Limited View Tomographic Reconstruction Using a Cascaded Residual Dense Spatial-Channel Attention Network With Projection Data Fidelity Layer

Bo Zhou, *Student Member, IEEE*, S. Kevin Zhou, *Fellow, IEEE*, James S. Duncan, *Life Fellow, IEEE*, and Chi Liu, *Senior Member, IEEE*

***Abstract*— Limited view tomographic reconstruction aims to reconstruct a tomographic image from a limited number of projection views arising from sparse view or limited angle acquisitions that reduce radiation dose or shorten scanning time. However, such a reconstruction suffers from severe artifacts due to the incompleteness of sinogram. To derive quality reconstruction, previous methods use UNet-like neural architectures to directly predict the full view reconstruction from limited view data; but these methods leave the deep network architecture issue largely intact and cannot guarantee the consistency between the sinogram of the reconstructed image and the acquired sinogram, leading to a non-ideal reconstruction. In this work, we propose a cascaded residual dense spatial-channel attention network consisting of residual dense spatial-channel attention networks and projection data fidelity layers. We evaluate our methods on two datasets. Our experimental results on AAPM Low Dose CT Grand Challenge datasets demonstrate that our algorithm achieves a consistent and substantial improvement over the existing neural network methods on both limited angle reconstruction and sparse view reconstruction. In addition, our experimental results on Deep Lesion datasets demonstrate that our method is able to generate high-quality reconstruction for 8 major lesion types.**

***Index Terms*— Tomographic reconstruction, cascaded network, projection data fidelity layer, RedSCAN, limited angle, sparse view.**

I. Introduction

TOMOGRAPHY imaging is a non-invasive projection-based imaging technique that visualizes an object's internal structures and hence finds wide applications in healthcare, security, and industrial settings [1]–[3]. In healthcare, tomography imaging techniques such as medical Computed Tomography (CT) based on x-ray projections, Positron Emission Tomography (PET), and Single-photon Emission Computed Tomography (SPECT) based on gamma-ray projections are indispensable imaging modalities for disease diagnosis and treatment planning. In the traditional CT setting, one assumes access to the measurements that are collected from a full range of view angles of an object. To reduce radiation dose and speed up acquisition, recently it is of increasing interest to develop methods that can recover images when a portion of the projection views is missing, namely limited view tomographic reconstruction. There are two notable sub-problems: limited angle (LA) reconstruction, i.e., when $\alpha \in [0, \alpha_{max}]$ with $\alpha_{max} < 180^\circ$ for equivalent parallel beam geometry, and sparse view (SV) reconstruction with a view interval larger than normal. Both LA and SV acquisitions can efficiently reduce radiation dose. Using LA acquisition, the scan time can also be drastically reduced by restricting the physical movement of the scan arc. Note that fast acquisition or high temporal resolution is paramount; even a slightly longer scan time can lead to appreciable motion blur and artifact in the image [4], [5].

There are two major factors, namely reconstruction quality and speed that need to be properly considered in designing a tomographic reconstruction algorithm. Currently, Filtered Back Projection (FBP) is widely used as the standard algorithm as it can reconstruct a high-quality image with a fast speed, following an analytical solution. However, FBP assumes the access to the measurements that are collected from a full range of views of an object. Reconstruction using FBP in both LA and SV conditions are highly ill-posed, yielding non-ideal image quality with severe artifacts and high noise. Previous algorithms for tomographic reconstruction under limited view conditions can be classified into two general categories: model-based iterative reconstruction (MBIR) and deep learning based reconstruction (DLR). MBIR can generate images with high quality by minimizing the predefined image domain regularizers and the sampled sinogram inconsistency in an iterative fashion. Common choices of the regularizer

Manuscript received January 27, 2021; revised February 25, 2021; accepted March 9, 2021. Date of publication March 17, 2021; date of current version June 30, 2021. This work was supported by the National Institutes of Health (NIH) under Grant R01EB025468. The work of Bo Zhou was supported by the Biomedical Engineering Ph.D. fellowship from Yale University. *(Corresponding authors: Bo Zhou; Chi Liu.)*

Bo Zhou is with the Department of Biomedical Engineering, Yale University, New Haven, CT 06511 USA (e-mail: bo.zhou@yale.edu).

S. Kevin Zhou is with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China.

James S. Duncan and Chi Liu are with the Department of Biomedical Engineering, Yale University, New Haven, CT 06511 USA, and also with the Department of Radiology and Biomedical Imaging, Yale University, New Haven, CT 06511 USA (e-mail: chi.liu@yale.edu).

Digital Object Identifier 10.1109/TMI.2021.3066318

include total variation [6], dictionary learning [7], and non-local patches [8]. However, MBIR methods are computationally heavy and time-consuming since they rely on repetitive forward- and back-projections. Moreover, using regularization solely based on prior assumptions requires careful hyper-parameter tuning and tends to bias the reconstruction results, especially when under-sampling rate is high.

Recently, deep learning techniques, such as convolutional neural networks (CNNs), have been widely adapted in tomography and demonstrated promising reconstruction performance [9]. Combining MBIR with deep learning, Gupta *et al.* [10] and We *et al.* [11] first proposed to model regularizer in MBIR frameworks with CNNs and Autoencoders. Adler et el. [12] unfolded the optimization procedure of MBIR to an N-stage network to balance the tradeoff between reconstruction and speed. Although improved over traditional MBIR methods, they still suffer from high computational cost with iterative procedures. As an alternative, DLR is often formulated as image post-processing. Jin *et al.* [13] and Chen *et al.* [14] proposed to use UNet [15] and Residual UNet to post-process the noise/artifacts in the sparse-view CT. In [16] and [17], adversarial loss and perceptual loss were used to reinforce the network's learning. Later, Zhang *et al.* [18] and Han *et al.* [19] proposed to incorporate dense block and wavelet decomposition into UNet for more robust feature learning for reconstruction. Direct sinogram inversion and sinogram completion strategies were also proposed. Lee *et al.* [20] found that synthesizing complete sinogram from sparse view sinogram and then using FBP can also reconstruct high-quality image. Although these methods can be easily applied to raw sinograms or corresponding FBP reconstructed images with relatively low computational cost and low design complexities, they either only applied on image domain that remove artifacts in already reconstructed image or synthesizing complete sinogram from sparse one, and cannot guarantee the sampled sinogram data are preserved. Note that the sampled sinogram data are the original sources that should be kept as identical as possible before and after reconstruction to ensure the high fidelity of reconstructed content. There are also recent ideas of replacing the already-sampled sinogram to the predicted sinogram during the test stage. Anurudh *et al.* [1] proposed to first use a sinogram-to-image auto encoder to predict an initial reconstruction. Then, during the test stage, the reconstruction's sinogram is partly replaced by the already-sampled sinogram to generate a final reconstruction. However, their method does not guarantee the continuity between the already-sampled sinogram and the predicted sinogram, which may further degrade the final reconstruction, and their method is limited to parallel-beam geometry. Similarly, Huang *et al.* [21] proposed to first use UNet [15] to predict an initial reconstruction. Then, during the test stage, the initial reconstruction is utilized in a TV reconstruction to help the projection data fidelity constraint of unmeasured projection data. However, the final reconstruction quality relies on a high-quality initial reconstruction from UNet's prediction. In addition, the projection data fidelity constraint of unmeasured projection data is not incorporated in the network design and used only in the separated test stage. On a different note, the network design issue is highly under-explored as a research topic and still limited to UNet-based or auto-encoder architectures [13], [14], [16], [17], [19], [20], [22]. In addition, none of previous works have evaluated the performance under both LA and SV scenarios, and reconstruction evaluation on CT scan with pathological finding are barely performed. While a k-space data consistency layer for MRI fast reconstruction is proposed in [23], [24], projection data consistency layer has not been systematically studied in tomographic reconstruction.

To tackle these limitations, we propose a **Cas**caded **Re**sidual **D**ense **S**patial-**C**hannel **A**ttention **N**etwork (CasRedSCAN) for tomographic reconstruction under limited view conditions. Our CasRedSCAN consisting of Residual Dense Spatial-Channel Attention Network (RedSCAN) and Projection Data Fidelity Layer (PDFL) closely resembles the iterative process in MBIR methods, which allows end-to-end optimization of the reconstruction. Specifically, RedSCAN is the backbone network that is used in each cascade block for de-aliasing the input image. PDFL is concatenated to the RedSCAN output to ensure the prediction's projection data fidelity while allowing gradient back-propagation. Experiments on limited angle and spare view scans using AAPM Low Dose CT Grand Challenge [25] and DeepLesion dataset [26] demonstrate that our CasRedSCAN can provide high-quality limited view tomographic reconstructions.

II. Problem Formulation

Let $I \in \mathbb{C}^N$ represent a 2D tomography image with a size of $N = N_x N_y$, and $Q \in \mathbb{C}^M$ represent its full-view sinogram with M projection views. Our problem is to reconstruct I from $Q_u \in \mathbb{C}^{M_u}$ ($M_u \ll M$), where Q_u is the under-sampled sinogram of limited views. Here, sinogram data is only measured for lines corresponding to a subset $\mathbf{\Omega} \subset \mathbb{A} \triangleq \{1, \cdots, M\}$, where $\mathbb{A}$ is the full projection set. Denoting $\mathcal{G}$ and $\mathcal{G}_u$ as the full-view and limited-view discretized forward projection operators, the full-view sinogram Q and limited-view sinogram Q_u are obtained via $Q = \mathcal{G}I$ and $Q_u = \mathcal{G}_u I$, respectively. While FBP provides stable numerical implementation of pseudo-inverse for Q, applying FBP to Q_u in the limited view conditions yields reconstructed I_u with severe artifacts.

Previous works of MBIR propose to solve I by

$$\min_I [\mathcal{T}(I) + \lambda ||\mathcal{G}_u I - Q_u||_n^n], \tag{1}$$

where $\mathcal{T}$ is the regularizer and $||\cdot||_n^n$ is the projection data fidelity constraint [6], [27]. Previous deep learning-based, post-processing methods utilize deep networks, denoted as $\mathcal{P}$ with parameters θ, to estimate the full-view reconstructed image $\mathcal{P}(I_u; \theta)$ by training $\mathcal{P}$ on (I_u, I_{gt}) pairs, where I_{gt} is the full-view reconstruction ground truth. However, these methods only consider a subsequent regularization of the initial solution I_u similar to the functionality of $\mathcal{T}(\cdot)$ in MBIR, and omit the projection data fidelity constraint of $||\mathcal{G}_u I - Q_u||_n^n$. One should force reconstruction I to be well-approximated by the CNN reconstruction and ensure the

consistency of acquired data in the projection domain by:

$$\min_{I}[||I - \mathcal{P}(I_u;\theta)||_2^2 + \lambda||\mathcal{G}_u I - Q_u||_2^2], \tag{2}$$

However, it is not feasible to directly optimize the above equation since the deep network reconstruction and the projection data fidelity terms are independent. Specifically, as deep network $\mathcal{P}$ only operates in the image domain, $\mathcal{P}$ is trained to reconstruct the full-view image without prior knowledge of the already acquired data in the projection domain. Similar to the MRI k-space data fidelity [23], given a portion of already acquired projection data from limited-view acquisitions, the deep network should be discouraged from changing the already acquired projection data up to the level of acquisition noise. Incorporating the projection data fidelity in the network design could potentially better preserve the image content and lead to a better reconstruction. In this work, we propose a projection data fidelity layer (PDFL) embedded in a cascade network for full-view reconstruction. With PDFL in our cascade network, the reconstruction output from our network is now conditioned on both network parameter θ and limited-view projection data Ω:

$$I_{rec} = \mathcal{P}(I_u;\theta,\Omega) \tag{3}$$

Then, given the training data pairs of (I_u, I_{gt}), we can train our network by minimizing the L2 loss function:

$$\mathcal{L} = ||\mathcal{P}(I_u;\theta,\Omega) - I_{gt}||_2^2 \tag{4}$$

Details of our PDFL and cascade network are explained in Section III and Section IV, respectively.

III. Projection Data Fidelity Layer

Let $\mathcal{G}$ and $\mathcal{G}_{fbp}$ be forward projection (FP) layer and filtered back-projection (FBP) layer, respectively. The projection data of the image reconstruction by a deep network can be formulated as: $S_{cnn} = \mathcal{G}I_{cnn} = \mathcal{G}\mathcal{P}(I_u;\theta)$, where $S_{cnn}(i)$ is the i-th projection data entry. Similarly, we denote the already acquired projection data as S_u, where S_u has identical size to S_{cnn} and the i-th projection data entry $S_u(i)$ is all zeros when $i \notin \Omega$. Then, we can write a closed-form solution for the second term in Eq.(2) as:

$$S_{rec}(i) = \begin{cases} \frac{\lambda S_{cnn}(i) + S_u(i)}{\lambda + 1} & \text{if } i \in \Omega \\ S_{cnn}(i) & \text{if } i \notin \Omega \end{cases} \tag{5}$$

where S_{rec} is the reconstructed sinogram, which is updated by the projection data fidelity. Then, the image can be reconstructed via filtered back projection, that is, $I_{rec} = \mathcal{G}_{fbp}S_{rec}$. To elaborate, when the i-th projection data is not acquired, we directly estimates the i-th projection data from the projection data of the deep network's output. Otherwise, the i-th projection data is a linear combination of the acquired projection data and projection data of the deep network's output, regularized by noise level parameter λ. Assuming noiseless sinogram acquisition, i.e. $\lambda = 0$, we simply replaces the i-th predicted projection data by the acquired projection data.

A. Forward Projection Layer

Our FP layer $\mathcal{G}$ is a differentiable layer implemented with fan-beam geometry, allowing gradient back-propagation while projecting the image into sinogram. In this work, we consider fan-beam geometry with arc detector [28]. Assuming the distance between x-ray source and the gantry rotation center as D, the forward pass of the FP layer can be written as:

$$S_{fan}(\gamma,\beta) = \iint_{\mathbb{R}^2} I(x,y)\delta[D\sin(\gamma) - x\sin(\beta-\gamma) - y\cos(\beta-\gamma)]dxdy \tag{6}$$

where a fan-beam sinogram $S_{fan}(\gamma,\beta)$ is generated. β means the detector rotation angle, and γ means the angle between central projection line and detector projection line. In the backward path of $\mathcal{G}$, the loss in the sinogram domain should be aggregated and back-projected to the image domain. Thus, we define the derivative of $\mathcal{G}$ with respect to the input image I as the filtered back-projection operation $\mathcal{G}_{fbp}$ (discussed in Section III-B).

B. Filtered Back-Projection Layer

Our FBP layer $\mathcal{G}_{fbp}$ is also a differentiable layer implemented with fan-beam geometry, allowing gradient back-propagation while reconstructing the image from sinogram. Similar to above, assuming the distance between x-ray source and the gantry rotation center as D, we have a fan-beam sinogram $S_{fan}(\gamma,\beta)$, where β is the detector rotation angle and γ is the angle between central projection line and detector projection line. Our FBP layer consists of three modules: i) parallel-beam conversion module, ii) filtering module, and iii) back-projection module.

Parallel-beam conversion module converts the fan-beam sinogram $S_{fan}(\gamma,\beta)$ to parallel-beam sinogram $S_{para}(\rho,\alpha)$ via:

$$\begin{cases} \alpha = \beta + \gamma, \\ \rho = D\sin\gamma. \end{cases} \tag{7}$$

where the change of variable is implemented by grid sampling[1] in (ρ,α), which allows gradient back-propagation.

Filtering module applies the filtering to the converted sinogram S_{para} in the Fourier domain:

$$\hat{S} = T_\rho^{-1}\{|\omega| \cdot T_\rho\{S_{para}(\rho,\alpha)\}\} \tag{8}$$

where T_ρ and T_ρ^{-1} are the discrete Fourier transform and inverse discrete Fourier transform along the detector dimension ρ, respectively.[2] ω is the window function and we used Ram-Lak in this work.

Back-projection module back-projects the filtered parallel-beam sinogram $\hat{S}$ to the image domain for every projection angle α via:

$$I(x,y) = \int_0^{2\pi} \hat{S}(x\cos\alpha + y\sin\alpha,\alpha)d\alpha \approx \Delta\alpha\sum_i \hat{S}(x\cos\alpha_i + y\sin\alpha_i,\alpha_i) \tag{9}$$

[1] implement with Pytorch using torch.nn.functional.grid_sample
[2] implement with Pytorch using torch.fft and torch.ifft

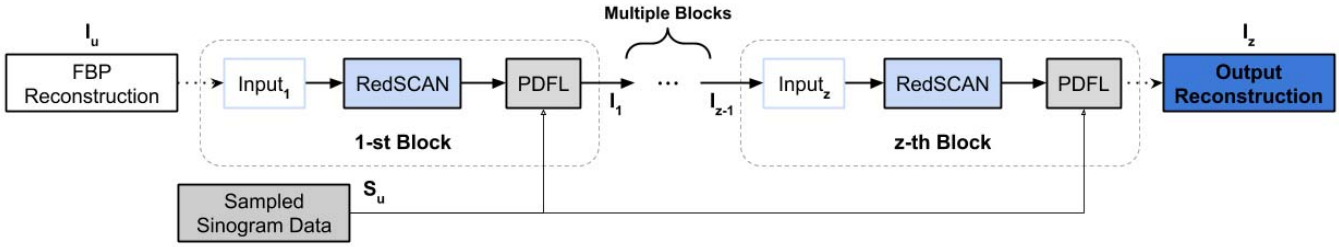

Fig. 1. The architecture of our CasRedSCAN. Each block consists of a RedSCAN (blue) and a PDFL (gray).

where we parallelize the back-projection operation,[3] such that the reconstruction can be efficiently computed. In the backward path of $\mathcal{G}_{fbp}$, the loss in the image domain should be aggregated and projected to the sinogram domain. Thus, we define the derivative of $\mathcal{G}_{fbp}$ with respect to the input sinogram S_{fan} as the forward projection operation $\mathcal{G}$ (discussed in Section III-A).

Here, we use pixel-driven algorithm for our implementation of forward projection and back-projection [29].

C. Forward and Backward Pass

Our Projection Data Fidelity Layer (PDFL) consists of three operations: i) forward project $\mathcal{G}$, ii) the projection data fidelity of Eq.(5), and iii) the FBP layer $\mathcal{G}_{fbp}$. The projection data fidelity of Eq.(5) can be formulated in matrix form as:

$$\mathcal{D}S_{cnn} + \frac{1}{\lambda+1}S_u \tag{10}$$

where $\mathcal{D} = diag(e_1, e_2, \cdots, e_M)$ with:

$$e_M = \begin{cases} \frac{\lambda}{1+\lambda}, & \text{when } i \in \Omega, \\ 1, & \text{when } i \notin \Omega \end{cases} \tag{11}$$

Then, our PDFL combines the three operations discussed above. Specifically, the forward pass of PDFL can be writtern as:

$$\begin{aligned} \mathcal{P}_{PDFL}(I_{cnn}, S_u) &= \mathcal{G}_{fbp}(\mathcal{D}\mathcal{G}I_{cnn} + \frac{1}{\lambda+1}S_u) \\ &= \mathcal{G}_{fbp}\mathcal{D}\mathcal{G}I_{cnn} + \frac{1}{\lambda+1}\mathcal{G}_{fbp}S_u \end{aligned} \tag{12}$$

where I_{cnn} is the image predicted from an image-domain deep network and is the input of our PDFL. The output of PDFL is an image with projection data fidelity from limited-view projection data S_u. Assuming low noise level, we set $\lambda = 0.001$ (analyzed in Section V-C.4). Given the forward pass of Eq.(12), the gradient of the PDFL with respect to the input I_{cnn} can thus be written as:

$$\frac{\partial \mathcal{P}_{PDFL}}{\partial I_{cnn}} = \mathcal{G}_{fbp}\mathcal{D}\mathcal{G} \tag{13}$$

which is defined for our PDFL's backward pass. There is no learnable parameter in our PDFL.

IV. Cascaded Residual Dense Spatial-Channel Attention Network

Previous MBIR methods solve the optimization problem in Eq.(1) for CT reconstruction by switching the de-aliasing step and the projection data fidelity step back and forth until convergence. However, in many previous deep-learning based reconstruction methods [13], [18], [19], they use single-step deep networks for de-aliasing and reconstruction. Unfortunately, a trained single-step network cannot be used for iterative de-aliasing, since iteratively applying single-step network de-aliasing does not guarantee to converge to a reasonable reconstruction. Moreover, single-step deep networks with limited de-alising capability are prone to issues, such as over-fitting. Therefore, it is desirable to have a network structure that is able to iteratively de-alias the image using a deep network with sufficient de-aliasing capability, while preserving the projection data fidelity. Here, we propose a cascaded network structure, called CasRedSCAN, with basic units of Residual Dense Spatial-Channel Attention Network (RedSCAN) and PDFL.

Similar to the process of MBIR that alternates between the de-aliasing step and the projection data fidelity step, our CasRedSCAN also alternates between the RedSCAN and PDFL, as illustrated in Figure 1. With the initial FBP reconstruction image inputted into the first RedSCAN, the de-aliasing output is fed into the first PDFL. Then, the PDFL output is fed into the second RedSCAN+PDFL block. The same procedure is iterated a fix number of times for a final reconstruction output I_z. The loss function can thus be formulated as:

$$\mathcal{L} = ||\mathcal{P}_{CasRedSCAN}(I_u; \theta, S_u) - I_{gt}||_2^2, \tag{14}$$

where I_u is initial FBP reconstruction. θ is the RedSCAN network parameters. S_u is the limited-view sinogram data. I_{gt} is the ground truth reconstruction from full-view sinogram data. The algorithm is summarized in Algorithm 1. In our implementation, all the RedSCAN shared the same network parameters in CasRedSCAN, thus maintaining nearly the same model size as compared to the single-step RedSCAN.

A. Residual Dense Spatial-Channel Attention Network

Our RedSCAN consists of three key components, including initial feature extraction (IFE) using two 3×3 convolution layers, multiple Residual Dense Spatial-Channel Attention Block (RedSCAB) followed by global feature fusion, and global residual learning. The network architecture is demonstrated in Figure 2.

Let $\mathcal{P}_{IFE_1}$ and $\mathcal{P}_{IFE_2}$ be the first and second convolutional operations in IFE, we first extract $F_{-1} = \mathcal{P}_{IFE_1}(I_u)$ for global residual learning, and $F_0 = \mathcal{P}_{IFE_2}(F_{-1})$ for feeding into

[3]implement with Pytorch's Custom C++ and CUDA extensions

Algorithm 1 Cascaded Residual Dense Spatial-Channel Attention Network

Input: $\Pi = \{(I_{u_i}, S_{u_i}, I_{gt_i})\}$, for $i \in \{1, \ldots, N\}$; ▷ training dataset
Initialize: $\theta_{init} :\leftarrow \mathcal{N}(0, 1)$; ▷ initialize weights
for $iter = 1$ *to* k **do**
  $T = \{(I_u, S_u, I_{gt})\} \leftarrow \Pi$; ▷ get training batch
  for $j = 1$ *to* z **do**
    if $j = 1$ **then**
      $input_j = I_u$;
    else
      $input_j = output_{j-1}$;
    $output_j \leftarrow \mathcal{P}_{RedSCAN}(input_j, \theta)$; ▷ unregularized output
    $output_j \leftarrow \mathcal{P}_{PDFL}(output_j, S_u)$; ▷ regularized output
  $I_z = output_z$;
  $\theta :\leftarrow min[\mathcal{L}(I_z, I_{gt})]$; ▷ optimization
Output θ ; ▷ return upon convergence

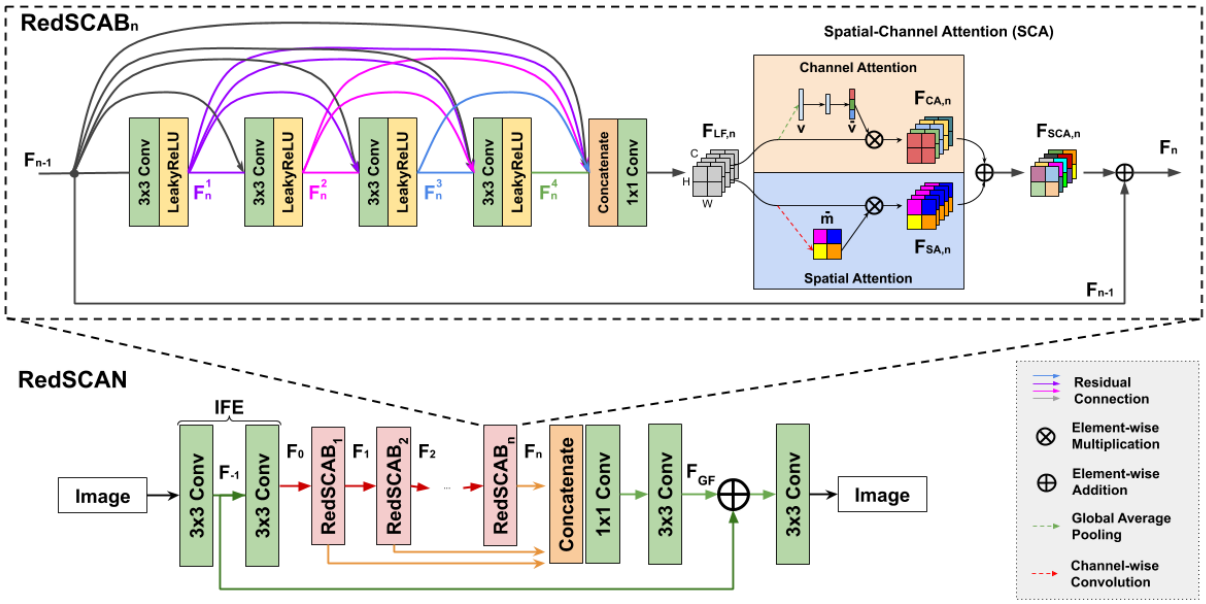

Fig. 2. The architecture of our residual dense spatial-channel attention network (RedSCAN), which are used in both the recurrent image reconstruction blocks in Figure 1.

RedSCAB. Assuming we have n RedSCABs, the n-th output F_n can thus be written as:

$$F_n = \mathcal{P}_{RedSCAB_n}(F_{n-1}), \tag{15}$$

where $\mathcal{P}_{RedSCAB_n}$ represents the n-th RedSCAB operation ($n \geq 1$). Given the extracted local features from a set of RedSCAB, we apply our global feature fusion (GFF) to extract the global feature:

$$F_{GF} = \mathcal{P}_{GFF}(\{F_1, F_2, \ldots, F_n\}), \tag{16}$$

where $\{\}$ means concatenation along feature channel and our global feature fusion function $\mathcal{P}_{GFF}$ consists of a 1×1 and 3×3 convolution layers to fuse the extracted local features from different levels of RedSCAB. The GFF output is used as input for our global residual learning:

$$I = \mathcal{P}_{final}(F_{GF} + F_{-1}), \tag{17}$$

The element-wise addition of global feature and initial feature are fed into our final 3×3 convolution layer for unregularized output. In our experiment, we set the size of IFE feature channel to 32.

Residual Dense Spatial-Channel Block contains four densely connected convolution layers, local feature fusion, local residual connection, and spatial-channel attention. In the n-th RedSCAB, the t-th convolution output is:

$$F_n^t = \mathcal{H}_n^t\{F_{n-1}, F_n^1, \ldots, F_n^{t-1}\}, \tag{18}$$

where $\mathcal{H}_n^t$ denotes the t-th convolution followed by Leaky-ReLU in the n-th RedSCAB, $\{\}$ means concatenation along feature channel, and the number of convolution $t \leq 4$. Then, we apply our local feature fusion (LFF), a 1×1 convolution layer, to fuse the output from the last RedSCAB and all convolution layers in current RedSCAB. Thus, the LFF output

TABLE I

QUANTITATIVE COMPARISON OF **LIMITED ANGLE RECONSTRUCTION** AND **SPARSE VIEW RECONSTRUCTION** RESULTS UNDER DIFFERENT LIMITED ANGLE AND SPARSE VIEW SETTINGS USING PSNR (DB), SSIM, AND RMSE ON AAPM DATASET. BEST RESULTS ARE MARKED IN RED

PSNR/SSIM/RMSE	Limited Angle Reconstruction			Sparse View Reconstruction			Time (ms)	Number of Parameters
	90°	120°	150°	1/6	1/4	1/2		
FBP	17.76/.555/388.7	21.76/.693/246.3	26.81/.782/138.5	26.85/.513/137.8	30.19/.648/94.1	39.02/.896/33.9	2.3	-
TV [6]	22.56/.762/230.3	30.67/.875/132.1	33.74/.898/69.7	30.91/.895/70.3	34.13/.911/41.3	35.83/.934/18.7	3096.3	-
FBPNet [13]	28.66/.887/111.7	35.14/.959/53.6	40.80/.982/28.2	34.73/.933/55.2	39.26/.962/32.8	46.61/.986/14.1	7.2	30M
DDNet [18]	31.03/.921/85.6	36.49/.965/45.9	41.45/.988/25.9	35.07/.933/53.3	39.60/.963/31.7	45.70/.984/15.6	5.1	0.56M
FUNet [19]	30.27/.903/93.5	35.87/.960/48.3	41.01/.985/26.7	35.01/.933/54.8	39.52/.962/32.0	46.64/.986/14.0	5.6	36M
CTNet [1]	29.05/.889/106.8	35.33/.962/52.4	40.97/.984/27.5	35.81/.936/48.8	39.80/.963/30.9	46.73/.987/13.8	10.3	31M
DCAR [21]	30.25/.900/94.4	37.94/.970/39.1	43.87/.989/21.7	36.99/.948/42.6	41.18/.973/26.3	47.01/.989/11.5	3187.6	30M
CasRedSCAN	34.74/.952/56.42	41.48/.983/26.1	48.23/.995/11.8	43.13/.979/21.1	46.43/.989/14.4	51.66/.996/7.8	148.2	0.51M

can be expressed as:

$$F_{LF,n} = \mathcal{P}_{LFF,n}(\{F_{n-1}, F_n^1, F_n^2, F_n^3, F_n^4\}), \tag{19}$$

where $\mathcal{P}_{LFF,n}$ denotes the LFF operation. Then, it is fed into our Spatial-Channel Attention (SCA) module with two branches to re-weigh channel-wise features and spatial-wise features, as illustrated in Figure 2. The channel attention output $F_{CA,n}$ and spatial attention output $F_{SA,n}$ are fused together via $F_{SCA,n} = F_{CA,n} + F_{SA,n}$. Finally, we apply the local residual learning to SCA output by adding the residual connection from RedSCAB input, generating the n-th RedSCAB output:

$$F_n = F_{SCA,n} + F_{n-1} \tag{20}$$

In our experiment, we set the number of RedSCAB to 5.

Spatial-Channel Attention contains two Squeeze-and-Excitation branches for Channel Attention (CA) and Spatial Attention (SA), respectively [30], [31]. Traditional CNNs treat channel-wise features and spatial-wise features equally. However, in an image reconstruction task, it is desirable to have the network focus more on informative features by acknowledging both the channel-wise feature interdependence and the spatial-wise contextual interdependence. The CA and SA structures are illustrated in orange and blue boxes in Figure 2, respectively.

For CA, similar to [30], we spatial-wise squeeze the input feature map using global average pooling, where the feature map is formulated as $F = [f_1, f_2, \ldots, f_C]$ here with $f_n \in \mathbb{R}^{H\times W}$ denoting the individual feature channel. We flatten the global average pooling output, generating $v \in \mathbb{R}^C$ with its z-th element:

$$v_z = \frac{1}{H \times W} \sum_i^H \sum_j^W f_z(i, j) \tag{21}$$

where vector v embeds the spatial-wise global information. Then, v is fed into two fully connected layers with weights of $w_1 \in \mathbb{R}^{\frac{C}{2}\times C}$ and $w_2 \in \mathbb{R}^{C\times\frac{C}{2}}$, producing the channel-wise calibration vector:

$$\hat{v} = \sigma(w_2\eta(w_1 v)) \tag{22}$$

where η and σ are the ReLU and Sigmoid activation function, respectively. The calibration vector is applied to the input feature map using channel-wise multiplication:

$$\hat{F}_{CA} = [f_1\hat{v}_1, f_2\hat{v}_2, \ldots, f_C\hat{v}_C] \tag{23}$$

where $\hat{v}_i$ indicates the importance of the i-th feature channel and lies in [0, 1]. With CA embedded into our network, the calibration vector adaptively learns to emphasize the important feature channels while plays down the others.

In SA, we formulate our feature map as $F = [f^{1,1}, \ldots, f^{i,j}, \ldots, f^{H,W}]$, where $f^{i,j} \in \mathbb{R}^C$ indicates the feature at spatial location (i, j) with $i \in \{1, \ldots, H\}$ and $j \in \{1, \ldots, W\}$. We channel-wise squeeze the input feature map using a convolutional kernel with weights of $w_3 \in \mathbb{R}^{1\times1\times C\times1}$, generating a volume tensor $m = w_3 \circledast F$ with $m \in \mathbb{R}^{H\times W}$. Each $f^{i,j}$ is a linear combination of all feature channels at spatial location (i, j). Then, the spatial-wise calibration volume that lies in [0, 1] can be written as:

$$\hat{m} = \sigma(m) = \sigma(w_3 \circledast F) \tag{24}$$

where σ is the sigmoid activation function. Applying the calibration volume to the input feature map, we have:

$$\hat{F}_{SA} = [f^{1,1}\hat{m}^{1,1}, \ldots, f^{i,j}\hat{m}^{i,j}, \ldots, f^{H,W}\hat{m}^{H,W}] \tag{25}$$

where the calibration parameter $\hat{m}^{i,j}$ provides the relative importance of a spatial information of a given feature map. Similarly, with SA embedded into our network, the calibration volume learns to stress the most important spatial locations while ignores the irrelevant ones.

Finally, channel-wise calibration and spatial-wise calibration are combined via element-wise addition operation $F_{SCA} = \hat{F}_{SA} + \hat{F}_{CA}$. With the two branch fusion, features at (i, j, c) possess high activation only when they receive high activation from both SA and CA. Our SCA encourages the networks to re-calibrate the feature map such that more accurate and relevant feature maps can be learned.

V. EXPERIMENTS AND RESULTS

A. Data Preparation and Training

We used two large-scale dataset for our experiments. In our first dataset, we collected 10 whole body CT scans from the AAPM Low Dose CT Grand Challenge [25]. Each 3D scan contains 318 ∼ 856 2D slices covering a range of anatomical regions from chest to abdomen to pelvis. From the AAPM dataset, the 2D dataset of 3397 images without lesion are split patient-wise into 1834 training images, 428 validation images, and 1135 test images. To evaluate the reconstruction

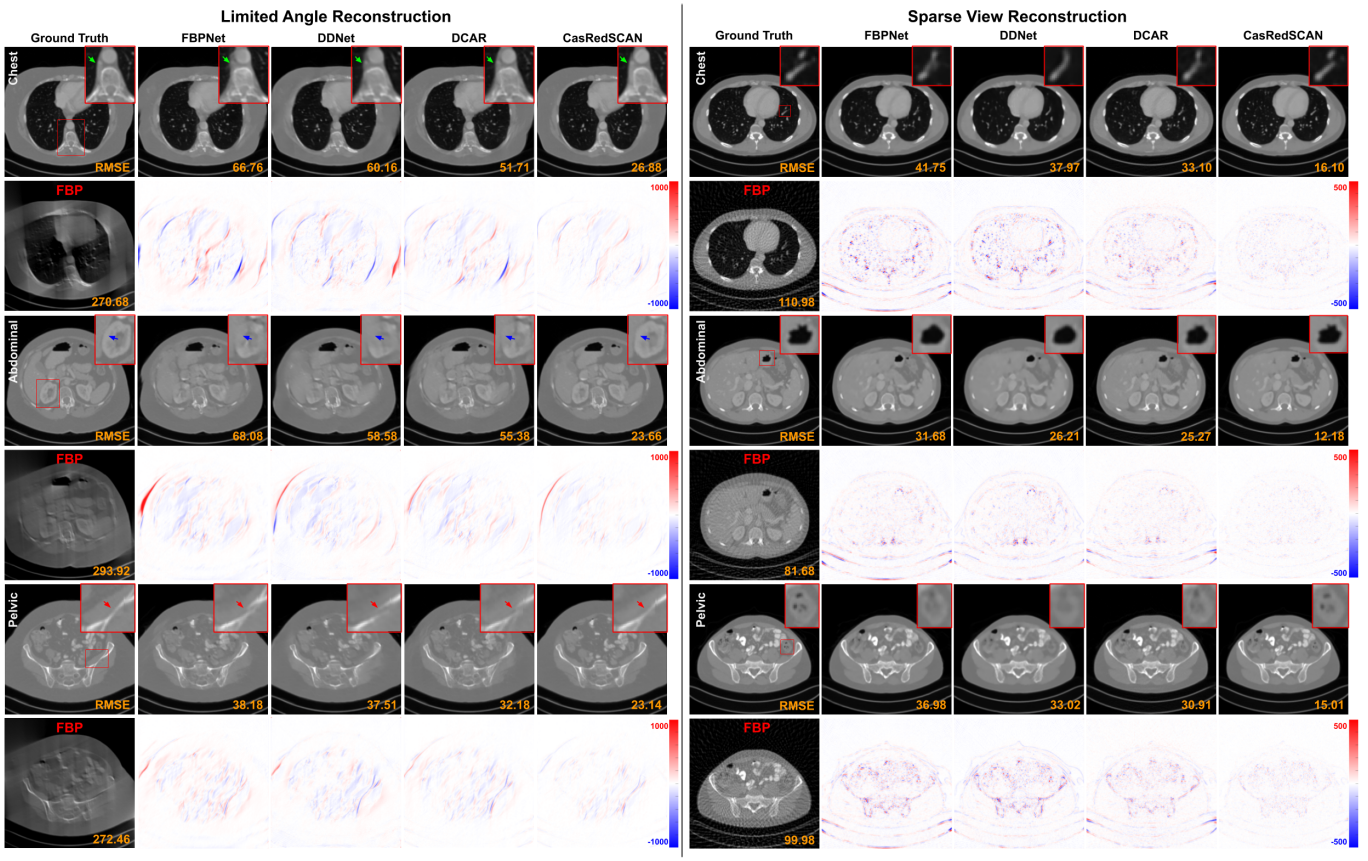

Fig. 3. Comparison of **limited angle reconstructions** (120° limited angle) and **sparse view reconstructions** (1/4 downsampling) in chest, abdominal, and pelvic CT scans along with error maps. In our LA chest reconstruction, important arterial structure (green arrows) is better preserved using our CasRedSCAN. Similarly for kidney boundary (blue arrows) in the abdominal reconstruction. The corresponding RMSE is indicated at the bottom. The display window is [−1000 1000] HU.

performance on CT image with important pathological findings, in our second dataset, we collected 2900 2D CT slices from the DeepLesion dataset [32], which consists of 8 different lesion types (bone:240, liver:380, lung:380, kidney:380, mediastinum:380, abdominal:380, pelvis:380, soft-tissue:380). We split the DeepLesion 2D dataset into 1960 training images (110 for bone, 250 for each of the rest lesion types), 300 validation images (50 slices for each lesion types), 640 test images (80 slices for each lesion types). All images are resized to 256×256. We combined two dataset for training and testing.

Similar to the CT projection simulation in [33], we assume an equi-angular fan-beam projection geometry. A 120 kVp polyenergetic x-ray source is simulated. To simulated Poisson noise in the sinogram, we assume the incident x-ray contains 2×10^7 photons. The distance between the x-ray source and the rotation center is set to 39.7 cm. There are 439 detector bins in a row and each image consists of 256×256 pixels. For each image, the fully sampled sinogram data S was generated via 360 projection views uniformly spaced between 0 and 360 degrees. In sparse view experiments, we uniformly sampled 180, 90, and 60 projection views from the 360 projection views to form S_u, mimicking 2, 4, and 6 fold radiation dose reduction. In limited angle experiments, we sampled 90, 120, and 150 (out of the 360 total) projection views that lies

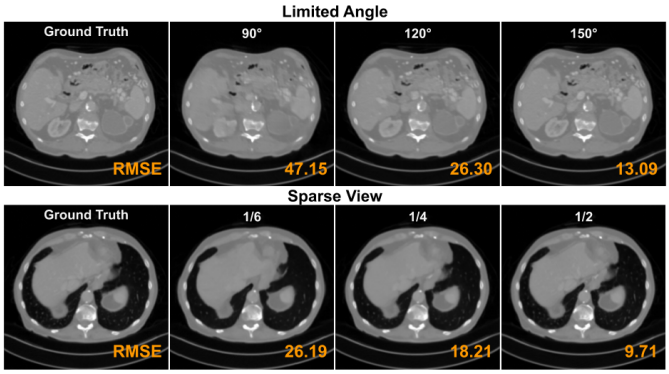

Fig. 4. **Limited angle reconstructions** and **sparse view reconstructions** at different limited angle settings and downsampling ratio settings. The display window is [−1000 1000] HU.

within $0-90$, $0-120$, and $0-150$ degrees for our S_u. The reconstructed image I and I_u were obtained by applying FBP to S and S_u, respectively.

We implemented our CasRedSCAN in Pytorch,[4] and trained it on an NVIDIA Quadro RTX 8000 GPU with 48G memory. The Adam solver [34] was used to optimize our models with

[4]http://pytorch.org/

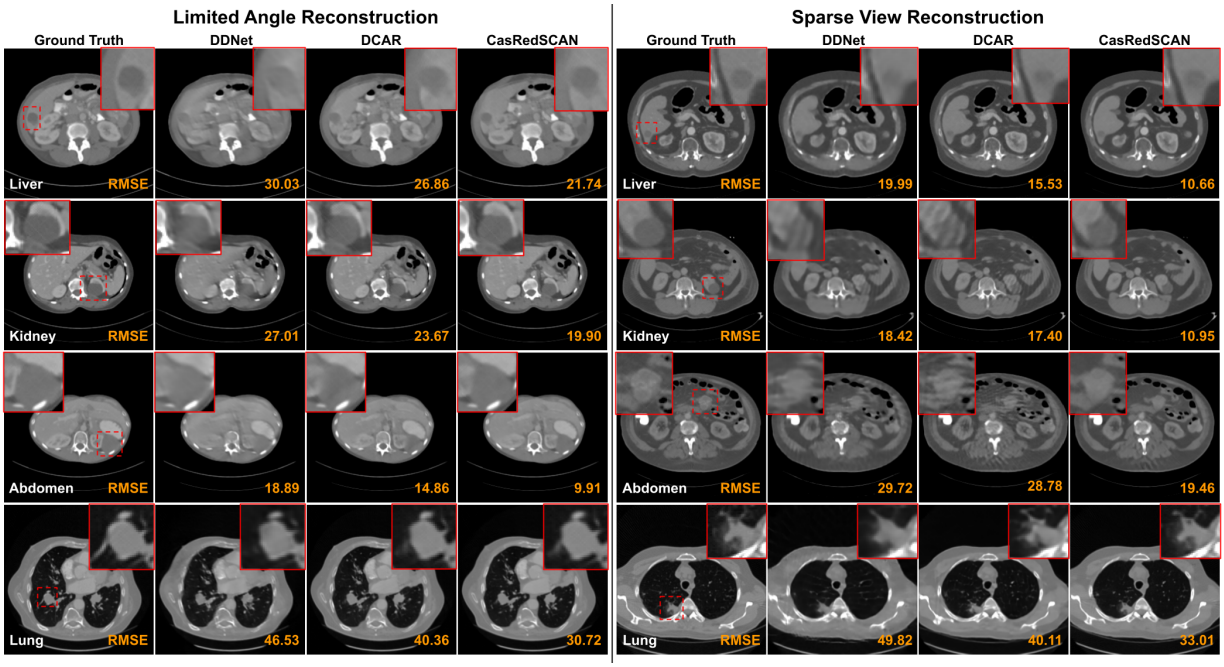

Fig. 5. Comparison of **limited angle reconstructions** (120° limited angle) and **sparse view reconstructions** (1/4 downsampling) in CT scans with lesions. The lesion region zoom-in views are shown on the top. The display window of liver, kidney, and abdomen CT is [−300 500] HU. The display window of lung CT is [−1000 1000] HU.

a momentum of 0.99 and a 0.0005 learning rate. We used a batch size of 4 during training.

B. Experimental Results

For quantitative evaluation, both SV and LA results were evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Root Mean Square Error (RMSE) by comparing the synthetic SV and LV reconstructions to the ground truth reconstruction from FBP of fully sampled sinogram. For comparative study, we compared our results on both SV and LA tasks against: 1) image-to-image translation-based methods, including the combination of Densenet and Deconvolution (DDNet) [18], Framing UNet (FUNet) [19], FBPNet [13], and 2) deep learning-based methods with projection data fidelity used in the test stage, including DCAR [21] and CTNet [1].

The qualitative comparison of different limited angle reconstruction methods with AAPM dataset is shown in Figure 3. As we can observe in chest region, previous methods have difficulties in reconstructing small anatomical structure, i.e. arteries. Similarly, with crowded organs in abdominal region, the organ boundaries are challenging to recover by previous methods along with additional patient boundary artifacts. Our CasRedSCAN with advanced network design and projection data fidelity constraint can provide superior limited angle reconstruction in terms of organ boundary recovery, small structure recovery, and boundary artifact elimination. Table VI outlines the quantitative comparison of different methods on limited angle reconstruction with AAPM dataset. Compared to the best previous method's performance of DCAR [21], we improve SSIM from 0.970 to 0.983 and reduce RMSE from 39.1 to 26.1 for 120° setup, respectively.

The qualitative comparison of different sparse view reconstruction methods with AAPM datset is also shown in Figure 3. Similar to the observations from limited angle experiments above, our CasRedSCAN yields high-quality reconstruction in crowded soft tissue area with fine details. As evidenced in Table VI, our CasRedSCAN achieves the best results among various previous methods. Compared to the best previous method's performance of DCAR [21], we improve SSIM from 0.973 to 0.989 and reduce RMSE 26.3 to 14.4 for 1/4 setup, respectively. Figure 4 shows the limited angle reconstructions and sparse view reconstructions from our CasRedSCAN at different settings.

As CT scan is often used for disease diagnosis, we also evaluated the reconstruction performance on CT images with 8 different lesion types. Figure 5 illustrates the qualitative comparison of various limited angle and sparse view reconstruction methods on 4 major lesion types. As we can observe, the liver lesion and kidney lesion are hard to recover by previous methods because these lesions have low contrast to the soft-tissue background, and their visualization are further degraded by the limited angle artifacts. Similarly, the lung lesion are also challenging to recover by previous methods due to their complex lesion texture. However, our CasRedSCAN can provide superior recovery of the shape and texture of the lesion even under these difficult conditions. For example, our liver and kidney reconstructions on the last column can provide clear lesion boundary which is critical for lesion progression assessment. The lung bronchi that originally diminished on FBP reconstruction can also be better recovered by our CasRedSCAN. Table II summarizes the reconstruction performance on CT images with 8 different lesion types. For 120° limited angle reconstruction, our CasRedSCAN achieves RMSE $<$ 30 HU across all 8 lesion types which consistently

TABLE II

QUANTITATIVE COMPARISON OF **LIMITED ANGLE RECONSTRUCTIONS** (120° LIMITED ANGLE) AND **SPARSE VIEW RECONSTRUCTIONS** (1/4 DOWNSAMPLING) RESULTS USING PSNR (DB), SSIM, AND RMSE. BEST RESULTS ARE MARKED IN RED

LA	Bone	Abdomen	Mediastinum	Liver	Lung	Kidney	Soft Tissue	Pelvis
FBP	22.29/.652/231.	21.83/.675/244.	22.54/.691/225.	21.73/.660/247.	22.18/.627/234.	21.91/.681/241.	22.81/.696/219.	22.75/.699/219.
TV [6]	30.67/.877/130.	30.18/.871/136.	30.83/.877/129.	30.27/.868/134.	30.64/.875/132.	30.34/.871/140.	31.03/.880/128.	30.83/.878/129.
FBPNet [13]	36.68/.945/45.6	39.99/.969/30.6	36.52/.956/45.7	38.47/.964/37.6	35.11/.932/53.9	40.47/.972/28.7	37.71/.961/41.0	38.01/.967/38.7
DDNet [18]	38.86/.971/36.9	41.87/.982/25.2	38.25/.972/38.7	40.67/.979/29.9	36.88/.963/44.5	42.97/.985/22.1	39.97/.977/33.0	40.15/.981/31.0
FUNet [19]	36.93/.948/43.2	40.14/.971/29.9	36.78/.960/42.9	38.93/.971/33.8	35.22/.943/49.1	41.06/.979/26.7	38.03/.968/39.9	38.04/.969/38.1
CTNet [1]	37.13/.949/40.8	40.11/.971/30.2	37.33/.962/41.7	38.97/.972/32.4	35.91/.952/46.2	41.17/.979/24.3	37.96/.963/40.2	38.00/.967/38.4
DCAR [21]	39.32/.977/32.5	42.92/.984/21.3	39.11/.976/33.8	40.86/.980/28.5	37.69/.970/41.3	43.13/.986/20.6	40.33/.980/30.5	40.84/.982/29.8
CasRedSCAN	42.08/.984/25.2	45.59/.990/16.2	41.83/.985/25.6	43.88/.987/20.7	40.72/.981/28.7	46.36/.991/14.7	43.34/.988/22.4	43.52/.989/20.8
SV	Bone	Abdomen	Mediastinum	Liver	Lung	Kidney	Soft Tissue	Pelvis
FBP	28.71/.591/112.	31.01/.676/85.3	29.50/.600/101.	30.83/.667/87.1	27.23/.538/132.	31.37/.680/81.7	29.94/.617/97.3	30.35/.636/91.8
TV [6]	32.31/.899/48.2	35.62/.919/39.4	33.94/.907/44.6	34.16/.911/41.8	31.54/.897/47.9	35.73/.918/39.5	33.98/.911/41.2	34.08/.910/42.5
FBPNet [13]	38.89/.952/35.5	42.28/.973/23.4	39.37/.961/32.9	41.91/.972/24.6	36.80/.931/45.6	42.56/.975/22.6	40.50/.968/29.3	41.61/.974/25.2
DDNet [18]	40.93/.960/28.5	44.67/.980/18.0	41.17/.968/27.1	44.23/.980/19.0	38.60/.941/37.2	45.12/.982/17.0	42.76/.974/22.9	44.13/.980/19.0
FUNet [19]	38.95/.956/33.1	42.83/.978/20.9	39.87/.966/31.2	42.37/.977/21.3	37.01/.938/42.5	42.88/.977/20.1	40.82/.970/27.8	41.68/.974/24.6
CTNet [1]	38.96/.956/32.8	42.88/.978/20.5	39.97/.968/29.7	42.20/.973/22.8	37.12/.940/40.2	42.99/.979/19.9	40.79/.968/28.7	41.73/.976/23.3
DCAR [21]	42.22/.972/23.6	45.13/.982/17.1	42.52/.973/23.5	45.79/.981/18.6	39.77/.958/32.9	45.54/.982/16.5	43.32/.977/19.2	45.67/.982/17.7
CasRedSCAN	46.00/.987/15.8	49.80/.994/9.8	46.71/.990/14.3	49.59/.994/10.2	43.90/.981/20.1	50.11/.994/9.5	48.24/.990/12.6	49.48/.994/10.2

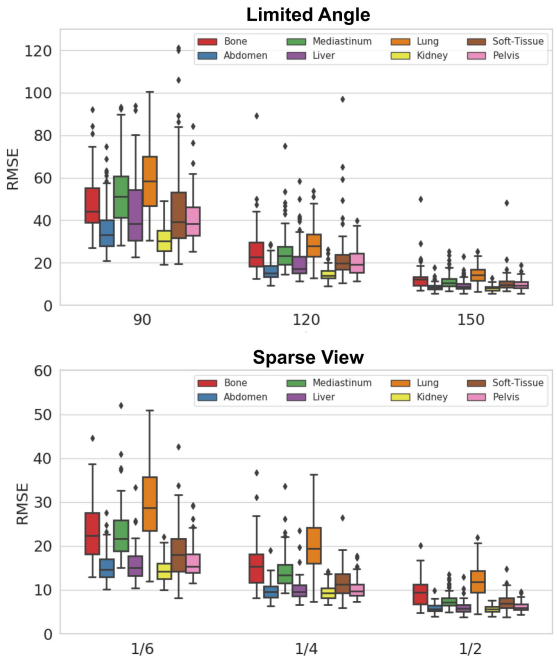

Fig. 6. Comparison of **limited angle** and **sparse view** results on CT images with 8 tumor types under different limited view settings.

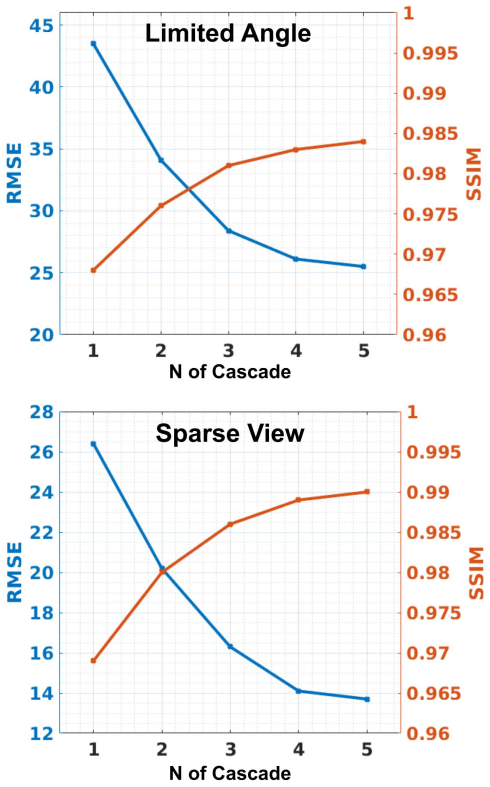

Fig. 7. The effect of increasing the number of cascade blocks (Z) in our CasRedSCAN for **limited angle reconstructions** (120° limited angle) and **sparse view reconstructions** (1/4 downsampling).

outperforms previous reconstruction methods. Similarly, for 1/4 sparse view reconstruction, our CasRedSCAN achieves the lowest RMSE across all 8 lesion types as compared to previous reconstruction methods. Performance comparison of our CasRedSCAN under different limited angle and sparse view settings on 8 different tumor types are illustrated in Figure 6. Our CasRedSCAN is able to keep the RMSE below 20 for limited angle reconstructions (150°) and sparse angle reconstructions (1/2) with different tumor types. However, the RMSE increases as the limited angle reduces or the sparse view undersampling rate increases.

C. Ablation Studies

1) Number of Cascade: The number of cascade block can be flexibly adjusted in our CasRedSCAN. We analyzed the effect of increasing the number of cascade blocks in our CasRedSCAN. The result is summarized in Figure 7 and evaluated using AAPM dataset. As we can observe, using more cascade blocks boosts the reconstruction performance, while the rate of improvement starts to converge after the number of blocks reaches 3. In LA, increasing the number of cascade from 4 to 5 only increase SSIM by less than 0.002 and reduce RMSE by less than 2 in average. Similar observation can be found in SV.

TABLE III

ATTENTION MECHANISM ANALYSIS USING PSNR, SSIM, RMSE. ✓ AND ✗ MEANS CHANNEL ATTENTION (CA) AND SPATIAL ATTENTION (SA) USED AND NOT USED IN OUR CASREDSCAN. THE OPTIMAL RESULTS ARE IN BOLD. $*$ MEANS THE DIFFERENCE COMPARED TO BASELINE WITHOUT SA AND CA ARE SIGNIFICANT AT $p < 0.1$, WHILE † MEANS NOT SIGNIFICANT

Task	CA	SA	PSNR	SSIM	RMSE
LA	✗	✗	39.61 ± 1.78	.973 ± .010	30.7 ± 4.3
	✓	✗	40.98 ± 1.62†	.979 ± .007†	28.8 ± 4.0†
	✗	✓	40.93 ± 1.63†	.978 ± .008†	28.6 ± 4.0†
	✓	✓	**41.48 ± 1.51***	**.983 ± .005***	**26.1 ± 3.8***
SV	✗	✗	44.01 ± 1.38	.979 ± .009	18.8 ± 2.8
	✓	✗	45.49 ± 1.23†	.983 ± .006†	16.9 ± 2.3†
	✗	✓	45.35 ± 1.24†	.981 ± .005†	16.7 ± 2.4†
	✓	✓	**46.43 ± 1.05***	**.989 ± .002***	**14.4 ± 1.7***

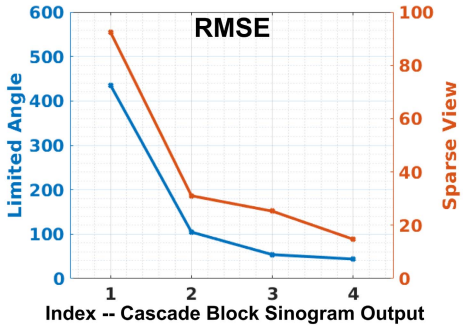

Fig. 8. Sinogram errors over the cascade block's output in our CasRedSCAN for **limited angle reconstructions** (120° limited angle) and **sparse view reconstructions** (1/4 downsampling).

2) Attention Mechanism: Two attention mechanisms are used and combined in our CasRedSCAN. We analyzed the effect of these two attention mechanisms in our CasRedSCAN. The result is illustrated in Table III and evaluated using AAPM dataset. We compared our CasRedSCAN's performance with or without channel attention or spatial attention. As we can observe, both channel attention and spatial attention can improve the reconstruction performance, and the combination of both attentions provides the best performance with the least variation, and significantly outperforms the baseline CasRedSCAN without both channel and spatial attentions.

3) Sinogram Evolution: With the number of cascade block set to 4 in our CasRedSCAN, we further analyzed how the generated sinogram evolves over the cascaded network. We computed the mean RMSE between each cascade block's sinogram outputs and the ground truth full view sinogram. The results for both LA and SV are plotted in Figure 8. As we can see, the sinogram errors gradually reduce as the generated data passes through the next cascaded block, while the rate of sinogram error reduction starts to converge after the first cascade block.

4) PDFL Parameter: In PDFL, λ is the noise level parameter that controls the linear combination of the acquired projection data and the projection data of RedSCAN's output. Assuming low noise x-ray acquisition as in our experiments, λ should be a small value as the impact of noise is minimal. We analyzed the impact of λ under both LA and SV conditions. The results are summarized in Figure 9. As we can observe, reconstruction without considering the noise, i.e. $\lambda = 0$, leads to degradation

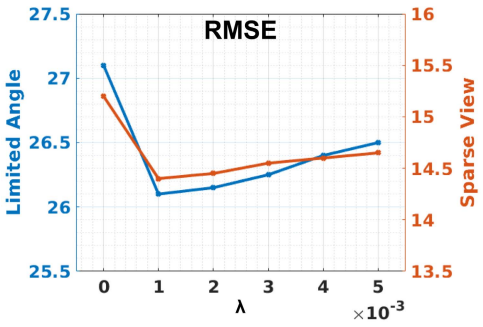

Fig. 9. Impact of λ in PDFL for **limited angle reconstructions** (120° limited angle) and **sparse view reconstructions** (1/4 downsampling).

TABLE IV

QUANTITATIVE COMPARISON OF **LIMITED ANGLE RECONSTRUCTION** (120°) AND **SPARSE VIEW RECONSTRUCTION** (1/4 DOWNSAMPLING) RESULTS USING DIFFERENT NETWORKS WITH AND WITHOUT OUR CASCADED FRAMEWORK

SSIM/RMSE-LA	FBPNet	FUNet	DDNet	Ours
Single	.959/53.6	.960/48.3	.965/45.9	.966/47.7
With Cascade	.970/39.9	.973/36.3	.978/32.5	.983/26.1
SSIM/RMSE-SV	**FBPNet**	**FUNet**	**DDNet**	**Ours**
Single	.962/32.8	.962/32.0	.963/31.7	.967/27.8
With Cascade	.977/23.8	.978/22.8	.981/19.3	.989/14.4

on reconstruction performance. Setting $\lambda = 0.001$ leads to the best reconstruction performance in our search range, while the RMSE difference is less than 1 between $\lambda = 0.001$ and $\lambda = 0.005$.

5) Embedded Networks: We embedded different previous image-to-image reconstruction networks [13], [18], [19] into our cascaded network and compared the performance with or without cascade. The qualitative results are visualized in Figure 10. The quantitative results are summarized in Table IV. The number of cascade is set to 4 in this study. As we can observed, embedding different previous image-to-image networks into our cascade design improves the reconstruction performance, while RedSCAN embedded into our cascade network achieves the best reconstruction performance.

VI. DISCUSSION

In this paper, a novel reconstruction framework, named CasRedSCAN, is proposed. Inspired by the recent advances in image super-resolution network designs and the projection data constraint in MBIR, we designed a customized RedSCAN as our backbone image reconstruction network, and we built a projection data fidelity layer that can be embedded in deep networks. First of all, our RedSCAN is developed based on image super-resolution network [35] with an addition of spatial-channel attention, which allows our RedSCAN to re-calibrate the channel attention and gives different levels of attention on recovering texture details at different spatial locations, as artifact distribution is not uniform in the image. In fact, Hu *et al.* [36] recently also demonstrated that spatial-channel attention can boost the image super-resolution performance. Then, we develop PDFL that can be concatenated to the RedSCAN's cascade outputs to ensure the projection data

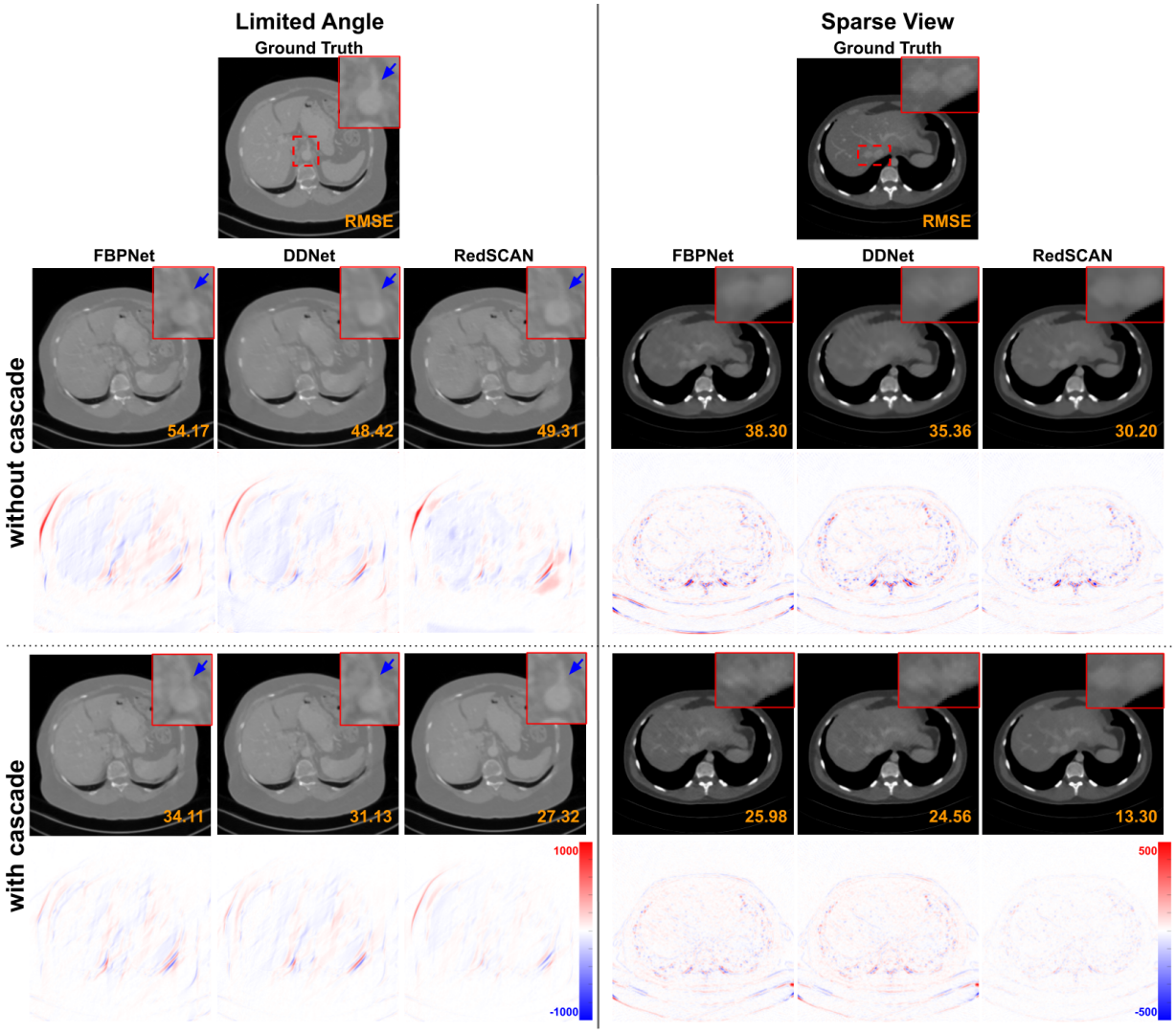

Fig. 10. Comparison of **limited angle reconstructions** (120° limited angle) and **sparse view reconstructions** (1/4 downsampling) with and without cascade framework using different basic networks. The display window of limited angle reconstruction is [−1000 1000] HU. The display window of sparse view reconstruction is [−300 800] HU.

fidelity at the sampled projection views. Our PDFL based on the analytical FBP solution with fan-beam geometry allows it to be embedded in a deep network and used during training and inference.

We demonstrate the feasibility of our CasRedSCAN on both LA and SV tomographic reconstruction tasks, as shown in the result section. Firstly, the LA acquisition is more difficult to reconstruct as compared to the SV acquisition since a range of projection angles are not covered in the LA acquisition. Severe image artifacts at these projection angles can be observed when using conventional FBP. As a result, the general performance of LA reconstructions are inferior to the SV reconstruction performance. For example, in 120° LA reconstruction, while previous methods can mitigate the artifacts and recover PSNR up to 37.94 and SSIM up to 0.970, they still have difficulties in recovering the organ boundaries that are critical for clinical diagnosis and treatment planning. Our CasRedSCAN provides superior reconstructions with clear organ boundaries and is able to improve the PSNR to 41.48 and SSIM to 0.983. In 1/4 SV reconstruction, while previous methods can generate visually plausible image content, the reconstruction prediction without projection data fidelity can result in artificial texture which is undesirable in clinical tasks. Our CasRedSCAN with PDFL can better preserve the image fidelity by incorporating the already-sampled projection data, resulting in best performance in terms of PSNR, SSIM, and RMSE.

Furthermore, we demonstrate the feasibility of our CasRedSCAN on CT lesion imaging under LA and SV conditions. Lesion is highly heterogeneous, and CT is one of the primary tool for diagnosis. Obtaining high-quality lesion region reconstruction under LA and SV is essential for disease diagnosis, staging, as well as planning and evaluation of treatment. While previous methods can reduce the reconstruction artifacts from the whole image perspective, the reconstruction in lesion region with high heterogeneity is still unsatisfying - the lesion boundary and texture are highly distorted by previous methods which will negatively impact the subsequent treatment options.

On the other hand, our CasRedSCAN can better preserve the lesion reconstruction even the lesions are highly heterogeneous. For example, the supplying vessels of LA lung lesion in Figure 5 are totally missed by previous methods, while our CasRedSCAN can better recover it. The complex interior texture of SV lung lesion in Figure 5 is highly distorted by previous methods, but our CasRedSCAN can still preserve the structure. In Figure 5, liver and kidney lesions embedded in soft-tissue background with low contrast are prone to smooth-out in SV and distorted in LA by previous methods, and our CasRedSCAN can better recover the boundary and the contrast of the lesions.

We believe there are several reasons that potentially lead to the superior performance of using RedSCAN in CasRedSCAN. First of all, our RedSCAN has no image down-sampling for abstraction, thus keeping the image restoration on original resolution. Second, convolutional layers in different depths have different sizes of receptive fields, resulting in hierarchical features. Image restoration should utilize all the hierarchical features, instead of only the last layer output. Our RedSCAN concatenating all the hierarchical features can potentially better learn the restoration. Thirdly, the hierarchical features are generated by our residual dense channel-spatial block that allows better feature learning at each hierarchical level. Moreover, the residual connection in each block also allows the gradient to be better passed to earlier layers, thus helping the training of our wide network design. As shown in Table , the design of our RedSCAN also provides a relatively smaller amount network parameter (0.51M) as compared to the previous method. Specifically, the RedSCANs in CasRedSCAN share the same network parameter and there is no learnable parameter in PDFL, thus the CasRedSCAN's parameter size remains the same as RedSCAN regardless of the number of cascading. In this case, our CasRedSCAN using the least amount of parameters achieves the best limited view reconstruction performance.

The presented work also has potential limitations. First of all, the inference time is longer compared to the previous deep learning based methods, as illustrated in Table VI. This is caused by the cascaded design with PDFL interleaved. On one hand, the iterative reconstruction prediction will increase the computation time. On the other hand, even though FBP is a fast analytic solution, the forward projection and FBP operations in PDFL still consume computation times. The combination of these two results in longer training and inference time. However, the inference time is about 150 ms which is acceptable and much faster than previous MBIR methods. Moreover, in our PDFL, we assume 360 degrees fan-beam projection combined from the already sampled sinogram and the predicted sinogram. The minimal complete sinogram with reduced number of projection could reduce the computation time of PDFL. However, additional step of sinogram weighting, such as Parker weighting [37], could be incorporated to address the data redundancy issue. Secondly, while increasing the number of cascade block in CasRedSCAN improves the performance, the memory consumption will increase along with longer training and inference time. As illustrated in Figure 7, the increase in performance starts to converge after $n = 3$. Thus, in this work, we set $n = 4$ to balance the memory consumption and inference time of our CasRedSCAN.

The architecture of our CasRedSCAN also suggests several interesting topics for future studies. The first one is combining the projection data fidelity layer with the deep learning based radon inversion techniques [38]. The cascaded framework with projection data fidelity can provide the projection domain constraint during the radon inversion via deep learning. It can potentially improve the inversion stability, yielding reconstruction with better data fidelity. Secondly, given the superior lesion region reconstruction performance demonstrated in the result sections, our framework could also potentially improve the projection data based Computer-Aided Diagnosis (CAD). Recently, there are increasing interests on combining limited-view reconstruction and CAD for a joint reconstruction-CAD network structure, and improved CAD performance is expected with such an end-to-end training strategy [39], [40]. We believe that our CasRedSCAN with high-quality lesion region reconstruction would provide new opportunities for these kinds of studies. Thirdly, CT metal artifact reduction (MAR) under limited-view acquisition is an important research direction. Current MAR techniques are mostly limited to full-view acquisition [41], [42]. The current state-of-the-art metal artifact reduction algorithm, such as DuDoNet [41], utilizes projection space and image space simultaneously which is similar to our CasRedSCAN design. Our CasRedSCAN could potentially integrated with current MAR network for MAR under limited view conditions. Fourthly, low-dose CT combined with limited-view acquisition may further reduce the radiation dose. As a matter of fact, Shan *et al.* [43] and Wu *et al.* [44] had proposed cascaded network structures with basic network of UNet [15] or sequential CNN layers, and demonstrated their efficiency in low-dose CT. As cascade network is also potentially efficient in low-dose CT, our CasRedSCAN could be adapted to limited-view low-dose CT that may further reduce the radiation dose and acquisition time. Lastly, we believe our CasRedSCAN could be adapted to other tomography imaging modalities with similar applications, such as SPECT, PET, and Cryo-ET [45]–[47].

VII. Conclusion

In this work, we proposed a cascaded network with RedSCAN and PDFL, a novel framework for limited view tomographic reconstruction. The proposed PDFL is interleaved in our cascaded network to ensure the sampled sinogram is consistent in sinogram domain with the network cascaded output. A customized image restoration network is used as the backbone in the cascaded network. Comprehensive evaluation demonstrates that our CasRedSCAN can provide high-quality limited angle and sparse view tomographic reconstruction while reducing radiation dose and shortening scanning time.

References

[1] R. Anirudh, H. Kim, J. J. Thiagarajan, K. A. Mohan, K. Champley, and T. Bremer, "Lose the views: Limited angle CT reconstruction via implicit sinogram completion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6343–6352.

[2] L. De Chiffre, S. Carmignato, J.-P. Kruth, R. Schmitt, and A. Weckenmann, "Industrial applications of computed tomography," *CIRP Ann.*, vol. 63, no. 2, pp. 655–677, 2014.

[3] B. Zhou, X. Lin, and B. Eck, "Limited angle tomography reconstruction: Synthetic reconstruction via unsupervised Sinogram adaptation," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Springer, 2019, pp. 141–152.

[4] J. Hwan Cho and J. A. Fessler, "Motion-compensated image reconstruction for cardiac CT with sinogram-based motion estimation," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2013, pp. 1–5.

[5] K. Aditya Mohan *et al.*, "TIMBIR: A method for time-space reconstruction from interlaced views," *IEEE Trans. Comput. Imag.*, vol. 1, no. 2, pp. 96–111, Jun. 2015.

[6] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Math.*, vol. 76, no. 2, pp. 167–188, Apr. 1997.

[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.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[8] H. Zhang *et al.*, "Iterative reconstruction for X-ray computed tomography using prior-image induced nonlocal regularization," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 9, pp. 2367–2378, Sep. 2014.

[9] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1289–1296, Jun. 2018.

[10] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, Jun. 2018.

[11] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "Iterative low-dose CT reconstruction with priors trained by artificial neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2479–2486, Dec. 2017.

[12] J. Adler and O. Oktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.

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

[14] H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.*, Springer, 2015, pp. 234–241.

[16] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[17] H. Liao, Z. Huo, W. J. Sehnert, S. K. Zhou, and J. Luo, "Adversarial sparse-view CBCT artifact reduction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2018, pp. 154–162.

[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 Trans. Med. Imag.*, vol. 37, no. 6, pp. 1407–1417, Jun. 2018.

[19] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.

[20] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 109–119, Mar. 2019.

[21] Y. Huang, A. Preuhs, G. Lauritsch, M. Manhart, X. Huang, and A. Maier, "Data consistent artifact reduction for limited angle tomography with deep learning prior," in *Proc. Int. Workshop Mach. Learn. Med. Image Reconstruction*. Springer, 2019, pp. 101–112.

[22] A. Kofler, M. Haltmeier, C. Kolbitsch, M. Kachelrieß, and M. Dewey, "A U-Nets cascade for sparse view computed tomography," in *Proc. Int. Workshop Mach. Learn. Med. Image Reconstruction*. Springer, 2018, pp. 91–99.

[23] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.

[24] B. Zhou and S. K. Zhou, "DuDoRNet: Learning a dual-domain recurrent network for fast MRI reconstruction with deep t1 prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 298–313.

[25] C. McCollough, "TU-FG-207A-04: Overview of the low dose CT grand challenge," *Med. Phys.*, vol. 43, pp. 3759–3760, Jun. 2016.

[26] K. Yan *et al.*, "Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9261–9270.

[27] C. Zhang, T. Zhang, M. Li, C. Peng, Z. Liu, and J. Zheng, "Low-dose CT reconstruction via L1 dictionary learning regularization using iteratively reweighted least-squares," *Biomed. Eng. OnLine*, vol. 15, no. 1, p. 66, Dec. 2016.

[28] A. C. Kak, M. Slaney, and G. Wang, "Principles of computerized tomographic imaging," *Med. Phys.*, vol. 29, no. 1, p. 107, 2002. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.1455742, doi: 10.1118/1.1455742.

[29] G. T. Herman, "Image reconstruction from projections," *Real-Time Imag.*, vol. 1, no. 1, pp. 3–18, Apr. 1995.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[31] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel 'squeeze and excitation' blocks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 540–549, Feb. 2019.

[32] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *J. Med. Imag.*, vol. 5, no. 3, 2018, Art. no. 036501.

[33] Y. Zhang and H. Yu, "Convolutional neural network based metal artifact reduction in X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1370–1381, Jun. 2018.

[34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[35] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jan. 21, 2020, doi: 10.1109/TPAMI.2020.2968521.

[36] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3911–3927, Nov. 2020.

[37] D. L. Parker, "Optimal short scan convolution reconstruction for fan beam CT," *Med. Phys.*, vol. 9, no. 2, pp. 254–257, Mar. 1982.

[38] J. He, Y. Wang, and J. Ma, "Radon inversion via deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2076–2087, Jun. 2020.

[39] Z. Wei, B. Liu, B. Dong, and L. Wei, "A joint reconstruction and segmentation method for limited-angle X-ray tomography," *IEEE Access*, vol. 6, pp. 7780–7791, 2018.

[40] J. Adler, S. Lunz, O. Verdier, C.-B. Schönlieb, and O. Öktem, "Task adapted reconstruction for inverse problems," 2018, *arXiv:1809.00948*. [Online]. Available: http://arxiv.org/abs/1809.00948

[41] W.-A. Lin *et al.*, "DuDoNet: Dual domain network for CT metal artifact reduction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10512–10521.

[42] M. Katsura, J. Sato, M. Akahane, A. Kunimatsu, and O. Abe, "Current and novel techniques for metal artifact reduction at CT: Practical guide for radiologists," *RadioGraphics*, vol. 38, no. 2, pp. 450–461, Mar. 2018.

[43] H. Shan *et al.*, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 269–276, Jun. 2019.

[44] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "A cascaded convolutional neural network for X-ray low-dose CT image denoising," 2017, *arXiv:1705.04267*. [Online]. Available: http://arxiv.org/abs/1705.04267

[45] L. Shi, J. A. Onofrey, H. Liu, Y.-H. Liu, and C. Liu, "Deep learning-based attenuation map generation for myocardial perfusion spect," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 47, pp. 2383–2395, Mar. 2020.

[46] B. Zhou, Y.-J. Tsai, and C. Liu, "Simultaneous denoising and motion estimation for low-dose gated pet using a siamese adversarial network with gate-to-gate consistency learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2020, pp. 743–752.

[47] B. Zhou, H. Yu, X. Zeng, X. Yang, J. Zhang, and M. Xu, "One-shot learning with attention-guided segmentation in cryo-electron tomography," *Frontiers Mol. Biosci.*, vol. 7, p. 473, Jan. 2021.