} & j<N \\ -x_{i,j} & j=N \end{cases}. \tag{A.2}$$

Using these two definitions, the spatial gradient ∂ can be written as

ACCEPTED MANUSCRIPT / CLEAN COPY

$$\partial x = \begin{pmatrix} \partial_c x \\ \partial_r x \end{pmatrix}. \tag{A.3}$$

The operator div is defined as $-\partial^T$. With the Equation (A.3), the spatial divergence $\partial \cdot$ becomes

$$\partial \cdot (x) = div(\partial x) = -\partial^T \partial x,$$

$$= \left((\partial_c x_{i,j} - \partial_c x_{i-1,j}) + (\partial_r x_{i,j} - \partial_r x_{i,j-1}) \right), \ i \in [1, M] \ \text{ and } \ j \in [1, N]. \tag{A.4}$$

The elements outside the image border are set to zero: $\partial_c x_{0,j} = 0$, $\partial_c x_{i,0} = 0$, $\partial_r x_{0,j} = 0$, and $\partial_r x_{i,0} = 0$.

a) Author to whom correspondence should be addressed. Electronic mail: gzgtg@163.com (Z. Gui) and chenyang.list@seu.edu.cn (Y. Chen)

REFERENCES

[1] L. A. Shepp, and B. F. Logan, "The Fourier reconstruction of a head section," IEEE Trans. Nucl. Sci. **21**, 218–227 (1974).

[2] K. J. Batenburg, and L. Plantagie, "Fast approximation of algebraic reconstruction methods for tomography," IEEE Trans. Image Process. **21**, 3648–3658 (2012).

[3] E. Y. Sidky, and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol. **53**, 4777–4807 (2008).

[4] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," Phys. Med. Biol. **57**, 3065–3091 (2012).

[5] S. Niu, Y. Gao, Z. Bian, J. Huang, W. Chen, G. Yu, Z. Liang, and J. Ma, "Sparse-view X-ray CT reconstruction via total generalized variation regularization," Phys. Med. Biol. **59**, 2997–3017 (2014).

[6] H. Zhang, J. Huang, J. Ma, Z. Bian, Q. Feng, H. Lu, Z. Liang, and W. Chen, "Iterative reconstruction for X-ray computed tomography using prior-image induced nonlocal regularization," IEEE Trans. Biomed. Eng. **61**, 2367–2378 (2014).

[7] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," IEEE Trans. Med. Imag. **31**, 1682–1697 (2012).

[8] S. Ye, S. Ravishankar, Y. Long, and J. A. Fessler, "SPULTRA: Low-dose CT image reconstruction with joint statistical and learned image models," IEEE Trans. Med. Imag. **39**, 729–741 (2020).

[9] Z. Li, S. Ravishankar, Y. Long, and J. A. Fessler, "DECT-MULTRA: Dual-energy CT image decomposition with learned mixed material models and efficient clustering," IEEE Trans. Med. Imag. **39**, 1223–1234 (2020).

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Found. Trends Mach. Learn. **3**, 1–122 (2011).

[11] A. Beck, and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imag. Sci. **2**, 183–202 (2009).

[12] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," Nat. Mach. Intell. **2**, 737–748 (2020).

[13] J. Adler, and O. Öktem, "Learned primal-dual reconstruction," IEEE Trans. Med. Imag. **37**, 1322–1332 (2018).

[14] 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 Trans. Med. Imag. **37**, 1333–1347 (2018).

[15] W. Xia, Z. Lu, Y. Huang, Z. Shi, Y. Liu, H. Chen, Y. Chen, J. Zhou, and Y. Zhang, "MAGIC: manifold and graph integrative convolutional network for low-dose CT reconstruction," IEEE Trans. Med. Imag. 40, 3459–3472 (2021).

[16] J. Liu, Y. Sun, W. Gan, X. Xu, B. Wohlberg, and U. S. Kamilov, "SGD-Net: efficient model-based deep learning with theoretical guarantees," IEEE Trans. Comput. Imag. **7**, 598–610 (2021).

[17] J. Wang, L. Zeng, C. Wang, and Y. Guo, "ADMM-based deep reconstruction for limited-angle CT," Phys. Med. Biol. **64**,

ACCEPTED MANUSCRIPT / CLEAN COPY

115011 (2019).

[18] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, "Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction," IEEE Trans. Med. Imag. **38**, 371–382 (2019).

[19] J. Xiang, Y. Dong, and Y. Yang, "FISTA-Net: learning a fast iterative shrinkage thresholding network for inverse problems in imaging," IEEE Trans. Med. Imag. **40**, 1329–1339 (2021).

[20] Y. Gao, Z. Liang, W. Moore, H. Zhang, M. Pomeroy, J. Ferretti, T. Bilfinger, J. Ma, and H. Lu, "A feasibility study of extracting tissue textures from a previous full-dose CT database as prior knowledge for Bayesian reconstruction of current low-dose CT images," IEEE Trans. Med. Imag. **38**, 1981–1992 (2019).

[21] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," Nature **555**, 487–492 (2018).

[22] F. Jiao, Z. Gui, K. Li, H. Shangguan, Y. Wang, Y. Liu, and P. Zhang, "A dual-domain CNN-based network for CT reconstruction," IEEE Access **9**, 71091–71103 (2021).

[23] A. Zheng, H. Gao, L. Zhang, and Y. Xing, "A dual-domain deep learning-based reconstruction method for fully 3D sparse data helical CT," Phys. Med. Biol. **65**, 245030 (2020).

[24] Y. Zhang, D. Hu, Q. Zhao, G. Quan, J. Liu, Q. Liu, Y. Zhang, G. Coatrieux, Y. Chen, H. Yu, "CLEAR: Comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose CT imaging," IEEE Trans. Med. Imag. **40**, 3089-3101 (2021).

[25] X. Yin, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, Y.Chen, H. Shu, L. Luo, J.-L. Coatrieux, "Domain progressive 3D residual convolution network to improve low dose CT imaging," IEEE Trans. Med. Imag. **38**, 2903-2913 (2019).

[26] D. Hu, Y. Zhang, J. Liu, S. Luo, Y. Chen, "DIOR: Deep iterative optimization-based residual-learning for limited-angle CT reconstruction," IEEE Trans. Med. Imag. **41**, 1778-1790 (2022).

[27] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable efficient deep learning for signal and image processing," IEEE Signal Process. Mag. **38**, 18–44 (2021).

[28] J. Zhang, and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1828-1837 (2018).

[29] H. Zhang, B. Liu, H. Yu, and B. Dong, "MetaInv-Net: Meta inversion network for sparse view CT image reconstruction," IEEE Trans. Med. Imag. **40**, 621-634 (2021).

[30] I. Y. Chun, Z. Huang, H. Lim, and J. Fessler, "Momentum-Net: Fast and convergent iterative neural network for inverse problems," IEEE Trans. Pattern Anal. Mach. Intell. (2020).

[31] X. Han, J. Bian, D. Eaker, T. Kline, E. Sidky, E. Ritman, and X. Pan, "Algorithm-enabled low-dose micro-CT imaging," IEEE Trans. Med. Imag. **30**, 606–620 (2011).

[32] A. Chambolle, and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," J. Math. Imag. Vis. **40**, 120–145 (2011).

[33] G. L. Zeng, and G. T. Gullberg, "Unmatched projector/backprojector pairs in an iterative reconstruction algorithm," IEEE Trans. Med. Imag. **19**, 548–555 (2000).

[34] Piglet Dataset. [Online]. Available: http://homepage.usask.ca/?xiy525/publication/sagan/

[35] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," NIPS 1–4 (2017).

[36] D. P. Kingma, and J. A. Ba, "Adam: A method for stochastic optimization," The 3rd International Conference on Learning Representations (ICLR), 1-15 (2015).

[37] C. Wang, K. Shang, H. Zhang, Q. Li, Y. Hui, and S. K. Zhou, "DuDoTrans: Dual-domain transformer provides more attention for sinogram restoration in sparse-view CT reconstruction," arXiv:2111.10790 (2021).

[38] T. Su, Z. Cui, J. Yang, Y. Zhang, J. Liu, J. Zhu, X. Gao, S. Fang, H. Zheng, Y. Ge, and D. Liang, "Generalized deep iterative reconstruction for sparse-view CT imaging." Phys. Med. Biol. **67**, 025005 (2022).

ACCEPTED MANUSCRIPT / CLEAN COPY

[39] A. Biguri, M. Dosanjh, S. Hancock, and M. Soleimani, "TIGRE: A MATLAB-GPU toolbox for CBCT image reconstruction," Biomed. Phys. Eng. Express **2**, 055010 (2016).

[40] C. You, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, Y. Zhang, W. Cong, and G. Wang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," IEEE Access **6**, 41839–41855 (2018).

[41] V. Antun, F. Renna, C. Poon, B. Adcock, A. C. Hansen, "On instabilities of deep learning in image reconstruction-Does AI come at a cost?" arXiv:1902.05300 (2019).

[42] S. Bhadra, V. A. Kelkar, F. J. Brooks, and M. A. Anastasio, "On hallucinations in tomographic image reconstruction," IEEE Trans. Med. Imag. **40**, 3249–3260 (2021).

[43] D. Gilton, G. Ongie, and R. Willett, "Model adaptation for inverse problems in imaging," IEEE Trans. Med. Imag. **7**, 661–674 (2021).

[44] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D. **60**, 259–268 (1992).

[45] J. Adler, and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," Inverse Problems **33**, 124007 (2017).

ACCEPTED MANUSCRIPT / CLEAN COPY