NUCL SCI TECH (2019) 30:59
https://doi.org/10.1007/s41365-019-0581-7

Slice-wise reconstruction for low-dose cone-beam CT using a deep residual convolutional neural network

Hong-Kai Yang[1,2] · Kai-Chao Liang[1] · Ke-Jun Kang[1,2] · Yu-Xiang Xing[1,2]

Received: 7 July 2018 / Revised: 30 September 2018 / Accepted: 31 October 2018 / Published online: 13 March 2019

Abstract Because of the growing concern over the radiation dose delivered to patients, X-ray cone-beam CT (CBCT) imaging of low dose is of great interest. It is difficult for traditional reconstruction methods such as Feldkamp to reduce noise and keep resolution at low doses. A typical method to solve this problem is using optimization-based methods with careful modeling of physics and additional constraints. However, it is computationally expensive and very time-consuming to reach an optimal solution. Recently, some pioneering work applying deep neural networks had some success in characterizing and removing artifacts from a low-dose data set. In this study, we incorporate imaging physics for a cone-beam CT into a residual convolutional neural network and propose a new end-to-end deep learning-based method for slice-wise reconstruction. By transferring 3D projection to a 2D problem with a noise reduction property, we can not only obtain reconstructions of high image quality, but also lower the computational complexity. The proposed network is composed of three serially connected sub-networks: a cone-to-fan transformation sub-network, a 2D analytical inversion sub-network, and an image refinement sub-network. This provides a comprehensive solution for end-to-end reconstruction for CBCT. The advantages of our method are that the network can simplify a 3D reconstruction problem to a 2D slice-wise reconstruction problem and can complete reconstruction in an end-to-end manner with the system matrix integrated into the network design. Furthermore, reconstruction can be less computationally expensive and easily parallelizable compared with iterative reconstruction methods.

Keywords Cone-beam CT · Slice-wise · Residual U-net · Low dose · Image denoising

This work was supported by the National Natural Science Foundation of China (Nos. 61771279, 11435007) and the National Key Research and Development Program of China (No. 2016YFF0101304).

✉ Yu-Xiang Xing
xingyx@mail.tsinghua.edu.cn

1 Department of Engineering Physics, Tsinghua University, Beijing 100084, China

2 Key Laboratory of Particle and Radiation Imaging, Ministry of Education, Tsinghua University, Beijing 100084, China

1 Introduction

CBCT is widely used in many fields such as clinic diagnosis, public safety inspection, and nondestructive testing. In the field of CBCT reconstruction, analytic methods such as FDK [1] are effective and practical to reconstruct 3D objects. To reduce the potential radiation risk, there are major efforts to lower X-ray radiation dose [2, 3] and to speed up CT scans [4]. Because computer hardware has greatly evolved, researchers have looked to optimization-based iterative methods to reduce noise [5, 6]. Iterative algorithms like algebraic reconstruction techniques (ART)-type [5], SIR [7], and MBIR [8] have been used to reconstruct objects when projection data is noisy. In the past few years, compressed sensing (CS) algorithms have been studied [9, 10]. These studies have shown that high-quality CT images can be reconstructed by iterative methods, which can incorporate both models of imaging physics and additional constraints [11–13]. A typical example is image denoising based on total variation (TV) regularization [14, 15]. However, optimization-based iterative reconstruction is normally computationally expensive. Therefore, it is almost impossible to use such iterative algorithms when real-time reconstruction is required.

Recently, deep learning and convolutional neural networks (CNN) have been widely used in image processing [16–18]. In the CT field, researchers have published some studies using different kinds of CNNs to gain better reconstruction images [19–21]. Researchers have shown that deep learning methods can be used to help with low-dose image denoising, metal artifact reduction, and sparse-data CT [22–25]. Recently, a U-net [26] structure for a large receptive field was also proposed to reduce noise in low-dose CT images [19]. The successful implementation of CNN in solving the problem of noisy data has shown that CT noise caused by the low-dose condition has characteristics that can be learned and reduced by CNNs. However, the networks mentioned above focused only on learning the characteristics in the image domain. Furthermore, not much study has been focused on the issue of cone-beam CT reconstruction. Considering the popular imaging modality of CBCT scan, we intended to catch the features in both the projection domain and the image domain, as well as employ the Radon transform relationship between these two domains. For a certain cross section of an object, we started our work by trying to implement a residual U-net structure. That structure would estimate fan-beam projection data using cone-beam projection data. In this way, we aimed to transfer the problem from 3D imaging to 2D imaging. In this work, we combine the processes of projection estimation, image reconstruction, and image refinement together into one network. Hence, we incorporate the complete imaging physics model of CBCT, as well as characterize data features in both the projection and image domains. An end-to-end reconstruction network is built accordingly.

2 Theory and method

In this section, we first introduce briefly the basic physics of a dental cone-beam CT system and address the problem of our interest. Then, a residual convolutional neural network-type neural network is specifically constructed according to the physics and reconstruction theories of X-ray CT. Detailed network architecture is presented and explained together with the step-by-step training method.

2.1 Physics of a dental cone-beam CT imaging

A typical cone-beam CT imaging system with a circular orbit is shown in Fig. 1. The rotation center is set as the origin O. The point D is the projection of O on the detector plane. As shown in the front view of the detector in Fig. 1, the position of D is denoted as $(c_\mathrm{D}, r_\mathrm{D})$. We denote the projection data as a matrix $\mathbf{P}^{3\mathrm{D}} \in \Re^{(CV)\times R}$, which is acquired with a detector array of R rows and C columns. In total, V projection views are acquired in one scan. To save cost, the detector is designed to cover half the FOV (field of view). For a certain slice of the phantom, an $M \times M$ cross-section attenuation map of the phantom can be denoted as $\boldsymbol{\mu}$. We assume a virtual fan-beam CT scan of $\boldsymbol{\mu}$, and with a C'-column detector covering the full projection of the FOV. The corresponding fan-beam CT projection can be denoted by $\mathbf{q}$, $\mathbf{q} \in \Re^{(VC')\times 1}$, with the same projection views as $\mathbf{P}^{3\mathrm{D}}$.

As shown in Fig. 1, in a cone-beam CT, the image $\boldsymbol{\mu}$ is projected onto as many as R^* rows on the detector, i.e., projections on these rows are related to $\boldsymbol{\mu}$.

2.2 Main architecture of the network

Our proposed reconstruction network is an end-to-end solution, i.e., the input is $\mathbf{P}^{3\mathrm{D}}$ and the output is $\hat{\boldsymbol{\mu}}$. Because $\mathbf{P}^{3\mathrm{D}}$ covers half of the FOV, the network will first estimate the corresponding half of the virtual fan-beam projection, which is denoted as $\hat{\mathbf{p}}$, using $\mathbf{P}^{3\mathrm{D}}$. Then, the missing half of the virtual fan-beam projection (denoted as $\hat{\mathbf{p}}'$) is estimated from $\hat{\mathbf{p}}$. Concatenating $\hat{\mathbf{p}}$ and $\hat{\mathbf{p}}'$ gives a complete set of fan-beam projections $\hat{\mathbf{q}}$. We set $\hat{\mathbf{p}}$, $\hat{\mathbf{p}}' \in \Re^{(VC)\times 1}$. A sub-block of the network executes the filtered back-projection (FBP) algorithm reconstruction. Finally, the network refines $\boldsymbol{\mu}_{\mathrm{FBP}}$ and outputs $\hat{\boldsymbol{\mu}}$. Details of our network are expressed as follows.

The whole network is composed of three sub-networks: the cone-to-fan transformation sub-network, the 2D analytical inversion sub-network, and the image refinement sub-network. Two residual U-nets are utilized in the cone-to-fan transformation and image refinement sub-networks. As shown in Fig. 2, the U-net structure we use consists of four stages connected by pooling layers in the first half and upsampling layers in the second half. The number of channels for the first convolution layer of the U-net is χ.

In the cone-to-fan transformation sub-network, $\mathbf{P}^{3\mathrm{D}}$ is the input. As shown in Fig. 2, in one branch, we reconstruct image $\boldsymbol{\mu}_0$ with $\mathbf{P}^{3\mathrm{D}}$ using an analytical reconstruction algorithm [27]. The simulated available half of the fan-beam projection of $\boldsymbol{\mu}_0$ is denoted as $\mathbf{p}_0$. In another branch, $\mathbf{P}^{3\mathrm{D}}$ is inputted into U-net A. U-net A is expected to learn the characteristics of $\mathbf{P}^{3\mathrm{D}}$ and output the residual between $\mathbf{p}_0$ and $\hat{\mathbf{p}}$. The red arrow denotes the missing half of the fan-beam projection $\hat{\mathbf{p}}'$ that is estimated with $\hat{\mathbf{p}}$, according to the symmetry over rotation of a fan-beam scan [28].

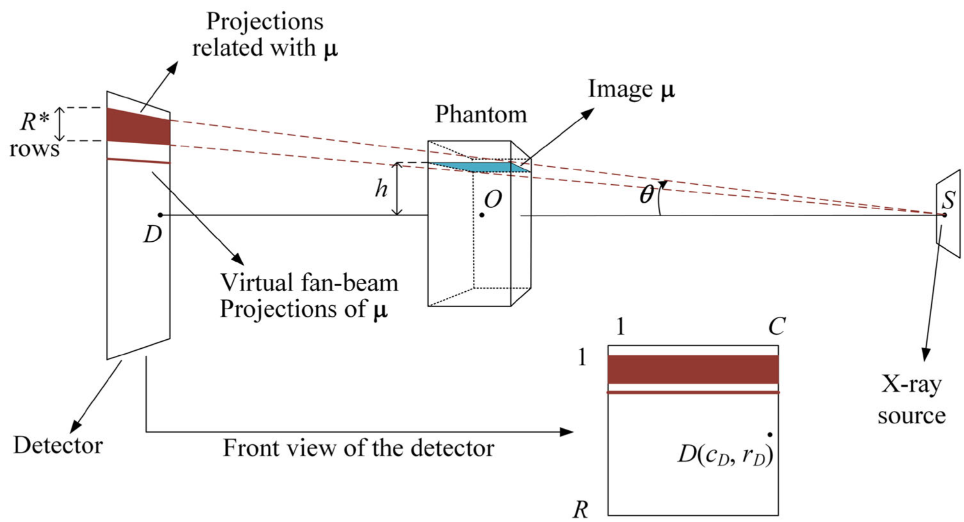

Fig. 1 Physical schematic diagram of a cone-beam CT imaging system (Color online)

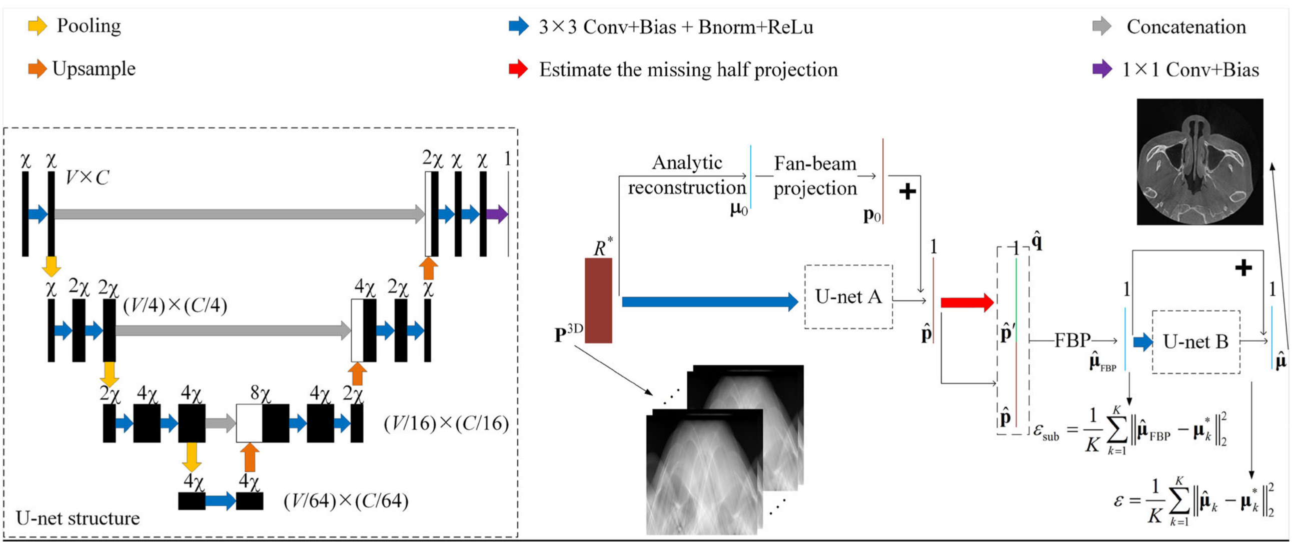

Fig. 2 Main structure of the dual-domain deep learning network (Color online)

According to our experience, CT reconstruction is sensitive to errors in projection data. It is hard for the cone-to-fan transformation sub-network to fully catch the characteristics of a sinogram. Hence, we configured a sinogram-to-image-domain transformation sub-network. We exactly follow the computation as an analytical reconstruction and hence refer to it as "2D analytical inversion sub-network." The procedure of this sub-network includes three steps: weighting, filtration, and back-projection. All three of these steps can be realized by matrix–vector multiplication. Therefore, we can formulate the 2D analytical inversion sub-network as FBP reconstruction steps:

$$\hat{\mu}_{\mathrm{FBP}} = \mathbf{H}_{\mathrm{w}}^{\mathrm{T}}\mathbf{F}\mathbf{W}\hat{\mathbf{q}} \tag{1}$$

here $\mathbf{W}$ is a diagonal matrix for ray-by-ray weighting, the matrix $\mathbf{F}$ represents a ramp filtration process in the detector axis for all views, and $\mathbf{H}_{\mathrm{w}}^{\mathrm{T}}$ is a weighted back-projection operator. The superscript T denotes a matrix transform. These three matrices are predetermined by the CT scanning geometry and can be pre-calculated. The 2D analytical inversion sub-network essentially executes a Radon inverse transform and completes the domain transformation. Subsequently, the image-domain loss can be easily fitted in. Because the inverse Radon transform is the physics model in CT imaging, this sub-network is deterministic and does not need to be trained, in theory. In other words, there is no learning in this sub-network.

The outputs of the 2D analytical inversion sub-network $\hat{\boldsymbol{\mu}}_{\mathrm{FBP}}$ are further fed forward into an image refinement sub-network. As shown in Fig. 2, the image refinement sub-network also has a residual U-net CNN structure. This sub-network (U-net B) further refines $\hat{\boldsymbol{\mu}}_{\mathrm{FBP}}$ and outputs the final reconstruction of the linear attenuation map $\hat{\boldsymbol{\mu}}$ of the slice of interest. According to our experience, a feasible choice is to set χ to 24 in U-net A and to 16 in U-net B. We refer to this whole proposed network as the dual-domain deep learning (DDL) reconstruction.

2.3 Network training

When training, we first train the cone-to-fan transformation sub-network separately by a loss of l2-norm, $\varepsilon_{\mathrm{sub}} = \frac{1}{K}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{\mu}}_{\mathrm{FBP}} - \boldsymbol{\mu}_k^*\right\|_2^2$, with $\boldsymbol{\mu}_k^*$ being the kth image label, and K being the number of images in a training set. Then, we train the whole network together. Similarly, the ultimate loss for this network is $\varepsilon = \frac{1}{K}\sum_{k=1}^{K}\left\|\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}_k^*\right\|_2^2$. The error backpropagation algorithm for the cone-to-fan transformation sub-network and the image refinement sub-network is not different from the commonly used stochastic gradient descent (SGD) in the deep learning field [26]. From Eq. (1), we can see that the chain rule for back-propagation errors through the 2D analytical inversion sub-network can simply be written as:

$$\frac{\partial \varepsilon}{\partial \hat{\mathbf{q}}} = \mathbf{W}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{H}_{\mathrm{w}}\frac{\partial \varepsilon}{\partial \hat{\boldsymbol{\mu}}_{\mathrm{FBP}}}. \tag{2}$$

3 Experiments and results

To examine the performance of the proposed method, we arranged our research to reconstruct a certain slice of a cone-beam dental CT.

3.1 Validating the network on simulated low-dose CT data

In total, 110 patients' normal-dose CT projection data were obtained from a cone-beam dental CT. Among them, 100 randomly chosen patients' data were used in training, while the other 10 patients' data were used for validation. All these 110 patients were chosen randomly from hospitals. The personal information of these patients was anonymized. The data were taken from the same dental CT system. Training data and validation data were independent. For the cone-beam scan system, the source-to-origin distance (l_{SO}) was 485 mm, and detector-to-origin distances was 250 (l_{DO}) mm. A flat-panel detector of 658 rows and 656 columns with bin size (b_{D}) 0.2^2 mm^2 was used, i.e., R = 658, C = 656. The position of Point D on the detector, $(c_{\mathrm{D}}, r_{\mathrm{D}})$ equaled (329.5, 637.5). During each scan, 600 projections were taken. The reconstruction area was on the 640 × 640 (M^2) grid with voxel size b_{I} of 0.25^2 mm^2. We reconstructed a slice 35.375 mm (i.e., h = 35.375 mm as shown in Fig. 1) away from the mid-plane and the maximum cone angle θ of the involved data was 5.02 degrees. As a result, R^* rows of detector bins were needed in this case. R^* can be calculated by:

$$R^* = \frac{(l_{\mathrm{DO}} + l_{\mathrm{SO}})b_{\mathrm{I}}M\tan\theta}{b_{\mathrm{D}}(l_{\mathrm{SO}} + b_{\mathrm{I}}M/2)}. \tag{3}$$

To achieve labels for our network, we collected data scanned with the X-ray source set to be 100 kV and 4 mA. Each projection was acquired in 20 ms. This was deemed a normal-dose situation, and the corresponding blank scan was denoted as $\mathbf{I}_{\mathrm{N}}$. We simulated the low-dose projections $\mathbf{P}^{3\mathrm{D}}_{\mathrm{L,\,train}}$ and $\mathbf{P}^{3\mathrm{D}}_{\mathrm{L,\,validation}}$, with blank scan $\mathbf{I}_{\mathrm{L}}$ randomly chosen according a Poisson distribution with mean equal to 20–25% of $\mathbf{I}_{\mathrm{N}}$. $\mathbf{P}^{3\mathrm{D}}_{\mathrm{L,\,train}}$, $\boldsymbol{\mu}_{\mathrm{N,train}}$ formed paired data for the training of the proposed network.

We denoted the projection data in the training and validation sets as $\mathbf{P}^{3\mathrm{D}}_{\mathrm{N,\,train}}$ and $\mathbf{P}^{3\mathrm{D}}_{\mathrm{N,\,validation}}$, respectively. A statistical image reconstruction algorithm using nonlocal mean (NLM) regularization [29] was applied to obtain the labels,

$$\hat{\boldsymbol{\mu}}^{3\mathrm{D}} = \arg\min_{\boldsymbol{\mu}^{3\mathrm{D}}}\left\{\left\|\mathbf{H}^{3\mathrm{D}}\boldsymbol{\mu}^{3\mathrm{D}} - \mathbf{P}^{3\mathrm{D}}_{\mathrm{N,\,train}}\right\|_2^2 + \beta\sum_{j=1}^{J} f_{\mathrm{NLM}}\left[\boldsymbol{\mu}^{(j)}\right]\right\}, \tag{4}$$

with $\boldsymbol{\mu}^{3\mathrm{D}}$ and $\mathbf{H}^{3\mathrm{D}}$ the 3D object and the corresponding projection matrix, respectively, $\boldsymbol{\mu}^{(j)}$ the jth slice of $\boldsymbol{\mu}^{3\mathrm{D}}$ with J slices in total, β a weighting coefficient, and $f_{\mathrm{NLM}}(\cdot)$ the slice-wise NLM cost function. We solved Eq. (4) iteratively using an analytical reconstruction from $\mathbf{P}^{3\mathrm{D}}_{\mathrm{N,\,train}}$ as an initial value. After $\hat{\boldsymbol{\mu}}^{3\mathrm{D}}$ was obtained, the slice that was 35.375 mm away from the mid-plane was extracted to be the label $\boldsymbol{\mu}_{\mathrm{N}}$ of our network.

When training the cone-to-fan transformation sub-network separately, the loss function of the network $\varepsilon_{\mathrm{sub}}$ reached its convergence (i.e., relative change of $\varepsilon_{\mathrm{sub}}$ was less than 0.01%) in about 250 epochs. Subsequently, the whole network was trained together. The relative change of the ultimate loss function ε was less than 0.01% after another 150 epochs. The convergence curve in the training process is plotted in Fig. 3. In validation, $\mathbf{P}^{3\mathrm{D}}_{\mathrm{L,\,validation}}$ was the input data with $\boldsymbol{\mu}_{\mathrm{N,validation}}$ used as a reference for performance evaluation.

As a demonstration, we show some intermediate results in Fig. 4. We can see that analytical reconstruction can provide high-frequency information for the slice of interest. The cone-to-fan transformation sub-network can recover the low-frequency information caused by the approximation in analytical reconstruction for the cone-beam problem. Details of intermediate results from the cone-to-fan

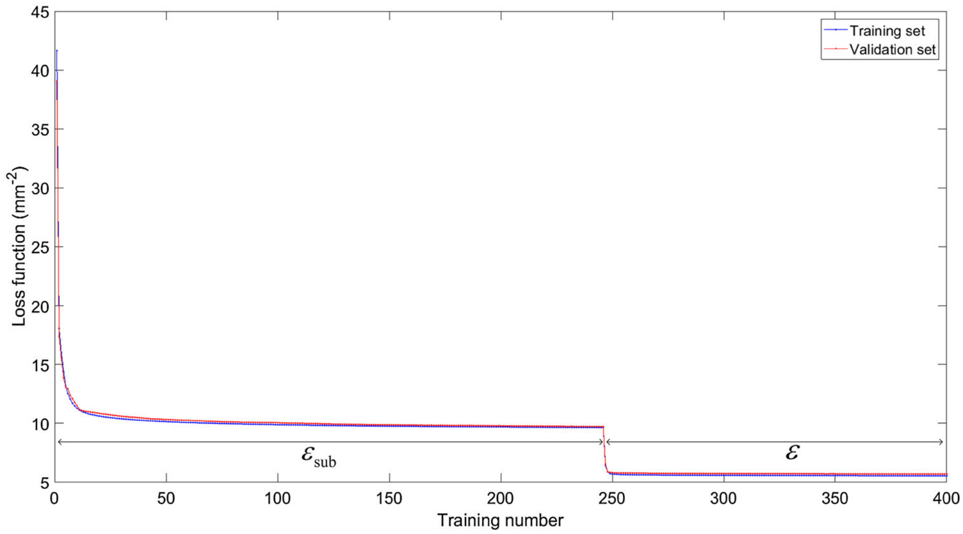

Fig. 3 Convergence curve in the training process (Color online)

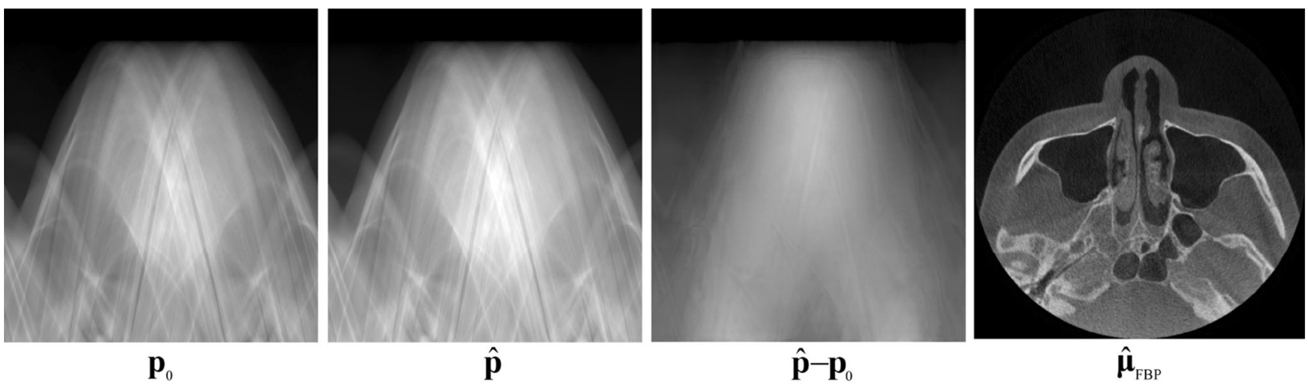

Fig. 4 Intermediate results of the proposed network. Display window for $\mathbf{p}_0$ and $\hat{\mathbf{p}}$: [0 4.0]. Display window for $\hat{\mathbf{p}} - \mathbf{p}_0$: [0 0.6]. Display window for $\hat{\boldsymbol{\mu}}_{FBP}$: [0 0.07]

Table 1 Quantitative comparison of reconstruction methods

Methods	RRMSE	SSIM	SNR
DDL with U-net	0.0589 ± 0.0016	0.9957 ± 0.0006	20.6768 ± 0.5722
Estimation of $\boldsymbol{\mu}_{FBP}$	0.0704 ± 0.0027	0.9939 ± 0.0007	19.1384 ± 0.5251
DDL with plain CNN	0.0718 ± 0.0017	0.9936 ± 0.0008	18.9579 ± 0.5572
Image-domain U-net	0.0768 ± 0.0017	0.9927 ± 0.0009	18.3803 ± 0.5373
Analytical reconstruction	0.1629 ± 0.0097	0.9682 ± 0.0040	11.8569 ± 0.6597

transformation sub-network ($\hat{\boldsymbol{\mu}}_{FBP}$) are also provided in Table 1.

Using low-dose data, we compared our method with U-net [19, 26] only in the image domain and with analytical reconstruction [27]. The image-domain U-net-based method also used $\boldsymbol{\mu}_N$ as labels and used $\boldsymbol{\mu}_0$, i.e., the analytical reconstructions [27] of $\mathbf{P}^{3D}_{L,\,train}$, as inputs. We also replaced the U-net A&B in our method with plain CNN, which has approximately the same computation as the U-net we used. U-net A was replaced by a nine-layer CNN, which contained 24, 24, 24, 24, 24, 12, 12, 12, and 1 kernels (3 × 3), respectively. U-net B was replaced with a five-layer CNN, which contained 16, 32, 16, 16, and 1 kernels (3 × 3), respectively. These results are included in our comparison.

Three cases in the validation data set are shown in Fig. 5 with zoom-in of the region of interest (ROI) in the blue boxes for demonstration. The horizontal profiles of the difference images between the reconstructions and the labels along the red line in Fig. 5 are plotted in Fig. 6. We can see that structural information is severely contaminated in analytical reconstructions for low-dose data. Image-

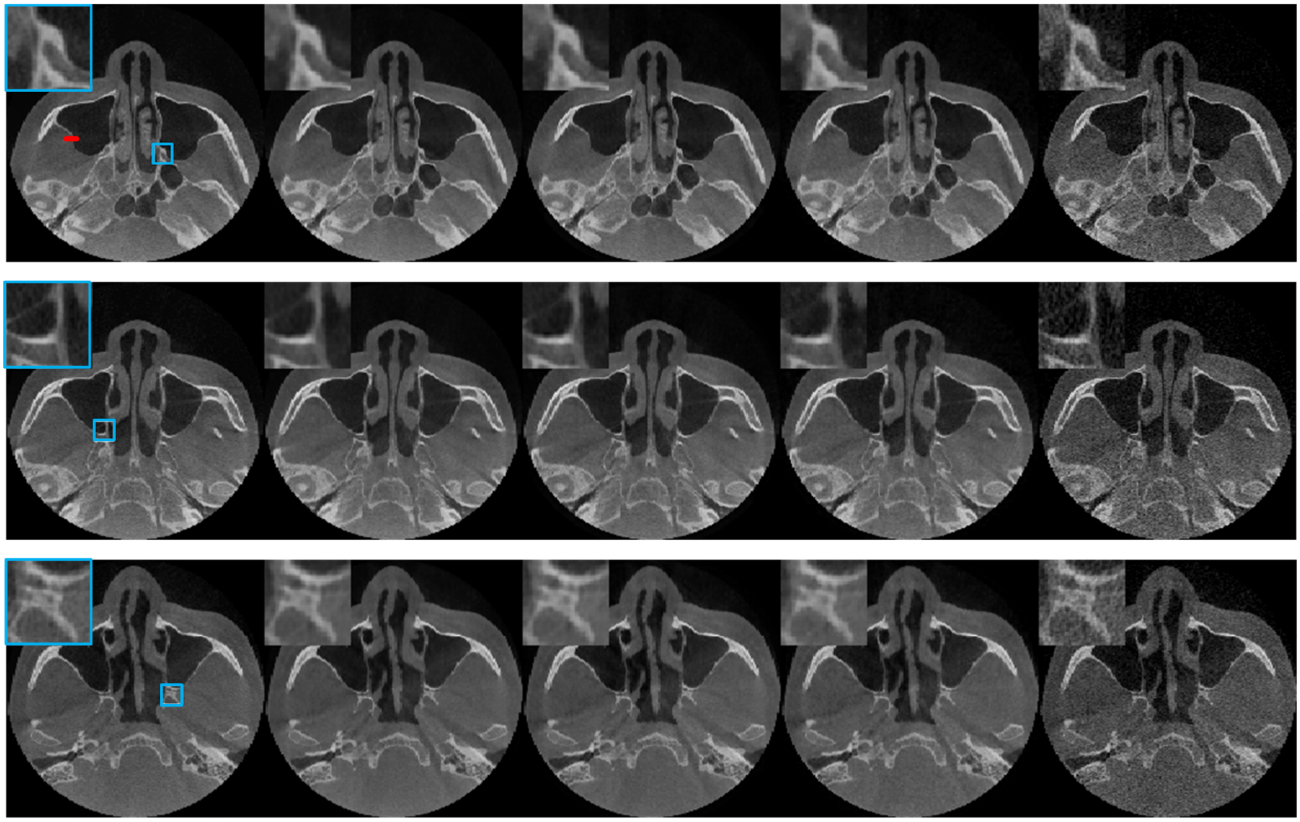

Fig. 5 Reconstructions in validation set. From left to right: normal-dose reconstruction (labels), low-dose reconstructions from DDL with U-net, DDL with plain CNN, image-domain U-net, analytical reconstruction. Display window: [0 0.07]

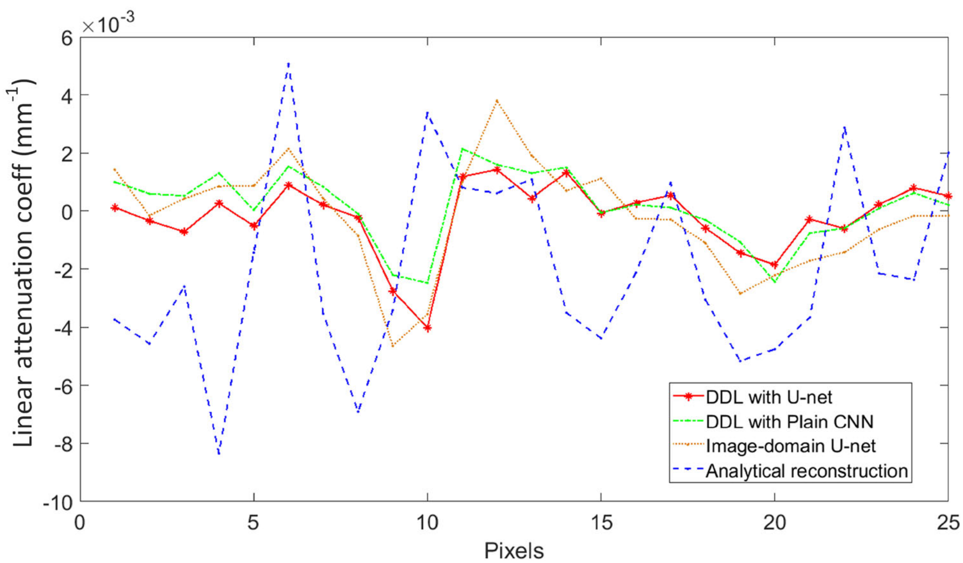

Fig. 6 Horizontal profiles of the difference image of the reconstructions and labels along the red line in Fig. 5. The red and green lines of DDL results show smaller differences from the labels than the other two lines, and the red line of DDL with U-net result is slightly better than the green line (Color online)

domain U-net-based methods can work well in reducing noise but the edges are a little blurred. The proposed network performs best in recovering structural details comparable to the normal-dose case.

Moreover, we quantitatively evaluated the image quality of all validation reconstructions in terms of relative root mean square error (RRMSE), the structural similarity (SSIM) index, and signal-to-noise ratio (SNR). The RRMSE index for each reconstruction image was computed as,

$$\mathrm{RRMSE} = \frac{\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\mathrm{N}}\|_2}{\|\boldsymbol{\mu}_{\mathrm{N}}\|_2}, \tag{5}$$

with $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}_{\mathrm{N}}$ denoting the output of the network and the corresponding label, respectively. The SSIM index for each reconstruction image was computed as,

$$\mathrm{SSIM} = \frac{\left(2\bar{\hat{\boldsymbol{\mu}}}\bar{\boldsymbol{\mu}}_{\mathrm{N}} + C_1\right)\left(2\sigma_{\hat{\boldsymbol{\mu}}\boldsymbol{\mu}_{\mathrm{N}}} + C_2\right)}{\left(\bar{\hat{\boldsymbol{\mu}}}^2 + \bar{\boldsymbol{\mu}}_{\mathrm{N}}^2 + C_1\right)\left(\sigma_{\hat{\boldsymbol{\mu}}}^2 + \sigma_{\boldsymbol{\mu}_{\mathrm{N}}}^2 + C_2\right)}, \tag{6}$$

where $\bar{\hat{\boldsymbol{\mu}}}$ and $\bar{\boldsymbol{\mu}}_{\mathrm{N}}$ are the means of $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}_{\mathrm{N}}$, respectively, $\sigma_{\hat{\boldsymbol{\mu}}}$ and $\sigma_{\boldsymbol{\mu}_{\mathrm{N}}}$ are standard deviations of $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}_{\mathrm{N}}$, and $\sigma_{\hat{\boldsymbol{\mu}}\boldsymbol{\mu}_{\mathrm{N}}}$ is the cross correlation. The constants C_1 and C_2 are stabilizers. The SNR was computed as,

$$\mathrm{SNR} = 10\log_{10}\left[\frac{\sum_j \left(\boldsymbol{\mu}_{\mathrm{N}-j} - \bar{\boldsymbol{\mu}}_{\mathrm{N}}\right)^2}{\sum_j \left(\hat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}_{\mathrm{N}-j}\right)^2}\right], \tag{7}$$

where j indexes the pixels. Results are shown in Table 1. It is shown that our DDL design is the most important factor for realizing high-quality reconstructions in this low-dose CT reconstruction problem. The RRMSE, SSIM, and SNR of $\hat{\boldsymbol{\mu}}_{\mathrm{FBP}}$ in Table 1 show that the cone-to-fan transformation sub-network can do a good job in the estimation of fan-beam projections, leading to rather good images from the subsequent analytical inversion sub-network. The additional image refinement sub-network further improves the image quality of the reconstructions.

Fig. 7 Phantom used in practical experiments

3.2 Validating the network on practical CT data

We also imaged a skull head phantom (shown in Fig. 7) for practical experiments. The low-dose scans were done on the same dental CT system used above. The voltage of the X-ray source was set to 80 kV and the current was set to 2 mA. Under this low-dose situation, the blank scan $\mathbf{I}_{\mathrm{L}}$ was about 25% of $\mathbf{I}_{\mathrm{N}}$. The reconstructions are shown in Fig. 8.

From these results, we can see that the proposed method can effectively reduce the noise in practical low-dose CT reconstructions. All main structures of the phantom are well reconstructed by the trained network.

4 Computational complexity

The computational complexity of the proposed method can be estimated from the computation load of the three sub-networks. The major computation load in both cone-to-fan transformation and image refinement sub-networks is in the convolution layers. If we count the multiplications only, there will be

$$X_l \times Y_l \times \text{kernel size} \times [\text{kernel}\,\#\,\text{of the}\,(l-1)\text{th layer}] \\ \times (\text{kernel}\,\#\,\text{of the}\,l\text{th layer})$$

multiplications for the lth convolution layer. Here, l is the layer index, with channels of dimension ($X_l \times Y_l$). Hence, the number of multiplications in the network will be approximately,

$$\sum_{l=2}^{L} X_l \times Y_l \times (\alpha_l \times \alpha_l) \times \kappa_{l-1} \times \kappa_l \approx 5 \text{ projections}.$$

Here, κ_l and $\alpha_l \times \alpha_l$ denote the channel number and the kernel size of the lth convolution layer, respectively, and L is the total number of convolution layers in the network. In Table 2, we list the multiplications in CNN layers of the dominant computation load. In our experiments, $R^* = 92$, $R = 658$, $C = 656$, $M = 640$, $\chi_A = 24$, and $\chi_B = 16$. The computation of the 2D analytical inversion sub-network is mainly a 2D back-projection. In addition to the computation in convolution layers, two back-projections and one projection are needed. For simplicity, we treat projection and back-projection as having the same amount of computational complexity. Therefore, we can see that the computational complexity is approximately $(5+3)/2 = 4$ 2D iterations in an iterative reconstruction method. This is much faster than many iterative methods in this field. We use an Intel core i7-5930 K CPU (3.5 GHz), and a GeForce GTX TITAN X GPU. With our method, the time–cost of one-slice reconstruction is about 5.2 s, while an NLM

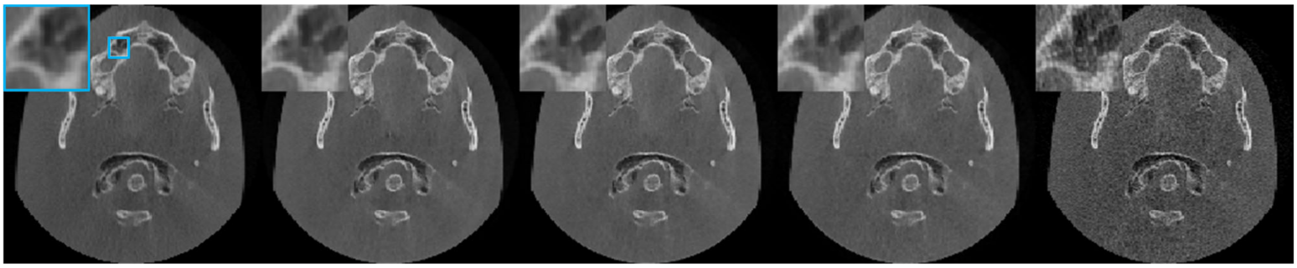

Fig. 8 Reconstructions of a skull head phantom. From left to right: low-dose reconstructions from statistical reconstruction using NLM regularization, DDL with U-net, DDL with plain CNN, image-domain U-net, and analytical reconstruction. Display window: [0 0.07]

Table 2 Multiplications in CNN layers of dominant computation load

Layer #	Multiplications in U-net A	Layer #	Multiplications in U-net B
2	$9RCR^{*}\chi_{\mathrm{A}}$	16	$9M^2\chi_{\mathrm{B}}$
3	$9RC\chi_{\mathrm{A}}^2$	17	$9M^2\chi_{\mathrm{B}}^2$
4	$(9/8)RC\chi_{\mathrm{A}}^2$	18	$(9/8)M^2\chi_{\mathrm{B}}^2$
5	$(9/4)RC\chi_{\mathrm{A}}^2$	19	$(9/4)M^2\chi_{\mathrm{B}}^2$
11	$(9/2)RC\chi_{\mathrm{A}}^2$	25	$(9/2)M^2\chi_{\mathrm{B}}^2$
12	$(9/8)RC\chi_{\mathrm{A}}^2$	26	$(9/8)M^2\chi_{\mathrm{B}}^2$
13	$18RC\chi_{\mathrm{A}}^2$	27	$18M^2\chi_{\mathrm{B}}^2$
14	$9RC\chi_{\mathrm{A}}^2$	28	$9M^2\chi_{\mathrm{B}}^2$

regularized iterative method takes about 260.2 s to converge.

5 Conclusion

We propose a new framework of X-ray CT reconstruction based on deep learning for slice-wise reconstruction in a cone-beam CT system. The proposed method utilizes a novel structure containing three parts, which were designed for cone-to-fan projection estimation, 2D analytical inversion transform, and image refinement, respectively. The cone-to-fan transformation and the image refinement subnetworks are both built using residual U-net structures. A 2D analytical inversion transformation sub-network completes the domain transformation from projection domain to image domain. The cone-to-fan transformation subnetwork is trained first. Then, the whole network is trained using ultimate image-domain loss. Our results with a realistic phantom show that the proposed method can effectively reduce noise and recover detailed structures in scanned objects. Reconstructions have higher image quality than commonly used low-dose cone-beam CT reconstructions.

It is worth pointing out that when the targeted slice is farther away from the mid-plane of a cone-beam CT, the projection of the slice is contained in multiple detector rows. The data of the slice of interest are mixed with many other slices, and this makes it more challenging to obtain the 2D projection of the slice. Because the cone angle of practical CT systems is usually within -5 to $+5$ degrees, we have researched the most difficult situation, where the cone angle is about 5 degrees, as an example. By building up similar branches for different slices, one could conveniently reconstruct a volume of interest or multiple inconsecutive slices.

This proposed network is initially designed to incorporate the imaging physics (modeled by a CT system matrix) in the network design so that it can learn the characteristics of both projection and image domains in an end-to-end mechanism. It combines the capability of physical models and information mining from big data sets. Moreover, this network simplifies the 3D imaging process by transferring it into a 2D form so that only a 2D system matrix is needed in the projection-to-image-domain transfer thereby reducing the memory requirement. By decoupling the 3D projection into an independent 2D problem, significant computation time can be saved compared with 3D projection and back-projection in iterative methods. Finally, the reconstruction using the trained network can be completed with good speed using currently available parallel-computing power. These advantages could be greatly beneficial to real-time applications.

In this work, we use dental CBCT data to confirm the effectiveness of our method of reconstructing a certain slice of a scanned object. We do not consider the issue of metal artifacts in this work. Our group is working on restraining metal artifacts as a separate problem [30]. For future work, we plan to combine our work together and further optimize the network for 3D volumes. We will extend the method to other CT scan geometries as well.

References

1. Y. Xing, L. Zhang, A free-geometry cone beam CT and its FDK-type reconstruction. J. X-ray Sci. Technol. **15**(3), 157–167 (2007)
2. K. Ozasa, Epidemiological research on radiation-induced cancer in atomic bomb survivors. J. Radiat. Res. **57**(Suppl 1), i112–i117 (2016). https://doi.org/10.1093/jrr/rrw005
3. D.L. Miglioretti, E. Johnson, A. Williams et al., The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk. Jama Pediatr. **167**(8), 700–707 (2013). https://doi.org/10.1001/jamapediatrics.2013.311
4. L.J.M. Kroft, J.J.H. Roelofs, J. Geleijns, Scan time and patient dose for thoracic imaging in neonates and small children using axial volumetric 320-detector row CT compared to helical 64-, 32-, and 16- detector row CT acquisitions. Pediatr. Radiol. **40**(3), 294–300 (2010). https://doi.org/10.1007/s00247-009-1436-x
5. A.C. Silva, H.J. Lawder, A. Hara et al., Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. AJR Am. J. Roentgenol. **194**(1), 191–199 (2010). https://doi.org/10.2214/AJR.09.2953
6. A.K. Hara, R.G. Paden, A.C. Silva et al., Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. AJR Am. J. Roentgenol. **193**(3), 764–771 (2009). https://doi.org/10.2214/AJR.09.2397
7. I.A. Elbakri, J.A. Fessler, Statistical image reconstruction for polyenergetic X-ray computed tomography. IEEE Trans. Med. Imaging **21**(2), 89–99 (2002). https://doi.org/10.1109/42.993128
8. K. Li, J. Tang, G.H. Chen, Statistical model based iterative reconstruction (MBIR) in clinical CT systems: experimental assessment of noise performance. Med. Phys. (2014). https://doi.org/10.1118/1.4867863
9. P.T. Lauzier, J. Tang, G. Chen, Prior image constrained compressed sensing: implementation and performance evaluation. Med. Phys. **39**(1), 66–80 (2012). https://doi.org/10.1118/1.3666946
10. G.H. Chen, J. Tang, S. Leng, Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. Med. Phys. **35**(2), 660–663 (2008). https://doi.org/10.1118/1.2836423
11. J. Liu, Y. Hu, J. Yang et al., 3D feature constrained reconstruction for low dose CT imaging. IEEE Trans. Circuits Syst. Video Technol. **28**(5), 1232–1247 (2016). https://doi.org/10.1109/TCSVT.2016.2643009
12. Y. Chen, L. Shi, Q. Feng et al., Artifact suppressed dictionary learning for low-dose CT image processing. IEEE Trans. Med. Imaging **33**(12), 2271–2292 (2014). https://doi.org/10.1109/TMI.2014.2336860
13. J. Liu, J. Ma, Y. Zhang et al., Discriminative feature representation to improve projection data inconsistency for low dose CT imaging. IEEE Trans. Med. Imaging **36**(12), 2499–2509 (2018). https://doi.org/10.1109/TMI.2017.2739841
14. E.Y. Sidky, X. Pan, Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys. Med. Biol. **53**(17), 4777–4807 (2008). https://doi.org/10.1088/0031-9155/53/17/021
15. X. Jia, Y. Lou, J. Lewis et al., GPU-based fast low-dose cone beam CT reconstruction via total variation. J. X-ray Sci. Technol. **78**(3), 139–154 (2010). https://doi.org/10.3233/XST-2011-0283
16. Y. Lecun, Y. Bengio, G. Hinton, Deep learning. Nature **521**(7553), 436 (2015). https://doi.org/10.1038/nature14539
17. T.Y. Lin, A. Roychowdhury, S. Maji, Bilinear CNN models for fine-grained visual recognition. IEEE Int. Conf. Comput. Vis. (2015). https://doi.org/10.1109/iccv.2015.170
18. A. Ramcharan, K. Baranowski, P. Mccloskey et al., Deep learning for image-based cassava disease detection. Front. Plant Sci. (2017). https://doi.org/10.3389/fpls.2017.01852
19. E. Kang, J. Min, J.C. Ye, A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med. Phys. **44**(10), e360–e375 (2017). https://doi.org/10.1002/mp.12344
20. H. Chen, Y. Zhang, M.K. Kalra et al., Low-dose CT with a residual encoder-decoder convolutional neural network (RED-CNN). IEEE Trans. Med. Imaging **99**, 1 (2017). https://doi.org/10.1109/TMI.2017.2715284
21. H. Chen, Y. Zhang, W. Zhang et al., Low-dose CT via convolutional neural network. Biomed. Opt. Express **8**(2), 679–694 (2017). https://doi.org/10.1364/BOE.8.000679
22. Q. Yang, P. Yan, Y. Zhang et al., Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans. Med. Imaging **37**(6), 1348–1357 (2018). https://doi.org/10.1109/TMI.2018.2827462
23. Y. Zhang, H. Yu, Convolutional neural network based metal artifact reduction in X-ray computed tomography. IEEE Trans. Med. Imaging **37**(6), 1370–1381 (2018). https://doi.org/10.1109/TMI.2018.2823083
24. Y. Han, J.C. Ye, Framing U-net via deep convolutional framelets: application to sparse-view CT. IEEE Trans. Med. Imaging **37**(6), 1418–1429 (2018). https://doi.org/10.1109/TMI.2018.2823768
25. H. Chen, Y. Zhang, Y. Chen et al., LEARN: learned experts' assessment-based reconstruction network for sparse-data CT. IEEE Trans. Med. Imaging **37**(6), 1333–1347 (2018). https://doi.org/10.1109/TMI.2018.2805692
26. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. Int. Conf. Med. Image Comput. Comput. Assist. Interv. **9351**, 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
27. J. Hao, L. Zhang, L. Li et al., A practical image reconstruction and processing method for symmetrically off-center detector CBCT system. Nucl. Sci. Technol. **24**(4), 17–22 (2013)
28. D.L. Parker, Optimal short scan convolution reconstruction for fan beam CT. Med. Phys. **9**(2), 254–257 (1982). https://doi.org/10.1118/1.595078
29. H. Zhang, J. Ma, J. Wang et al., Statistical image reconstruction for low-dose CT using nonlocal means-based regularization. Comput. Med. Imaging Graph. **38**(6), 423–435 (2014). https://doi.org/10.1016/j.compmedimag.2014.05.002
30. K. Liang, L. Zhang, H. Yang, et al. Optimize interpolation-based MAR for practical dental CT with deep learning, in *The 5th International Conference on Image Formation in X-ray Computed Tomography* (*CT meeting 2018*) (2018), pp. 423–425