IPEM Institute of Physics and Engineering in Medicine

Artifact removal using a hybrid-domain convolutional neural network for limited-angle computed tomography imaging

Artifact Removal Using a Hybrid-Domain Convolutional Neural Network for Limited-Angle Computed Tomography Imaging

Qiyang Zhang[1,2,3]**, Zhanli Hu**[2]**, Changhui Jiang**[1,2,3]**, Hairong Zheng**[2]**, Yongshuai Ge**[1,2,4] **and Dong Liang**[1,2,4]

[1] Research Center for Medical Artificial Intelligence, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

[2] Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China

[3] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China

[4] Author to whom any correspondence should be addressed

E-mail: ys.ge@siat.ac.cn and dong.liang@siat.ac.cn

Abstract. The suppression of streak artifacts in computed tomography with a limited-angle configuration is challenging. Conventional analytical algorithms, such as filtered backprojection (FBP), are not successful due to incomplete projection data. Moreover, model-based iterative total variation (TV) algorithms effectively reduce small streaks but do not work well at eliminating large streaks. In contrast, FBP mapping networks and deep-learning-based postprocessing networks are outstanding at removing large streak artifacts; however, these methods perform processing in separate domains, and the advantages of multiple deep learning algorithms operating in different domains have not been simultaneously explored. In this paper, we present a hybrid-domain convolutional neural network (hdNet) for the reduction of streak artifacts in limited-angle computed tomography. The network consists of three components: the first component is a convolutional neural network operating in the sinogram domain, the second is a domain transformation operation, and the last is a convolutional neural network operating in the CT image domain. After training the network, we can obtain artifact-suppressed CT images directly from the sinogram domain. Verification results based on numerical, experimental and clinical data confirm that the proposed method can significantly reduce serious artifacts.

Keywords: computed tomography (CT), convolutional neural network (CNN), domain transformation, limited-angle, streak artifacts

1. Introduction

X-ray computed tomography (CT) produces clear, high-quality images and plays a very important role in clinical diagnostics and image-guided interventions. X-ray CT can help radiologists identify and diagnose infectious illnesses, musculoskeletal conditions,

2

cardiovascular ailments, trauma and even certain kinds of cancers. However, in some practical applications of CT, such as mammography (Wang et al. 2011) and large-pitch helical CT (Li et al. 2006), the data acquisition angle is often limited by the size of the object and the flexibility of the scan. These factors lead to incomplete data in the Radon domain, an issue known as the limited-angle problem, which presents a serious challenge in CT imaging tasks. Generally, when the projection data cover a scan angle range of less than 180 degrees (parallel-beam tomography), serious artifacts may arise in the reconstructed image, and thus, high-quality reconstruction cannot be achieved (Tuy 1983, Li et al. 2008). Recently, however, researchers have reported significant successes in the suppression of image artifacts with regard to the limited-angle problem. One approach is to improve the reconstruction algorithm; examples of such methods include sinogram extrapolation, singular value decomposition, wavelet decomposition, iterative reconstruction-reprojection and projections onto convex sets (POCS) (Nassi et al. 1982, Prince & Willsky 1990, Natterer 2001, Noo et al. 2002, Candes & Romberg 2005, Barrett & Myers 2013, Choi et al. 2016). Another technique for achieving such suppression is to employ prior information as additional constraints; the information that can be used in this way includes the boundaries, shapes, and density ranges of the objects of interest and the total variation (TV) (Chen et al. 2008, Chen et al. 2013, Wu et al. 2013, Heußer et al. 2014, Sidky et al. 2006, Sen Sharma et al. 2013, Huang et al. 2018, Zhang et al. 2018, Luo et al. 2018). Unfortunately, while all of these methods, especially TV-based methods, can achieve favorable results in reducing small streaks, these techniques are not effective at eliminating large streaks (Huang et al. 2018).

In recent years, deep learning has demonstrated exciting advantages over traditional methods in fields such as signal processing, image processing, natural language processing, and pattern recognition (LeCun et al. 2015, Schmidhuber 2015, Ronneberger et al. 2015, Goodfellow et al. 2016, Dong et al. 2016, Arjovsky et al. 2017, Mao et al. 2017, Xu et al. 2018, Cao et al. 2019). The success of deep learning in these fields has also led to its increasing use in the field of X-ray CT, especially for noise reduction, artifact removal, and image reconstruction (Chen et al. 2017, Chen et al. 2018, Hammernik et al. 2018, Kobler et al. 2018, Zhang & Yu 2018, Han & Ye 2018a, Ge et al. 2018, Würfl et al. 2018, Zhu et al. 2018, Li et al. 2019, Yin et al. 2019).

For example, Chen *et al.* proposed a noise reduction method for low-dose CT in the image domain via a deep convolutional neural network (CNN) (Chen et al. 2017); they argued that the CNN method has great potential for artifact reduction and structure preservation. Subsequently, Chen *et al.* further proposed a learned experts' assessment-based reconstruction network (LEARN) for sparse-data CT, which unrolls an iterative algorithm into a fixed N-step iterative algorithm (Chen et al. 2018). Kobler *et al.* embedded a variational model into an unrolled gradient decent scheme for compressed-sensing-based CT reconstruction (Hammernik et al. 2018); this model has also been successfully applied to magnetic resonance imaging (MRI) reconstruction (Kobler et al. 2018). The main characteristic of this unrolled network is that all the

3

parameters and regularization schemes used in the iterative reconstruction process are learned from training data. Furthermore, to improve the quality of low-dose CT imaging, Yin *et al.* proposed a domain progressive 3D residual convolution network with three components: a sinogram-domain network, a filtered backprojection (FBP) stage, and an image-domain network (Yin et al. 2019). In addition, Ge *et al.* proposed a network to learn the deconvolution of images after direct backprojection (Ge et al. 2018). Zhang *et al.* (Zhang & Yu 2018) developed an open metal artifact reduction (MAR) framework based on a CNN that combines an original image with corrected image information to suppress artifacts. Han *et al.* (Han & Ye 2018*a*) revealed the limitations of U-Net and proposed a new multiresolution deep learning solution for sparse-view CT; they pointed out that alternative U-Net variants, such as dual-frame and tight-frame U-Nets, are more efficient than the original version of U-Net for recovering high-frequency edges in sparse-view CT. Han *et al.* also studied the region of interest (ROI) problem and proposed two types of neural networks that could be applied to ROIs of any size (Han & Ye 2018*b*). Zhu *et al.* (Zhu et al. 2018) presented a manifold learning method for image reconstruction called the automated transform by manifold approximation (AUTOMAP) framework, which can learn a manifold mapping between the sensor and image domains; this framework is very powerful and can address almost all medical image reconstruction issues. However, because of the inclusion of fully connected layers, this framework has very large computational storage requirements when processing large quantities of data, such as clinical CT images. As an alternative, Wurfl *et al.* (Würfl et al. 2018) proposed a method for mapping an FBP algorithm onto a neural network, in which they used an analytical backprojection algorithm instead of a fully connected layer to minimize the memory requirements; the resulting network can automatically learn strategies such as compensation weights and apodization windows. In a more recent article, Li *et al.* (Li et al. 2019) proposed a deep learning method called iCT-Net, which also maps an FBP algorithm. The authors pointed out that when the sparse-view reconstruction problem becomes entangled with the classic internal tomography problem, their proposed method can achieve accurate reconstruction. However, because the above methods all involve the mapping of FBP algorithms, these techniques will, to some extent, exhibit the characteristics of FBP algorithms; consequently, when addressing limited-angle problems, the suppression or removal of artifacts in areas with missing data will be insufficient.

Inspired by the research successes summarized above, we propose a hybrid-domain deep convolutional neural network (hdNet) to reduce artifacts in limited-angle CT images. The network performs artifact removal not only in the sinogram domain but also in the CT image domain. The sinogram-domain network learns a weighted and filtered sinogram, which is then fed into a domain transfer operation to obtain a CT-domain image; finally, the network can further reduce artifacts and reconstruct textures within this CT image. In this network, we use an analytical algorithm to exchange data between the two domains; this ensures that the network focuses on artifact removal instead of demanding excessive computational and storage resources to learn the domain

4

transformation, which can be regarded as prior knowledge that does not contribute to artifact reduction. The novel contribution of this method is its hybrid-domain approach to artifact reduction, which sets it apart from other deep learning methods currently used for limited-angle CT reconstruction problems, such as model-based unrolling methods, FBP mapping methods and postprocessing methods. Results based on numerical, experimental and clinical data indicate that the proposed network exhibits excellent performance in artifact suppression and detail retention and can recover high-quality images in limited-angle CT problems. The remainder of this paper is organized as follows. In section 2, we derive and explain our proposed hdNet framework. In section 3, we describe the experimental design, training data, preparation of testing data and training process. Then, we analyze the results in section 4. Finally, the relevant issues are discussed in section 5, and our conclusions are presented accordingly.

2. Methods

2.1. Artifact reduction model

We model the limited-angle artifact reduction problem as follows. Let $y \in R^{p \times q}$ be a limited-angle sinogram, and let $x \in R^{m \times n}$ be the corresponding CT image without artifacts. The network model is constructed to draw on some prior physical knowledge during the CT reconstruction process. In explicit analysis algorithms, methods for weighting sinogram data, such as Parker weighting and its variants, are commonly used to reduce artifacts during the reconstruction of CT images under limited-angle conditions. Naturally, we assume that there exist a weighting function W and a filtering function F, which can be learned through a neural network (Würfl et al. 2018). These two functions operate in the sinogram domain. Next, the processed sinogram is fed into a domain transformation operation T_{bp}, following which the processed image-domain data $s \in R^{m \times n}$ are obtained:

$$s = T_{bp}(W \circ F(y)) \tag{1}$$

where $\circ$ is the composition operator and T_{bp} is the conventional backprojection operator.

In the image domain, we use a modified convolutional encoder-decoder architecture to further reduce artifacts. Such encoder-decoder architectures are widely used in inverse problems (Ronneberger et al. 2015, Hyun et al. 2018, Häggström et al. 2019). Accordingly, encoder-decoder theory has been analyzed in many works (Ye et al. 2018, Ye & Sung 2019). Here, the processed data $s \in R^{m \times n}$ are mapped by the encoder network into the feature space $z \in R^{u \times v} \subset R^{m \times n}$, $z = E(s)$, after which the decoder network takes the resulting feature map as its input and produces an output $\hat{x} \in R^{m \times n}$:

$$\hat{x} = D(z) = D(E(s)) \quad s.t. \ \hat{x} \geq 0 \tag{2}$$

where E is the encoder function and D is the decoder function, both of which can also be learned through a neural network. Consequently, the process of generating x from y

can be written as a sequence of function evaluations:

$$\hat{x} = D(E(T_{bp}(W \circ F(y)))) \quad s.t. \ \hat{x} \geq 0 \tag{3}$$

Therefore, the main purpose of our proposed neural network is to find the optimal W^*, F^*, E^* and D^* functions by minimizing the squared Frobenius norm of x and $\hat{x}$ for a given training dataset $\{(x_i, y_i)\}_{i=1}^N$:

$$D^*, E^*, W^*, F^* = \underset{D,E,W,F}{\arg\min} \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|_F^2 \tag{4}$$

$$= \underset{D,E,W,F}{\arg\min} \frac{1}{N} \sum_{i=1}^N \|x_i - D(E(T_{bp}(W \circ F(y_i))))\|_F^2 \quad s.t. \ \hat{x} \geq 0 \tag{5}$$

2.2. Network architecture

The neural network we propose has three components that operate in two domains. The first component, which is executed in the sinogram domain, weights and filters the input raw sinogram data (via the functions W and F). The second component is the domain transformation operation, which transforms the preprocessed sinogram data into the image domain by means of the conventional backprojection operator (T_{bp}). The last component is a modified encoder-decoder network with multiple different levels of skip connections; this network further suppresses limited-angle artifacts (via the functions E and D). The proposed hdNet architecture is illustrated in figure 1.

The first component, named CNN-Prefix, is a four-layer CNN in which all the kernels operate along the dimensions of the detector elements. In each box shown in figure 1, (p, q) represents the number of acquisition angles and the number of detector elements, while the following number is the number of output feature maps. The L1 layer has 16 convolution kernels, each with dimensions of 1×55. The L2 and L3 layers also each have 16 convolution kernels; the dimensions of each L2 convolution kernel are 1×45, and the dimensions of each L3 convolution kernel are 1×21. The L4 layer has one convolution kernel with dimensions of 1×11. For all layers in the CNN-Prefix component, the hyperbolic tangent ($tanh$) function is used as the activation function.

The second component is a domain transformation operation (DTO), which transforms the data from the sinogram domain into the image domain. We implement this function using a pixel-driven analytical backprojection algorithm (Peters 1981, Zhuang et al. 1994) with a gradient backpropagation function for two types of fan projections: equidistant and equiangular. The algorithm is written in C++ using the TensorFlow framework, and the pseudocode is provided in algorithms 1 and 2. An analytical algorithm is utilized instead of a deep learning algorithm to implement the DTO because in CT reconstruction, the acquired signal and the reconstructed image have very clear geometric meanings and a natural relationship (Kak & Slaney n.d., Hsieh et al. 2009). Therefore, there is no need to consume a large amount of computational resources to learn this explicit conversion relationship. Consequently, by means of a

6
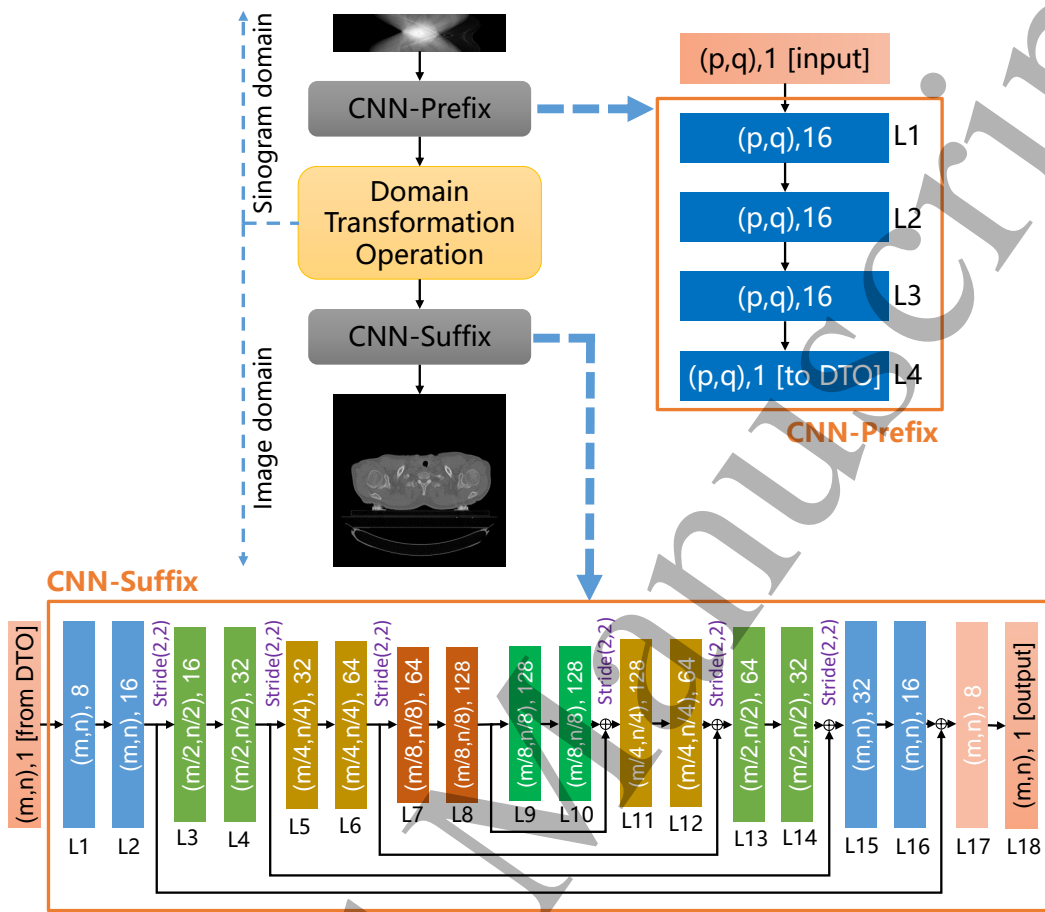
Figure 1. The architecture of the proposed hybrid-domain convolutional neural network (hdNet). The proposed neural network has three components. The first component (CNN-Prefix), which operates in the sinogram domain, weights and filters the input raw sinogram data. The second component is the domain transformation operation (DTO), which transforms the preprocessed sinogram data into the image domain. The last component (CNN-Suffix) is a modified encoder-decoder network with skip connections that further reduces limited-angle artifacts.

same-scale deep learning model, the proposed network can focus on learning only the most important information, whereas this clearly known mapping relationship can be frozen into the network as prior knowledge.

The last component, named CNN-Suffix, is a modified encoder-decoder CNN network with 18 layers and four different levels of skip connections. This network structure has been widely used in image restoration, super-resolution imaging, and denoising (Mao et al. 2016, Chen et al. 2017) and has been demonstrated to be very effective. In each box shown in figure 1, (m, n) represents the dimensions of the output feature maps, while the following number is the number of output feature maps. The convolution kernels in all layers have dimensions of 5×5, and the leaky rectified linear unit (ReLU) activation function is employed in all layers (except the last layer), with

7

the parameter alpha = 0.2. The input to the 1st layer is an image-domain feature map that has been transformed by the DTO from the output of the first component. The last layer is the output layer, which produces the final CT image from the feature maps with the ReLU activation function subject to $\hat{x} \geq 0$.

Algorithm 1: Domain transformation operation (DTO) algorithm: the forward propagation algorithm in the deep learning framework

---

Input: $S(t, j)$ ($t \in \{1, ..., P\}$, $j \in \{1, ..., Q\}$)
Output: $I(x, y)$ ($x \in \{1, ..., M\}$, $y \in \{1, ..., N\}$)
Function: $A(x, y, t)$, the projection of the point (x, y) onto the detector array at scan angle t to obtain the projected point position

1 Set all elements of $I(x, y)$ to zero: $I = 0$;
2 **for** $(x, y) = (1, 1)$ **to** (M, N) **do**
3   **for** $t = 1$ **to** P **do**
4          $c = A(x, y, t)$;
5          $c_b = max(1, floor(c))$, $c_e = min(Q, c_b + 1)$;
6          $\alpha_b = c_e - c$, $\alpha_e = 1 - \alpha_b$;
7          $I(x, y) = I(x, y) + \alpha_b \times S(t, c_b) + \alpha_e \times S(t, c_e)$;
8   **end**
9 **end**

2.3. Backpropagation in hdNet

To perform the backpropagation procedure in the proposed hdNet framework, the gradients in each component need to be calculated. Most of the gradient calculations are similar to those in other well-known CNN models; however, special care is needed for the second component, that is, the DTO (T_{bp}). As mentioned above, the second component of the network is a pixel-driven analytical backprojection algorithm (Peters 1981, Zhuang et al. 1994). Thus, the forward propagation process of this component can be expressed as follows:

$$I(x, y) = \sum_{t=1}^{P} S(t, c) \mid_{c=A(x,y,t)} \tag{6}$$

where $I(x, y)$ denotes the feature image output from T_{bp}; $S(t, c)$ denotes the input to T_{bp}, which is the output of the first component; and $A(x, y, t)$ is a function that calculates the position of the point (x, y) projected onto the detector array at scan angle t, $t \in 1, 2, ..., P$. The position c is usually not an integer and cannot be located by a single detector unit; therefore, we use linear interpolation to estimate the value of $S(t, c)$:

8

$$S(t,c) = \alpha_b S(t,c_b) + \alpha_e S(t,c_e) \Big|_{\substack{c_b=floor(c);\ c_e=c_b+1 \\ \alpha_b=c_e-c;\ \alpha_e=1-\alpha_b}} \tag{7}$$

Equations 6 and 7 demonstrate that the calculation parameters involved in the transformation are fixed values. Therefore, these parameters do not need to be learned during gradient backpropagation and can be frozen into the algorithm. The pseudocode for the forward propagation algorithm is given in algorithm 1. Let L denote the loss function; then, the associated gradient $\dfrac{\partial L}{\partial S}$ can be obtained through backpropagation as follows:

$$\frac{\partial L}{\partial S} = \frac{\partial I}{\partial S}\frac{\partial L}{\partial I} \tag{8}$$

Here, $\dfrac{\partial L}{\partial I}$ can be calculated using the standard backpropagation procedure for CNN-Suffix, which is a classic CNN (Goodfellow et al. 2016). Since only a few points in I are projected onto the same point in S after linear interpolation, only these points need to be considered:

$$\frac{\partial L}{\partial S(t,c)} = \sum_{x,y\in\Omega_{t,c}} \frac{\partial I(x,y)}{\partial S(t,c)}\frac{\partial L}{\partial I(x,y)} \tag{9}$$

where $\Omega_{t,c}$ represents all points in I projected onto the same point (t,c) in S after linear interpolation. The pseudocode for the backpropagation algorithm for the DTO component is shown in algorithm 2. For the first component, CNN-Prefix, the gradient can be calculated via standard backpropagation (Goodfellow et al. 2016).

Algorithm 2: Backpropagation algorithm for the DTO

    Input: $\nabla_{I(x,y)}L$ ($x \in \{1,...,M\}$, $y \in \{1,...,N\}$), the partial derivative of the loss function L with respect to the input $I(x,y)$ to the next layer

    Output: $\nabla_{S(t,j)}L$ ($t \in \{1,...,P\}$, $j \in \{1,...,Q\}$), the partial derivative of the loss function L with respect to the input $S(t,j)$ to the DTO

    Function: $A(x,y,t)$, the projection of the point (x,y) onto the detector array at scan angle t to obtain the projected point position

1 Set all elements of $\nabla_{S(t,j)}L$ to zero: $\nabla_S L = 0$;

2 for $(x,y) = (1,1)$ **to** (M,N) **do**

3   **for** $t = 1$ **to** P **do**

4 $c = A(x,y,t)$;

5 $c_b = max(1, floor(c))$, $c_e = min(Q, c_b+1)$;

6 $\alpha_b = c_e - c$, $\alpha_e = 1 - \alpha_b$;

7 $\nabla_{S(t,c_b)} = \nabla_{S(t,c_b)} + \alpha_b$, $\nabla_{S(t,c_e)} = \nabla_{S(t,c_e)} + \alpha_e$;

8   **end**

9 end

9

3. Experimental setup

In this work, two types of fan projections, equidistant and equiangular, are considered to verify the feasibility of the network. The geometric parameters for the equidistant fan-beam acquisition were set based on the CT imaging platform used in our laboratory, while those for the equiangular fan-beam acquisition were set based on a clinical CT scanner (SOMATOM Definition AS+, Siemens Healthcare, Forchheim, Germany) that was used in the Low-Dose CT Grand Challenge (AAPM 2016). Table 1 shows the detailed parameters of the two fan-beam projection geometries. Then, corresponding numerical, experimental and clinical data were used to verify the effectiveness of hdNet. To further verify the adaptability of the proposed network, we consider different projection coverage areas for the experimental and clinical data. The starting angles and projection coverage areas are shown in figure 2.

(a)   **(b)**
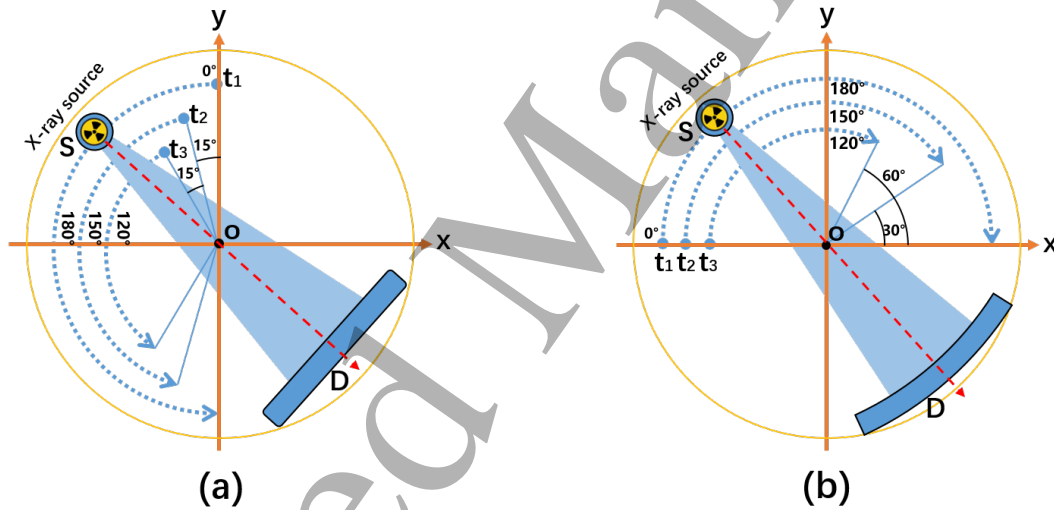
Figure 2. Equidistant (a) and equiangular (b) fan-beam scanning geometries for limited-angle CT. S denotes the X-ray source, D denotes the detector array, o denotes the center of rotation of the X-ray source, and t_1, t_2, and t_3 denote the starting angles for the three scan angle ranges of 180, 150, and 120 degrees, respectively.

3.1. Training data preparation

A total of 50,000 anonymous diagnostic CT images were downloaded from The Cancer Imaging Archive (TCIA) to train the network on the two types of projections. Anatomically, these images encompass the typical human body structure from the head to the pelvis, and many of them include pathological structures. All images have dimensions of 512×512 and are saved in DICOM format. The pixel value of a DICOM-format image is expressed in Hounsfield units (HU), which we converted into the linear attenuation coefficient μ $(1/mm)$ as follows:

$$\mu = \mu_w \times (HU/1000 + 1) \tag{10}$$

10

Table 1. Parameters of the two fan-beam projection geometries

Parameter	Equidistant	Equiangular
Distance between the source and detector center (DSD) (mm)	1500	1085.6
Distance between the source and isocenter (DSO) (mm)	1000	595
Sampling interval between two adjacent projection views (deg)	1.0	0.1563
Number of detector elements $(number)$	5120	736
Sampling interval between two adjacent detector elements $(mm$ & $deg)$	0.1	0.0679
Size of each detector element (mm)	0.1	1.2858
Detector center offset $(element\ number)$	2	1.625
Voxel size of the object (mm)	0.5	1.0
Reconstructed image dimensions $(pixels)$	512×512	512×512

where μ_w is the linear attenuation coefficient of water. Here, we used $\mu_w = 0.02$, which is the linear attenuation coefficient of water near 60 keV. The original 50,000 images were then augmented to obtain a total of 200,000 images by right/left flipping and by rotating the images by 90 degrees. Finally, 160,000 of the images were employed for training, 20,000 were used for validation, and the remaining 20,000 were used for testing. The images used for training, validation and testing were grouped by individual patients. Specifically, the patients were divided into three groups, and the data of the first group of patients were used for training, the data of the second patient group were used for verification, and the data of the last patient group were used for testing. This was done to ensure the independence of the training, verification and test data.

Using the parameters listed in table 1, a standard ray-driven forward projection procedure was applied to generate limited-view sinogram data for the equidistant and equiangular fan-beam geometries (Zhao & Reader 2003). To ensure that the simulations resembled reality, the mean number of entering photons was set to 1×10^6 per ray. Furthermore, 5% Poisson noise and zero-mean Gaussian noise with a variance of 10 were added to the projection data (Xu & Tsui 2009, Ma et al. 2012). In this work, we are interested in solving three different limited-angle problems, involving acquisition geometries limited to 180 degrees, 150 degrees, and 120 degrees, as shown in figure 2; accordingly, each of these cases required a corresponding training dataset, validation dataset, and test dataset. Considering these requirements, to avoid wasting storage resources by storing multiple copies of data, we used an online training data generation method to train the network. A corresponding sinogram was generated online for each case, where each training data pair for the network consisted of a sinogram and its corresponding labeled image.

3.2. Test data preparation

To evaluate the general capability of hdNet to suppress artifacts in a real-world environment, it is important to assess its reconstruction performance for objects of different sizes under different data acquisition conditions. Therefore, in this work, numerical data, experimental phantom data, clinical patient data, and two different

11

fan-beam projection conditions were employed. For the numerical simulation data, we used data from the testing dataset, which was applied to train the experimental data reconstruction network.

Experiments were performed on our laboratory's CT imaging platform, which uses an equidistant projection geometry. The geometric parameters are listed in table 1. The X-ray tube was operated at a tube voltage of 120 kV and a current of 11 mA (Varex G-242, Varex Imaging Corporation, UT, USA). A bowtie filter and an additional 0.50 mm copper filter were used. The X-ray detector was an energy-resolving photon counting detector (XC-Hydra FX50, XCounter AB, Sweden) with an imaging area of 512 $mm \times 6$ mm; the native component size was 0.1 $mm \times 0.1$ mm. The projection datasets were acquired using a 1×1 binning mode, and 6×6 rebinning was performed during postprocessing. We extracted the middle two rows and averaged them to obtain the sinogram. In this work, there was no need to distinguish the photon energy, so experiments were performed using only one energy threshold (10 keV). Catphan 700 phantoms (The Phantom Laboratory Inc., Salem, NY, USA) were employed in this work. The phantoms were rotated through 360 degrees at angular intervals of 1 degree. To obtain data with a low noise level, the phantoms were scanned 10 times in succession to average the scanned data. We then extracted three scans, one each at 180 degrees, 150 degrees, and 120 degrees, as shown in figure 2(a), to evaluate the proposed hdNet framework. An image reconstructed with the standard FBP algorithm using the full scan angle range of 360 degrees was used as the reference. Neither a beam-hardening correction nor a scattering correction was applied during the experiment.

The clinical patient data were acquired using the Siemens SOMATOM Definition AS+ CT system in helical mode. These data were established and authorized by the Mayo Clinic for the "2016 NIH-AAPM-Mayo Clinic Low-Dose CT Challenge" (refer to (AAPM 2016) for more details about this dataset). In this work, the patient dataset labeled L067 was used. The patient's liver displays hemangiomas, and the $CTDI_{vol}$ values in the DICOM-format file where the lesions are located are approximately 9.4 mGy. The raw projection dataset is composed of projection data comprising 2,304 views per scan. The geometric parameters are listed in table 1, and the patient's position is feet-first supine (FFS). The helical data were rebinned into an equiangular fan-beam geometry using the single-slice rebinning method (Noo et al. 1999). Scans from angles of 180 degrees (1,152 projection views), 150 degrees (960 projection views), and 120 degrees (768 projection views), which are different from the experimental data, were extracted to evaluate the proposed hdNet framework. Here, we used the same starting angle for each scan, as shown in figure 2(b), making it easy to implement the tests on clinical equipment. The reference images were generated from all 2,304 projection views using the standard FBP algorithm with a Hanning window (Natterer 2001).

12

3.3. Training process

The networks for both types of fan-beam projection artifact suppression were trained under the same training conditions. The proposed network was trained using the Adam algorithm (Kingma & Ba 2014) with an initial learning rate of 3×10^{-5}, which exponentially decayed by a factor of 0.98 after every 1000 steps. Each minibatch had a size of 4, and the number of epochs was 200. The network was implemented using the TensorFlow deep learning framework (ver. 1.13.1), and we employed a single NVIDIA GeForce GTX 1080 Ti GPU for training.

4. Results

4.1. Evaluation metrics

We use the peak signal-to-noise ratio (PSNR) and the structural similarity index metric (SSIM) to quantitatively assess the advantages of the proposed network. The PSNR is defined as follows:

$$PSNR(\hat{x}, x) = \frac{10}{log10}log \frac{MAX^2(x)}{\frac{1}{M \times N} \sum_{i,j}^{M,N} (\hat{x}(i,j) - x(i,j))^2} \tag{11}$$

where x is the reference image, with dimensions of $M \times N$, and $\hat{x}$ is the image to be measured. A large PSNR suggests little difference from the reference image. The SSIM is defined as follows:

$$SSIM(\hat{x}, x) = \frac{(2\mu_x \mu_{\hat{x}} + a_1)(2\sigma_{x,\hat{x}} + a_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + a_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + a_2)} \tag{12}$$

where $\mu_{\hat{x}}$ and $\sigma_{\hat{x}}^2$ denote the mean value and variance, respectively, of $\hat{x}$; similar properties μ_x and σ_x^2 are defined for the reference image; $\sigma_{x,\hat{x}}$ represents the covariance of $\hat{x}$ and x; and a_1 and a_2 are two constants used to ensure stable division with a weak denominator, with values of 1×10^{-6} and 3×10^{-6}, respectively, adopted here. We also use linear profiles and image residuals to validate the accuracy of the network in reconstructing CT images under different limited-angle conditions. In addition, to analyze whether the results of the proposed network show significant differences with respect to the results of other methods, one-way analysis of variance (ANOVA) is used.

4.2. Reference methods

We compare our hdNet framework with four methods, namely, the FBP algorithm with the Ram-Lak filter, the iterative adaptive-weighted total variation (AwTV) algorithm (Liu et al. 2012), the iterative non-local means (NLM) algorithm (Huang et al. 2011) and the FBPConvNet framework (Jin et al. 2017). The iterative AwTV algorithm and the iterative NLM algorithm have both been widely used in various fields of image processing, and their advantages have been thoroughly explained (Liu et al. 2012, Luo et al. 2018, Huang et al. 2018, Shi et al. 2014, Köhler et al. 2016, Huang

13

et al. 2011, Qi et al. 2016). Based on previous work, in this study, we specifically choose the AwTV-POCS algorithm (Liu et al. 2012) and NLM-POCS algorithm (Huang et al. 2011) for comparison. In addition, the FBPConvNet framework is a well-known deep learning reconstruction network (an image-domain data-driven postprocessing-based method) applied to sparse-view reconstruction problems. The core component of FBPConvNet is a modified U-Net that processes images reconstructed via FBP. To facilitate fair comparisons, the parameters of AwTV-POCS and NLM-POCS are derived from the settings specified in (Liu et al. 2012) and (Huang et al. 2011) and are optimized to minimize the SSIM. For the FBPConvNet algorithm, we use the same training strategy used in (Jin et al. 2017), and the training dataset for this network is the same as the training dataset for hdNet.

4.3. Numerical simulation results

Numerical validations are presented in figure 3. The first, second, and third rows of images show the results for the scan angle ranges of 180 degrees, 150 degrees, and 120 degrees, respectively, for comparison. The image in the first column in each row is the reference image, and the subsequent columns show (in succession) the images reconstructed by the hdNet, FBPConvNet, AwTV-POCS, NLM-POCS and FBP algorithms. These comparisons reveal that our method effectively eliminates large streak artifacts, achieves a favorable reconstruction quality and preserves fine image details in comparison with the other methods. The red arrow in each figure highlights the location of some obvious artifacts. As shown in figure 3, the FBP-reconstructed images exhibit large streak artifacts due to the lack of projection data. In contrast, the AwTV method effectively suppresses small streak artifacts because it considers edge anisotropy among neighboring image voxels; however, this technique cannot effectively suppress large streak artifacts. The NLM method, which accounts for self-similarity characteristics in the imagery, also behaves as described above; the NLM algorithm can effectively remove small streak artifacts but cannot strongly mitigate large-sized artifacts. Moreover, the FBPConvNet framework, a postprocessing-based deep learning reconstruction method, relies heavily on the FBP reconstruction results, so it is also rather ineffective at suppressing large streaks. Compared with the above four methods, the proposed method performs artifact correction in two domains and can suppress large streak artifacts very effectively while preserving edge information and structural details.
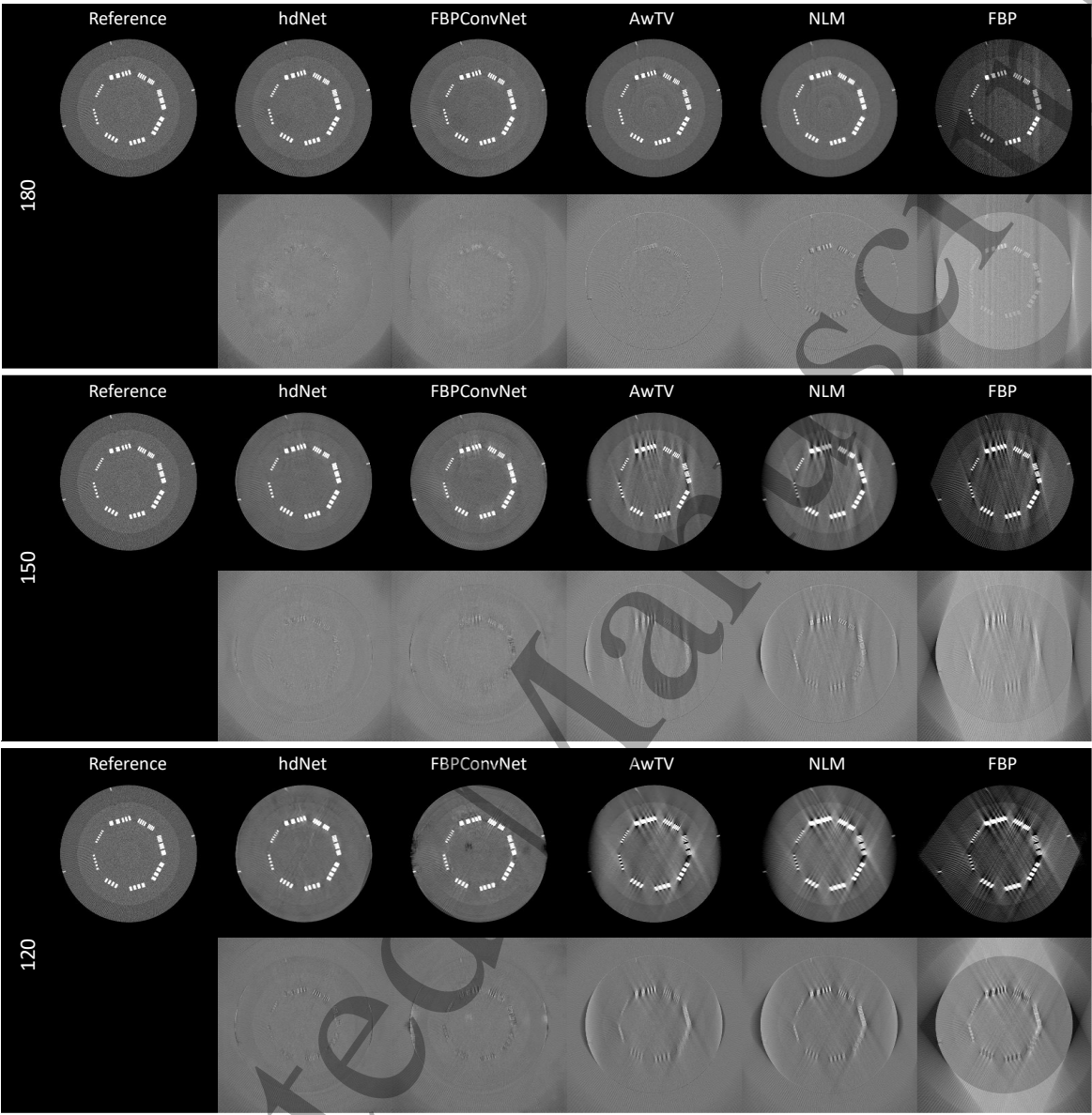
4.4. Experimental data results

The experimental data results are presented in figure 4. The first, third and fifth rows of images display the results for the scan angle ranges of 180 degrees, 150 degrees, and 120 degrees, respectively, for comparison. The image in the first column in each of these rows is the reference image, and the subsequent columns show (in succession) the images reconstructed by the hdNet, FBPConvNet, AwTV-POCS, NLM-POCS and FBP algorithms. The images in the second row are the residual images between the reference

14
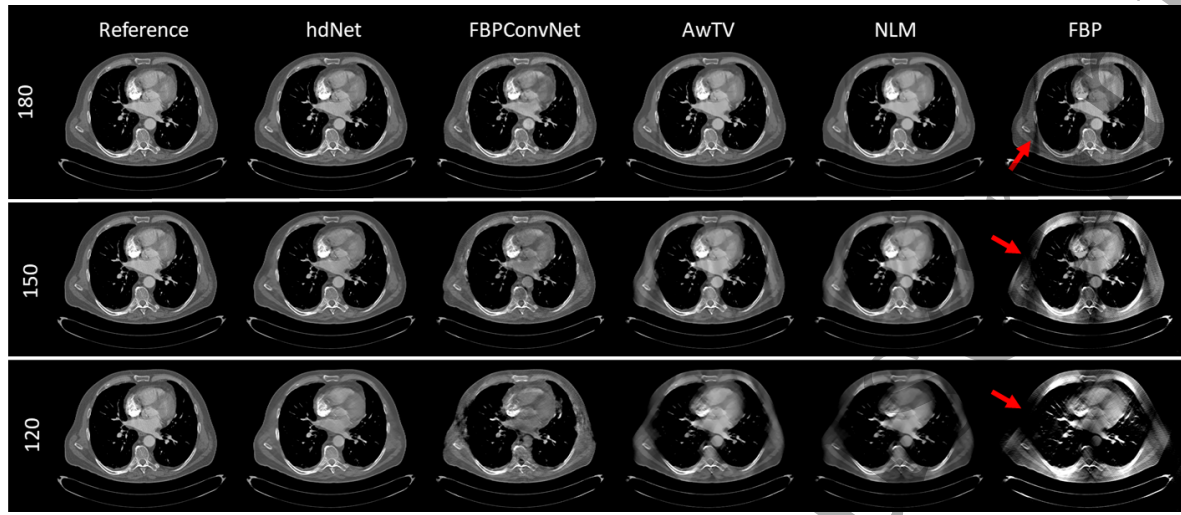
Figure 3. Numerical simulation results. The images in the first, second, and third rows represent comparisons for scan angle ranges of 180 degrees, 150 degrees, and 120 degrees, respectively. The image in the first column in each row is the reference image, and the subsequent columns show the images reconstructed by the hdNet (2nd column), FBPConvNet (3rd column), AwTV-POCS (4th column), NLM-POCS (5th column) and FBP (6th column) algorithms. The display window is [-400 600] HU (W/L=1000/100 HU).

image and the images reconstructed by the different reconstruction algorithms for the scan angle range of 180 degrees. Similarly, the fourth and sixth rows show the residual images for scan angle ranges of 150 degrees and 120 degrees, respectively. These residual images suggest that the deviation between the left and right parts of the image is large; this discrepancy is related to the limited projection coverage area (Natterer 2001). With an angular coverage of 180 degrees, all methods except the FBP algorithm can obtain suitable results because the streak artifacts in this case are not serious; furthermore, the AwTV and NLM methods can effectively eliminate artifacts based on structural information. As the angular coverage decreases, however, the large streak artifacts become more severe. As a result, the AwTV and NLM algorithms experience difficulty in removing these large artifacts, while the deep-learning-based methods (FBPConvNet and hdNet) show better performance in this respect; nevertheless, because FBPConvNet is a postprocessing method, it is limited by the FBP reconstruction results. When the angular coverage is 120 degrees, the artifacts clearly affect the image reconstruction results of the FBP algorithm. By contrast, the hdNet framework, which uses a hybrid-domain CNN, can effectively avoid these effects because hdNet learns the optimal filter and weights in the sinogram domain and then applies the optimal postprocessing function in the image domain during training, as described in section 2.

15

Figure 4. Experimental data results. The images in the first, third and fifth rows represent comparisons for scan angle ranges of 180 degrees, 150 degrees, and 120 degrees, respectively. The image in the first column in each of these rows is the reference image reconstructed using the standard FBP algorithm with the full scan angle range of 360 degrees, while the subsequent columns show the images reconstructed by the hdNet (2nd column), FBPConvNet (3rd column), AwTV-POCS (4th column), NLM-POCS (5th column) and FBP (6th column) algorithms. Below each reconstruction result, the residual image obtained by subtracting the corresponding reconstructed image from the reference image is shown. The display window is [-1000 1000] HU (W/L=2000/0 HU for the reconstructed images and [-400 400] HU (W/L=800/0 HU) for the residual images).

4.5. Clinical data results

The clinical data results are presented in figures 5, 6, and 7 and in tables 2 and 3. In figure 5, the first, third and fifth rows of images display the results for
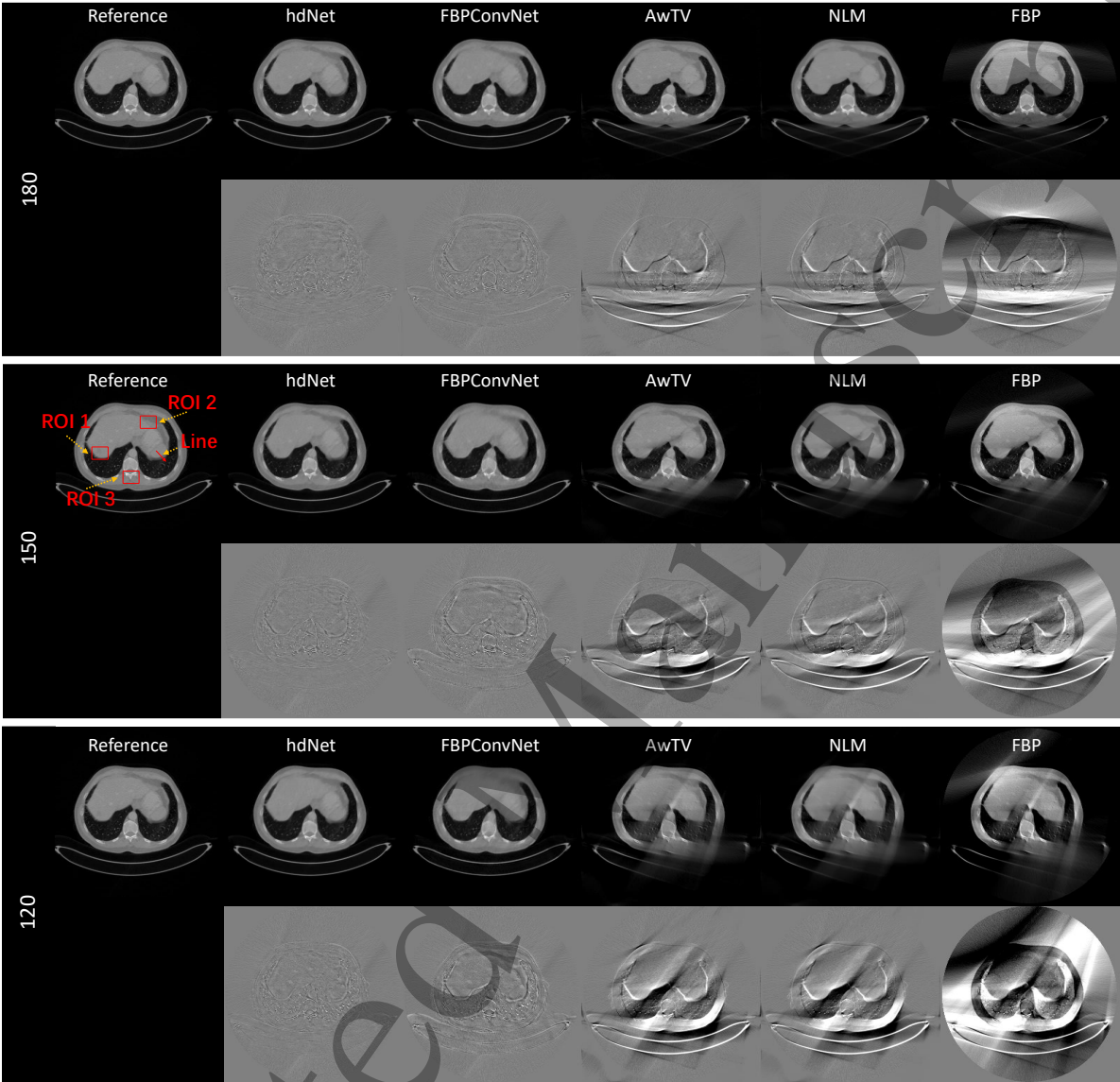
16

the scan angle ranges of 180 degrees, 150 degrees, and 120 degrees, respectively, for comparison. The image in the first column in each of these rows is the reference image, and the subsequent columns show (in succession) the images reconstructed by the hdNet, FBPConvNet, AwTV-POCS, NLM-POCS and FBP algorithms. Below each reconstruction result, the residual image obtained by subtracting the corresponding reconstructed image from the reference image is presented. Similar to the experimental data results, the residual images exhibit obvious structures related to the projection coverage area (Natterer 2001). Figure 6 shows magnified views of the results in the red box labeled ROI 1 in figure 5 (see the left-most image in the third row). ROI 1 represents a liver region containing a hemangioma. The red arrows highlight some clinically important textural details. The artifacts in the FBP results are severe and obscure all information contained in this region of the image. The AwTV-POCS and NLM-POCS methods can retain some structural details, but as the angular coverage decreases, the artifacts become increasingly severe. FBPConvNet attempts to recover the original image from the FBP results, but the presence of severe artifacts disrupts the reconstruction. By contrast, our hdNet framework retains better texture information than the other methods do and allows the hemangioma to be recognized for all scan angle ranges.

To further demonstrate the ability of our method to effectively preserve the structures in an image, the profile along the red line in figure 5 (see the left-most image in the third row) is plotted for each reconstructed image in figure 7. These plots clearly demonstrate that the profiles of the hdNet results are the most consistent with the reference image profile at the edges and in the approximately homogeneous region. Tables 2 and 3 list the quantitative evaluation results for reconstruction using the different methods under different angular coverage conditions. The proposed hdNet framework achieves the best results in terms of both metrics, consistent with our visual observations. With an angular coverage of 120 degrees, the PSNRs of our model in ROI 1 and ROI 2 reach 16.25 dB and 16.32 dB, respectively; in contrast, both of these ROIs exhibit severe artifacts in the FBP results. Our results are also impressive in terms of the SSIM. To evaluate whether the results of the proposed method are significantly different from those of the other methods, we used these methods to reconstruct 100 different images (50 lung images and 50 abdominal images) and calculated their respective SSIM values and then applied one-way ANOVA for statistical analysis based on these results. The results are summarized in table 4, which shows that $P < 0.001$ in all cases. Therefore, the structural similarity between the reconstructed image and the reference image is significantly higher with our method than it is with the other methods.

5. Discussion and conclusion

Deep-learning-based methods are able to automatically learn strategies such as compensation weights and apodized ramp filters in the sinogram domain of FBP mapping algorithms (Würfl et al. 2018, Li et al. 2019), and deep-learning-based image-
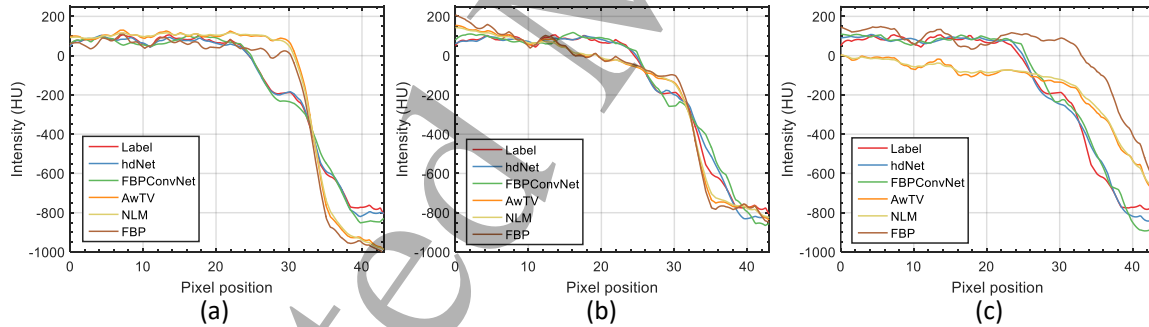
Figure 5. Clinical data results. The images in the first, third and fifth rows represent comparisons for scan angle ranges of 180 degrees, 150 degrees, and 120 degrees, respectively. The image in the first column in each of these rows is the reference image reconstructed by the standard FBP algorithm using the full scan angle range of 360 degrees. The subsequent columns show the images reconstructed by the hdNet (2nd column), FBPConvNet (3rd column), AwTV-POCS (4th column), NLM-POCS (5th column) and FBP (6th column) algorithms. Below each reconstruction result, the residual image obtained by subtracting the corresponding reconstructed image from the reference image is shown. The display window is [-1000 1000] HU (W/L=2000/0 HU for the reconstructed images and [-400 400] HU (W/L=800/0 HU) for the residual images).

domain postprocessing methods have also achieved excellent performance in CT imaging (Jin et al. 2017, Chen et al. 2017, Han & Ye 2018*a*, Han & Ye 2018*b*). All of these methods have demonstrated superiority over traditional methods in terms of

18
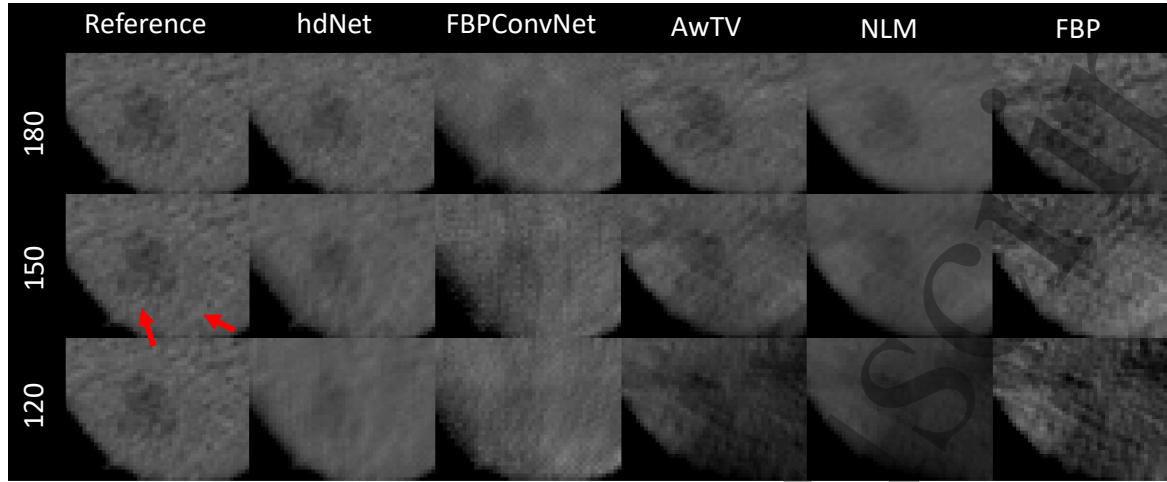
Figure 6. Magnified views of ROI 1 (within the corresponding red box in figure 5). The reference image is presented for comparison with the images reconstructed using the hdNet, FBPConvNet, AwTV-POCS, NLM-POCS and FBP algorithms with different scan angle ranges. The dark area in the middle of the ROI is the hemangioma. The red arrows indicate some important textural details, which are reasonably maintained in the hdNet reconstruction results but not in the results of the other methods. The display window is [-200 600] HU (W/L=800/200 HU).

Figure 7. Profiles along the red line in figure 5 for the reference image and the images reconstructed using the hdNet, FBPConvNet, AwTV-POCS, NLM-POCS and FBP algorithms with different scan angle ranges. (a), (b) and (c) show the profile comparisons for scan angle ranges of 180 degrees, 150 degrees and 120 degrees, respectively.

Table 2. SSIM values of the reconstructed images obtained by the different methods over different scan angle ranges from the clinical data.

Method	180				150				120			
	Whole	ROI 1	ROI 2	ROI 3	Whole	ROI 1	ROI 2	ROI 3	Whole	ROI 1	ROI 2	ROI 3
hdNet	**0.9559**	**0.8628**	**0.8198**	**0.7767**	**0.9231**	**0.7454**	**0.6765**	**0.662**	**0.8786**	**0.5934**	**0.5778**	**0.5602**
FBPConvNet	0.8582	0.4614	0.4296	0.4919	0.822	0.281	0.4884	0.3472	0.7742	0.2416	0.1947	0.3029
AwTV	0.8444	0.5655	0.725	0.3086	0.8049	0.3052	0.6016	0.2367	0.7534	0.1991	0.3085	0.1762
NLM	0.8205	0.4698	0.6387	0.304	0.7851	0.2322	0.545	0.2333	0.7442	0.1517	0.2931	0.1716
FBP	0.7802	0.2622	0.157	0.2833	0.7577	0.2909	0.1502	0.2146	0.7212	0.105	0.2936	0.1495

Table 3. PSNRs of the reconstructed images obtained by the different methods over different scan angle ranges from the clinical data.

Method	180				150				120			
	Whole	ROI 1	ROI 2	ROI 3	Whole	ROI 1	ROI 2	ROI 3	Whole	ROI 1	ROI 2	ROI 3
hdNet	**39.78**	**20.57**	**20.24**	**29.97**	**36.88**	**15.79**	**17.46**	**26.51**	**34.78**	**16.25**	**16.32**	**25.13**
FBPConvNet	32.89	12.67	14.53	22.57	30.46	10.1	12.64	20.28	29.83	10.06	9.954	20.08
AwTV	24.05	4.395	16.75	18.69	21.19	2.226	14.72	15.87	18.6	1.498	7.59	14.51
NLM	23.94	4.627	16.52	18.49	21.07	2.272	15.28	15.74	18.53	1.37	7.992	14.3
FBP	18.37	2.191	3.283	12.46	17.36	0.4364	2.135	12.32	14.15	3.947	3.81	11.05

Table 4. Statistical significance analysis between the results of the hdNet framework and the FBPConvNet, AwTV, NLM, and FBP methods.

Range	SSIM, mean±SD					P-value			
	hdNet	FBPConvNet	AwTV	NLM	FBP	vs. FBPConvNet	vs. AwTV	vs. NLM	vs. FBP
180	0.948±0.014	0.892±0.015	0.871±0.014	0.843±0.012	0.783±0.016	< 0.001	< 0.001	< 0.001	< 0.001
150	0.912±0.018	0.853±0.022	0.824±0.017	0.804±0.014	0.746±0.020	< 0.001	< 0.001	< 0.001	< 0.001
120	0.871±0.21	0.790±0.022	0.751±0.022	0.75±0.023	0.700±0.023	< 0.001	< 0.001	< 0.001	< 0.001

reconstruction accuracy and computational efficiency. Inspired by these results, in this work, we use a deep learning neural network operating in two domains to suppress artifacts in limited-angle CT problems. According to our hypothesis and the results described in (Würfl et al. 2018), the proposed network can learn the weights and filtered sinograms in the sinogram domain and then feed the processed sinograms to a DTO to obtain CT-domain images, in which the network can further reduce the remaining artifacts. Results obtained on numerical, experimental and clinical data show the correctness of our hypothesis. To obtain these results, we used models trained on two different types of projections, namely, equiangular and equidistant projections, and with three finite scan angle ranges with different starting angles. All of the results are satisfactory and confirm that the proposed network has strong generalizability.

In this work, an analytical algorithm is employed to implement the DTO because applying deep learning to learn the domain transformation would demand immense quantities of computing resources and storage space. Thus, it would be unrealistic to reconstruct images suitable for the sizes of clinical CT imagery because of an intrinsic property of the transformation algorithm, namely, that generating a single element in one domain requires an operation involving all elements of the other domain. Notably, when the image dimensions are small (128 *pixels* × 128 *pixels*), the DTO can be implemented by means of a fully connected layer (Zhu et al. 2018). However, even with such a small image size, 2 GB of storage space is already required for the weight variables, and of course, far more storage space would be needed for an image with dimensions of 512 *pixels* × 512 *pixels*. Additionally, the DTO from the sinogram domain to the image domain in the field of CT reconstruction is a natural mapping process. Regardless of what reconstruction method is used (i.e., an analytical algorithm or an iterative algorithm), some domain transformation process must be implemented, which can be achieved by means of either multiplication with the system adjoint matrix or backprojection. Therefore, there is no need to force the network to learn this known

20

transformation process. Instead, the network should be allowed to direct sufficient computational resources toward learning more complex nonlinear mappings that are more difficult to determine. In the CNN-Suffix component of the proposed network, skip connections are used to connect symmetric early and late layers to effectively restore the details of the image and prevent the image from being excessively smoothed. Networks similar to the CNN-Suffix component have been widely used in image restoration and denoising (Mao et al. 2016, Chen et al. 2017); these works have proven that skip connections can effectively retain details and prevent undesired image smoothing. The residual images presented in the results of this work also illustrate this point.

One limitation of the current work is that when the geometric parameters (such as the DSD, DSO, and detector element size) change, the network needs to be retrained. Fortunately, if the projection type is the same, we can apply the transfer learning method to an existing model to train new models within a relatively short time. Another limitation is that we have studied only cases involving a fan-beam geometry. However, by specifically designing the DTO and network structure appropriately, the proposed method can also be easily applied to other situations, such as a cone-beam geometry.

To conclude, the proposed deep-learning-based hdNet framework provides a new paradigm for reconstructing images from data obtained within limited scan angle ranges while suppressing artifacts. Numerical, experimental and clinical results show that compared with model-based iterative reconstruction methods and deep-learning-based image-domain postprocessing methods, the proposed method shows a superior ability to reduce artifacts.

Acknowledgments

References

AAPM 2016 'Low-dose ct grand challenge' https://www.aapm.org/grandchallenge/.

21

Arjovsky M, Chintala S & Bottou L 2017 *in* 'International Conference on Machine Learning' pp. 214–223.

Barrett H H & Myers K J 2013 *Foundations of image science* John Wiley & Sons.

Candes E J & Romberg J K 2005 *in* 'Computational Imaging III' Vol. 5674 International Society for Optics and Photonics pp. 76–87.

Cao Y J, Jia L L, Chen Y X, Lin N, Yang C, Zhang B, Liu Z, Li X X & Dai H H 2019 *IEEE Access* **7**, 14985–15006.

Chen G H, Tang J & Leng S 2008 *Medical physics* **35**(2), 660–663.

Chen H, Zhang Y, Chen Y, Zhang J, Zhang W, Sun H, Lv Y, Liao P, Zhou J & Wang G 2018 *IEEE transactions on medical imaging* **37**(6), 1333–1347.

Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J & Wang G 2017 *Biomedical optics express* **8**(2), 679–694.

Chen Z, Jin X, Li L & Wang G 2013 *Physics in Medicine & Biology* **58**(7), 2119.

Choi J K, Dong B & Zhang X 2016 *J. Comput. Math* **34**(6), 575–589.

Dong C, Loy C C, He K & Tang X 2016 *IEEE transactions on pattern analysis and machine intelligence* **38**(2), 295–307.

Ge Y, Zhang Q, Hu Z, Chen J, Shi W, Zheng H & Liang D 2018 *arXiv preprint arXiv:1807.01833* .

Goodfellow I, Bengio Y & Courville A 2016 *Deep learning* MIT press.

Häggström I, Schmidtlein C R, Campanella G & Fuchs T J 2019 *Medical Image Analysis* .

Hammernik K, Klatzer T, Kobler E, Recht M P, Sodickson D K, Pock T & Knoll F 2018 *Magnetic resonance in medicine* **79**(6), 3055–3071.

Han Y & Ye J C 2018*a* *IEEE transactions on medical imaging* **37**(6), 1418–1429.

Han Y & Ye J C 2018*b* *arXiv preprint arXiv:1810.00500* .

Heußer T, Brehm M, Ritschl L, Sawall S & Kachelrieß M 2014 *Medical physics* **41**(2).

Hsieh J et al. 2009 SPIE Bellingham, WA.

Huang J, Ma J, Liu N, Zhang H, Bian Z, Feng Y, Feng Q & Chen W 2011 *Computers in Biology and Medicine* **41**(4), 195–205.

Huang Y, Taubmann O, Huang X, Haase V, Lauritsch G & Maier A 2018 *IEEE Transactions on Radiation and Plasma Medical Sciences* **2**(4), 307–314.

Hyun C M, Kim H P, Lee S M, Lee S & Seo J K 2018 *Physics in Medicine & Biology* **63**(13), 135007.

Jin K H, McCann M T, Froustey E & Unser M 2017 *IEEE Transactions on Image Processing* **26**(9), 4509–4522.

Kak A C & Slaney M n.d. *IEEE, New York* .

Kingma D P & Ba J 2014 *arXiv preprint arXiv:1412.6980* .

Kobler E, Muckley M, Chen B, Knoll F, Hammernik K, Pock T, Sodickson D & Otazo R 2018 *in* '2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)' IEEE pp. 6687–6691.

Köhler T, Huang X, Schebesch F, Aichert A, Maier A & Hornegger J 2016 *IEEE Transactions on Computational Imaging* **2**(1), 42–58.

LeCun Y, Bengio Y & Hinton G 2015 *nature* **521**(7553), 436.

Li L, Chen Z, Zhang L & Kang K 2006 *Journal of X-ray Science and Technology* **14**(2), 109–118.

Li L, Kang K, Chen Z, Zhang L, Xing Y, Yu H & Wang G 2008 *Journal of X-Ray Science and Technology* **16**(1), 43–49.

Li Y, Li K, Zhang C, Montoya J & Chen G H 2019 *IEEE transactions on medical imaging* .

Liu Y, Ma J, Fan Y & Liang Z 2012 *Physics in Medicine & Biology* **57**(23), 7923.

Luo F, Li W, Tu W & Wu W 2018 *IEEE Access* **6**, 64225–64236.

Ma J, Liang Z, Fan Y, Liu Y, Huang J, Chen W & Lu H 2012 *Medical physics* **39**(7Part1), 4051–4065.

Mao X, Li Q, Xie H, Lau R Y, Wang Z & Paul Smolley S 2017 *in* 'Proceedings of the IEEE International Conference on Computer Vision' pp. 2794–2802.

Mao X, Shen C & Yang Y B 2016 *in* 'Advances in neural information processing systems' pp. 2802–2810.

Nassi M, Brody W R, Medoff B P & Macovski A 1982 *IEEE Transactions on biomedical engineering*

22

(5), 333–341.

Natterer F 2001 *The mathematics of computerized tomography* SIAM.

Noo F, Defrise M & Clackdoyle R 1999 *Physics in Medicine & Biology* **44**(2), 561–570.

Noo F, Defrise M, Clackdoyle R & Kudo H 2002 *Physics in medicine & biology* **47**(14), 2525.

Peters T 1981 *IEEE transactions on nuclear science* **28**(4), 3641–3647.

Prince J L & Willsky A S 1990 *Optical Engineering* **29**(5), 535–545.

Qi H, Chen Z, Wu S, Xu Y & Zhou L 2016 *Physica Medica* **32**(9), 1041–1051.

Ronneberger O, Fischer P & Brox T 2015 *in* 'International Conference on Medical image computing and computer-assisted intervention' Springer pp. 234–241.

Schmidhuber J 2015 *Neural networks* **61**, 85–117.

Sen Sharma K, Jin X, Holzner C, Narayanan S, Liu B, Wang D, Agah M, Wang L, Yu H & Wang G 2013 *Journal of X-ray Science and Technology* **21**(1), 25–42.

Shi J, Liu F, Pu H, Zuo S, Luo J & Bai J 2014 *Biomedical optics express* **5**(11), 4039–4052.

Sidky E Y, Kao C M & Pan X 2006 *Journal of X-ray Science and Technology* **14**(2), 119–139.

Tuy H K 1983 *SIAM Journal on Applied Mathematics* **43**(3), 546–552.

Wang Z, Huang Z, Chen Z, Zhang L, Jiang X, Kang K, Yin H, Wang Z & Stampanoni M 2011 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **635**(1), 103–107.

Wu D, Li L & Zhang L 2013 *Physics in Medicine & Biology* **58**(12), 4047.

Würfl T, Hoffmann M, Christlein V, Breininger K, Huang Y, Unberath M & Maier A K 2018 *IEEE transactions on medical imaging* **37**(6), 1454–1463.

Xu J & Tsui B M 2009 *IEEE Transactions on Image Processing* **18**(6), 1228–1238.

Xu T, Zhang P, Huang Q, Zhang H, Gan Z, Huang X & He X 2018 *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition' pp. 1316–1324.

Ye J C, Han Y & Cha E 2018 *SIAM Journal on Imaging Sciences* **11**(2), 991–1048.

Ye J C & Sung W K 2019 *arXiv preprint arXiv:1901.07647* .

Yin X, Coatrieux J L, Zhao Q, Liu J, Yang W, Yang J, Quan G, Chen Y, Shu H & Luo L 2019 *IEEE Transactions on Medical Imaging* **38**(12), 2903–2913.

Zhang W, Zhang H, Wang L, Cai A, Li L & Yan B 2018 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **880**, 107–117.

Zhang Y & Yu H 2018 *IEEE transactions on medical imaging* **37**(6), 1370–1381.

Zhao H & Reader A J 2003 *in* '2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515)' Vol. 4 IEEE pp. 2808–2812.

Zhu B, Liu J Z, Cauley S F, Rosen B R & Rosen M S 2018 *Nature* **555**(7697), 487.

Zhuang W, Gopal S & Hebert T 1994 *IEEE transactions on nuclear science* **41**(4), 1660–1665.