Improving the Quality of Sparse-view Cone-Beam Computed Tomography via Reconstruction-Friendly Interpolation Network

Yanli Wang, Lianying Chao, Wenqi Shan, Haobo Zhang, Zhiwei Wang(✉), and Qiang Li(✉)

Wuhan National Laboratory for Optoelectronics,
Huazhong University of Science and Technology, Wuhan, China
{zwwang,liqiang8}@hust.edu.cn

Abstract. Reconstructing cone-beam computed tomography (CBCT) typically utilizes a Feldkamp-Davis-Kress (FDK) algorithm to 'translate' hundreds of 2D X-ray projections on different angles into a 3D CT image. For minimizing the X-ray induced ionizing radiation, sparse-view CBCT takes fewer projections by a wider-angle interval, but suffers from an inferior CT reconstruction quality. To solve this, the recent solutions mainly resort to synthesizing missing projections, and force the synthesized projections to be as realistic as those actual ones, which is extremely difficult due to X-ray's tissue superimposing. In this paper, we argue that the synthetic projections should restore FDK-required information as much as possible, while the visual fidelity is the secondary importance. Inspired by a simple fact that FDK only relies on frequency information after ramp-filtering for reconstruction, we develop a Reconstruction-Friendly Interpolation Network (RFI-Net), which first utilizes a 3D-2D attention network to learn inter-projection relations for synthesizing missing projections, and then introduces a novel Ramp-Filter loss to constrain a frequency consistency between the synthesized and real projections after ramp-filtering. By doing so, RFI-Net's energy can be forcibly devoted to restoring more CT-reconstruction useful information in projection synthesis. We build a complete reconstruction framework consisting of our developed RFI-Net, FDK and a commonly-used CT post-refinement. Experimental results on reconstruction from only one-eighth projections demonstrate that using RFI-Net restored full-view projections can significantly improve the reconstruction quality by increasing PSNR by 2.59 dB and 2.03 dB on the walnut and patient CBCT datasets, respectively, comparing with using those restored by other state-of-the-arts.

Keywords: Sparse-view computed tomography (SVCT) reconstruction · FDK algorithm · Reconstruction-friendly projections · Interpolation · Ramp-filter loss

Y. Wang and L. Chao—Co-first authors.

L. Wang et al. (Eds.): ACCV 2022, LNCS 13846, pp. 86–100, 2023.
https://doi.org/10.1007/978-3-031-26351-4_6

1 Introduction

Cone Beam Computed Tomography (CBCT) is one of the key imaging techniques for various clinical applications, e.g., cancer diagnosis [11], image-guided surgery [5] and so on. The principle of reconstructing CBCT is first to take hundreds of 2D X-ray projections at regular intervals within a certain angle range, e.g., 360 °C, and then to utilize a Feldkamp-Davis-Kress (FDK) [24] algorithm to reconstruct a 3D CT image from those projections. Although CBCT enjoys lots of merits such as fast speed and large range, its brought ionizing radiation [4] is harmful to patients, which hinders long-term intensive usage [3]. Using fewer projections, that is, widening the sampling angle interval, is a crucial means of lowering CBCT's radiation dose, which is known as sparse-view CBCT [2]. However, sparse-view CBCT's dose reduction comes with a price of lost structures and streaking artifacts in CT images, which severely degrades reconstruction quality.

To improve the quality of sparse-view CBCT, several post-refinement methods [14,26,32] have been studied, and mainly focused on developing deep learning approaches to refine those FDK-reconstructed sparse-view CBCT images. Their objective is typically minimizing a voxel-wise L_2 distance between the refined sparse-view and original full-view CBCT images, however, this often yields over-smoothed refinement results. To overcome the over-smoothness problem [8,29], Liao et al. [19] used a VGG-based perceptual loss to minimize the L_2 distance between features extracted from the refined sparse-view CBCT image and its full-view counterpart, which was claimed to well preserve the high-frequency information in CT images. However, the VGG-network used in [19] was trained for natural image classification but not for CT image refinement, which may impair the perceptual loss's capability in CT data. Therefore, Li et al. [18] utilized a customized perceptual loss, which is based on a trained self-supervised auto-encoder network using CT data. This expanded the network's representational ability for further improving the post-refinement performance. Although some progress has been made, such CT post-refinement is disengaged from the CBCT imaging device/system, thus may ignore the valuable information contained in those raw X-ray projections. That is, the lost structures or streaking artifacts in sparse-view CBCT are hard to be rectified solely by a post-refinement without touching the raw projections.

Recently, a joint strategy has been proposed by two works [6,15], and it successively processes X-ray projections and CT images before and after FDK-reconstruction. Specifically, Hu et al. [15] developed a hybrid-domain neural network (HDNet), which, in the projection domain, first utilizes a non-parametric linear interpolator to restore the full-view projections, and then introduces a convolutional neural network (CNN) as a pixel-to-pixel translator to refine those linearly interpolated projections. However, the linear interpolator hardly handles rotating, and the caused interpolation errors are too difficult to be corrected. By comparison, Chao et al. [6] developed a DualCNN which utilized a projection domain interpolation CNN (PDCNN) to learn a direct interpolation of those missing projections from the sparse ones. PDCNN restores the full-view projections in a multi-step manner, i.e., the number of projections doubles in each step by synthesizing the middle in every two consecutive ones. Both HDNet and PDCNN bother forcing an identical appearance between the restored and original full-view X-ray projections by minimizing the voxel-wise distance. Thanks to the rotating projection and X-ray's tissue superimposing, such objective is extremely difficult to achieve.

If those restored full-view projections are not necessarily perfect, but just accurate to contain FDK-required information for reconstruction, the learning difficulty could be significantly alleviated, and the CNN efficacy is thus maximized. To this end, a straightforward idea is to have FDK differentiable, making CT reconstruction errors be back-propagated to guide the projection synthesis. However, this end-to-end manner involves concurrently processing hundreds of 2D projections and high-dimensional 3D CT images, bringing an unbearable huge computing cost inevitably.

In this paper, we develop a Reconstruction-Friendly Interpolation Network (RFI-Net) to trade off the computational efficiency, and introduce a novel Ramp-Filter Loss (RF-Loss) to have RFI-Net focus on learning the FDK-required projection information. Specifically, RFI-Net is implemented with a 3D-2D attention network architecture, which includes a 3D feature extractor, an inter-projection fusion module, and a 2D projection generator, as shown in Fig. 2. First, sparse 2D projections in a wide-angle range are stacked as a 3D volume that is encoded into a 3D feature map via the 3D feature extractor. Then, the inter-projection fusion module integrates features along the angle dimension for converting the 3D feature map to a 2D feature map. Lastly, the 2D projection generator decodes the 2D map into projections on missing angles.

RF-Loss is motivated by the FDK principle that only information of projections filtered by a ramp-filter will be used for reconstruction. Therefore, RF-Loss computes and minimizes the frequency difference between the restored and actual projections after ramp-filtering, and the ramp-filter in RF-Loss is set to be identical with that in FDK. With no need to optimize the CT reconstruction, RFI-Net is still aware of generating reconstruction-friendly full-view projections, rather than just mimicking the superficial appearance.

The main contributions of this work are as follows:

– We develop a CNN-based interpolator named RFI-Net, which can capture wide-angle range inter-projection relations, and synthesize reconstruction-friendly full-view projections, improving the quality of reconstructed sparse-view CBCT.
– We introduce a novel RF-Loss, which encourages RFI-Net's synthesized projections to contain FDK-required frequency information, without resorting to a computation-intensive end-to-end learning fashion.
– We build a complete reconstruction framework, which consists of a trained RFI-Net, FDK, and a commonly-adopted CT post-refinement. By enjoying those reconstruction-friendly projections synthesized by RFI-Net, our framework is experimentally demonstrated to increase PSNR by 2.59 dB and 2.03

dB on sparse-view CBCT reconstructions for the walnut and patient CBCT datasets under one-eighth dose, respectively, comparing with other state-of-the-arts.

2 Methods

Figure 1 presents the built complete CBCT reconstruction framework. First, RFI-Net synthesizes reconstruction-friendly full-view 2D projections from those sparse ones. Then, FDK reconstructs a 3D CT image from the synthesized full-view projections. Finally, a post-refinement network (Post-Net) is employed to further refine the FDK-reconstructed CT image. In the following, we detail RFI-Net and Post-Net, and explain our proposed RF-Loss for training.

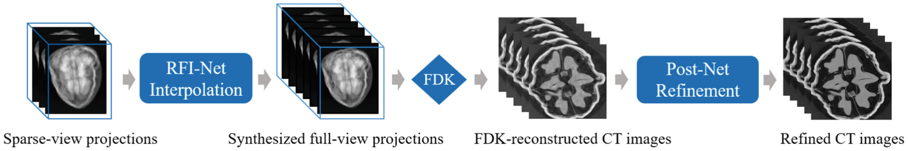

Fig. 1. Our built complete framework for sparse-view CBCT reconstruction.

2.1 RFI-Net Architecture

As shown in Fig. 2, RFI-Net is implemented with a 3D-2D attention network, which contains three components, i.e., a 3D feature extractor, an inter-projection fusion module and a 2D projection generator. Given full-view projections with the total number of N, the *quarter* dose sparse-view projections can be constructed as $\{P_{4i-3}|i = 1, 2, \ldots, N/4\}$, where N is 600 as the number of full-view projections and P_1 can be also denoted as P_{N+1}. Our goal is to restore the three missing projections $\{P_{4i-2}, P_{4i-1}, P_{4i}\}$ between every two adjacent sparse-view projections, i.e., P_{4i-3} and P_{4i+1}.

The 3D feature extractor stacks consecutive projections $\{P_{4i-3}, \ldots, P_{4(i+D-1)}\}$ sampled within a wide-angle range into a 3D volume, and encodes it into a 3D feature map F_{3D} with the size of $C \times H \times W \times D$, where C and D are the number of feature channels, input sparse-view projections, respectively, and we set D to 4 in this work, where H and W represent the size of X-ray projection. Specifically, the 3D feature extractor employs 3D ResUNet [31] as the backbone and has three main advantages: (i) the residual path can avoid gradient vanishing in the training phase; (ii) the skip connection [25] between encoder and decoder can fuse low-level features and high-level features to well express the projection information; (iii) the 3D kernels jointly capture the angle and spatial information of projections.

The inter-projection fusion module consists of a reshape operation, a 2D convolution layer and a channel attention module, which bridges the 3D feature

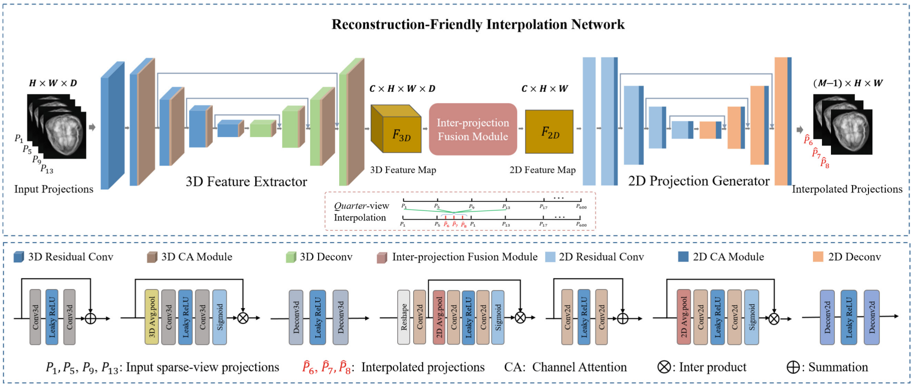

Fig. 2. The RFI-Net architecture. For example, sparse-view projections P_1, P_5, P_9, P_{13} are stacked and input into RFI-Net, and the missing projections $\hat{P}_6$, $\hat{P}_7$, $\hat{P}_8$ between P_5 and P_9 are synthesized.

extractor and the 2D projection generator. This fusion module first reshapes the 3D feature map F_{3D} to a 2D map with the size of $CD \times H \times W$ by merging the angle and channel dimensions, and then employs a 2D convolution layer to generate the final 2D feature map F_{2D} with size of $C \times H \times W$. The channel attention module finally measures the inter-dependencies between the feature channels and allows the module to focus on useful ones.

The 2D projection generator has the similar architecture with the 3D feature extractor, and replaces the 3D kernels with the 2D ones. The output size of the generator is $(M-1) \times H \times W$, where $M-1$ represents the number of missing projections for $1/M$ dose-level. We adjust M according to different dose-levels, e.g., $M = 4$ for *quarter* dose and $M = 8$ for *one-eighth* dose. Finally, the result of the generator is sliced along channel into individual synthesized projections with the total number of $M-1$.

Note that, channel attention (CA) module [9] is embedded into each layer of the 3D-2D network architecture. In the CA module, global average pooling [20] is first used to enlarge the receptive field by compressing the spatial features. Then the following two convolution layers are utilized to capture the non-linear inter-channel relationships. Finally, sigmoid activation function is used to introduce non-linearity.

2.2 RF-Loss of RFI-Net

In FDK algorithm, projections are not directly back-projected to reconstruct CT images, but filtered by the ramp-filter in the frequency domain beforehand. The ramp-filter is a correction filter that can redistribute the frequency information by suppressing the low-frequency information but encouraging the pass of high-frequency information [22]. Motivated by this, we design RF-Loss to make RFI-Net mainly focus on the ramp-filtered frequency information in projections.

RF-Loss $\mathcal{L}_{RF}$ minimizes the frequency-wise error between the synthesized $\{\hat{P}_t|t = 1, 2, ..., M-1\}$ and ground-truth $\{P_t^{GT}|t = 1, 2, ..., M-1\}$ projections, which can be formulated as:

$$\mathcal{L}_{RF} = \frac{1}{M-1}\sum_{t=1}^{M-1}\left|RF(\hat{P}_t) - RF(P_t^{GT})\right| \tag{1}$$

where $RF(*)$ represents the frequency representations after ramp-filtering. Specifically, we first convert a projection P_t into its frequency representation $F_t(u,v)$ by calculating the 2D Fast Fourier Transform (FFT) as follows:

$$F_t(u,v) = \sum_{x=0}^{H-1}\sum_{y=0}^{W-1} P_t(x,y) \times \left[\cos 2\pi(\frac{ux}{H} + \frac{vy}{W}) - i\sin 2\pi(\frac{ux}{H} + \frac{vy}{W})\right] \tag{2}$$

where $u = 0, 1, ..., H-1, v = 0, 1, ..., W-1$, H and W are the height and width of the projection, and (x,y) denotes the position in the spatial domain. $P_t(x,y)$ is the pixel intensity at position (x,y). In the frequency domain, the projection is decomposed into cosine and orthogonal functions for constituting the real and imaginary parts of the complex frequency value. After applying the 2D FFT, a ramp-filter weight matrix $\alpha(u,v)$ [30] is used to multiply with the complex frequency value $F_t(u,v)$ as follows:

$$RF(P_t) = |\alpha(u,v) \times F_t(u,v)| \tag{3}$$

$\mathcal{L}_1$ loss [28] is also used to minimize the pixel-wise error between the interpolated projections and the references. The total loss $\mathcal{L}_{RFINet}$ can be formulated as:

$$\mathcal{L}_{RFINet} = \mathcal{L}_1 + \gamma\mathcal{L}_{RF} \tag{4}$$

where γ is the weighting parameter to balance the two losses and is set to 0.1 in our experiments.

2.3 Post-Net Architecture

In the reconstruction domain, we modified a simple 3D UNet [1] as the post-processing network to further refine the pre-processed CT images.

As shown in Fig. 3, Post-Net contains four convolutional blocks to extract rich features and four deconvolution blocks to restore the image contents. The artifacts distribution is generated through the last layer, and the final high-quality CT images are produced by subtracting the predicted artifacts distribution from the pre-processed CT images. The size of kernels used in Post-Net is $3 \times 3 \times 3$. A joint loss [7] that includes a perceptual loss and a SSIM loss is used for optimizing the Post-Net. Perceptual loss [16] makes Post-Net retain the high-frequency information in CBCT images to avoid the over-smoothness problem. SSIM loss [33] makes Post-Net well preserve the delicate structures in CT images. The joint loss $\mathcal{L}_{PostNet}$ can be formulated as:

$$\mathcal{L}_{PostNet} = \beta\mathcal{L}_{perc} - \mathcal{L}_{ssim} \tag{5}$$

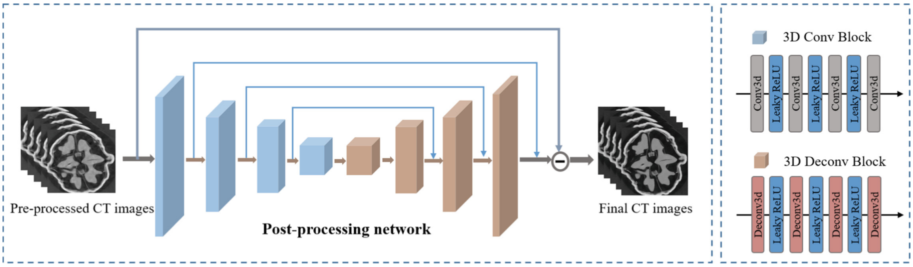

Fig. 3. Post-processing network architecture.

where $\mathcal{L}_{perc}$ and $\mathcal{L}_{ssim}$ denote the perceptual loss and SSIM loss, respectively. β is the weighting parameter to balance the two losses. $\mathcal{L}_{perc}$ and $\mathcal{L}_{ssim}$ are detailed in reference [7]. In the training of Post-Net, the input includes nine consecutive slices and the weighting parameter β in Eq. 5 is set to 200/3.

3 Experiment

3.1 Dataset

Our experiments are validated on two CBCT datasets:

(i) **Walnut dataset**: This is a public CBCT dataset provided for deep learning development [10]. In this dataset, 42 walnuts were scanned with a special laboratory X-ray CBCT scanner, with raw projection data. For each walnut, we take 600 projections evenly over a full circle as the full-view projections, and explore sparse-view CBCT reconstruction under the *quarter* and *one-eighth* dose by using only 150 and 75 projections, respectively. The CBCT reconstruction uses the FDK algorithm in ASTRA toolbox [27]. For each walnut, the size of the projection image is 972×768, and the reconstructed CT volume is $448 \times 448 \times 448$. The first five walnuts are used for test, the sixth for validation, and the rest for training.

(ii) **Patient dataset**: This dataset includes 38 real normal-dose 3D CT images provided by the TCIA data library [21], and the corresponding full- and sparse-view 2D projections are simulated by forward-projecting the CT images using the ASTRA toolbox. The number of full-view projections is 600, and the size of simulated projection is 972×768, and the size of the reconstructed CT volume is $n \times 512 \times 512$, where n is the slice number for each patient. Five patients are used for test, one patient for validation, and the rest for training.

3.2 Implementation Details

Our networks are implemented with Pytorch [23] ver. 1.9, Python ver. 3.7, and CUDA ver. 11.2. The Adam [17] solver with momentum parameters $\beta_1 = 0.9$

Table 1. Abbreviations of different comparison methods.

Abbreviations	Descriptions
FVCT	Full-view + FDK (as ground truth)
SVCT	Sparse-view + FDK (as baseline)
CNCL [13]	Sparse-view + FDK + Post-refinement by CNCL
SI-UNet [12]	Linear interpolation and 2D U-Net enhancement + FDK
HDNet [15]	Linear interpolation and 3D U-Net enhancement + FDK + Post-refinement by 3D U-Net
DualCNN [6]	PDCNN interpolation + FDK + Post-refinement by IDCNN
Ours	RFI-Net interpolation + FDK + Post-refinement by Post-Net

and $\beta_2 = 0.99$ is used to optimize RFI-Net with the learning rate of 2e−4. The number of training epochs and batch size are set to 150 and 1, respectively. All networks are trained using a NVIDIA GPU GTX3090 of 24 GB memory.

3.3 Evaluation Metrics and Comparison Methods

We use three evaluation metrics, including the root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM), to quantify the difference or similarity between the full-view CBCT and improved sparse-view CBCT by different methods. The comparison methods are listed in Table 1. CNCL and DualCNN released their source codes. SI-UNet and HDNet were reimplemented by us by following their methodological descriptions.

3.4 Comparison with State-of-the-Arts

We first evaluate all methods on reconstructing sparse-view CBCT by using the quarter number of projections, i.e., 1/4 dose, and the evaluation results of the three metrics on the walnut and patient datasets are shown in Table 2.

endtable

From this table, we can have two major observations:

(i) The sparse-view CBCT reconstructed by our method has the highest quality compared with those by other state-of-the-arts. For walnut data, our method decreases RMSE by approximately 16%, increases PSNR by approximately 6%, and increases SSIM by approximately 11%, compared with the second-best method DualCNN. For patient data, the three quality metrics are improved by approximately 18%, 5% and 2%, respectively;

(ii) The reconstruction quality only using RFI-Net's restored full-view projections without any post-refinement is consistently better than those by projection interpolation only (SI-UNet, HDNet*, PDCNN) or post-refinement only (CNCL) methods, and even comparable to those by the two methods (HDNet, DualCNN), which jointly consider both projection interpolation and post-refinement.

Table 2. Quantitative quality on the walnut and patient datasets of **1/4** dose sparse-view CBCT with respect to the full-view CBCT. Best results are in bold. Methods with the marker '*' discard the post-refinement.

Methods	Walnut dataset			Patient dataset		
	RMSE	PSNR(dB)	SSIM	RMSE	PSNR(dB)	SSIM
SVCT	0.128	16.517	0.207	0.025	28.493	0.610
SI-UNet	0.039	26.374	0.662	0.016	31.218	0.870
HDNet*	0.035	27.236	0.686	0.018	30.246	0.804
DualCNN*(PDCNN)	0.034	27.568	0.713	0.012	33.764	0.900
Ours*(RFI-Net)	0.032	28.305	0.790	0.010	35.594	0.924
CNCL	0.038	26.444	0.548	0.014	33.442	0.874
HDNet	0.032	27.854	0.671	0.014	33.016	0.874
DualCNN	0.031	28.347	0.722	0.011	34.683	0.916
Ours	**0.026**	**30.148**	**0.799**	**0.009**	**36.413**	**0.934**

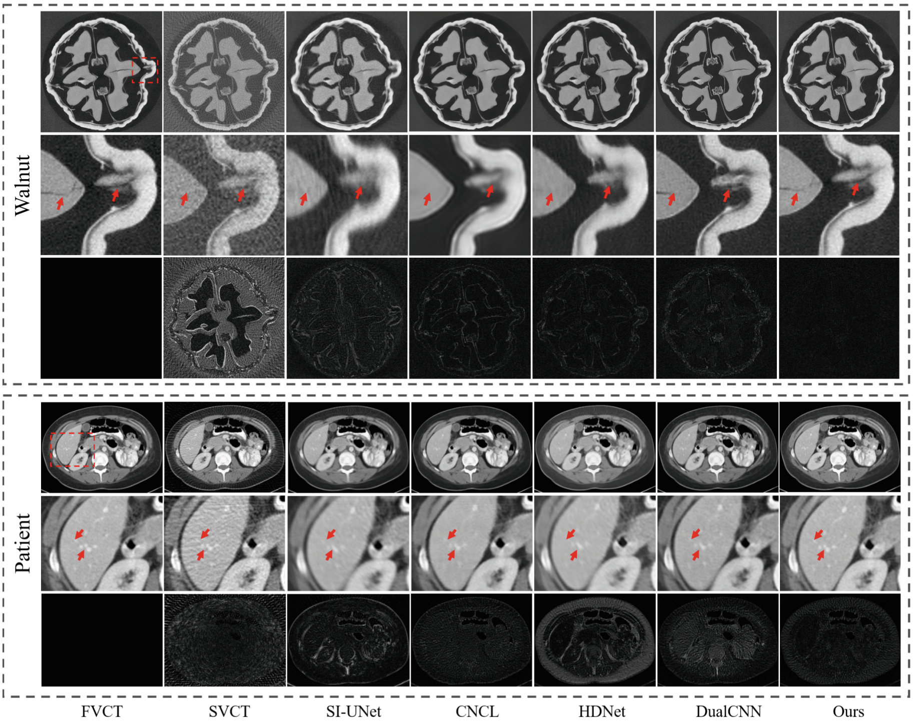

Fig. 4. Visual effects for the first walnut and the patient labeled L273 CBCT reconstruction with **1/4** dose sparse-view projections. For each dataset, the first row shows the representative slices; the second row shows the corresponding magnified ROIs; the third row shows the absolute difference images between the optimized slices and full-view slice.

Figure 4 visualizes the reconstructed sparse-view CBCT of the walnut and patient datasets under 1/4 dose by different methods. As can be seen, SVCT without any improvement contains severe streaking artifacts and hazy structures. SI-UNet, which considers projection interpolation only, suppresses the artifacts to some extent, but still suffers from the lost structures. CNCL, which considers post-refinement only, excessively smooths the reconstructed CBCT image, losing inner textures. HDNet and DualCNN achieve a relatively better visual quality compared with the projection interpolation only and post-refinement only methods, but an unsatisfactory reconstruction result on those delicate structures in Fig. 4. In comparison, our method accurately restores structures without losing inner textures, especially for those tiny structures indicated by the red arrows in the walnut and patient, and thus yields the cleanest difference map with respect to FVCT, as shown in the third rows of the walnut and patient in Fig. 4.

We further explore a more extreme case that reconstructs sparse-view CBCT using only one-eighth projections, i.e., 1/8 dose. Table 3 presents quantitative assessments. Figure 5 presents the visual examples of the walnut and patient datasets for different methods.

Table 3. Quantitative quality on the walnut and patient datasets of **1/8** dose sparse-view CBCT with respect to the full-view CBCT. Best results are in bold. Methods with the marker '*' discard the post-refinement.

Methods	Walnut dataset			Patient dataset		
	RMSE	PSNR (dB)	SSIM	RMSE	PSNR (dB)	SSIM
SVCT	0.184	13.623	0.131	0.048	22.805	0.437
SI-UNet	0.046	24.400	0.555	0.024	27.169	0.774
HDNet*	0.044	24.956	0.576	0.023	27.806	0.773
DualCNN*(PDCNN)	0.040	25.906	0.620	0.018	29.749	0.833
Ours*(RFI-Net)	0.037	27.074	0.727	0.015	32.194	0.873
CNCL	0.048	23.580	0.424	0.019	30.045	0.818
HDNet	0.038	26.202	0.582	0.018	29.979	0.831
DualCNN	0.037	26.224	0.633	0.017	30.829	0.865
Ours	**0.029**	**28.814**	**0.742**	**0.014**	**32.863**	**0.891**

Besides the consistently superior performance of our method over others just like the case of 1/4 dose, we can also have two new observations:

(i) Our method shows more improvement at 1/8 dose compared with 1/4 dose. On the walnut data, PSNR is improved by 2.59 dB at 1/8 dose and 1.80 dB at 1/4 dose comparing with the second-best method DualCNN. On the patient data, PSNR is improved 2.03 dB and 1.73 dB under the 1/8 and 1/4 dose, respectively. This suggests that our method may be useful for the ultra-sparse-view acquisition;

(ii) Comparing with SI-UNet and HDNet*, which only consider projection interpolation, RFI-Net can achieve a better reconstruction quality even with the dose halved. Specially, on the walnut data, with only 1/8 dose, the quality of sparse-view CBCT reconstructed by our method is still comparable to the 1/4 dose sparse-view CBCT reconstructed by HDNet and DualCNN in terms of RMSE and PSNR, and even better in terms of SSIM.

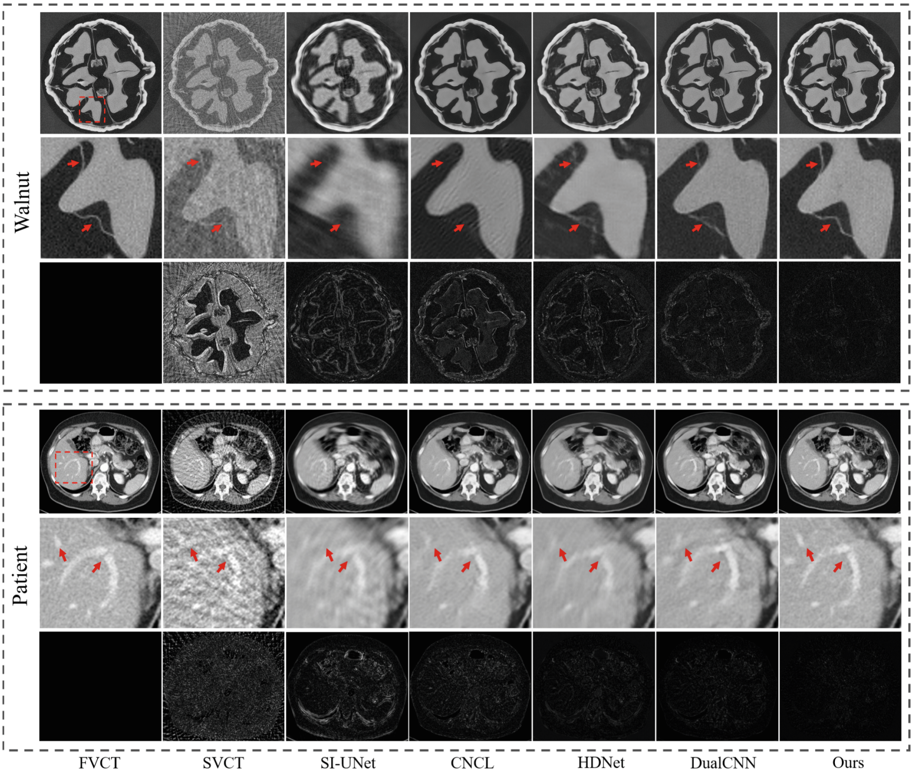

Fig. 5. Visual effects for the first walnut and the patient labeled L299 CBCT reconstruction with **1/8** dose sparse-view projections. For each dataset, the first row shows the representative slices; the second row shows the corresponding magnified ROIs; the third row shows the absolute difference images between the optimized slices and full-view slice.

3.5 Ablation Study

In this section, we investigate the effectiveness of RFI-Net in two aspects: (1) using different training losses, and (2) considering interpolation only, post-

Table 4. The investigation results by (1) using different training losses, and (2) different processing domains. Best results are in bold.

	Settings	Test images	RMSE	PSNR(dB)	SSIM
Loss	$\mathcal{L}_1$	Projections	**0.008**	39.320	0.955
	$\mathcal{L}_1$+0.1*$\mathcal{L}_{RF}$		**0.008**	**39.680**	**0.958**
	$\mathcal{L}_1$+1.0*$\mathcal{L}_{RF}$		0.009	38.553	**0.958**
	$\mathcal{L}_1$+10.0*$\mathcal{L}_{RF}$		0.010	37.859	0.957
	$\mathcal{L}_{RF}$		0.010	37.869	**0.958**
	$\mathcal{L}_1$	CT images	0.038	26.664	0.692
	$\mathcal{L}_1$+0.1*$\mathcal{L}_{RF}$		**0.037**	**27.074**	**0.727**
	$\mathcal{L}_1$+1.0*$\mathcal{L}_{RF}$		**0.037**	26.956	0.725
	$\mathcal{L}_1$+10.0*$\mathcal{L}_{RF}$		0.038	26.794	0.724
	$\mathcal{L}_{RF}$		**0.037**	27.026	0.726
Domain	Post-Net	CT images	0.048	22.463	0.487
	RFI-Net		0.037	27.074	0.727
	RFI-Net+Post-Net		**0.029**	**28.814**	**0.742**

refinement only, and both of them. All investigations are performed on the 1/8 dose sparse-view CBCT reconstruction of walnuts. The results are shown in Table 4.

Ablation Study on RF-Loss. To validate the effectiveness of the proposed RF-Loss, we train five versions of RFI-Net: one only using $\mathcal{L}_1$ loss, one only using $\mathcal{L}_{RF}$ loss and the remaining three using different coefficients on $\mathcal{L}_{RF}$. We assess the interpolated projections shown in the first five rows, and their reconstructed CT images shown in the 6^{th}-10^{th} rows of Table 4.

As can be seen, our method achieves the best performance on both projections and CT images when the coefficient of $\mathcal{L}_{RF}$ is set to 0.1. Besides, increasing the coefficient of $\mathcal{L}_{RF}$ or using single $\mathcal{L}_{RF}$ somewhat slightly degrades the projection quality in terms of RMSE and PSNR metrics, but improves the CT image quality reconstructed from these projections in terms of all the three metrics, especially for SSIM. These results indicate that RF-Loss can indeed interpolate the missing projections containing the FDK-required information. This is exactly what we expected since we are concerned with the high-quality CT images rather than visually similar projections.

Ablation Study on Domain. Furthermore, we build two new complete reconstruction frameworks which use either RFI-Net only or Post-Net only. The quantitative comparison results are presented in the last three rows of Table 4.

As can be seen, without those reconstruction-friendly projections interpolated by RFI-Net, Post-Net produces the reconstructed CT images with the lowest quality especially in terms of SSIM. This verifies that CT post-refinement has limited capability of further reducing the radiation dose, because it is difficult to restore the structures already lost in the reconstruction process. Only using those RFI-Net interpolated projections for reconstruction has already achieved a promising reconstruction quality by improving the SSIM value to that over 70%, while the quality can be further improved with the assistance of a post-refinement by Post-Net.

4 Conclusion

In this paper, we build a sparse-view CBCT reconstruction framework, which can be deeply embedded in the low-dose CBCT systems. This framework consists of three parts: (i) our developed RFI-Net to restore reconstruction-friendly projections from those sparse ones; (ii) FDK to translate 2D X-ray projections into a 3D CT image; (iii) a post-refinement network named Post-Net to further refine the quality of the reconstructed CT image. We also carefully design a novel loss named RF-Loss to help RFI-Net focus on learning FDK-required information of projections. Therefore, our method is expected to significantly improve the quality of sparse-view CBCT with no need to train the entire framework in an end-to-end manner. Experimental results demonstrate that no matter reducing the dose by four or eight times, the sparse-view CBCT reconstructed by our method has the highest quality in all comparison methods, with well persevering delicate structures and presenting the closest quality to that of the full-view CBCT images.

Acknowledgements. This work was supported in part by National Key R&D Program of China (Grant No. 2022YFE0200600), National Natural Science Foundation of China (Grant No. 62202189), Fundamental Research Funds for the Central Universities (2021XXJS033), Science Fund for Creative Research Group of China (Grant No. 61721092), Director Fund of WNLO, Research grants from United Imaging Healthcare Inc.

References

1. Baid, U., et al.: Deep Learning Radiomics Algorithm for Gliomas (DRAG) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11384, pp. 369–379. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_33
2. Bian, J., Siewerdsen, J.H., Han, X., Sidky, E.Y., Prince, J.L., Pelizzari, C.A., Pan, X.: Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. Phys. Med. Biol. **55**(22), 6575 (2010)
3. Brenner, D.J., Hall, E.J.: Computed tomography-an increasing source of radiation exposure. N. Engl. J. Med. **357**(22), 2277–2284 (2007)
4. Callahan, M.J., MacDougall, R.D., Bixby, S.D., Voss, S.D., Robertson, R.L., Cravero, J.P.: Ionizing radiation from computed tomography versus anesthesia for magnetic resonance imaging in infants and children: patient safety considerations. Pediatr. Radiol. **48**(1), 21–30 (2018)

5. Casal, R.F., et al.: Cone beam computed tomography-guided thin/ultrathin bronchoscopy for diagnosis of peripheral lung nodules: a prospective pilot study. J. Thorac. Dis. **10**(12), 6950 (2018)
6. Chao, L., Wang, Z., Zhang, H., Xu, W., Zhang, P., Li, Q.: Sparse-view cone beam CT reconstruction using dual CNNs in projection domain and image domain. Neurocomputing **493**, 536–547 (2022)
7. Chao, L., Zhang, P., Wang, Y., Wang, Z., Xu, W., Li, Q.: Dual-domain attention-guided convolutional neural network for low-dose cone-beam computed tomography reconstruction. Knowledge-Based Systems, p. 109295 (2022)
8. Chen, Z., Qi, H., Wu, S., Xu, Y., Zhou, L.: Few-view CT reconstruction via a novel non-local means algorithm. Physica Med. **32**(10), 1276–1283 (2016)
9. Choi, M., Kim, H., Han, B., Xu, N., Lee, K.M.: Channel attention is all you need for video frame interpolation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 10663–10671 (2020)
10. Der Sarkissian, H., Lucka, F., van Eijnatten, M., Colacicco, G., Coban, S.B., Batenburg, K.J.: A cone-beam x-ray computed tomography data collection designed for machine learning. Sci. Data **6**(1), 1–8 (2019)
11. Ding, A., Gu, J., Trofimov, A.V., Xu, X.G.: Monte Carlo calculation of imaging doses from diagnostic multidetector CT and kilovoltage cone-beam CT as part of prostate cancer treatment plans. Med. Phys. **37**(12), 6199–6204 (2010)
12. Dong, X., Vekhande, S., Cao, G.: Sinogram interpolation for sparse-view micro-CT with deep learning neural network. In: Medical Imaging 2019: Physics of Medical Imaging. vol. 10948, pp. 692–698. SPIE (2019)
13. Geng, M., et al.: Content-noise complementary learning for medical image denoising. IEEE Trans. Med. Imaging **41**(2), 407–419 (2021)
14. Han, Y., Ye, J.C.: Framing u-net via deep convolutional framelets: application to sparse-view CT. IEEE Trans. Med. Imaging **37**(6), 1418–1429 (2018)
15. Hu, D., et al.: Hybrid-domain neural network processing for sparse-view CT reconstruction. IEEE Trans. Radiation Plasma Med. Sci. **5**(1), 88–98 (2020)
16. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
17. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
18. Li, M., Hsu, W., Xie, X., Cong, J., Gao, W.: Sacnn: self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. IEEE Trans. Med. Imaging **39**(7), 2289–2301 (2020)
19. Liao, H., Huo, Z., Sehnert, W.J., Zhou, S.K., Luo, J.: Adversarial sparse-view CBCT artifact reduction. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 154–162. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_18
20. Lin, M., Chen, Q., Yan, S.: Network in network. arXiv preprint arXiv:1312.4400 (2013)
21. McCollough, C., et al.: Low dose CT image and projection data [data set]. The Cancer Imaging Archive (2020)
22. Pan, X., Sidky, E.Y., Vannier, M.: Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? Inverse Prob. **25**(12), 123009 (2009)
23. Paszke, P.: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst (32), 8026

24. Rodet, T., Noo, F., Defrise, M.: The cone-beam algorithm of feldkamp, davis, and kress preserves oblique line integrals. Med. Phys. **31**(7), 1972–1975 (2004)
25. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
26. Shen, T., Li, X., Zhong, Z., Wu, J., Lin, Z.: R^2-Net: recurrent and recursive network for sparse-view CT artifacts removal. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 319–327. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_36
27. Van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Batenburg, K.J., Sijbers, J.: Fast and flexible x-ray tomography using the astra toolbox. Opt. Express **24**(22), 25129–25147 (2016)
28. Wang, Q., Ma, Y., Zhao, K., Tian, Y.: A comprehensive survey of loss functions in machine learning. Annal. Data Sci. **9**(2), 187–212 (2022)
29. Yang, Q., Yan, P., Kalra, M., Wang, G.: Ct image denoising with perceptive deep neural networks. arxiv 2017. arXiv preprint arXiv:1702.07019 (2017)
30. Zeng, G.L.: Revisit of the ramp filter. In: 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–6. IEEE (2014)
31. Zhang, Y., et al.: Clear: comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose CT imaging. IEEE Trans. Med. Imaging **40**(11), 3089–3101 (2021)
32. Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G.: A sparse-view CT reconstruction method based on combination of densenet and deconvolution. IEEE Trans. Med. Imaging **37**(6), 1407–1417 (2018)
33. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging **3**(1), 47–57 (2016)