IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED SIGNAL PROCESSING METHODS IN MEDICAL IMAGING

Received July 23, 2017, accepted October 16, 2017, date of publication October 25, 2017,
date of current version November 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2766438

Improving Low-Dose CT Image Using Residual Convolutional Network

WEI YANG[1], HUIJUAN ZHANG[2,3,4,5], JIAN YANG[6], JIASONG WU[2,3,4,5], XIANGRUI YIN[2,3,4,5], YANG CHEN[2,3,4,5], HUAZHONG SHU[2,3,4,5], LIMIN LUO[2,3,4,5], GOUENOU COATRIEUX[7], ZHIGUO GUI[8], AND QIANJIN FENG[1]

[1]Guangdong Provincial Key Laboratory of Medical Image Processing, Department of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China
[3]Centre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), 35000 Rennes, France
[4]Key Laboratory of Computer Network and Information Integration, Southeast University, Ministry of Education, Nanjing 210096, China
[5]International Joint Research Laboratory of Information Display and Visualization, Southeast University, Ministry of Education, Nanjing 210096, China
[6]Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China
[7]Institut Mines-Telecom, Telecom Bretagne, INSERM U1101 LaTIM, 29238 Brest, France
[8]Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan 030051, China

Corresponding author: Yang Chen (chenyang.list@seu.edu.cn)

This work was supported in part by the State's Key Project of Research and Development Plan under Grant 2017YFC0107900, Grant 2017YFC0109202, and Grant 2017YFA0104302, in part by the National Natural Science Foundation under Grant 81370040 and Grant 81530060, in part by the Fundamental Research Funds for the Central Universities, and in part by the Qing Lan Project in Jiangsu Province. *(Wei Yang and Huijuan Zhang contributed equally to this work.)*

ABSTRACT Low-dose CT is an effective solution to alleviate radiation risk to patients, it also introduces additional noise and streak artifacts. In order to maintain a high image quality for low-dose scanned CT data, we propose a post-processing method based on deep learning and using 2-D and 3-D residual convolutional networks. Experimental results and comparisons with other competing methods show that the proposed approach can effectively reduce the low-dose noise and artifacts while preserving tissue details. It is also pointed out that the 3-D model can achieve better performance in both edge-preservation and noise-artifact suppression. Factors that may influence the model performance, such as model width, depth, and dropout, are also examined.

INDEX TERMS Low-dose CT, convolution neural network, residual learning, 3D convolution.

I. INTRODUCTION

X-ray Computed Tomography (CT) provides major anatomical and pathological information of the human body for medical diagnosis and treatment. However, repetitive clinical CT examinations require reducing the radiation dose. This issue has become an urgent and challengeable focus in the CT research field. Three main approaches have been considered to improve the quality of low-dose CT images: projection data filtering, iterative reconstruction algorithms and CT images post-processing. The easier access to Filtered Back Projection (FBP) reconstructed CT images and to the projection data obtained from the manufacturers opens the way for post-processing methods which offer good implementability and expansibility to existing CT scanner equipment in hospitals [1]. Chen *et al.* proposed a large-scale non-local mean algorithm to remove the low-dose CT (LDCT) [2] artifacts via a nonlinear large scale filter correction [2]. Loubele *et al.* [2] reported an effective processing of abdomen LDCT images based on a sparse representation using a pre-trained dictionary. Deep learning techniques have recently been considered to tackle this problem. A Multi-layer Perceptron (MLP) machine based method was introduced to learn the mapping from the noisy images to the corresponding noise-free images and has shown an impressive performance in image restoration [5]. However, the application of MLP with fully connected layers is often limited by the requirement of fixed input/output size and the weight parameter explosion in network training. In [7], a residual convolutional network architecture was designed to build the relationship between the wavelet coefficients of low-dose and high-dose CT images. Han *et al.* [7] proposed a U-net structured architecture with residual learning to predict the artifacts in sparse-angle reconstructed CT image. Although the Convolutional Neural Networks (CNN)-based denoising model has achieved desirable noise and artifact removal, the limitation is that they are all based on two-dimensional convolution operations and have ignored the 3D structure continuity, leading so to tissue detail loss (vessel for instance).

2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This paper explores the application of residual network in LDCT image processing. The proposed residual network (ResNet) performs noise-artifact removal by predicting the residual component (mainly composed of noise-artifact component) between the LDCT images and the corresponding high-dose CT (HDCT) images included in the training dataset. 2D and 3D ResNet versions are evaluated using these clinical data. The structure of this paper goes as follows. The residual network architecture is introduced in section II. In section III, the 2D image-to-image residual network to learn the mapping from the LDCT images to the residual component containing noise and artifacts is explained in detail. The factors that may influence the model performance, such as network depth, width, dropout are analyzed. In section III, the proposed model is extended to 3D. The conducted experiments are described in section IV and their results compared to other methods. The overall research is summarized in section V.

II. RESIDUAL NETWORK ARCHITECTURE

Commonly used network architectures include plain network, with connections only between adjacent layers, and multi-branch network structured as directed acyclic graph. Deep network layers (i.e. deep CNN) provide an improved representation accuracy of image features. Some examples of plain networks are AlexNet [10] to VGGNet [10]. However, a gradient diffusion tends to occur when the depth is increased, which might result in training failure. This gradient diffusion problem can be solved by ReLU and Batch Normalization (BN) [11] to some extent. Indeed, as the network depth continues to increase, both the training error and the testing error surprisingly increase. ResNet [12], a typical representative of multi-branch networks, has been introduced to solve this difficulty. ResNet tries to learn the local and global features via skip connections combining different levels, and so, to overcome the incapability of integrating different level features found in plain networks. The most notable difference between residual and plain networks is the residue estimation strategy. They share however some basic elements such as the convolution operator, the activation layer, the loss function, and the optimization method.

A. RESIDUAL NETWORK

Compared to the plain network in Fig. 1(a), a residual network [12] consists of an ensemble of basic residual unit, containing two stacked convolution layers, as shown in Fig. 1(b). Here x_l and x_{l+1} are the input and output of the l-th residual unit respectively, $F(x)$ denotes the residue mapping of the stacked convolution layers. W_l represents the convolution weight of the l-th layer. If the dimensions of the input x_l and the mapping $F(x_l)$ match in the element-wise addition, their relation can be expressed by

$$x_{l+1} = x_l + F(x_l, \{W_l\}) \tag{1}$$

Otherwise, a dimension mapping convolution layer is needed in the skip connection, whose convolution kernel is represented as W_s, thus

$$x_{l+1} = F(x_l, \{W_l\}) + W_s x_l \tag{2}$$

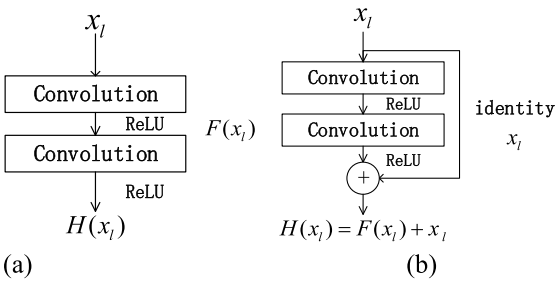

FIGURE 1. Plain network and residual network: (a) plain network, and (b) residual network.

Recursively, from the *0-th* layer to the *L-th* layer, Eq. (2) can be expressed as $x_L = x_0 + \sum_{i=0}^{L-1} F(x, \{W_i\})$. Such residual network appears as a straight addition operation between layers except for the activation and BN layers.

He *et al.* [14] suggest that the form of residue should be as simple as possible, with a shortcut connection minimum train error and test error. But provided that the residue is $0.5x$, convolution or dropout will block the forward and backward propagation, leading to an increased error [14]. Therefore, this paper adopts the shortcut connection, and the convolution layer, which aims at realizing the dimension mapping, is added only when the dimensions do not match.

Using ResNet with the skip connection, a simple identity mapping directly connects the input and output layers. It has the same computational complexity as the plain network with the same depth because no extra parameters need to be learned and the only computation required is the gradient of loss with respect to the input.

B. BATCH NORMALIZATION

Batch Normalization [11] can be used to solve the internal covariate shift, which is caused by the change of distribution of each layer's input after the convolution and activation layers during training. With the batch normalization, the input of each layer is normalized to zero mean and unit variance, then scaled and shifted to restore the distribution. It was pointed out in [11] that the dependence on dropout can be reduced due to the regularization role of BN. Another merit of BN is that it can significantly accelerate the CNN training because the inputs of each layer have a similar distribution.

C. LOSS FUNCTION

The strategy of residual learning [7] is adopted to learn the residue in the ResNet processing. The benefit of learning residue is that it can avoid building complicated regression model for mapping LDCT images to HDCT images due to the inherently rich details in CT images. To measure the similarity between the predicted residue N' and the real residue$N$,

which can be obtained by subtracting HDCT image Q from LDCT image P in the training dataset, we can build the loss function $J(W, b)$ as follows:

$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m} J\left(W,b;x^{(i)},y^{(i)}\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2 \quad (3)$$

where m is the sample number in current batch, and $J(W, b)$ is the loss of these samples. (x^i, y^i) is the i-th sample. $J\left(W, b; x^{(i)}, y^{(i)}\right)$ is the mean square error (MSE) of the i-th sample, which is defined as $J\left(W, b; x^{(i)}, y^{(i)}\right) = \frac{1}{RC}\sum_{j=0}^{C-1}\sum_{k=0}^{R-1}\left\|N'(j,k) - N(j,k)\right\|^2$, where $N = P - Q$. R and C are respectively the width and height of the sample. The second term in Eq. (3) is the regularization term called weight decay. n_l is the number of convolution layers. s_l is the number of nodes in the $l-th$ layer, and s_{l+1} is the number of nodes in the $l+1-th$ layer. The weight can be constrained by setting different weight decay parameter λ. A smaller λ will lead to a wider weight range.

The Adam optimization method [15] can dynamically update the learning rate of parameters using the unbiased estimation of the gradient's first moment m and the second moment v during backward propagation. This Adam optimization method is used to minimize the MSE in this study.

III. LOW-DOSE CT IMAGE POST-PROCESSING

In this paper, we construct 2D and 3D CNN residual convolution network models with skip connection to process low-dose CT images.

A. 2D NETWORK

The randomly sampled and cropped 128 × 128 low-dose CT and the corresponding noise patches, obtained by subtracting high-dose CT images from the corresponding low-dose CT images, are used as the training set. We design an end-to-end architecture as depicted in Fig. 2. Here the Adam method is used to minimize the MSE between the output of the last convolution layer and the actual residual images (low-dose noise). In [14], it was pointed out that a simple zero padding strategy does not generate boundary artifacts. This technique was used here in order to ensure that the size of the output image is equal to the original input size. For each layer, the numbers of convolution kernels are set to 64, the size of the convolution kernel are set as 3 × 3 and the convolution stride to 1. Then, the entire CT slides are input into the trained model to estimate the residual data contained in the LDCT images, since our model is independent of the input size by the fact that a fully convolution layer is used instead of a fully connection layer. Experiments were conducted to explore the factors that might influence the model performance, e.g. model width, depth and dropout options. We analyzed three different depths by setting n (the number of basic units for plain network or residual network), leading to a total of 12, 16, 24 convolution layers (the layer number is $2n + 6$), in the plain network and the residual network.

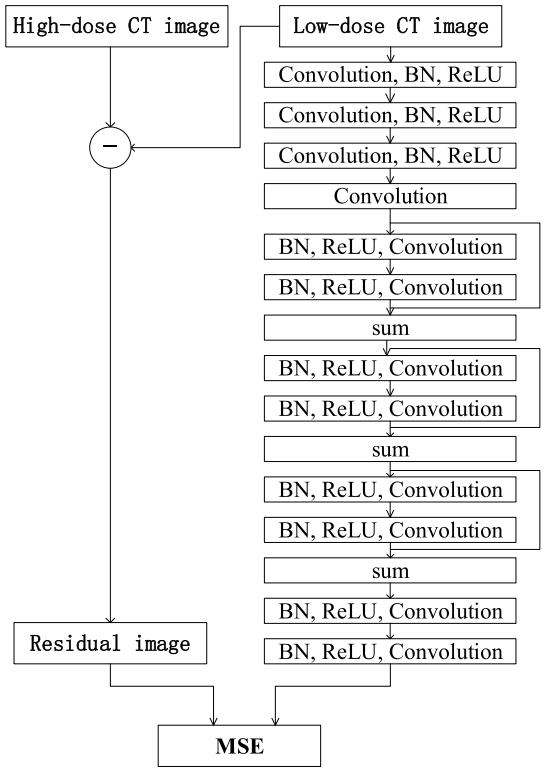

FIGURE 2. 2D ResNet based Low-dose CT post-processing architecture.

TABLE 1. 2D residual network with increased width.

layer	kernel size	channel number
C1_conv, C1_bn, C1_relu	3 × 3	32
C2_conv, C2_bn, C2_relu	3 × 3	64
C3_conv, C3_bn, C3_relu	3 × 3	96
C4_conv	3 × 3	96
resBlock1_bn, resBlock1_relu resBlock1_conv	$\begin{bmatrix}3\times3\\3\times3\end{bmatrix}\times3$	128
resBlock2_bn, resBlock2_relu resBlock2_conv		128
res_sum		
res_bn, res_relu		
C10_conv C10_bn, C10_relu	3 × 3	64
C11_conv	3 × 3	1

In order to study the influence of the convolution kernel number in each network layer, we compared the 2D-Resnet-9 network with a broader network, whose parameters are listed in table 1. The cell across multiple step lines of the table represents the skip connection in the residual network. Here, the suffix of layer name 'conv' represents the

TABLE 2. 3D residual network.

layer	kernel size
C1_conv	$3\times3\times3$
C2_conv	$3\times3\times1$
C3_conv	$3\times3\times3$
C4_conv	$3\times3\times1$
resBlock1_conv	$\begin{bmatrix}3\times3\times3\\3\times3\times1\end{bmatrix}\times3$
resBlock2_conv	
res_sum	
C10_conv	$3\times3\times3$
C11_ conv	$3\times3\times1$

convolution, besides, 'bn' is the BN and 'relu' represents the ReLU activation function. The suffix 'sum' presents the element-wise summation. The prefix 'C#' denotes the #-th convolution layer. 'res' indicates the layer related to the residual blocks. 'resBlocki' denotes the i-th convolution layer in the residual network. 3×3 denotes the convolution kernel size and the channel number is the number of kernels in the current convolution layer.

B. 3D ARCHITECTURE

The 2D network model merely extracts features into single slides and thus it does not take into account the spatial continuity of the tissues and organs. Therefore, keeping the same model, we extend our 2D architecture to 3D. The 3D training blocks are so small that the layer-wise zero padding may lead to a loss of the original details in CT blocks. 3D convolution is just applied in the valid respective field, which represents image field without padding, of the former input layer and, in such a way, the output size of the model is smaller than the input. The LDCT images are randomly sampled and cropped into 44×44 patch from the adjacent 24 CT slices. These $44 \times 44 \times 24$ blocks constitute the residue images used as the training set. The blocks is preprocessed into zero mean and unit variance. The stride is equal to 1 and the number of feature map is 64 in the convolution layer. The considered kernel sizes are listed in table 2. A non-padding convolution layer for dimension mapping is added in the shortcut connection because the dimensions of the two inputs of the element-wise summation are not identical. The 3D model is also independent of the input size, and the several adjacent CT slices are directly fed into the network to obtain the prediction.

IV. RESULTS AND EVALUATION

The data set for evaluation was provided by AAPM Low-Dose CT Grand Challenge thanks to the Mayo Clinic [15] which consists of low-dose and high-dose CT images from 10 patients, with 512×512 resolution. The utilization of the real projection data is permitted with signed agreement authorized by the Mayo clinic. The high-dose scanning voltage is 100kV or 120kV and the X-ray tube current varies from 200 mA to 500mA. The slice thickness is 1 mm. Poisson noise was inserted into the high-dose projection data to simulate the corresponding 25% of the high-dose data. The reconstructed CT images using both the high-dose and the low-dose projection data are provided in the challenge dataset. We use 9 patient CT images as the training data set (5,080 CT slices in total). The remaining patient data set is used as the test data to validate the algorithm performance. We randomly sample small patches over the whole training set to allow more images to be included in the training process in single batch. Mean value subtraction and variance normalization were carried out on each patch to obtain training samples with an approximate Gaussian distribution.

The computer platform was configured as follows: CPU is Intel(R) Core(TM) i7-4790K 4.00GHz; GPU is NVIDIA gtx1080 with 8G memory. We used the MatConvNet deep learning framework [17], the Matlab version R2015b, and the MSRA [18] method to initialize the weights of convolution layers. The learning rate is set to 0.001 (halved every 40 epochs), the weight decay to 10^{-5}. The moments for the Adam algorithm were respectively set as default values 0.9 and 0.999. The batch size was set respectively to 64 and 2 for 2D and 3D networks to fully exploiting GPU memory. Following the principle that stacked small convolution kernels can achieve the same receptive field size as the large kernels, whereas saving memory, we adopted the small convolution kernel size 3×3 for all networks. The model training has last 150 epochs.

The final results are obtained by subtracting the LDCT images to the residue components estimated using the trained residual network, including the comparison with iterative total-variation (TV) reconstruction [20]. They are illustrated in Fig. 3. Sagittal views of the lung and liver are also depicted in V and Fig. 5, respectively. It can be clearly seen in the abdomen window illustrations (Fig. 3 and Fig. 4) that the 3D ResNet model have the best performance in preserving the inherent details and textures in CT images (see the red arrows in Fig. 3 and Fig. 4). Similar results can be observed for the results for the lung window illustrations (Fig. 5) for both the 2D and 3D ResNet models.

The average values of PSNR, SSIM, RMSE and MAE (refer for definitions to [21]) for the different methods are listed in Table 3. It can be seen in this table that 3D-Resnet-3 model leads to the best results for all evaluation measures. The 2D-Resnet-3 model is slightly better than the 2D-plain-3. When compared to the BM3D [22], Non-Local Means (NLM) [2], Dictionary Learning (DL) [4] and Discriminative Feature Representation (DFR) [22], the Resnet-based model provides better image quality. To give fair comparisons, the parameters of these competing methods were optimized to get the best results in terms of PSNR values.

It can be visually observed in Fig. 3 the improvements brought by the residual CNN model. The 3D model better preserves the continuous structures like the vessels and organ edges (see the features indicated by red arrows

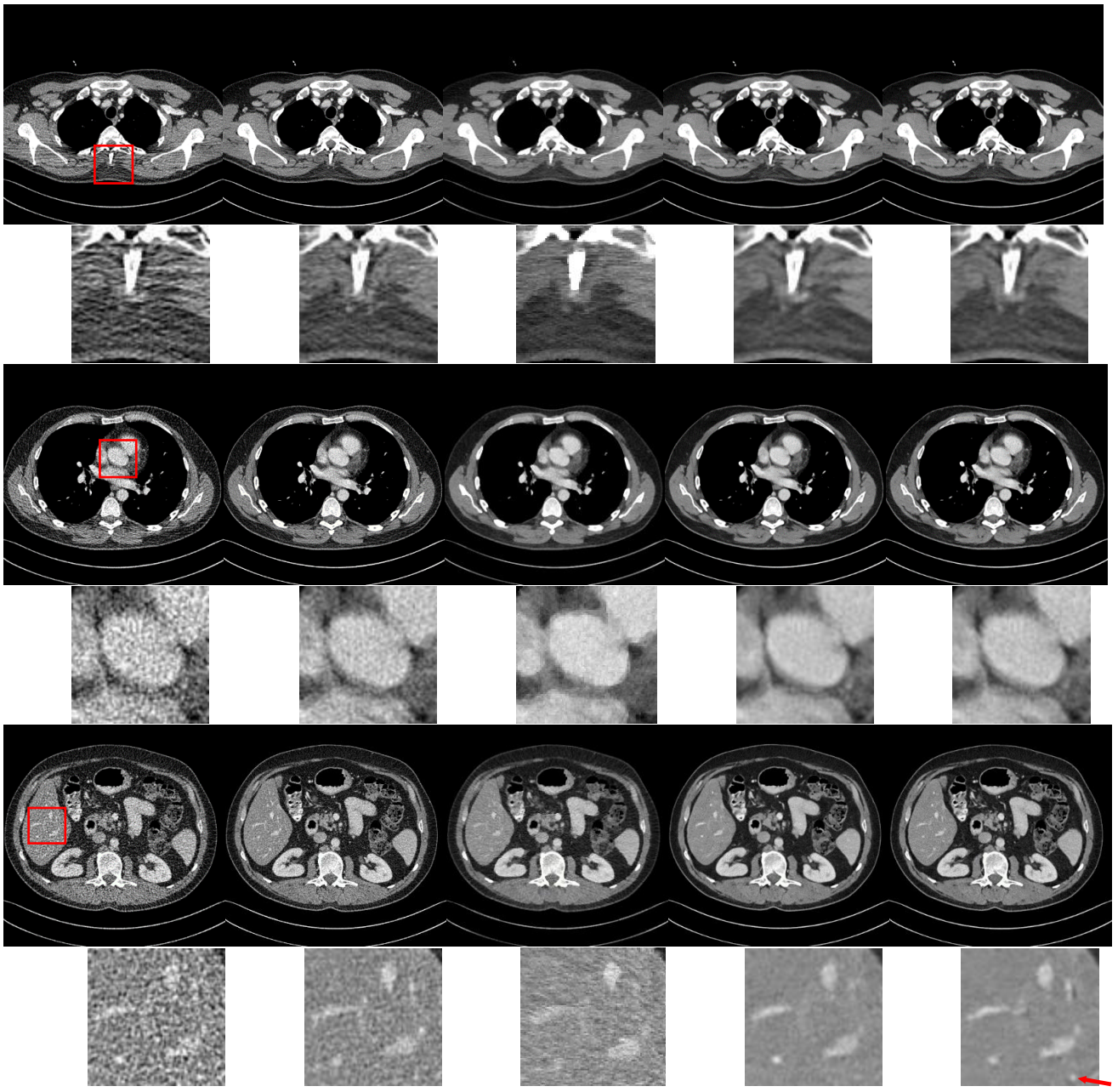

FIGURE 3. Selected axial views of the 2D and 3D post-processing results. The first column is original LDCT image; the second column is the reference HDCT image; the third column is the iterative TV reconstruction results; the fourth and fifth columns are the 2D and 3D results, respectively.

in Fig.3-Fig.5). Table 3 shows that the 3D model has reached substantially higher PSNR and SSIM and lower MSE values than the 2D models. The computation time on a single GPU for each CT image using the 2D model is 0.3s while for the 3D model is 0.25s. The results in Table 3 also show that the residual networks have better performance than plain networks when the same depth is used. It can be noted in Fig. 6 that training error and testing error decrease in both plain networks and residual networks when the depth increases. The extremely deep 2D-resnet_deep-20 model ($n = 20$, 46 layers in total) performs also very well.

Training MSE values of the original and the broaden networks (the cost is an increased computation time and more memory resource) are depicted in Fig. 7. We can see that the broader network with more convolution kernels at each layer can lead to a better suppression of noise and artifacts in the training phase while a worse performance is observed in the testing phase due to the overfitting.

The BN operation can reduce the dependency with dropout [23] due to its regularization effect. In order to explore the influence of dropout on the CT post-processing performance, the original 2D-Resnet-9 (only BN) and

TABLE 3. Quantitative evaluation of 2D and 3D processing with the residual network.

	Quality Evaluation	PSNR	SSIM	RMSE	MAE
Single 2D models	2D-plain-3	39.4846	0.9427	21.8932	13.3266
	2D-plain-5	39.4928	0.9431	21.7116	13.2390
	2D-plain-9	39.5220	0.9436	21.6387	13.1992
	2D-resnet-3	39.4204	0.9430	21.7321	13.2589
	2D-resnet-5	39.4950	0.9432	21.7061	13.2262
	2D-resnet-9	39.5177	**0.9437**	21.6495	**13.1690**
	2D-resnet_broaden-9	39.228	0.9413	22.3837	13.5412
	2D-resnet_deep-20	39.5293	0.9432	21.6205	13.2049
	2D-resnet_dropout-9	**39.5362**	0.9435	**21.6033**	13.2083
Coupled 2D models	2D-resnet-3 + resnet-5	39.5236	0.9433	21.6347	13.2045
	2D-plain-3 + resnet-3	39.5208	0.9433	21.6416	13.2113
	2D-resnet_deep-20 + 2D-resnet_dropout-9	**39.5902**	**0.9438**	**21.4695**	**13.1391**
	2D-resnet_broden-9 + 2D-resnet_dropout-9	39.5126	0.9434	21.6620	13.2217
Single 3D model	3D-resnet-3	**39.8329**	**0.9454**	**20.8779**	**12.9163**
Competing methods	BM3D[20]	37.3669	0.9168	27.7319	16.5695
	NLM[3]	37.2940	0.9143	27.9660	16.9213
	DFR[21]	37.7377	0.9274	26.5733	15.4111

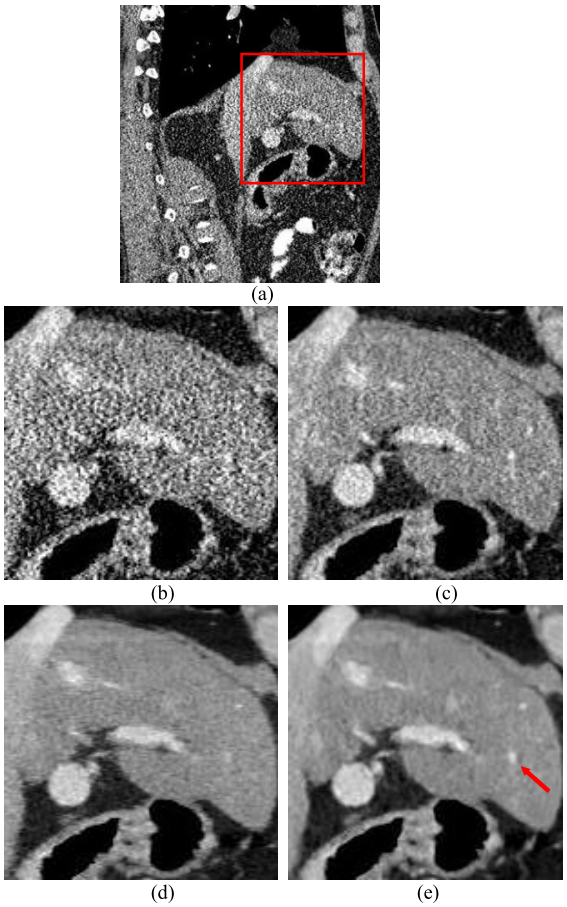

FIGURE 4. Selected sagittal view of 2D and 3D post-processing results in abdomen window. (b), (c), (d), (e) are the zoomed views of the original LDCT images, the original HDCT images, and the corresponding results of 2D resnet-3 and 3D resnet-3 networks. (a) The sagittal view. (b) LDCT. (c) HDCT. (d) 2D resnet-3. (e) 3D resnet-3.

2D-Resnet_dropout-9 (containing BN+dropout) networks are compared in Fig. 8. It appears that the network with dropout added before the last convolution layer has improved convergence. However, we can see that the Resnet-9 with the

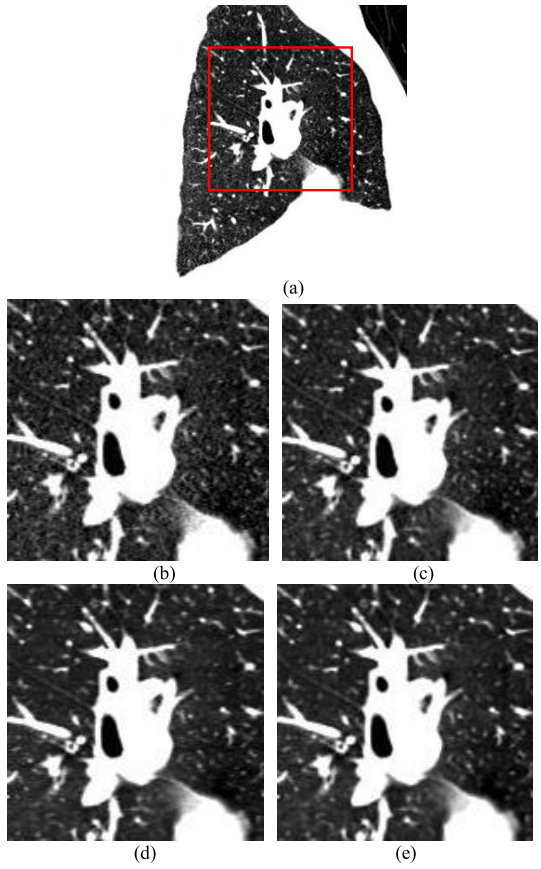

FIGURE 5. Selected sagittal view of 2D and 3D post-processing results in lung window. (b), (c), (d), (e) are the zoomed views of the original LDCT images, the original HDCT images, and the corresponding results of 2D resnet-3 and 3D resnet-3 networks. (a) The sagittal view. (b) LDCT. (c) HDCT. (d) 2D resnet-3. (e) 3D resnet-3.

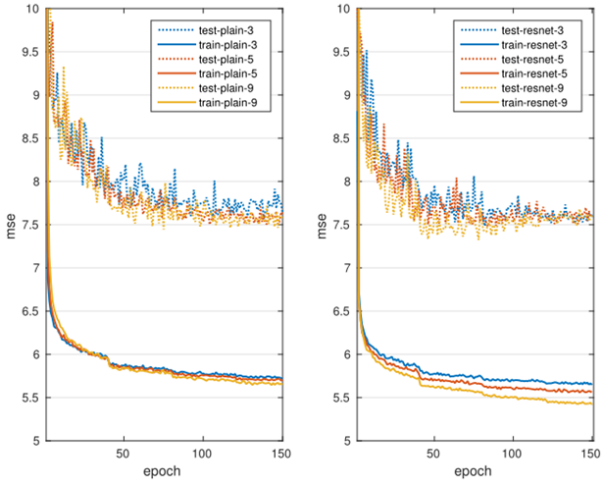

FIGURE 6. Solid lines denote training loss, and dashed lines are testing loss. Left: plain networks. Right: Residual networks. (the legend is formatted 'A-B-n', 'A' denotes the training and testing phase; 'B' is the network type, 'plain' means the plain networks, 'resnet' is the residual networks; 'n' represents the number of basic element blocks, here n is selected from the values in $\{3, 5, 9\}$).

dropout operation leads to lower convergence loss in the testing stage, indicating that the dropout boosts the generalization ability.

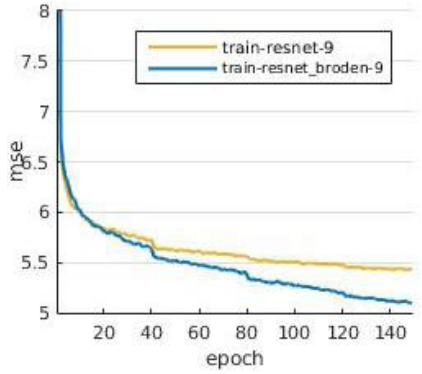

FIGURE 7. Training error of original 2D-resnet-9 and broaden 2D-resnet-9. (the legend is formatted 'A-B-9', 'A' denotes the training phase; 'B' is the network type, 'resnet' is the original residual networks, 'resnet_broden' is the residual networks with increased width; '9' represents the number of basic element blocks is 9, namely n=9.

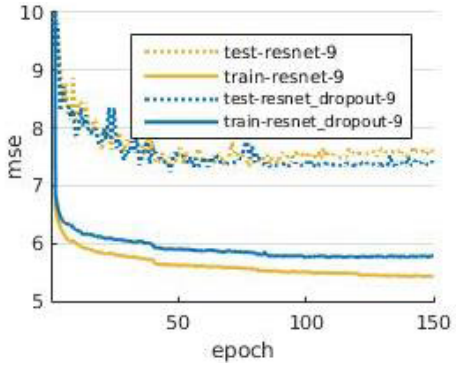

FIGURE 8. Training and testing error of original BN only 2D-resnet-9 and 2D BN+dropout 2D-resnet-9.(the legend is formatted 'A-B-9', 'A' denotes the training and testing phase; 'B' is the network type, 'resnet' is the only BN resnet, 'resnet_dropout' is the resnet with dropout added; '9' represents the the number of basic element blocks is 9, namely $n = 9$).

The effect of jointly trained 2D denoising models has also been explored. As illustrated in Table 3, the two coupled models can improve slightly the post-processing performance. As an example, the association of 2D-Resnet_deep-20 with 2D-Resnet_dropout-9 shows better results than any other single 2D models.

V. CONCLUSIONS AND PERSPECTIVES

Deep learning approaches have made breakthrough in many applications in the field of visual vision due to their powerful capability on feature representation. This work addressed low-dose CT post-processing based on residual network. Experimental results have shown that the proposed 2D or 3D networks have good performance on preserving image details and removing noise-artifact structure as well. Especially, it has been also found that the 3D network architecture is able to yield better results than the equivalent 2D model due to its ability to account for the three-dimensional tissue structures. The 2D model is thus suggested for CT scans with sick thickness, which implies poor structure continuity between neighboring slices.

Our work in progress will focus on methodological issues (such as compensating the contrast loss after processing) and extended clinical assessment. Further work will also include incorporating the deep learning strategy into iterative reconstruction framework to obtain reconstruction with higher image quality.

ACKNOWLEDGMENT

The authors would like to thank Dr. Cynthia McCollough (the Mayo Clinic, USA) for providing clinical projection data of Somatom Definition AS+ CT.

REFERENCES

[1] Y. Zhu, M. Zhao, Y. Zhao, H. Li, and P. Zhang, "Noise reduction with low dose CT data based on a modified ROF model," *Opt. Exp.*, vol. 20, no. 16, pp. 17987–18004, 2012.
[2] M. Loubele *et al.*, "Radiation dose vs. image quality for low-dose CT protocols of the head for maxillofacial surgery and oral implant planning," *Radiat. Protection Dosimetry*, vol. 117, nos. 1–3, pp. 211–216, 2005.
[3] Y. Chen *et al.*, "Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means," *Phys. Med. Biol.*, vol. 57, no. 9, pp. 2667–2688, 2012.
[4] Y. Chen *et al.*, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, no. 16, pp. 5803–5820, 2013.
[5] H. C. Burger, C. J. Schuler, and S. Harmeling. (Nov. 2012). "Image denoising with multi-layer perceptrons, part 1: Comparison with existing algorithms and with bounds." [Online]. Available: https://arxiv.org/abs/1211.1544
[6] H. C. Burger, C. J. Schuler, and S. Harmeling. (Nov. 2012). "Image denoising with multi-layer perceptrons, part 2: Training trade-offs and analysis of their mechanisms." [Online]. Available: https://arxiv.org/abs/1211.1552
[7] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2392–2399.
[8] E. Kang, J. Min, and J. C. Ye. (Oct. 2016). "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction." [Online]. Available: https://arxiv.org/abs/1610.09736
[9] Y. S. Han, J. Yoo, and J. C. Ye. (Nov. 2016). "Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis." [Online]. Available: https://arxiv.org/abs/1611.06391
[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[11] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[12] S. Ioffe and C. Szegedy. (Feb. 2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167
[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.
[15] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[16] D. Kingma and J. Ba. (Dec. 2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980
[17] *Low Dose CT Grand Challenge*. Accessed: Apr. 6, 2017. [Online]. Available: http://www.aapm.org/GrandChallenge/LowDoseCT/
[18] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2012, pp. 689–692.
[19] K. He, X. Zhang, S. Ren, and J. Sun. (Feb. 2015). "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification." [Online]. Available: https://arxiv.org/abs/1502.01852
[20] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, Aug. 2008.
[21] G.-H. Chen, C.-L. Yang, and S.-L. Xie, "Gradient-based structural similarity for image quality assessment," in *Proc. Int. Conf. Image Process.*, Atlanta, GA, USA, 2006, pp. 2929–2932.

[22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[23] Y. Chen *et al.*, "Discriminative feature representation: An effective post-processing solution to low dose CT imaging," *Phys. Med. Biol.*, vol. 62, no. 6, pp. 2103–2131, 2017.
[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

WEI YANG received the B.Sc. degree in automation from the Wuhan University of Science and Technology, Wuhan, China, in 2001, the M.Sc. degree in control theory and control engineering from Xiamen University, Xiamen, China, in 2005, and the Ph.D. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009. He is currently a Professor with the School of Biomedical Engineering, Southern Medical University, Guangzhou, China. His main areas include medical image analysis, machine learning, and computerized-aid diagnosis.

HUIJUAN ZHANG received the M.S. degree from Southeast University in 2017. His interests are in image processing and medical image analysis.

JIAN YANG received the Ph.D. degree in optical engineering from the Beijing Institute of Technology in 2007. He was a Post-Doctoral Research Fellow with the Mouse Imaging Centre, Hospital for Sick Children, Toronto, ON, Canada, from 2007 to 2009. He is currently a Professor with the School of Optoelectronics, Beijing Institute of Technology, China. His research interests include medical image processing, computer vision, virtual reality, and augmented reality.

JIASONG WU received the Ph.D. degree in signal and image processing from the University of Rennes, Rennes, France, in 2011. He is a Lecturer in the School of Computer Science and Engineering, Southeast University, Nanjing, China. His research interests are mainly focused on medical image processing and machine learning.

XIANGRUI YIN is currently pursuing the master's degree with Southeast University. His interests are in image processing and medical image analysis.

YANG CHEN received the M.S. and Ph.D. degrees in biomedical engineering from First Military Medical University, China, in 2004 and 2007, respectively. Since 2008, he has been a Faculty Member of the Department of Computer Science and Engineering, Southeast University, China. His recent work concentrates on the medical image reconstruction, image analysis, pattern recognition, and computerized-aided diagnosis.

HUAZHONG SHU received the B.S. degree in applied mathematics from Wuhan University, China, in 1987, and the Ph.D. degree in numerical analysis from the University of Rennes, Rennes, France, in 1992. He is currently a Professor with the Department of Computer Science and Engineering, Southeast University, China. His recent work concentrates on the image analysis, pattern recognition, and fast algorithms of digital signal processing.

LIMIN LUO received the Ph.D. degree from the University of Rennes, Rennes, France, in 1986. He is currently a Professor with the Department of Computer Science and Engineering, Southeast University, Nanjing, China. His current research interests include medical imaging, image analysis, computer- assisted systems for diagnosis and therapy in medicine, and computer vision.

GOUENOU COATRIEUX received the Ph.D. degree in signal processing and telecommunication from the University of Rennes 1, Rennes, France, in collaboration with Ecole Nationale Supérieure des Télécommunications, Paris, France, in 2002. He is a Professor with the Information and Image Processing Department, Institut Mines-Télécom, Telecom Bretagne, Brest, France. His research is conducted in the LaTIM Laboratory, INSERM U1101, Brest. His research interests include medical information system security, watermarking, secure processing of outsourced data, digital forensics in medical imaging, electronic patient records.

ZHIGUO GUI received the Ph.D. degree in signal and information processing from the North University of China, Taiyuan, China, in 2004. He is currently a Professor with the North University of China. His research interests include signal and information processing, image processing and recognition, and image reconstruction.

QIANJIN FENG received the M.S. and Ph.D. degrees in biomedical engineering from First Military Medical University, China, in 2000 and 2003, respectively. From 2003 to 2004, he was a Faculty Member with the School of Biomedical Engineering, First Military Medical University, China. Since 2004, he has been with Southern Medical University, China, where he is a Professor and the Dean of the School of Biomedical Engineering. His research interests are in medical image analysis, pattern recognition, and computerized-aided diagnosis.