IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 3, 2020, accepted June 9, 2020, date of publication June 15, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002534

Image Restoration for Low-Dose CT via Transfer Learning and Residual Network

ANNI ZHONG[1], BIN LI[2], NING LUO[3], YUAN XU[1], LINGHONG ZHOU[1], AND XIN ZHEN[1]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Department of Radiation Oncology, Sun Yat-sen University Cancer Center, Guangzhou 510060, China
[3]Department of Radiation Oncology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China

Corresponding authors: Linghong Zhou (smart@smu.edu.cn) and Xin Zhen (xinzhen@smu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFA0202003 and Grant 2017YFC0112900, in part by the National Natural Science Foundation of China under Grant 81874216 and Grant 61971463, and in part by the Guangzhou Science and Technology Plan Project under Grant 202002030385.

ABSTRACT Deep learning has recently been extensively investigated to remove artifacts in low-dose computed tomography (LDCT). However, the power of transfer learning for medical image denoising tasks has not been fully explored. In this work, we proposed a transfer learning residual convolutional neural network (TLR-CNN) to restore LDCT images at single and blind noise levels. A residual network was implemented to effectively estimate the difference between denoised image and its original map, and a noise-free image was obtained by subtracting the residual map from the LDCT image. The results were compared to competing baseline denoising methods in terms of quantitative metrics including the PSNR, RMSE, SSIM and FSIM. For the single noise level, the proposed method demonstrated better denoising performance than the other algorithms for both simulation data and clinical data. As for the blind denoising, the image qualities were improved for all noise levels for all the quantitative metrics, but such improvements were decreasing as the noise level decrease (higher mAs). Comparative experiments suggested that the proposed network could effectively suppress artifacts and preserve image details with faster converge rate and reduced computational time.

INDEX TERMS LDCT, image denoising, CNN, transfer learning, residual network.

I. INTRODUCTION

X-ray computed tomography (CT) is widely used to help doctors diagnose diseases and guide surgery and radiotherapy through imaging [1], [2]. Due to the health risk of radiation exposure to patients, however, researchers continue to seek ways to reduce the radiation dose by decreasing the number of projection views or lowering the tube mAs [3]. However, low-dose techniques cause streaking artifacts and degrade image quality, both of which affect clinical diagnosis significantly. To cope with the problems associated with low-dose CT (LDCT), a number of algorithms have been designed; these algorithms can be roughly divided into three categories, namely, (a) sinogram filtering techniques, (b) iterative reconstruction, and (c) postprocessing methods.

Sinogram domain filtering compensates for the missing sinogram of raw data before image reconstruction, such as filtered backprojection (FBP), is applied. Representative traditional methods to denoise in sinogram domain mainly includes bilateral filtering [4], structural adaptive filtering [5] and penalized likelihood methods [6], [7]. These approaches, however, suffer from the difficulty of raw data acquisition in clinical practices, as well as the potential loss of image edges over-smoothed by filters. Over the past decades, iterative reconstruction algorithms have been adopted in LDCT denoising. These approaches try to integrate the prior knowledge into the objective function as the penalty term and smooth out noise in the image. Commonly used image priors includes total variation [8]–[10], nonlocal means (NLM) [11], [12], dictionary learning [13], tight wavelet frames [14], and low-rank matrix decomposition [15]. Nevertheless, these methods require excessive computational times because of the back-and-forth iteration between the projection and

The associate editor coordinating the review of this manuscript and approving it for publication was Hengyong Yu.

backprojection, which limits the clinical use of these algorithms.

Unlike the previous techniques, postprocessing algorithms focus on reducing artifacts in the image domain instead of raw projection data so that they can easily be incorporated into the clinical CT imaging workflow. Although image-based noise does not obey a regular distribution, numerous efforts have been proposed for CT noise suppression. For example, block-matching 3D (BM3D) is such a popular postprocessing approach which has shown remarkable improvements in noise reduction and spatial resolution in several CT-imaging tasks [16]–[18]. By exploring similarities between the central pixels and pixels in a voxel neighborhood, Chen *et al.* applied a large-scale NLM method to LDCT reconstruction images to suppress noise and artifacts [19]. Similarly, Li *et al.* and Green *et al.* improved the NLM method via noise map estimation and a database of high-SNR CT patches, respectively [20], [21]. Based on sparse representation theory, Chen *et al.* applied a k-means singular value decomposition (K-SVD) approach [22] to a fast dictionary learning and made it effectively suppress mottled noise and streak artifacts [23].

Recently, deep learning (DL) [24] has been successfully applied to many areas, such as speech recognition, visual object recognition, and object detection, thus inspiring researchers to expand its utility into the image-denoising field. For LDCT image denoising tasks, researchers have proposed many effective convolutional neural networks (CNNs) [25] which were proved to be superior than commercial iterative reconstruction algorithms [26]. Recently, generative adversarial networks (GANs) were witnessed applications in LDCT denoising [27]–[29]. In particular, it is proven that residual images have a simpler topological structure so that complex artifacts from CT reconstruction images can be learned with greater ease than the original artifact-free images [30]. Various residual learning architectures [31], [32] have been successfully applied for medical image denoising. For example, Kang *et al.* introduced a deep residual learning CNN architecture using a directional wavelet to extract the directional component of artifacts and remove them from LDCT images [33]. Chen *et al.* combined the autoencoder, deconvolution and shortcut connections to a residual encoder–decoder CNN (RED-CNN) for LDCT noise suppression and structural preservation [34]. Wolterink *et al.* trained a CNN to predict the noise and then merged it into GAN architecture and claimed to obtain improved denoising ability [35]. Wu *et al.* proposed a cascaded training network for residue artifacts estimation and removal [36]. However, the obstacle for current studies were the limited sample size available for training of a relatively large CNNs. As the CNN architecture is getting deeper, larger dataset is needed. But access to raw and clear medical images is often limited for security and privacy.

Transfer learning technique has shed new light on the data scarcity issue on CNNs training. Transfer learning hypothesizes that the source and target domains are related and similar in sample distribution, while the target domain often has fewer high-quality training data and labels than the source domain. Although natural and medical images differ significantly, massive works for transferring shared knowledge from the former to the latter have shown considerable success. For example, Zhen *et al.* learned transferred knowledge from VGG-16 CNN to address data limitations in a rectum toxicity classification model [37]. Jiang *et al.* explored transfer learning by pre-training GoogLeNet and AlexNet on a large-scale visual dataset to tackle the problem of limited image samples in breast mass classification tasks [38]. Moreover, ImageNet-pretrained CNNs have been widely used for object detection and segmentation, such as chest pathology identification [39]–[42]. Such tremendous progress showed that fine-tuning pre-trained CNN models from natural image dataset to medical image tasks was an effective method to avoid overfitting and had been proved to achieve satisfactory performance. As for LDCT denoising, Shan *et al.* proposed an initial 3D CPCE denoising model, which can be directly obtained by extending a trained 2D CNN, and then fine-tuned to incorporate the 3D spatial information from adjacent slices [43]. They also proposed a transfer learning framework that first trained several basic CNN units that were fused in a cascaded fashion for the denoising task [44]. Gong *et al.* introduced a parameter-constrained GAN with Wasserstein distance and perceptual loss for low-dose PET image denoising [45]. A transfer learning strategy was developed to train the proposed model with parameters being constrained in order to make the training process more efficient and thus with improved denoising capability. These reported methods all employed the transfer learning techniques and claimed to be able to effectively suppress image noise while preserving image details.

Inspired by these efforts, we explored the feasibility of applying transfer learning to a residual neural network (TLR-CNN) for LDCT image denoising. We also applied the model to perform denoising with unknown noise level (i.e., blind denoising). For noise estimation, we designed a CNN architecture with a residual network to achieve matching from LDCT to a noise map. We pre-trained this model using natural images and then fine-tuned its shared parameters with limited LDCT images in order to remove the artifacts in LDCT. Finally, we confirmed our hypotheses using a series of comparative experiments.

The rest of this paper is organized as follows. In Section II, we introduce the workflow of our proposed TLR-CNN and the network architecture. In Section III, we provide a detailed description of the implementation of the method and the configuration of the parameters in the model. Results are also shown and evaluated by comparison with several state-of-the-art denoising methods. Finally, Section IV draws discussions and conclusion.

II. METHODS

A. NETWORK ARCHITECTURE

Our proposed TLR-CNN is designed to predict the noise in LDCT, the difference between the original and noise-free

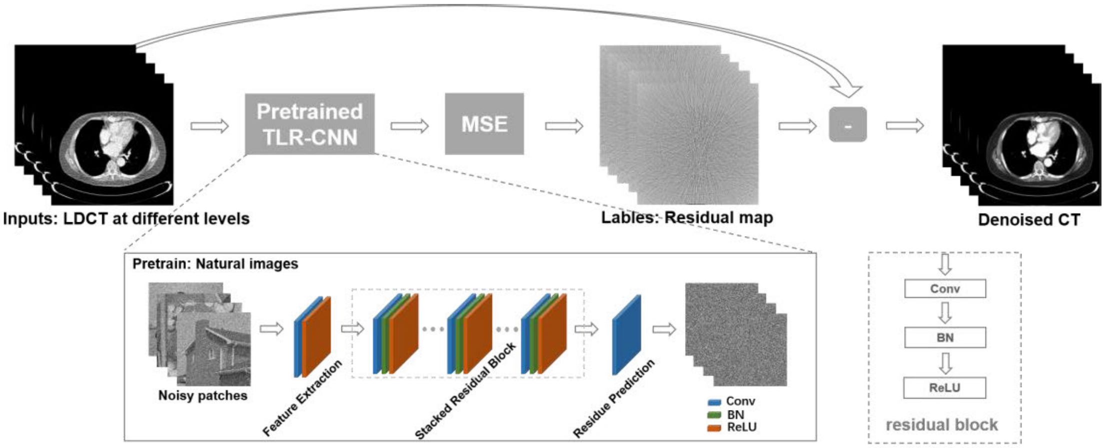

FIGURE 1. **Architecture of the proposed TLR-CNN.**

maps. Residual learning was utilized in our work to improve the effectiveness of the training process.

Let us assume $y \in R^{m \times n}$ to be a LDCT image contaminated with noise, and $x \in R^{m \times n}$ is the corresponding normal-dose image. Their relationship is formulated as $y = \Im(x)$, where $\Im$ denotes the corrupting process that contaminates normal-dose CT image due to the quantum noise. For TLR-CNN, we adopted residual learning formulation to train a residual mapping $\Re(y)$ such that $x = y - \Re(y)$. The noise reduction problem can be solved by minimizing the mean squared error (MSE) between the desired residual maps and estimated ones from noisy image

$$\zeta(\Omega) = \arg\min_{\Omega} \left\| \Re(y_i; \Omega) - (y_i - x_i) \right\|^2, \tag{1}$$

which is served as the objective function of TLR-CNN by learning the trainable parameters Ω. Here, $\{(y_i; x_i)\}_{i=1}^{N}$ represents N noisy-clean training pairs. In our work, Adam [46] were utilized to optimize the above objective function.

As shown in Figure 1, the proposed LDCT image blind denoising TLR-CNN mainly includes three types of computation layers: feature extraction, stacked residual blocks, and residue prediction, which are detailed below.

1) FEATURE EXTRACTION

The first layer was set to extract the feature maps for the following residual learning, including convolution (Conv) and nonlinear activation, formulated as

$$L_0(y) = \mathrm{ReLU}(\mathrm{Conv}(y, W, b)), \tag{2}$$

where W and b denote the weights and biases of the first layer. Here we used rectified linear unit, ReLU for short, to be the nonlinear activation function, formulated as

$$\mathrm{ReLU} = \max(0, \kappa), \tag{3}$$

where κ represents the convolutional kernel, formulated as

$$\kappa = \sum_i W_i * y_i + b_i \quad i = 1, 2, \ldots, n, \tag{4}$$

where $*$ represents the convolution operator, and n is the number of convolutional kernel. Here, W can be considered as n_0 convolutional kernel with a size of $s_0 \times s_0 \times c_0$, and b is a feature vector with size of $n_0 \times 1$. Parameter c_0 is the channel of feature maps in the layer and set to 1 for grayscale images.

2) STACKED RESIDUAL BLOCKS

In this study, batch normalization (BN) [47] was used to boost the denoising performance with input layer normalization by adjusting and scaling the activation. Previous experiments have shown that the integration of residual learning and BN demonstrated superior performance among a series of relative experiments [48].

Each residual block in the proposed TLR-CNN was composed of three operations, connected by Conv, BN, and ReLU. Unlike the residual building block in He's work [31], our residual block has no bypass connection such that the residual images are fed into the next Conv layer directly. The next connected layers are stacked by a single residual block, Conv + BN + ReLU, formulated as:

$$L_i(y) = \mathrm{ReLU}(\mathrm{BN}(\sum_i W_i * L_{i-1}(y) + b_i)) \quad i = 1, 2, \ldots, l, \tag{5}$$

where l is the number of stacked residual blocks, and W_i with a size of $s_i \times s_i \times c_i$ and b_i with a size of $n_i \times 1$ are the weights and bias of the Conv operator in i^{th} residual block. Parameter c_i is the number of convolutional kernels in the previous convolutional layer, equal to n_{i-1}.

3) RESIDUE PREDICTION

The final layer of our network was a single convolution layer to predict the residue map from multiple channels of the previous layer, and could be formulated same as Formula (4). What is different is, there is only one convolutional kernel with size of $s' \times s' \times c'$, to compound the residue prediction of the model in only one channel. We then obtained the noise-free image by subtracting the residue prediction from the LDCT image.

B. TRANSFER LEARNING

We utilized the transfer learning technique in our method to avoid over-fitting and reduce the training time. Initializing the parameters randomly in the CNN model could easily lead to overfitting and even gradient explosion [49], [50]. Options to utilize the pre-trained parameters in the model include: 1) "shallow tuning": fine-tune only the last few fully connected layers; and 2) "deep tuning": fine-tune all the network layers. As suggestion by Choi *et al.* [51], fine-tuning the last layer may not maximally extract the knowledge from a pre-trained network, it's better to activate all intermediate layers to find the most effective representation. Diamant *et al.* [52] implemented a general-image trained network as initialization of the network by all the layers, trained on ImageNet data to categorize the chest radiograph data. Similarly, Yuan *et al.* [53] enhance the resolution of hyperspectral images by exploiting the knowledge from a whole pre-trained CNN developed by natural images, and achieved comparable performance. Inspired by these work and considering the substantial difference between the source application (natural image) and target application (LDCT image), we opted to fine-tune all layers in our proposed TLR-CNN.

C. DATA PREPARATION

The proposed TLR-CNN was first pre-trained by using the natural noisy-clean image pairs, and then was fine-tuned by the LDCT image dataset from end to end.

1) NATURAL IMAGES

We adopted a dataset including 400 natural images [54] to pre-train the TLR-CNN for blind Gaussian denoising. The images were cut and tailored into patches of 40 × 40 pixels, and a data augmentation strategy, including rotation, flipping, and scaling transformation, was applied to generate additional data samples for the training process [55]–[57], as shown in Figure 2. Then, we simulated the noisy images by adding Gaussian noise to the clean natural image patches to generate noisy-clean training patch pairs, i.e.,

$$\delta(\gamma) = \gamma + \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(\gamma-\mu)^2}{2\sigma^2}), \tag{6}$$

where γ represents the original image patch, μ represents the mean value of Gaussian noise, and σ represents the Gaussian noise level ranging within [1, 50] in current study. Finally, we generated 217,856 blind noisy-clean patch pairs for the network pre-training.

2) LDCT IMAGES

The CT image dataset employed included 5,761 normal-dose CT images of pixels from 60 patients obtained from The Cancer Image Archive (TCIA). These images covered different sites of the human body, including the lung, colon and head–neck, etc. Among the 5,761 images, 5,311 images were randomly selected for model training, while the remaining 450 images were used for testing. We followed Zeng's method [58] to inject Poisson and Gaussian noise into the normal-dose CT images in the sinogram domain to simulate LDCT images.

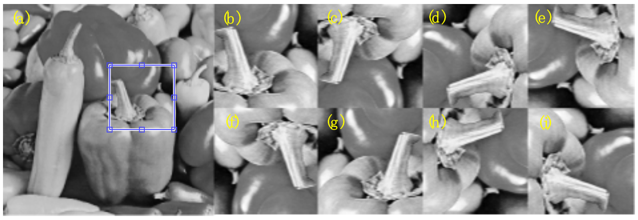

FIGURE 2. **Data augmentation of the natural-image dataset: (A) Original natural image; (B)–(I) Augmented images by rotation and flipping.**

First, the CT values of the normal-dose images $I_{nd,HU}$ were transformed to attenuation coefficients $I_{nd,coff}$ by using the following formula:

$$I_{nd,coff} = (\frac{I_{nd,HU}}{1000} + 1) \times \nu_{water}, \tag{7}$$

where ν represents the linear attenuation coefficient of the scanned substances. At a tube voltage of 120 kVp, ν_{water} is equal to 0.2154. Second, we used the GPU-accelerated forward projection operation in the Astra toolbox [59] to simulate the corresponding normal-dose CT sinogram $S_{nd,coff}$ by:

$$S_{nd,coff} = \mathrm{P}I_{nd,coff}, \tag{8}$$

in which P is the projection matrix and is defined by both the volume geometry and the projection geometry. Here, we set a 2D parallel beam geometry with detector numbers of 1024, and projection angles of $2\pi/1024$.

Third, the LDCT sinogram $S_{ld,coff}$ at different tube current ranges between [10:10:100] mAs was simulated by using the method introduced by Zeng *et al.* using the following formula:

$$S_{ld,coff} = S_{nd,coff} + \hat{I}_i\{\mathrm{Poisson}(\lambda) + \mathrm{Gaussian}(m_e + \sigma_e^2)\} \\ i = 10, 20, \ldots, 100mAs, \tag{9}$$

where $\hat{I}_i$ is the measured noisy transmission datum, λ is the mean number of photon passing through the patient, and m_e and σ_e^2 are the mean and variance of electronic noise, which are determined by the specific mAs. Here, Poisson and Gaussian represent the injected noise, which obey Poisson distribution and uniform distribution, respectively [58].

Finally, we used the analytical reconstruction method FBP [60] in Astra toolbox to reconstruct the LDCT attenuated coefficients $I_{ld,coff}$, accelerated with GPU. Then we use reverse transform of Formula (7) to generate the CT value images $I_{ld,HU}$.

Three typical examples of the simulated normal-dose CT and LDCT images were shown in Figure 3. Finally, we extracted image patches from the normal-dose

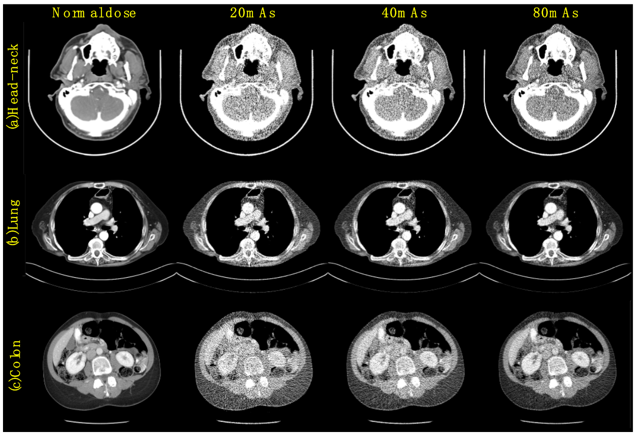

FIGURE 3. Typical CT images of the (a) head-neck, (b) lung, (c) colon, and their simulated LDCT images at 20, 40 and 80 mas, displayed in [−160, 240]HU.

CT images $I_{nd,HU}$ and LDCT images $I_{ld,HU}$ using a similar procedure for processing the natural images as described above.

III. EXPERIMENTAL DESIGN AND RESULTS

A. PARAMETER CONFIGURATION

For the neural network architecture, the sizes of convolutional filters were set to 3 × 3. The numbers of filters in the stacked residual blocks were set to 64, and the first and last layers numbered 64 and 1, respectively. At the beginning, weights in the convolution operation were randomly initialized to satisfy the Gaussian distribution with zero mean and a standard deviation of 0.01. The other hyper-parameters of the model were initialized with a weight decay of 0.0005 and momentum of 0.9 [61]. Different number of stacked residual blocks ranging from 5 to 30 were used and the corresponding performances were compared. We also used gradient clipping to limit the magnitude of the gradient to avoid gradient explosion. The gradient was clipped to the predefined range of [−0.005, 0.005].

In the pre-training process, the learning rate was initialized to 1e-3 and then decayed to 1e-4 when half number of the epochs was reached. The patch size was 40, the batch size was 128, and the number of epochs was 50. As for the fine-tuning process, a smaller learning rate [1e-4, 1e-5] was set to avoid the shared parameters in the pre-trained model to be destroyed. And in fine-tuning, the patch size was 128, the batch size was 16 and the model could converge within only one or two epochs from end to end.

B. RESULTS

We compared our model with state-of-the-art denoisers, namely, BM3D [62], KSVD [63], WNNM [64], FFDNet [65] and WavResNet [66], where the BM3D, KSVD, and WNNM are traditional image-based methods, whereas FFDNet and WavResNet are DL-based models. For quantitative assessments, we utilized four evaluation indices, namely, PSNR [67], RMSE [68], SSIM [69], and FSIM [70], where the PSNR and RMSE are metrics to quantify the image quality, while SSIM and FSIM are methods to assess the image similarity based on human visual neurobiology and perception.

1) SINGLE DENOISING MODEL

We first evaluated the denoising capability of the proposed TLR-CNN models with 15 residual blocks that trained with LDCT images at known noise levels of 30, 50, and 70 mAs, respectively. Figure 4 shows the comparison results of the proposed method and the results of other baseline methods. To show the image details, the ROI areas (in the red box) indicated in Figure 4 were enlarged in Figure 5. Visually, the results of the proposed TLR-CNN are better than those competing methods. For quantitative comparison, we measured all the 450 LDCT images in the testing set compared with different methods in terms of means ± SDs, as shown in Table 1. It can be seen that all the denoising methods were able to suppress image noise, while TLR-CNN scored the highest quantitatively, followed by the WavResNet, BM3D, FFDNet, KSVD and WNNM methods.

2) BLIND DENOISING MODEL

To obtain a generalizable model for different noise levels, we randomly mixed LDCT images at [10:10:100] mAs to the training set and trained the pre-trained model within only one epoch. The number of stacked residual blocks was 15. After training, we selected two slices to test TLR-CNN and compared the results with those of BM3D, K-SVD, WNNM, FFDNet and WavResNet. Figures 6 and 7 show the observation results of LDCT at 10 and 40 mAs, respectively. We can observe the severe streaking artifacts caused by LDCT. All of the denoising methods could remove the noise to a certain degree. Our proposed TLR-CNN could not only remove the artifacts but also preserve the image structure better, while other methods retained some streaking artifacts and showed over-smoothing. Figure 8 shows the pseudo-color results of Figures 6 and 7, the difference between the NDCT and the LDCT, BM3D, K-SVD, WNNM, FFDNet, WavResNet and TLR-CNN images. We can observe that our proposed method could push the output distribution closer to the ground truth. Table 2 shows the quantitative results of the detailed ROIs in Figures 6 and 7. For visualized results of 10 and 40 mAs, our proposed method clearly outperforms the other algorithms.

In this experiment, the blind TLR-CNN could be generalized to different noise levels even when the actual level is unknown. To test the blind denoising ability of the proposed model, we used ten sets of different levels of noise from 10mAs to 100mAs for a total of 450 LDCT test images. The PSNR, RMSE, SSIM and FSIM values of each group were tested with the blind TLR-CNN model, and the average values were calculated and compared with the other five denoising methods, as shown in Table 3. The results indicate that all the methods with these ten sets of noise can

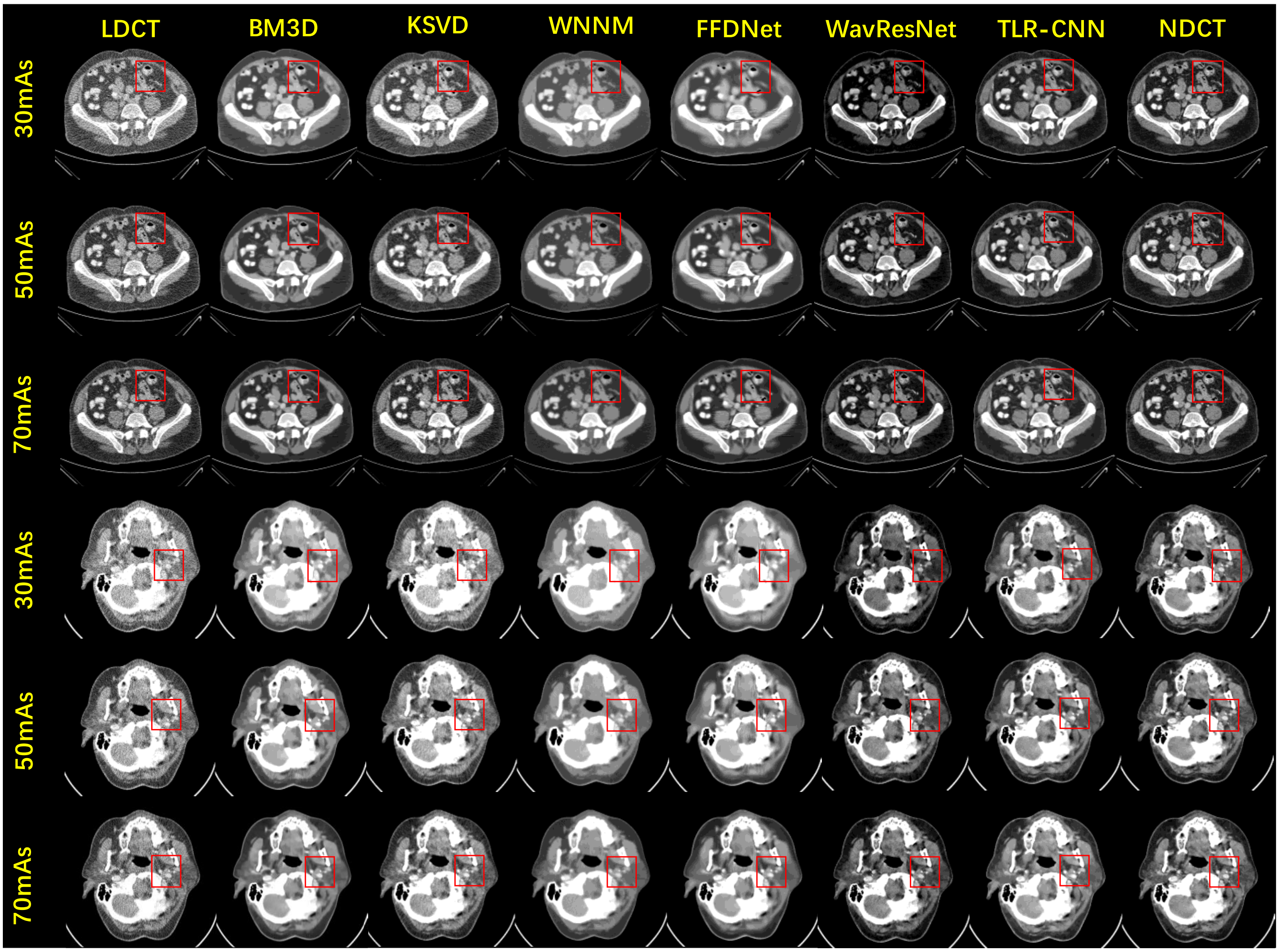

FIGURE 4. **Results of denoising models for different signal noise levels: 30, 50, and 70 mAs. From the first to the last column: LDCT, BM3D, KSVD, WNNM, FFDNet, WavResNet, TLR-CNN, and NDCT. Red rectangles indicate ROIs as shown in Figure 5. Images are displayed in [−160,240] HU.**

improve the image quality when compared with the LDCT. We calculated the increasing rate of the quantitative values, the ratio of the difference between LDCT and the denoised image by TLR-CNN. As shown in Figure 10, the maximum absolute increases in PSNR, RMSE, SSIM and FSIM were 30.81%, 65.05%, 94.45% and 9.90%, respectively. The image qualities were improved for all noise levels for all the PSNR, SSIM, FSIM and RSME, but such improvements were decreasing as the noise level decrease (higher mAs). In fact, in Table 3, we can see that the TLR-CNN scored the highest quantitative values in the range of 20 to 60 mAs, however, this superiority was not observed at other noise levels.

C. MODEL AND PERFORMANCE TRADE-OFF

In this subsection, we explored several factors that will affect the performance of TLR-CNN, including transfer learning, number of residual blocks and optimization of the model.

1) TRANSFER LEARNING

In our proposed method, we first pre-trained TLR-CNN using natural images with blind Gaussian noise and then fine-tuned the model by blind LDCT images; the combined process is known as the transfer learning strategy. In order to verify the necessity of pre-training, a comparative experiment between three models was conducted. The first model was trained by LDCT images from end to end directly. The second model was double-trained with the LDCT image by firstly pre-trained with the LDCT images (with comparable size of natural images) and then further fine-tuned with the LDCT images. While the third model was pre-trained with the natural images and then fine-tuned with the LDCT images. The comparative results were shown in Table 4. Compared with the original LDCT, the model without pre-training could hardly achieve satisfactory denoising capability, while the models with pre-training (either with LDCT or natural images) demonstrated superior denoising performances, and the model pre-trained with natural image and fine-tuned with LDCT performed the best with respect to all the evaluation metrics. These observations were consistent with the previous studies [71], [72] which reported that the model pre-trained in natural image domain seemly achieved better performance than that pre-trained with CT images.

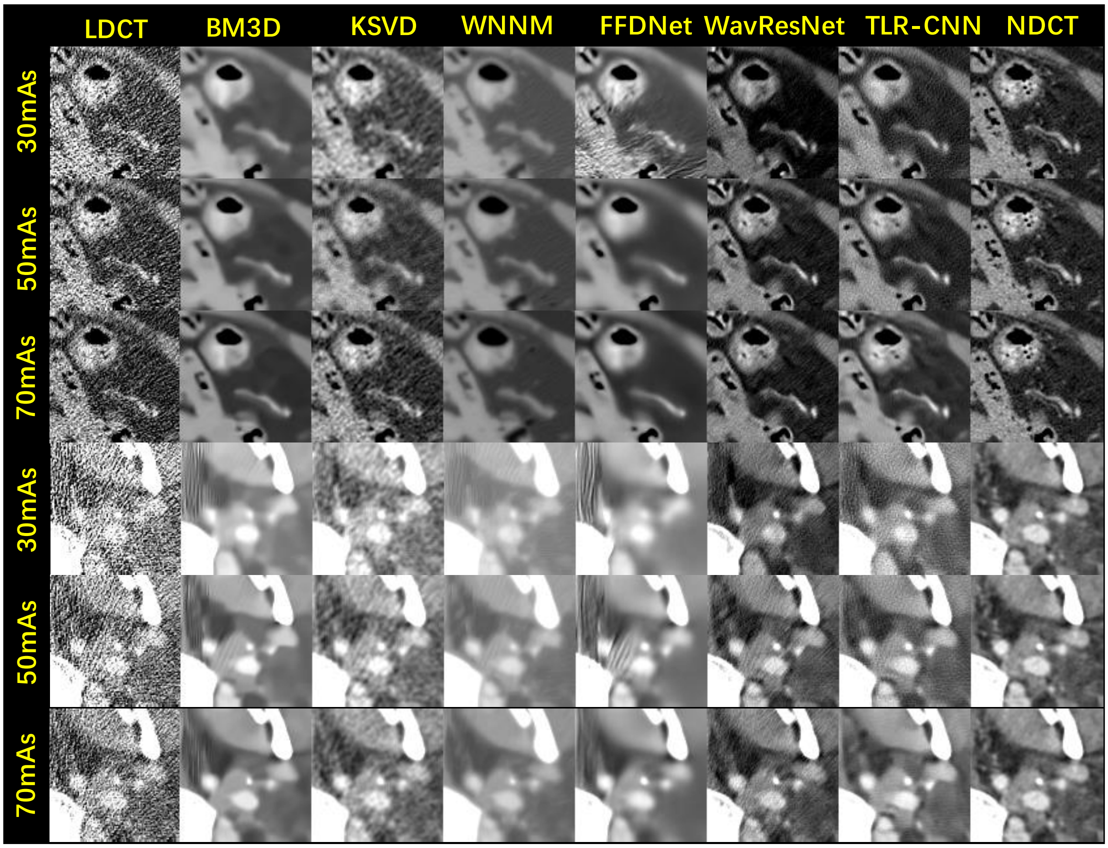

FIGURE 5. ROIs marked by red in the results of denoising models for different signal noise levels in Figure 4. Displayed in [−160,240] HU.

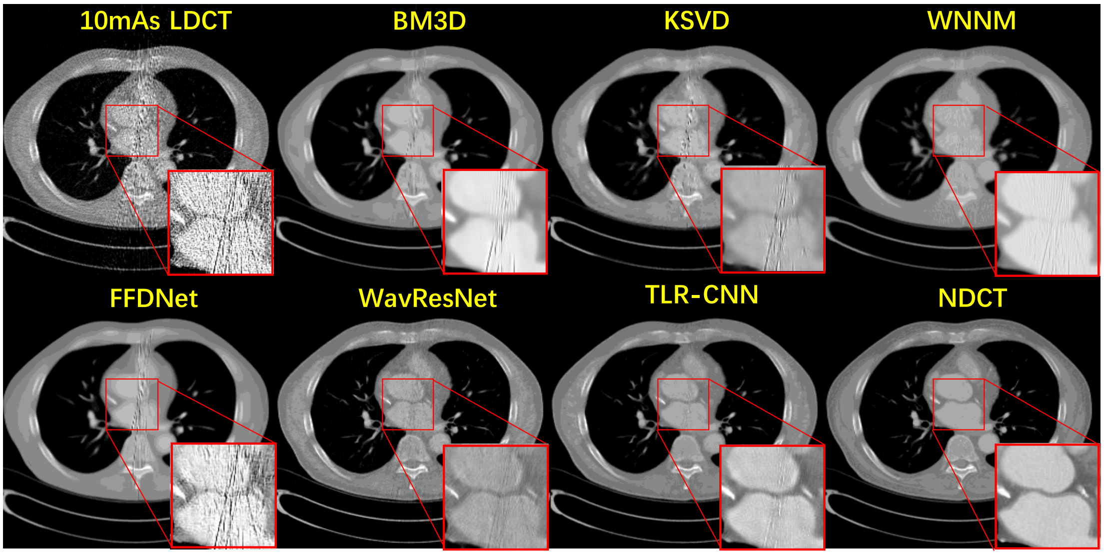

FIGURE 6. Results from the lung image for comparison: 10mAs LDCT, BM3D, K-SVD, WNNM, FFDNet, WavResNet, TLR-CNN and NDCT. The ROIs are marked by red and enlarged in the images. Displayed in [−400, 400] HU.

This might ascribe to the fact that natural images usually show more complex textures which are beneficial for learning of low-level features of medical images. Besides, pre-training using additional data from different domains would tend to lead to better generalization and thus improved denoising performance as observed in current study.

TABLE 1. Quantitative results (mean±SDs) associated with different algorithms and dosage for the testing dataset.

Noise	Method	PSNR(dB)	RMSE	SSIM	FSIM
30 mAs	LDCT	32.05 ± 1.83	6.5113 ± 1.3681	0.6176 ± 0.0906	0.9631 ± 0.9915
	BM3D	35.98 ± 1.14	4.0870 ± 0.5183	0.8104 ± 0.0446	0.9827 ± 0.0030
	KSVD	35.19 ± 0.95	4.4645 ± 0.4694	0.8016 ± 0.0440	0.9731 ± 0.0085
	WNNM	34.66 ± 1.08	4.7498 ± 0.5973	0.7993 ± 0.0424	0.9642 ± 0.0062
	FFDNet	35.61 ± 1.00	4.2564 ± 0.4820	0.7978 ± 0.0452	0.9697 ± 0.0071
	WavResNet	36.48 ± 0.89	3.8457 ± 0.3841	0.9147 ± 0.0251	0.9889 ± 0.0025
	TLRCNN	**42.56 ± 1.32**	**1.9229 ± 0.3516**	**0.9494 ± 0.0224**	**0.9915 ± 0.0034**
50 mAs	LDCT	34.90 ± 1.93	4.6985 ± 1.0420	0.7384 ± 0.0773	0.9700 ± 0.0169
	BM3D	39.51 ± 1.07	2.7178 ± 0.3207	0.8801 ± 0.0329	0.9863 ± 0.0028
	KSVD	37.95 ± 1.04	3.2521 ± 0.3765	0.8708 ± 0.0326	0.9782 ± 0.0062
	WNNM	36.97 ± 1.26	3.6531 ± 0.5533	0.8658 ± 0.0304	0.9667 ± 0.0060
	FFDNet	38.88 ± 0.80	2.9126 ± 0.2649	0.8647 ± 0.0333	0.9733 ± 0.0071
	WavResNet	40.30 ± 0.90	2.4770 ± 0.2514	0.9362 ± 0.0197	0.9923 ± 0.0018
	TLRCNN	**44.46 ± 1.14**	**1.5402 ± 0.2433**	**0.9645 ± 0.0143**	**0.9937 ± 0.0025**
70 mAs	LDCT	36.99 ± 2.05	3.7072 ± 0.8804	0.8203 ± 0.0660	0.9782 ± 0.0126
	BM3D	42.77 ± 1.22	1.8714 ± 0.2517	0.9425 ± 0.0155	0.9897 ± 0.0022
	KSVD	40.09 ± 1.57	2.5651 ± 0.4435	0.9348 ± 0.0146	0.9804 ± 0.0055
	WNNM	38.52 ± 1.64	3.0800 ± 0.6071	0.9252 ± 0.0147	0.9680 ± 0.0060
	FFDNet	42.08 ± 1.06	2.0227 ± 0.2447	0.9312 ± 0.0152	0.9825 ± 0.0046
	WavResNet	42.63 ± 1.89	1.8925 ± 0.1849	0.9316 ± 0.0195	0.9938 ± 0.0018
	TLRCNN	**46.04 ± 1.23**	**1.2864 ± 0.2105**	**0.9739 ± 0.0085**	**0.9949 ± 0.0021**

TABLE 2. Quantitative results of detailed ROIs in Figs. 6 and 7.

	Fig. 6				Fig. 7			
	PSNR(dB)	RMSE	SSIM	FSIM	PSNR(dB)	RMSE	SSIM	FSIM
LDCT	24.44	15.2926	0.3043	0.5941	30.10	7.9756	0.5599	0.7720
BM3D	31.76	6.5812	0.8935	0.9305	37.95	3.2301	0.9629	0.9441
K-SVD	32.42	6.1051	0.8911	0.9145	37.76	3.3011	0.9534	0.9635
WNNM	32.72	5.8957	0.9296	0.9364	36.62	3.7635	0.9517	0.9474
FFDNet	30.36	7.7350	0.8250	0.8731	37.32	3.4708	0.9500	0.9394
WavResNet	36.66	3.7460	0.8732	0.9251	39.15	2.8136	0.9344	0.9609
TLR-CNN	**40.11**	**2.5189**	**0.9411**	**0.9656**	**41.51**	**2.1423**	**0.9582**	**0.9699**

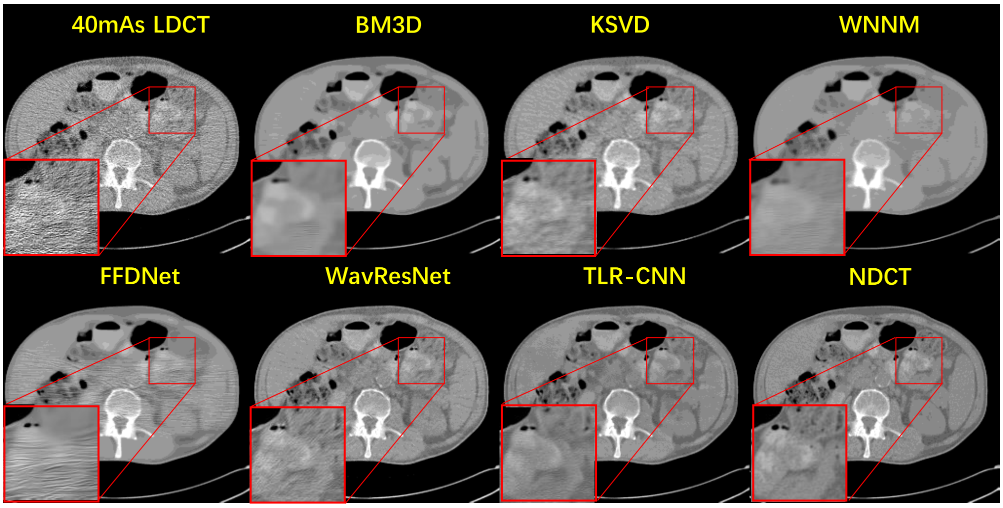

FIGURE 7. Results from the lung image for comparison: 40mAs LDCT, BM3D, K-SVD, WNNM, FFDNet, WavResNet, TLR-CNN and NDCT. The ROIs are marked by red and enlarged in the images. Displayed in [−400, 400] HU.

2) NUMBER OF RESIDUAL BLOCKS

The deeper the neural network is, the better performance will achieve by the model. However, limited training data and the hardware context could restrict the depth of the model. We tested different numbers of residual blocks using the same training dataset to investigate the trade-off between

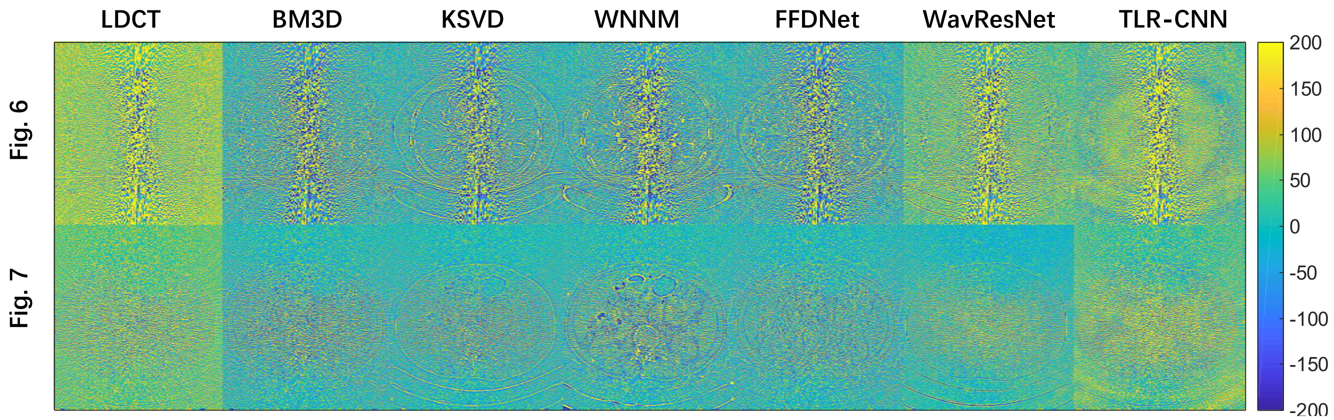

FIGURE 8. The difference between the NDCT and the LDCT, BM3D, K-SVD, WNNM, FFDNet, WavResNet and TLR-CNN images, rendered in pseudocolor for Fig. 6 and 7.

TABLE 3. Quantitative results (mean value) associated with different algorithms for different noise levels.

Index	Method	10mAs	20mAs	30mAs	40mAs	50mAs	60mAs	70mAs	80mAs	90mAs	100mAs
PSNR (dB)	LDCT	26.81	30.07	32.05	33.60	34.90	36.03	36.99	37.82	38.50	39.06
	BM3D	31.00	33.89	35.98	37.80	39.51	40.06	42.78	43.78	**44.74**	**45.27**
	KSVD	30.80	33.36	35.19	36.91	38.42	39.77	41.09	41.92	42.66	43.13
	WNNM	30.67	33.40	34.66	35.94	36.97	37.79	38.40	38.88	39.16	39.29
	FFDNet	30.64	33.57	35.61	37.36	38.93	40.37	42.08	42.96	43.89	44.05
	WavResNet	**35.71**	37.61	39.14	40.50	41.72	42.74	**43.74**	**43.99**	44.16	44.04
	TLR-CNN	35.07	**38.94**	**41.08**	**42.35**	**43.06**	**43.35**	43.66	42.96	42.46	41.88
RMSE	LDCT	11.9887	8.1783	6.5113	5.4512	4.6985	4.1365	3.7072	3.3763	3.1226	2.9316
	BM3D	7.2570	5.1960	4.0870	3.3079	2.7178	2.2715	1.8678	1.6683	**1.4990**	**1.4125**
	KSVD	7.4168	5.5139	4.4645	3.6692	3.0830	2.6415	2.2721	2.0712	1.9082	1.8123
	WNNM	7.5286	5.7252	4.7489	4.1055	3.6531	3.3342	3.1177	2.9621	2.8756	2.8377
	FFDNet	7.5669	5.3814	4.2564	3.4782	2.8977	2.4580	2.0227	1.8368	1.6659	1.6512
	WavResNet	**4.2194**	3.3874	2.8370	2.4203	2.1057	1.8756	**1.6766**	**1.6362**	1.6099	1.6321
	TLR-CNN	4.6400	**2.9263**	**2.2757**	**1.9674**	**1.8138**	**1.7584**	1.6980	1.8381	1.9451	2.0736
SSIM	LDCT	0.3911	0.5295	0.6176	0.6843	0.7384	0.7833	0.8203	0.8503	0.8737	0.8908
	BM3D	0.7014	0.7694	0.8104	0.8469	0.8801	0.9080	0.9312	0.9523	**0.9639**	**0.9700**
	KSVD	0.6891	0.7570	0.8016	0.8238	0.8605	0.8916	0.9282	0.9379	0.9531	0.9633
	WNNM	0.6975	0.7584	0.7993	0.8343	0.8658	0.8931	0.9152	0.9368	0.9480	0.9550
	FFDNet	0.6910	0.7590	0.7978	0.8333	0.8649	0.8918	0.9312	0.9385	0.9541	0.9574
	WavResNet	**0.7747**	0.8231	0.8567	0.8843	0.9075	0.9267	0.9537	**0.9544**	0.9628	0.9680
	TLR-CNN	0.7683	**0.8626**	**0.9055**	**0.9309**	**0.9465**	**0.9547**	**0.9583**	0.9538	0.9454	0.9334
FSIM	LDCT	0.8862	0.9346	0.9531	0.9634	0.9700	0.9747	0.9782	0.9808	0.9829	0.9846
	BM3D	0.9646	0.9775	0.9827	0.9848	0.9863	0.9872	0.9883	0.9889	0.9892	0.9902
	KSVD	0.9470	0.9664	0.9728	0.9768	0.9805	0.9831	0.9845	0.9863	0.9873	0.9882
	WNNM	0.9547	0.9616	0.9656	0.9667	0.9647	0.9647	0.9678	0.9684	0.9681	0.9686
	FFDNet	0.9373	0.9584	0.9697	0.9743	0.9764	0.9791	0.9825	0.9849	0.9874	0.9895
	WavResNet	**0.9765**	0.9845	**0.9895**	0.9907	0.9907	0.9917	0.9923	0.9928	0.9931	0.9934
	TLR-CNN	0.9739	**0.9864**	0.9893	**0.9908**	**0.9918**	**0.9925**	**0.9933**	**0.9934**	**0.9936**	**0.9937**

TABLE 4. Quantitative results for evaluation of transfer learning (Mean±SD).

	PSNR(dB)	RMSE	SSIM	FSIM
LDCT	34.70 ± 4.27	5.3357 ± 2.9835	0.7217 ± 0.1724	0.9614 ± 0.0376
Without pretraining	35.37 ± 3.01	4.8841 ± 2.2138	0.8120 ± 0.1071	0.9631 ± 0.0289
Pretrained with LDCT	38.65 ± 3.01	3.1538 ± 1.0565	0.8837 ± 0.0656	0.9855 ± 0.0103
Pretrained with natural image	**41.44** ± **2.84**	**2.3000** ± **0.9696**	**0.9172** ± **0.0649**	**0.9900** ± **0.0088**

performance and network depth. We tested 5, 10, 15, 20, 25 and 30 residual blocks for single noise levels of 30, 50 and 70 mAs. The models with 5, 10, 15, 20, 25 and 30 residual blocks featured 1.9e+5, 3.7e+5, 5.6e+5, 7.4e+5, 9.3e+5 and 1.1e+6 parameters, respectively.

Figure 9 shows the testing results of the models in terms of four image quality indices and the average value of the 450 testing images. Though model of 15 residual blocks performs the best at noise level of 30 mAs, blocks number of 10, 15 and 20 are similar at noise level of 50 and 70 mAs. To statistically evaluate the sensibility of residual blocks, one-way ANOVA test with $p<0.05$ was performed. The statistical results are in Table 5. There is no significant difference between 10, 15 and 20 residual blocks at noise

TABLE 5. Statistical analysis of FSIM for different residual blocks at different noise levels (Mean±SD).

V1	V2	30mAs	50mAs	70mAs
15 blocks	LDCT	0.9531 ± 0.0252*	0.9700 ± 0.0169*	0.9782 ± 0.0126*
	5 blocks	0.9894 ± 0.0055*	0.9894 ± 0.0044*	0.9917 ± 0.0033*
	10 blocks	0.9857 ± 0.0060*	0.9938 ± 0.0020	0.9948 ± 0.0018
	20 blocks	0.9887 ± 0.0045*	0.9940 ± 0.0021	0.9951 ± 0.0020
	25 blocks	0.9850 ± 0.0052*	0.9894 ± 0.0031*	0.9920 ± 0.0025*
	30 blocks	0.9721 ± 0.0100*	0.9908 ± 0.0030*	0.9918 ± 0.0026*

* indicates $P < 0.05$, which means significantly different.

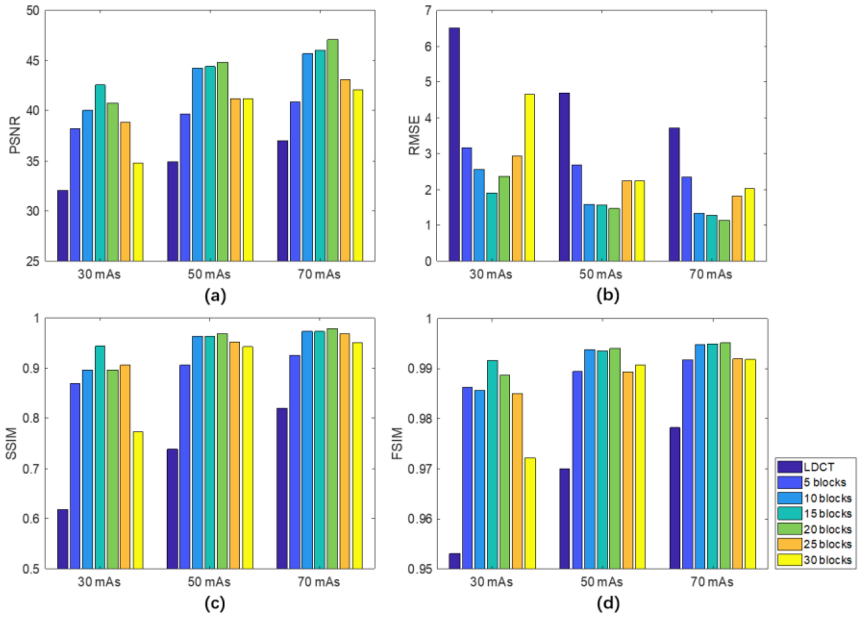

FIGURE 9. Results of models with different numbers of residual blocks. (a) PSNR; (b) RMSE; (c) SSIM; (d) FSIM.

TABLE 6. Results of models optimized with different solvers (Mean±SD).

	PSNR(dB)	RMSE	SSIM	FSIM
LDCT	34.70±4.27	5.3357±2.9835	0.7217±0.1724	0.9614±0.0376
SGD	40.36±3.29	2.6557±1.2516	0.8885±0.0781	0.9859±0.0110
Adam	**41.44±2.84**	**2.3000±0.9696**	**0.9172±0.0649**	**0.9900±0.0088**

level of 50 and 70 mAs, except 30 mAs. At noise level of 30 mAs, the best model was the one with 15 residual blocks. What's more, experiments show that a simple network was also difficult to learn the characteristics, while a deep network would cause overfitting by the limited training data, as the models with 5 and 30 residual blocks shows. Therefore, considering the balance between the performance and the computation time, we chose 15 as the number of residual blocks in our studies.

3) OPTIMIZATION

As mentioned earlier, the Adam solver was utilized to optimize the parameters in our TLR-CNN. Here, we drew a gloomy contrast with the SGD solver. Table 6 shows the results, indicating that using the Adam solver was better than using the SGD solver overall.

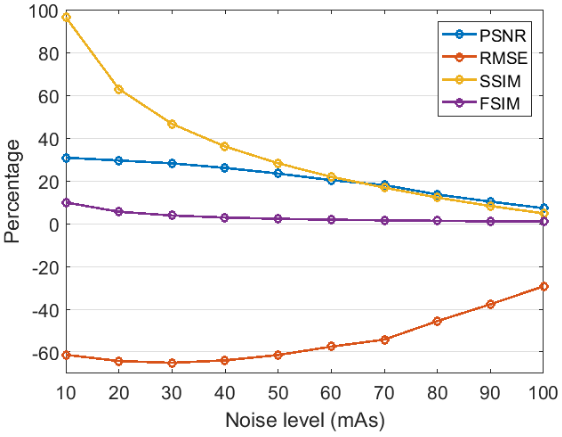

FIGURE 10. Increasing rate of different noise levels by the blind denoising model compared with original LDCT.

4) IMPLEMENTATION PLATFORM AND EFFICIENCY

Our implementation was derived from the MatConvNet toolbox for CNNs of version 1.0-beta25 and coded on Matlab 2017a. Equipped with an NVIDIA GeForce GTX TITAN Xp GPU, we used CUDA8.0 and cudnn5.1 to accelerate the training processing.

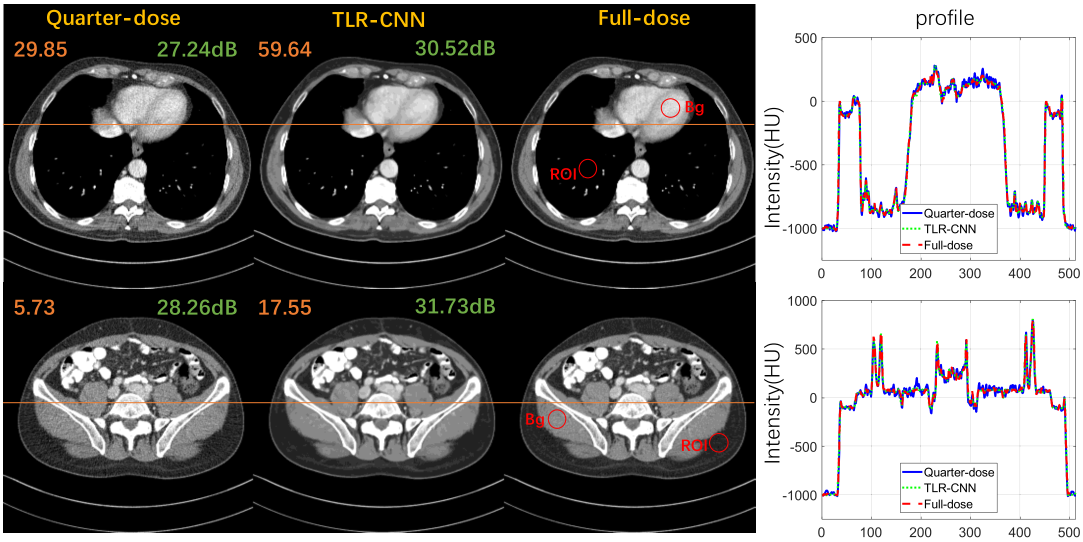

FIGURE 11. Denoising results from the test data. The first to third columns represent quarter-dose CT, denoised images by TLR-CNN and the full-dose CT, displayed in [−160, 240]. And the profiles of the corresponding positions are plotted in the fourth column.

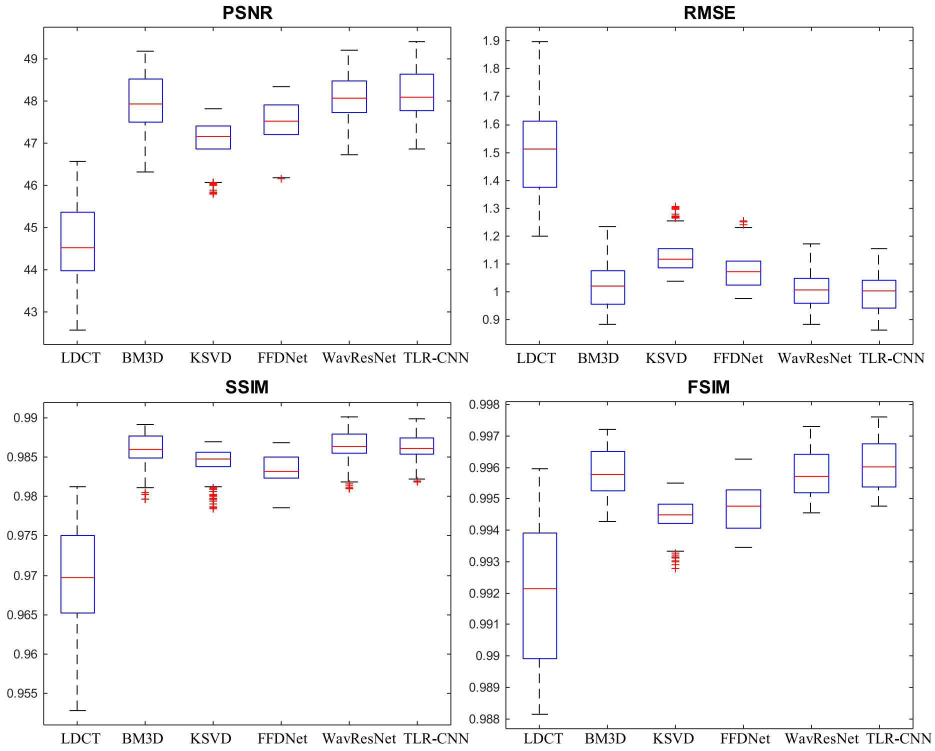

FIGURE 12. Quantitative results of LDCT, BM3D, KSVD, FFDNet, WavResNet and TLR-CNN (boxplot): PSNR, RMSE, SSIM and FSIM.

In our experiments, approximately 1.8, 3.3, 5.4, 7.5, 9.5 and 13.0 h were needed to pre-train models with 5, 10, 15, 20, 25 and 30 residual blocks, respectively. During the fine-tuning process, we generated approximately 107 patches from the LDCT dataset, which required approximately 4.6, 5.6, 8.1, 11.7, 19.5 and 29.5 h to train models from end to end with 5, 10, 15, 20, 25 and 30 residual blocks, respectively. A deeper model would require a longer training time because it uses larger parameters. In this study, the execution times for BM3D, K-SVD, WNNM, FFDNet, WavResNet and TLR-CNN were 1.49, 2.96, 180.99, 0.093 17.66 and 0.10 s, respectively. Among these methods, DL methods, TLR-CNN and FFDNet, were the most effective ways for denoising LDCT images and outperformed the traditional denoising algorithms.

D. CLINICAL DATA TEST

To validate the clinical performance of TLR-CNN, we utilized a real data set, provided by Mayo Clinic for 2016 Low Dose CT Grand Challenge. This training data includes ten patient cases, with totally 2,378 3mm thickness full-dose collected at 120 kV and quarter-dose images pairs. We chose the blind denoising model as the pre-trained model, and then fine-tuned it with 9 patients' data within only 2 epochs. The remaining one patient was used as the test set. Five-fold cross-validation was also performed to select the optimal model. And it took an average of 6.92 hours to train a model. As shown in Fig. 11, we gave two cases to indicate the denoising results and found that the noise in quarter-dose images could be removed by our proposed method very well. In addition, the profiles of the horizontal line across to the corresponding images were drawn to demonstrate that the results by TLR-CNN were close to the ground truth.

Moreover, to quantitatively illustrate the results, we restricted two areas including regions of interest (ROI) and background to calculate the value of contrast-to-noise ratio (CNR). It is a critical measurement to determine image quality for physicians' diagnostic, calculated by:

$$\mathrm{CNR} = \frac{|\mu_{ROI} - \mu_{Bg}|}{\sigma_{Bg}}, \tag{10}$$

where μ_{ROI} and μ_{Bg} is the mean value of the intensity of ROI and the background, and σ_{Bg} stands for the standard deviation [73], [74]. Also, signal-to-noise ratio (SNR) is analyzed by the whole images of LDCT and the result of our method. Higher CNR and SNR indicates better spatial resolution and higher quality of the images. As shown in Figure 11, the yellow number written in the images is the CNR value calculated from the red correspond circle regions, and the green number is the SNR value. Compared with the LDCT, the values of CNR and SNR of the result increased, which confirms that our proposed method is helpful to remove the artifact in LDCT and increase its spatial resolution. Figure 12 shows the quantitative results of quarter-dose CT, BM3D, KSVD, FFDNet, WavResNet and TLRCNN. The results show that the average values of our proposed method are 48.16 ± 0.57, 0.9993±0.0652, 0.9862±0.0016 and 0.9916±0.0008, which are superior to the other baseline methods.

IV. DISCUSSION AND CONCLUSION

In our study, the model was built based on the residual network, which utilized the integration advantages of batch normalization and residual learning to boost the denoising performance of the model. For the training process, we utilized transfer learning strategy and divided it into two steps, including pre-training and fine-tuning, which was proved to be able to help prevent overfitting by data scarcity. The datasets for these two steps were from two different domains, and the experimental evaluations demonstrated that the model pre-trained inclusive of the natural images was better than the model pre-trained merely with the CT images.

In conclusion, we proposed a deep convolutional network TLR-CNN for LDCT image restoration by transferring the knowledge from natural domain to medical image. This network comprises of residual blocks and can estimate noise maps directly and hence facilitate the process of LDCT artifact reduction. To verify the performance of our method, we performed comparisons with state-of-the-art methods and achieved superior results. The proposed TLR-CNN was proved to be able to remove image artifacts while preserving texture details for LDCT at a low computational cost. In the future, we plan to add a human reader study into clinical practice for subjective image quality evaluation. Furthermore, we can optimize our method and extend it to other CT image-processing tasks, such as CT image reconstruction, CBCT scatter correction, and geometric correction.

ACKNOWLEDGMENT

(Anni Zhong and Bin Li contributed equally to this work.)

REFERENCES

[1] F. Alam, S. Rahman, S. Ullah, A. Khalil, and A. Uddin, "A review on extrinsic registration methods for medical images," *Tech. J. Univ. Eng. Technol. Taxila*, vol. 21, pp. 110–119, Jul. 2016.

[2] N. P. Nguyen, A. Chi, M. Betz, F. Almeida, P. Vos, R. Davis, B. Slane, M. Ceizyk, D. Abraham, L. Smith-Raymond, M. Stevie, S. Jang, S. Gelumbauskas, and V. Vinh-Hung, "Feasibility of intensity-modulated and image-guided radiotherapy for functional organ preservation in locally advanced laryngeal cancer," *PLoS ONE*, vol. 7, no. 8, Aug. 2012, Art. no. e42729.

[3] M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M. A. Blake, J.-A. Shepard, and S. Saini, "Strategies for CT radiation dose optimization," *Radiology*, vol. 230, no. 3, pp. 619–628, Mar. 2004.

[4] A. Manduca, L. Yu, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough, and J. G. Fletcher, "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Med. Phys.*, vol. 36, no. 11, pp. 4911–4919, Oct. 2009.

[5] M. Balda, J. Hornegger, and B. Heismann, "Ray contribution masks for structure adaptive sinogram filtering," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1228–1239, Jun. 2012.

[6] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

[7] S. Tang and X. Tang, "Statistical CT noise reduction with multiscale decomposition and penalized weighted least squares in the projection domain," *Med. Phys.*, vol. 39, no. 9, pp. 5498–5512, Aug. 2012.

[8] 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, p. 4777, Sep. 2008.

[9] Y. Zhang, W. Zhang, Y. Lei, and J. Zhou, "Few-view image reconstruction with fractional-order total variation," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 31, no. 5, pp. 981–995, May 2014.

[10] Y. Zhang, Y. Wang, W. Zhang, F. Lin, Y. Pu, and J. Zhou, "Statistical iterative reconstruction using adaptive fractional order regularization," *Biomed. Opt. Express*, vol. 7, no. 3, pp. 1015–1029, 2016.

[11] Z. Tian, X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "Low-dose 4DCT reconstruction via temporal nonlocal means," *Med. Phys.*, vol. 38, no. 3, pp. 1359–1365, Feb. 2011.

[12] Y. Chen, D. Gao, C. Nie, L. Luo, W. Chen, X. Yin, and Y. Lin, "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Comput. Med. Imag. Graph.*, vol. 33, no. 7, pp. 495–500, Oct. 2009.

[13] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[14] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse Problems*, vol. 27, no. 11, pp. 115012–115033, 2011.

[15] J.-F. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, and H. Zhao, "Cine cone beam CT reconstruction using low-rank matrix factorization: Algorithm and a Proof-of-Principle study," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1581–1591, Aug. 2014.

[16] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Blocking matching 3D random noise filtering for absorption optical projection tomography," *Phys. Med. Biol.*, vol. 55, p. 5401, 2010.

[17] D. Kang, P. Slomka, R. Nakazato, J. Woo, D. S. Berman, C.-C. J. Kuo, and D. Dey, "Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm," *Proc. SPIE*, vol. 8669, Mar. 2013, Art. no. 86692G.

[18] K. Sheng, S. Gou, J. Wu, and S. X. Qi, "Denoised and texture enhanced MVCT to improve soft tissue conspicuity," *Med. Phys.*, vol. 41, no. 10, Oct. 2014, Art. no. 101916.

[19] Y. Chen, Z. Yang, Y. Hu, G. Yang, Y. Zhu, Y. Li, L. Luo, W. Chen, and C. Toumoulin, "Thoracic low-dose CT image processing using an artifact suppressed large-scale nonlocal means," *Phys. Med. Biol.*, vol. 57, no. 9, p. 2667, 2012.

[20] Z. Li, L. Yu, J. D. Trzasko, D. S. Lake, D. J. Blezek, J. G. Fletcher, C. H. McCollough, and A. Manduca, "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Med. Phys.*, vol. 41, no. 1, Dec. 2013, Art. no. 011908.

[21] M. Green, E. M. Marom, N. Kiryati, E. Konen, and A. Mayer, "Efficient low-dose CT denoising by locally-consistent non-local means (LC-NLM)," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 423–431.

[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[23] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.-L. Coatrieux, and C. Toumoulin, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, no. 16, pp. 5803–5820, Aug. 2013.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[25] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT via convolutional neural network," *Biomed. Opt. Express*, vol. 8, no. 2, pp. 679–694, 2017.

[26] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Mach. Intell.*, vol. 1, no. 6, pp. 269–276, Jun. 2019.

[27] X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *J. Digit. Imag.*, vol. 31, no. 5, pp. 655–669, Oct. 2018.

[28] C. You, W. Cong, G. Wang, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, and Y. Zhang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, vol. 6, pp. 41839–41855, 2018.

[29] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[30] Y. S. Han, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis," 2016, *arXiv:1611.06391*. [Online]. Available: https://arxiv.org/abs/1611.06391

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[32] X. Yin, J.-L. Coatrieux, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, Y. Chen, H. Shu, and L. Luo, "Domain progressive 3D residual convolution network to improve low-dose CT imaging," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2903–2913, Dec. 2019.

[33] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, 2017.

[34] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[35] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proc. SPIE*, vol. 6064, Feb. 2006, Art. no. 606414.

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

[37] X. Zhen, J. Chen, Z. Zhong, B. Hrycushko, L. Zhou, S. Jiang, K. Albuquerque, and X. Gu, "Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study," *Phys. Med. Biol.*, vol. 62, no. 21, p. 8246, 2017.

[38] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in *Proc. 5th Int. Conf. Bioinf. Comput. Biol. (ICBCB)*, 2017, pp. 59–62.

[39] B. van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 286–289.

[40] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 294–297.

[41] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, and B. V. Ginneken, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Image Anal.*, vol. 26, no. 1, pp. 195–202, Dec. 2015.

[42] S. Hoochang, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[43] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[44] H. Shan, U. Kruger, and G. Wang, "A novel transfer learning framework for low-dose CT," in *Proc. 15th Int. Meeting Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.*, May 2019, Art. no. 110722.

[45] Y. Gong, Y. Teng, H. Shan, T. Xiao, M. Li, G. Liang, G. Wang, and S. Wang, "Parameter constrained transfer learning for low dose PET image denoising," 2019, *arXiv:1910.06749*. [Online]. Available: http://arxiv.org/abs/1910.06749

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

[47] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[48] 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.

[49] D. Balduzzi, M. Frean, L. Leary, J. Lewis, K. W.-D. Ma, and B. McWilliams, "The shattered gradients problem: If resnets are the answer, then what is the question?" in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 342–350.

[50] G. Philipp, D. Song, and J. G. Garbonell, "Gradients explode-deep networks are shallow-resnet explained," in *Proc. 6th Int. Conf. Represent. ICLR Workshop Track*, Vanvouver, BC, Canada, 2018. [Online]. Available: https://openreview.net/forum?id=rJjcdFkPM

[51] K. Choi, G. Fazekas, M. Sandler, and K. Cho, "Transfer learning for music classification and regression tasks," 2017, *arXiv:1703.09179*. [Online]. Available: http://arxiv.org/abs/1703.09179

[52] I. Diamant, Y. Bar, O. Geva, L. Wolf, G. Zimmerman, S. Lieberman, E. Konen, and H. Greenspan, "Chest radiograph pathology categorization via transfer learning," in *Deep Learning for Medical Image Analysis*, ed. Amsterdam, The Netherlands: Elsevier, 2017, pp. 299–320.

[53] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.

[54] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2017.

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

[56] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[57] D. Han, Q. Liu, and W. Fan, "A new image classification method using CNN transfer learning and Web data augmentation," *Expert Syst. Appl.*, vol. 95, pp. 43–56, Apr. 2018.

[58] D. Zeng, J. Huang, Z. Bian, S. Niu, H. Zhang, Q. Feng, Z. Liang, and J. Ma, "A simple low-dose X-ray CT simulation from high-dose scan," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 5, pp. 2226–2233, Oct. 2015.

[59] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25129–25147, 2016.

[60] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Philadelphia, PA, USA: SIAM, 2001.

[61] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of the Trade*, vol. 7700. Berlin, Germany: Springer, 2012, pp. 437–478.

[62] 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.

[63] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[64] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

[65] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[66] E. Kang, j. Min, and J. Chul Ye, "Wavelet domain residual network (WavResNet) for low-dose X-ray CT reconstruction," 2017, *arXiv:1703.01383*. [Online]. Available: http://arxiv.org/abs/1703.01383

[67] Y. Wang, Y. Shao, Q. Zhang, Y. Liu, Y. Chen, W. Chen, and Z. Gui, "Noise removal of low-dose CT images using modified smooth patch ordering," *IEEE Access*, vol. 5, pp. 26092–26103, 2017.

[68] M. Diwakar and M. Kumar, "A review on CT image noise and its denoising," *Biomed. Signal Process. Control*, vol. 42, pp. 73–88, Apr. 2018.

[69] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[70] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[71] T. Schlegl, J. Ofner, and G. Langs, "Unsupervised pre-training across image domains improves lung tissue classification," in *Proc. Int. MICCAI Workshop Med. Comput. Vis.*, 2014, pp. 82–93.

[72] A. Menegola, M. Fornaciali, R. Pires, F. V. Bittencourt, S. Avila, and E. Valle, "Knowledge transfer for melanoma screening with deep learning," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 297–300.

[73] B. Li, C. Shen, Y. Chi, M. Yang, Y. Lou, L. Zhou, and X. Jia, "Multienergy cone-beam computed tomography reconstruction with a spatial spectral nonlocal means algorithm," *SIAM J. Imag. Sci.*, vol. 11, no. 2, pp. 1205–1229, Jan. 2018.

[74] N. Desai, A. Singh, and D. J. Valentino, "Practical evaluation of image quality in computed radiographic (CR) imaging systems," *Proc. SPIE*, vol. 7622, Mar. 2010, Art. no. 76224Q.

ANNI ZHONG received the B.E. degree in biomedical engineering from Southern Medical University, China, in 2017, where she is currently pursuing the master's degree with the School of Biomedical Engineering. Her two abstracts have been accepted as presentations for the 60th Annual Meeting and Exhibition of the American Association of Physicists in Medicine, in 2019. Her research interests include X-ray imaging, artificial intelligence, computed tomography denoising, and CBCT scatter correction.

BIN LI received the B.E. degree in biomedical engineering and the Ph.D. degree from Southern Medical University, China, in 2013 and 2018, respectively. From 2015 to 2017, he was with the Department of Radiation Oncology, UT Southwestern Medical Center. He currently works as a Medical Physicist with the Department of Radiation Oncology, Sun Yat-sen University Cancer Center, China. His main research interests include medical image reconstruction, image guided radiation therapy, and artificial intelligence in medical imaging.

NING LUO received the B.E. degree in biomedical engineering and the M.E. degree from Southern Medical University, China, in 2016 and 2019, respectively. She currently works as a Medical Physicist with the Department of Radiation Oncology, The First Affiliated Hospital, Sun Yat-sen University, China. Her research interests include X-ray imaging and image guided radiation therapy.

YUAN XU received the B.E. degree in electronic and information engineering, the master's degree in biomedical engineering, and the Ph.D. degree in biomedical engineering from Southern Medical University, China. He is currently a Lecturer with the School of Biomedical Engineering, Southern Medical University. He has published a number of research articles in image guided radiation therapy. His research interests include Monte Carlo (MC) simulation, X-ray imaging, and image guided radiation therapy.

LINGHONG ZHOU received the B.E. degree from the Hefei University of Technology and the master's degree from Tsinghua University. He is currently a Professor with the Institute of Medical Instrument, School of Biomedical Engineering, Southern Medical University. He is also an Executive Director of the Medical Physics Division, Tumor Radiation Branch, Chinese Medical Association. He is also a Standing Director of the Medical Instrument Branch, China Instrument and Control Society. He actively supervises many students at all level of study-bachelor's, master's, and Ph.D. His research interests include Monte Carlo (MC) simulation, image guided radiation therapy, and adaptive radiation therapy. He received one Second Class Prizes of the State Scientific and Technological Progress Award, and seven the Ministerial and Provincial-Level Science and Technology Awards.

XIN ZHEN was born in Taishan, Guangdong, China, in 1981. He received the B.S. degree in biomedical engineering, in 2005, and the Ph.D. degree in biomedical engineering from Southern Medical University, Guangzhou, China, in 2013. From 2011 to 2012, he was a Visiting Student with the Moores Cancer Center, University of California at San Diego, San Diego. From 2017 to 2018, he was a Visiting Assistant Professor with the Department of Radiation Oncology, University of Texas Southwestern Medical Center. He is currently an Associate Professor with the Department of Biomedical Engineering, Southern Medical University. His research interests include medical image processing, treatment outcome prediction, motion tracking, and applications of artificial intelligence in radiation oncology.