

Image Denoising for Low-Dose CT via Convolutional Dictionary Learning and Neural Network

Rongbiao Yan , Yi Liu , Yuhang Liu , Lei Wang , Rongge Zhao, Yunjiao Bai, and Zhiguo Gui 

Abstract—Removing noise and artifacts from low-dose computed tomography (LDCT) is a challenging task, and most existing image-based algorithms tend to blur the results. To improve the resolution of denoising results, we combine convolutional dictionary learning and convolutional neural network (CNN), and propose a transfer learning densely connected convolutional dictionary learning (TLD-CDL) framework. In detail, we first introduce the dense connections and multi-scale Inception structure to the network, and train the pre-model on the natural image dataset, then fit the model to the post-processing of LDCT images in the way of transfer learning. In addition, considering that a single pixel-level loss is difficult to achieve satisfactory results both in the index and visual perception, we use the compound loss function of L1 loss and SSIM loss to guide the training. The experimental result shows that TLD-CDL has a good balance between noise reduction and the preservation of details, and acquires inspiring effectiveness in terms of qualitative and quantitative perspective.

Index Terms—LDCT, convolutional dictionary learning, CNN, transfer learning.

I. INTRODUCTION

X-RAY computed tomography (CT) is one of the most important imaging modalities in modern hospitals which overcomes the shortcomings of traditional X-ray images, such as overlapping tissue structure and low density resolution. In view of the fact that CT may cause genetic damage and induce cancer in patients, LDCT has aroused considerable interest in the field of medical imaging [1]. However, reducing the radiation dose will cause noise and artifacts which seriously degrade the

quality of image, may result in misdiagnosis of radiologists. To cope with the problem associated with LDCT, the corresponding methods can be broadly classified into three categories: (a) sinogram filtering techniques, (b) iterative reconstruction (IR), and (c) image post-processing after reconstruction.

Representative traditional methods in sinogram domain mainly include penalized weighted least-squares (PWLS) algorithm [2], bilateral filtering [3] and structural adaptive filtering [4], which may result in blurred edges and loss of detail structure in the reconstructed CT image. The main principle of IR is to optimize an objective function that incorporates system model, statistical noise model and prior information about the image [5]. Commonly image priors include total variation (TV) and its variants [6], [7], Markov random field (MRF) [8] and dictionary learning [9], [10]. Nonetheless, both sinogram filtering techniques and iterative reconstruction need to access the raw projection data. In contrast, image post-processing can be directly performed on LDCT images and integrated into clinical CT system. In [11], the non-local means (NLM) was utilized to take advantage of the feature similarity within a large neighborhood in a reconstructed image. Kang et al. [12] optimized an adaptive algorithm based on classical block-matching 3D (BM3D) [13], which performed certain denoising effectiveness for low-radiation dose coronary CTA. Chen et al. [14] proposed an improved K-SVD algorithm [15] to reduce artifacts in CT images. However, the K-SVD model ignores the structural dependence between different blocks, which makes the sparse representation of the whole image highly redundant. In order to solve the problem, convolutional dictionary learning [16] is considered as an alternative method that can replace the linear representation with a sum of convolutions.

With the rapid development of deep learning, the methods based on CNN have been successfully applied in many fields, such as image classification, object detection, image denoising and image fusion. Also, a growing number of studies [17], [18], [19], [20] have shown the effectiveness of CNN in medical image processing. Chen et al. applied deep learning to the field of LDCT [21] and proposed a residual encoder-decoder convolutional neural network (RED-CNN) [1], which suppressed noise to some extent but blurred details. By extracting the directional component of artifact via the directional wavelet transform and utilizing the intra- and inter-band correlations, the network designed in [22] suppressed CT-specific noise, but some of the textures are not fully recovered. To recover the texture, wavelet domain residual network [23] (WavResNet) obtained wavelet coefficients through the operation of subtracting estimated noise

Manuscript received 17 December 2021; revised 29 October 2022; accepted 26 January 2023. Date of publication 3 February 2023; date of current version 13 February 2023. This work was supported in part by the National Nature Science Foundation of China under Grant 61801438, in part by the Science and Technology Innovation Project of Colleges and Universities of Shanxi Province under Grants 2020L0282 and 2020L0595, and in part by the Top Young Academic leader Project of North University of China under Grant QX201801. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Webster Stayman. (*Corresponding author: Yi Liu.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Biological and Medical Ethics Committee of North University of China. Application No. NUC-20200827-003, and performed in line with the Research on artifact suppression algorithms for low-dose CT.

Rongbiao Yan, Yuhang Liu, Lei Wang, and Rongge Zhao are with the Shanxi Provincial Key Laboratory for Biomedical Imaging and Big Data, North University of China, Taiyuan 030051, China (e-mail: s2005140@st.nuc.edu.cn; yuhangliu98@163.com; b1905047@st.nuc.edu.cn; zrg0920@163.com).

Yunjiao Bai is with the Department of Mechanics, Jinzhong University, Jinzhong 030619, China (e-mail: byjtengfei@126.com).

Yi Liu and Zhiguo Gui are with the State Key Laboratory of Dynamic Testing Technology, North University of China, Taiyuan 030051, China (e-mail: liuyi@nuc.edu.cn; gzgtg@163.com).

Digital Object Identifier 10.1109/TCI.2023.3241546

from the input wavelet transform bands, and got improvements in both qualitative and quantitative analysis. However, the above deep learning methods still have certain limitations such as low denoising performance and vague structure. To solve the problem of blur in processed LDCT images, a new denoising method [24] (WGAN-VGG) based on the generative adversarial network with Wasserstein distance and perceptual similarity was used for low-dose CT image denoising, and made some progress in detail preservation and visual perception. The edge enhancement based densely connected convolutional neural network (EDCNN) [25] was designed to use trainable Sobel convolution. By introducing a compound loss that combines MSE loss and multi-scales perceptual loss to solve the over-smoothed problem, the quality of LDCT processed by EDCNN was improved obviously. Fan et al. [26] adopted quadratic neurons to construct an encoder-decoder structure, referred as the quadratic autoencoder, and effectively removed noise and artifacts in LDCT denoising. Zhong et al. [27] proposed a transfer learning residual convolutional neural network (TLR-CNN) to restore LDCT images at single and blind noise levels.

Besides, Wang et al [28] constructed the first pure Transformer architecture (CTformer) for LDCT denoising, and outperforms the state-of-the-art denoising methods. Nonetheless, most of the CNN-based image restoration methods learn the mapping relation between noisy images and their clean ground truth, which lack intuitive interpretability, so that many researchers try to make algorithms have both inspiring performance and clear physical meaning. The representative strategy is combining dictionary learning with CNN, these methods [29], [30] are often called deep unfolding, whose usual step is first unfolding the traditional models in certain algorithms and then parameterizing the model by CNN. Fu et.al [31] utilized a deep convolutional sparse coding network to represent JPEG artifacts, and generated comparable or better de-blocking results. Wang [32] also used the strategy of deep unfolding to represent metal artifacts in CT images. The deep convolutional dictionary learning (Dcdicl) proposed by Zheng et al. [33] follows the representation model of dictionary learning strictly and utilizes CNN to learn the priors for both dictionary and representation coefficient, which achieved intuitive interpretability in the task of removing Gaussian noise from natural images. However, Dcdicl is still not satisfactory in processing LDCT images, since the noise model in LDCT image is complex and difficult to be described. Inspired by Dcdicl, we propose the TLD-CDL framework, which improves the quality of CT images with complex noise and artifacts. The main contributions of this paper are as follows.

- 1) Convolutional dictionary learning is applied to the post-processing of LDCT denoising. As far as we know, there are no relevant papers relating convolutional dictionary learning to the post-processing of LDCT.
- 2) To strengthen the feature extraction ability of the network, we incorporate a multi-scale structure and dense shortcut connections into the proposed framework. In order to achieve satisfactory results not only in indicators but also in visual perception, we introduce a compound loss that combines L1 loss and SSIM loss to guide the training of the model.

TABLE I
THE CONFIGURATION OF THE PROPOSED MODEL (F: KERNEL SIZE, S: STRIDE,
P: PADDING, D: CHANNEL DIM)

	Layers	Parameters	Output size
<i>InitNet_X</i>	Conv	f=5;p=2;s=1;d=64	128x128x64
	Conv	f=3;p=1;s=1;d=64	128x128x64
<i>HypaNet_(t)</i>	Conv	f=1; s=1;d=256	1x1x256
	Conv	f=1; s=1;d=4	1x1x4
<i>Net_X</i>	Softplus	beta=1;threshold=20	1x1x4
	ResBlock	{f=3;p=1;s=1;d=65} x 4	128x128x65
	Conv	f=2; s=2;d=128	64x64x128
	ResBlock	{f=3;p=1;s=1;d=128}x 4	64x64x128
	Conv	f=2; s=2;d=256	32x32x256
	ResBlock	{f=3;p=1;s=1;d=256}x 4	32x32x256
	Conv	f=2; s=2;d=512	16x16x512
	ResBlock	{f=3;p=1;s=1;d=512}x 4	16x16x512
	TConv	f=2; s=2;d=256	32x32x256
	ResBlock	{f=3;p=1;s=1;d=256}x 4	32x32x256
	TConv	f=2; s=2;d=128	64x64x128
	ResBlock	{f=3;p=1;s=1;d=128}x 4	64x64x128
	TConv	f=2; s=2;d=64	128x128x64
	ResBlock	{f=3;p=1;s=1;d=64} x 4	128x128x64
	Conv	f=3;p=1;s=1;d=64	128x128x64
	Conv	{f=3;p=1;s=1;d=64} x 4	5x5x64
	Inception	-	5x5x64
<i>Net_D</i>			

- 3) To evaluate the denoising effectiveness of the proposed model, our method is compared with the previous LDCT denoising methods and achieves leading performance in terms of indicators and visual perception. The ablation experiment confirms that every improvement of TLD-CDL relative to Dcdicl has a positive contribution.

II. RELATED WORKS

A. Convolutional Dictionary Learning

Convolutional dictionary learning is proposed to decompose the whole image by using the convolution operation to replace the matrix multiplication, which reduces the redundancy of block representation and maintains the details of the image. Most convolutional dictionary learning models adopt a universal dictionary to represent different images, and its objective function can be expressed as:

$$\min_{\mathbf{D}, \{\mathbf{X}_i\}} \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|\mathbf{D} * \mathbf{X}_i - Y_i\|_2^2 + \lambda_X \phi(\mathbf{X}_i) \quad (1)$$

where Y_i is the i -th training sample with the size of $h \times w$; $\phi(\mathbf{X}_i)$ is a penalty term, such as $\|\mathbf{X}_i\|_1$; λ_X is a regularization parameter balancing sparsity and fidelity; $\mathbf{D} = \{D_1, D_2, \dots, D_c\}$ represents a convolutional dictionary with c channels, each channel denotes a dictionary filter with the size of $k \times k$; $\mathbf{X}_i = \{X_{[i,1]}, X_{[i,2]}, \dots, X_{[i,c]}\}$ is a set of the coefficient map that represents the sample Y_i , and the size of each coefficient map is also $h \times w$; $*$ denotes convolution; $\mathbf{D} * \mathbf{X}_i = \sum_{j=1}^c D_j * X_{[i,j]}$.

The Dcdicl framework employs adaptive \mathbf{D} on each image, and its loss function can be written as:

$$\min \frac{1}{N} \sum_{i=1}^N L(\mathbf{D}_i * \mathbf{X}_i, Y_i^{gt}) \quad (2a)$$

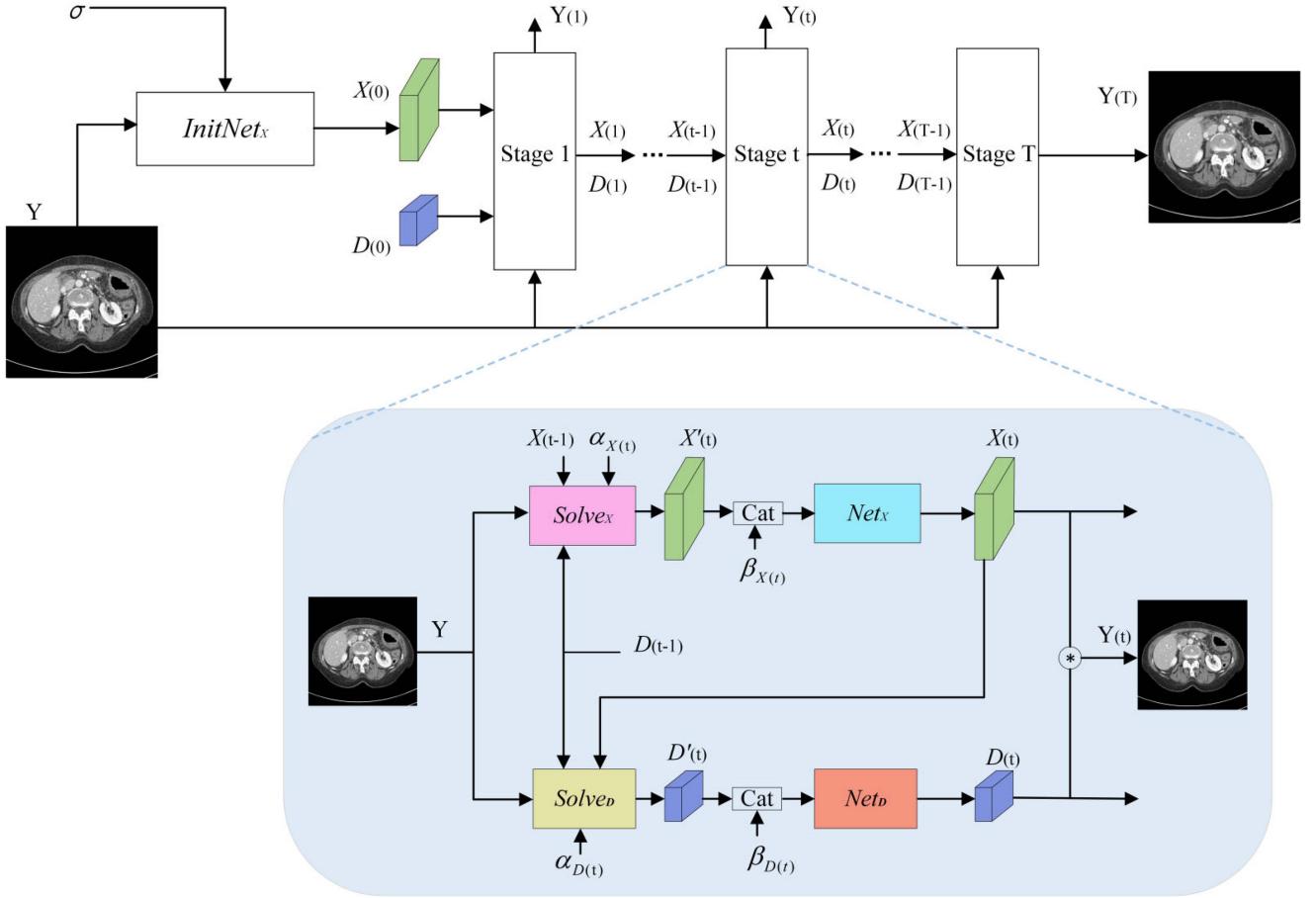


Fig. 1. The overall architecture of TLD-CDL.

$$\begin{aligned} \text{s.t. } \{\mathbf{D}_i, \mathbf{X}_i\} = & \arg \min_{D, X} \frac{1}{2\sigma_i^2} \|D * X - Y_i\|_2^2 \\ & + \lambda_X \phi(X) + \lambda_D \varphi(D) \end{aligned} \quad (2b)$$

where Y_i^{gt} represents the ground truth of noisy image Y_i ; σ_i is the noise level of image Y_i ; To unfold the objective in (2b) with Half Quadratic Splitting (HQS) algorithm, it is necessary to introduce the auxiliary variables \mathbf{X}'_i and \mathbf{D}'_i of \mathbf{X}_i and \mathbf{D}_i , then optimization problem is equivalent to minimize the following two formulas:

$$\min_{X_i, X'_i} \frac{1}{2\sigma_i^2} \|\mathbf{D}_i * \mathbf{X}'_i - Y_i\|_2^2 + \lambda_X \phi(\mathbf{X}_i) + \frac{\mu_{X_i}}{2} \|\mathbf{X}_i - \mathbf{X}'_i\|_2^2 \quad (3)$$

$$\min_{D_i, D'_i} \frac{1}{2\sigma_i^2} \|\mathbf{D}'_i * \mathbf{X}_i - Y_i\|_2^2 + \lambda_D \varphi(\mathbf{D}_i) + \frac{\mu_{D_i}}{2} \|\mathbf{D}_i - \mathbf{D}'_i\|_2^2 \quad (4)$$

where μ is the penalty coefficient, (3) and (4) can be solved iteratively. In order to make it easier to express, the subscript “ i ” in the following development is omitted. In the t -th iteration (stage):

$$\mathbf{X}'_{(t)} = \arg \min_{X^\Theta} \frac{1}{2} \|\mathbf{D}_{(t-1)} * X^\Theta - Y\|_2^2$$

$$+ \frac{\alpha_X}{2} \|X^\Theta - \mathbf{X}_{(t-1)}\|_2^2 \quad (5)$$

$$\mathbf{X}_{(t)} = \arg \min_{X^\Theta} \phi(X^\Theta) + \frac{\beta_X}{2} \|\mathbf{X}'_{(t)} - X^\Theta\|_2^2 \quad (6)$$

$$\begin{aligned} \mathbf{D}'_{(t)} = & \arg \min_{D^\Theta} \frac{1}{2} \|D^\Theta * \mathbf{X}_{(t)} - Y\|_2^2 \\ & + \frac{\alpha_D}{2} \|D^\Theta - \mathbf{D}_{(t-1)}\|_2^2 \end{aligned} \quad (7)$$

$$\mathbf{D}_{(t)} = \arg \min_{D^\Theta} \varphi(D^\Theta) + \frac{\beta_D}{2} \|\mathbf{D}'_{(t)} - D^\Theta\|_2^2 \quad (8)$$

where $\{\alpha_X, \alpha_D, \beta_X, \beta_D\} = \{\mu_X \sigma^2, \mu_D \sigma^2, \mu_X / \lambda_X, \mu_D / \lambda_D\}$; the solving process of \mathbf{X}' , \mathbf{X} , \mathbf{D}' and \mathbf{D} is briefly recorded as $Solve_X$, Net_X , $Solve_D$ and Net_D . According to [33], $Solve_X$ is based on Fast Fourier Transform (FFT) to find the closed-form solution of \mathbf{X}' . Based on the least squares method, $Solve_D$ unfolds (7) from the convolution form into the matrix multiplication form, and utilizes the modern least squares solvers to obtain the solution of \mathbf{D}' . Besides, Net_X and Net_D update the corresponding variables via neural networks.

When Dcdicl is directly applied to the LDCT denoising task, there are three problems as listed below: 1) Dcdicl is tasked to remove Gaussian noise, whilst LDCT image is mainly accompanied by streak artifacts whose distribution cannot be

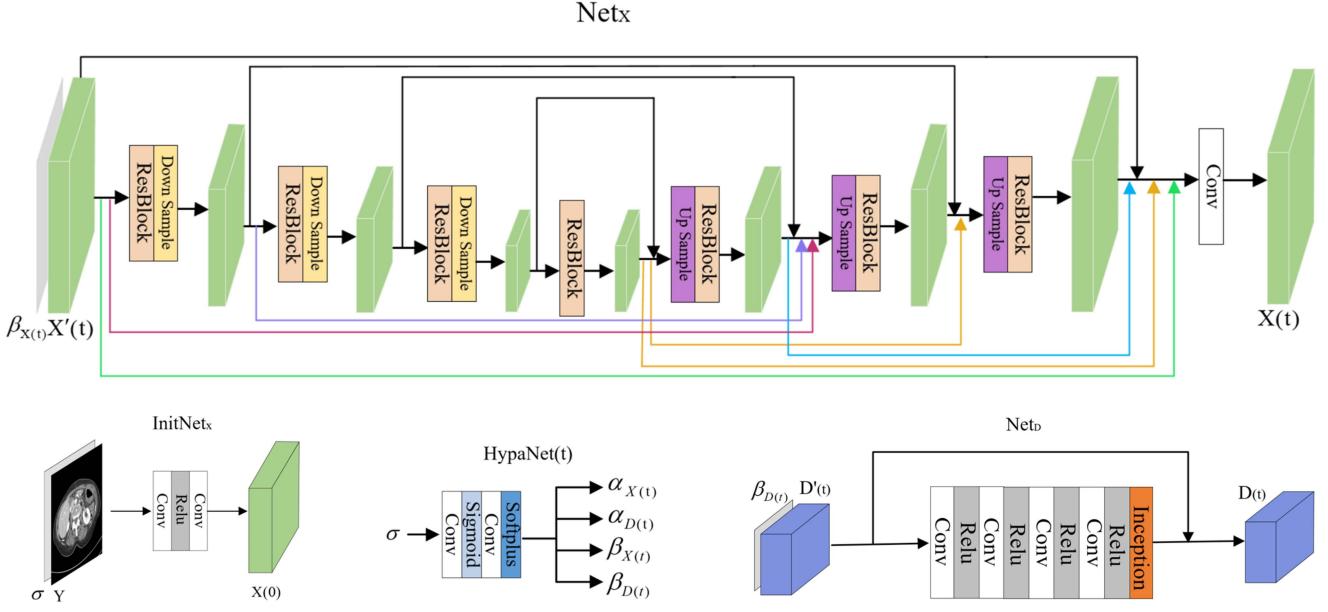


Fig. 2. The sub-network architectures of TLD-CDL.

described by math model. Thus, Dcdicl shows limited ability when performing Dcdicl directly on LDCT images. 2) Dcdicl requires a large number of paired data sets, but paired CT images are often scarce, and the structure of Net_X and Net_D is difficult to obtain multi-dimensional information. Definitely, different scales of feature maps have arisen different concerns, what captures rich spatial information while getting location information is actually a superposition of different scales of feature. In general, radiologists are most interested in low contrast lesions in CT images which need multi-scale structure to extract the contours and details. Therefore, Dcdicl will show obvious limitations when dealing with texture-rich CT images.

3) Without transfer learning, it is difficult for the Dcdicl to quickly and steadily train its model and remove complex noise in LDCT denoising.

B. Network Design

To solve the limitations of Dcdicl in LDCT denoising, we utilized transfer learning and improved Dcdicl by introducing dense connections and multi-scale Inception structure, here we call it TLD-CDL. Fig. 1 is the overall framework of the proposed TLD-CDL, whose standard deviation σ is calculated by the difference between LDCT and NDCT. As same as Dcdicl, TLD-CDL also contains four sub-networks but introduces 6 bypass connections and multi-scale Inception structure in Net_X and Net_D respectively, whose architectures are illustrated in Fig. 2. The coefficient map X is initialized with $InitNet_X$ and the hyper-parameters are predicted by $HypaNet$, Net_X and Net_D can update the corresponding variables via neural networks. The

$HypaNet$ takes noise level σ as inputs and predicts the hyper-parameters for each stage, with SoftPlus [33] as the activate function which ensures all hyper-parameters are positive. It is worth noting that in order to prevent the image from losing too

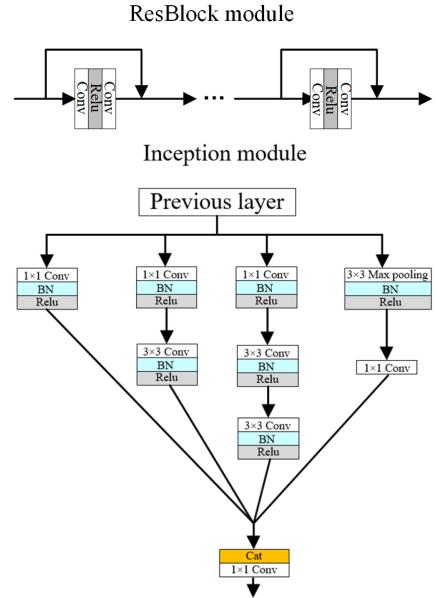


Fig. 3. Inception module and ResBlock module.

much detail after down-sampling, Net_X adopts 2×2 convolutions with step size 2 for down-sampling feature maps instead of pooling layers. According to the experimental experience, the elements of the convolutional dictionary are set to 5×5 , a shallow network can provide sufficient receptive field. Recent works of DenseNet [34] and U-net 3+ [35] have shown that convolutional networks can be substantially more accurate and efficient to extract features if they contain shorter connections between layers close to the input and those close to the output. Inspired by U-net 3+, we add dense connections to the original Net_X for back-propagation to further improve the feature reused ability, and the last Conv layer is used to fuse the information between different layers.

The structures of ResBlock module and Inception module are shown in Fig. 3, the 2 Conv layers at the end of the original Net_D are replaced by a multi-scale Inception module for the purpose of improving the learning ability from full-scale features maps. Here, a 1×1 Conv can be seen as a linear transformation of the input channel. In addition, two 3×3 Conv are used instead of 5×5 Conv to further reduce the parameters while achieving the same receptive field. The BN layers behind the convolution layers are utilized to prevent over-fitting. The more detailed configuration of the proposed model is described in Table I.

III. EXPERIMENTS

A. Implementation Details

Due to the effectiveness of transfer learning [27], [36], we used transfer learning to overcome the scarcity of CT data sets and stabilize the training process. In the pre-training process, we adopted WED [37] as the training set. Then we simulated noisy image by adding additive Gaussian noise of standard deviation σ to the clean natural image, which is set to $[0, 50]$.

$$Y = Y^{gt} + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Y^{gt} - \mu)^2}{2\sigma^2}\right) \quad (9)$$

where μ represents the mean value of Gaussian noise, and Y is generated noisy image. As for the fine-tuning process, we selected the dataset of the 2016 NIH AAPM-Mayo Clinic Low-Dose CT Grand Challenge [38], which contains paired 3mm normal-dose CT (NDCT) images and synthetic quarter-dose CT (LDCT) images with a size of 512×512 , collected from 10 patients. In our experiments, we utilized 760 pairs of CT images as our training inputs and labels, 35 pairs (selected from patient L004) are used as the test set and 50 image pairs are utilized as validation. In order to simulate the noise distribution of LDCT images, we calculated standard deviation σ of the difference (the result of LDCT minus NDCT). During the training, we applied a data augmentation strategy that crops 128×128 images randomly. In addition, we carried out a robustness verification experiment, for example, trained on Mayo dataset, and testing on clinic dataset, which was approved by the hospital ethics committee. Through trial and error experiments, the standard deviation $\sigma = 0.12$ in the clinical study.

TLD-CDL applies InitNetx to construct the initial X and simply uses zero initialization for D , then it updates X and D through multi-stage learning. In the process of optimization, we utilized the Adam optimizer with a default configuration to update the learnable parameters. The loss function of each stage chose the compound loss:

$$\ell^{Compound} = \eta \ell^{L1} + (1 - \eta) \ell^{SSIM} \quad (10)$$

where η was set as 0.85. The total reconstruction loss is weighted by the loss of each stage, and the output Y of the last stage is the final reconstructed image. Furthermore, we set the weight of the loss imposed on the last stage as 1, and the loss weight of the rest ($T-1$) stages are $1/(T-1)$. This work achieved pre-training process with mini-batch size of 4, the learning rate started from $4e-5$ and

slowly decreased down to $1e-5$ after 30 epochs. In the fine-tuning process, a smaller learning rate [$1e-6, 5e-7$] was set to avoid the shared parameter in the pre-trained model being destroyed, and conducted 14 epochs to make the model converge. Since the sizes of the testing images are often different from the training, regularization intensity of D needs to be adjusted, so we scaled α_D by $(h_{test} \times w_{test}) / (128 \times 128)$.

All the experiments were implemented in Python with the Pytorch library. Equipped with a Windows computer (Intel Core i7-9700K CPU, RAM 32G, NVIDIA GeForce RTX 2080 SUPER), we used NVIDIA CUDA11.1 to accelerate the training processing.

B. Comparison With Others

To verify the effectiveness of the proposed method, we compared our model with the classic K-SVD [14], BM3D [12], WGAN-VGG [24], WavResNet [23], RED-CNN [1], EDCNN [25], and CTformer [28] in terms of visual effects and quantitative analysis. For quantitative assessments, we utilized six evaluation indices, namely, peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [39], gradient magnitude similarity deviation (GMSD) [40], feature similarity (FSIM) [41], visual information fidelity (VIF) [42] and its multi-scale pixel domain implementation (VIFs) [42].

1) Mayo Experiment: The denoising effects of different denoising methods on two representative slices (denoted as Case1 and Case2) are shown in Figs. 4–5. All CT images in axial view are displayed in the window [-160, 240]HU. To demonstrate the detail preservation performance, we enlarged the ROI (region of interest) areas from red and blue boxes. We can observe normal human tissue structure (shown by the ROI3 in Fig. 5) and lesions (such as ROI1 in Fig. 4) clearly in NDCT images. However, the LDCT image is contaminated by visible streak artifacts (shown in ROI2 in Fig. 4) and speckle noise. Compared with the LDCT images, all algorithms mentioned above have denoising capabilities at different levels. KSVD and BM3D eliminated noise dramatically, but both algorithms introduced additional artifacts, such as the region marked with green arrows. In contrast, the denoising results of RED-CNN did not introduce additional artifacts, but they appeared blurred visual effects. WavResNet and WGAN-VGG made some progress in detail preservation, while there still exists speckle noise in Figs. 4(f) and 5(e). As shown, EDCNN is better than that of above methods in finely preserving the image details, but the additional shadow (marked with yellow arrows in Fig. 4(h)) and streak artifacts (such as ROI2 in Fig. 4(h)) are still visible. Images (i) in Figs. 4–5 show that CTformer can significantly suppress noise. From the perceptual performance, the proposed TLD-CDL removed a wide range of noise levels while maintaining the edge information. Furthermore, the complete structure of lesion and the edges of ROI regions are clear, while the shadow and the streak artifacts are also the shallowest of all the contrast methods. As we all know, the texture information of CT images affects the diagnosis of radiologists. The result of the proposed method had a prominent effect in suppressing artifacts and restored the textures and details closer to NDCT.

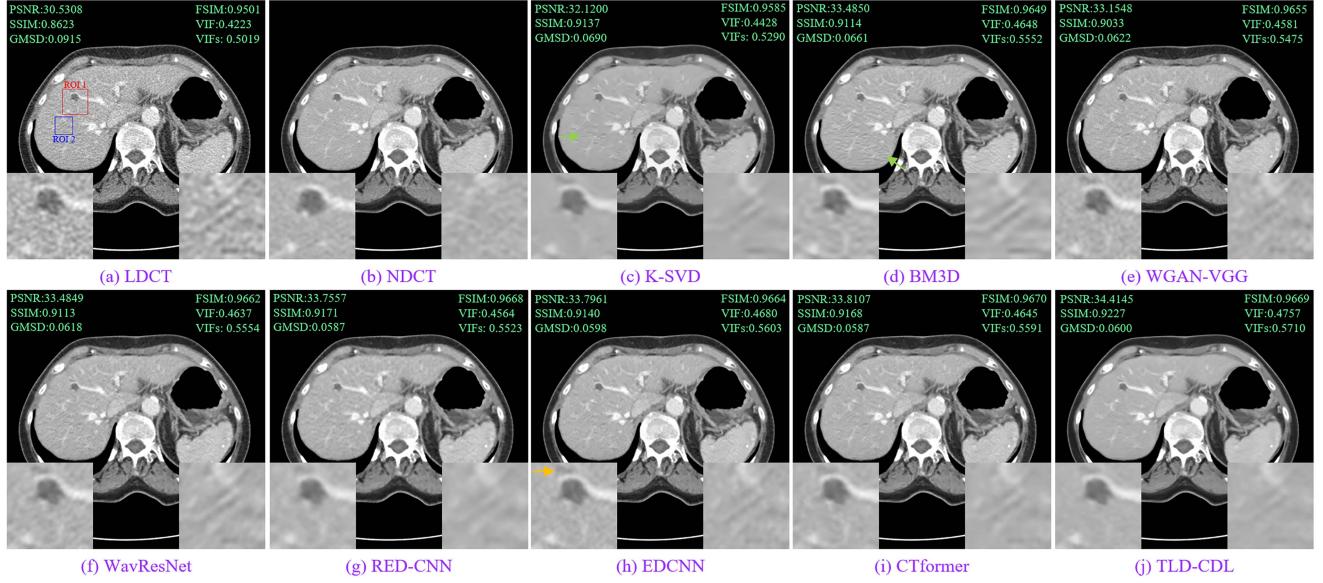


Fig. 4. Comparison of denoising results by different methods for Case1.

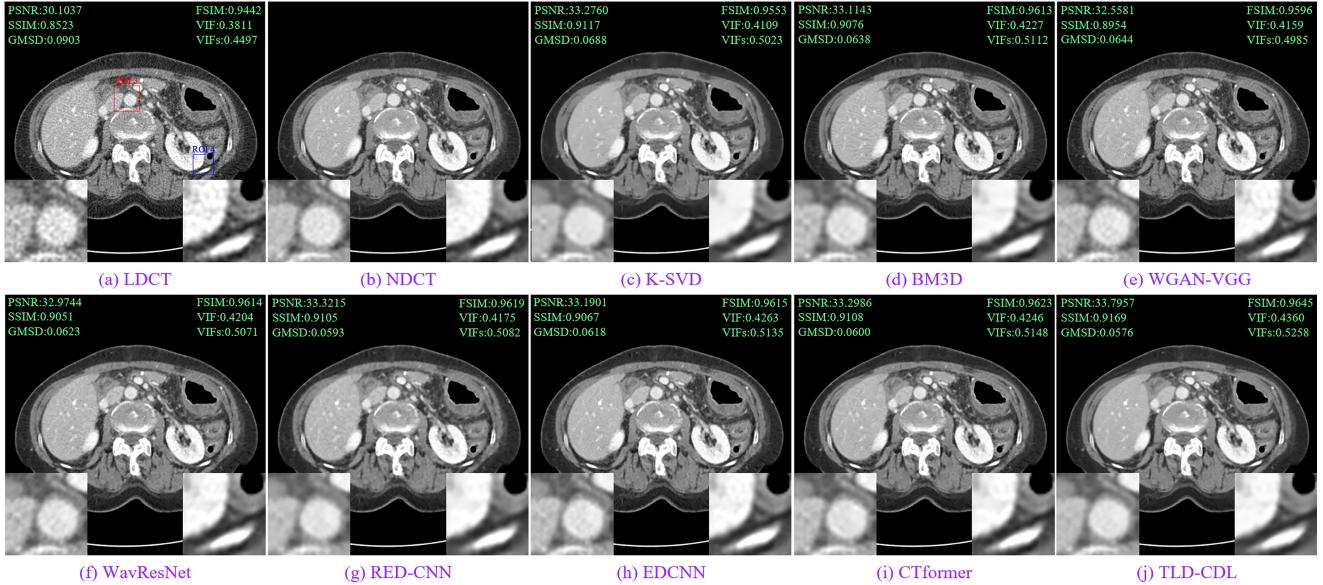


Fig. 5. Comparison of denoising results by different methods for Case2.

For quantitative evaluation, the average values of all the tested CT images were utilized to measure the effectiveness of different methods, and the profiles of all results are shown in Table II (the red font represents the best and the second is marked with blue font). Besides, the values of six metrics are tagged in the left-upper and right-upper corners in the images in Figs. 4–5. It can be seen that our method has the best quantitative performance in all objective indicators, and the improvement of PSNR and VIF is remarkable. Consistent with the visual judgment, WGAN-VGG and WavResNet are inferior in quantitative values as they cannot remove artifacts effectively. In general, there are no significant differences in the index values of K-SVD, BM3D, RED-CNN EDCNN, and CTformer. EDCNN had a sub-optimal performance in VIF, but there is a significant gap

of 0.6dB in PSNR with TLD-CDL. CTformer is sub-optimal in most metrics, which is why transformer architecture is popular. Considering that doctors tend to focus on local regions of interest in actual clinical applications, as shown in Fig. 6, we quantified the performance of ROIs in the results obtained by different methods. Comparing the PSNR of each method in the same ROIs, TLD-CDL always has the leading quantitative performance, while BM3D and WGAN show the worst numerical value. TLD-CDL and RED-CNN has a significant lead in SSIM scores on ROI1 and ROI2. Furthermore, there is no obvious gap (both PSNR and SSIM) between the remaining algorithms in ROI3 and ROI4. Judging from the quantitative performance and visual perception, our algorithm acquired the most optimal effectiveness in LDCT denoising task.

TABLE II
QUANTITATIVE RESULTS OF THE DIFFERENT ALGORITHMS (MEAN \pm SD)

	PSNR	SSIM	GMSD	FSIM	VIF	VIFs
LDCT	30.2805 \pm 0.5197	0.8585 \pm 0.0124	0.0876 \pm 0.0063	0.9452 \pm 0.0039	0.3931 \pm 0.0161	0.4643 \pm 0.0157
K-SVD	33.2760 \pm 0.3758	0.9117 \pm 0.0064	0.0688 \pm 0.0014	0.9553 \pm 0.0025	0.4214 \pm 0.0123	0.5023 \pm 0.0112
BM3D	33.2346 \pm 0.4626	0.9110 \pm 0.0090	0.0625 \pm 0.0044	0.9616 \pm 0.0024	0.4328 \pm 0.0139	0.5225 \pm 0.0129
WGAN-VGG	32.7117 \pm 0.4116	0.8999 \pm 0.0088	0.0624 \pm 0.0624	0.9607 \pm 0.0025	0.4259 \pm 0.0149	0.5113 \pm 0.0140
WavResNet	33.0984 \pm 0.4664	0.9084 \pm 0.0087	0.0606 \pm 0.0038	0.9619 \pm 0.0026	0.4323 \pm 0.0154	0.5197 \pm 0.0145
RED-CNN	33.3665 \pm 0.3933	0.9132 \pm 0.0074	0.0584 \pm 0.0030	0.9621 \pm 0.0023	0.4271 \pm 0.0139	0.5193 \pm 0.0132
EDCNN	33.3322 \pm 0.4042	0.9103 \pm 0.0079	0.0600 \pm 0.0032	0.9622 \pm 0.0024	0.4370 \pm 0.0145	0.5256 \pm 0.0135
CTformer	33.3942 \pm 0.3801	0.9141 \pm 0.0075	0.0589 \pm 0.0030	0.9625 \pm 0.0023	0.4352 \pm 0.0147	0.5259 \pm 0.0131
TLD-CDL	33.9225 \pm 0.3422	0.9201 \pm 0.0068	0.0579 \pm 0.0036	0.9636 \pm 0.0022	0.4467 \pm 0.0136	0.5377 \pm 0.0123

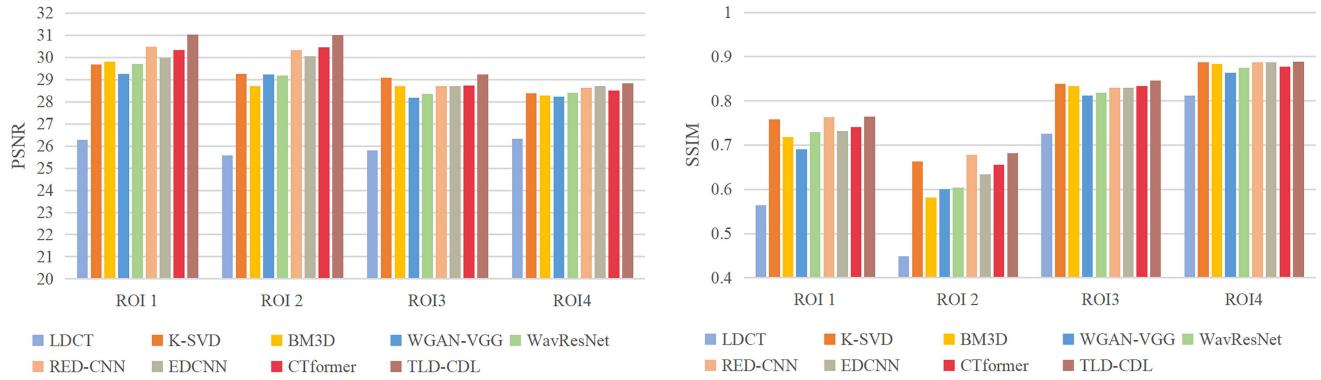


Fig. 6. Quantitative performance of different ROIs in Figs. 4-5.

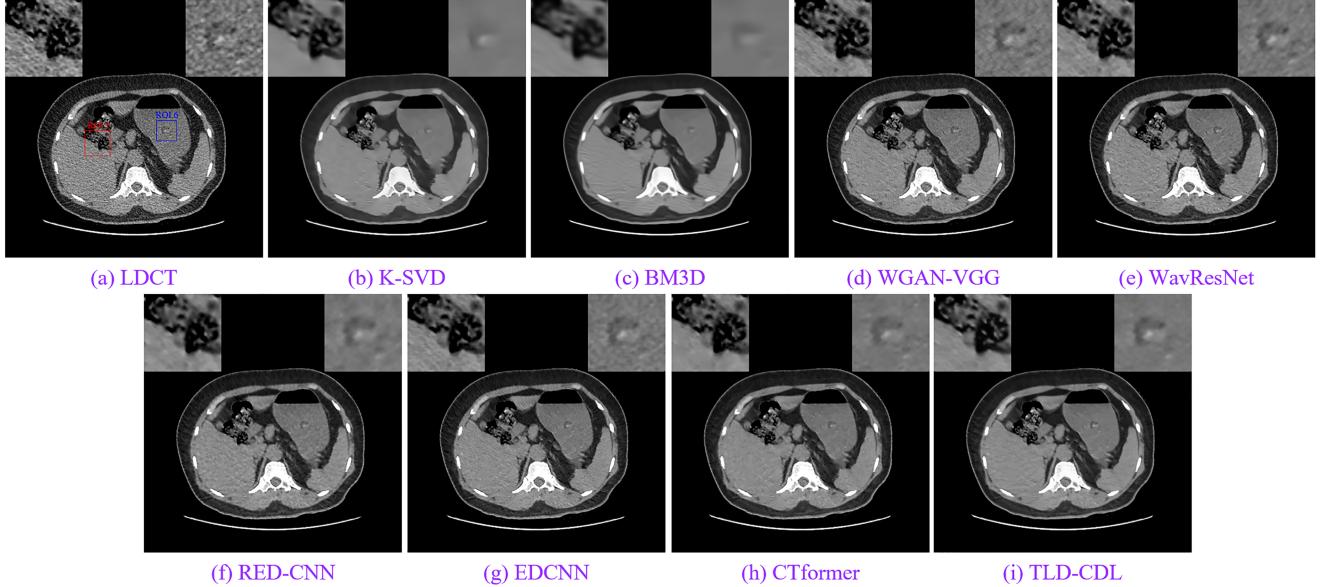


Fig. 7. The denoised results for Case3. The display window is [-160,240]HU.

2) *Real Clinic Experiment*: We carried out a robustness verification experiment on the real clinic CT images, and showed two representative slices (denoted as Case3 and Case4) in Figs. 7-8. We also selected four ROI areas (marked by rectangles in Figs. 7(a) and Fig. 8(a)) for better comparison, and put them on the left-upper and right-upper corners. It can be seen that K-SVD and BM3D acquire over-smoothing results (see the ROIs in Figs. 7(b)-(c) and 8(b)-(c)). WGAN-VGG, WavResNet,

and EDCNN retain edges and details, see ROI5 in Fig. 7, but they still exist speckle noise and streak artifacts (as shown in ROI6-ROI8). The results of RED-CNN blur the lesion (see ROI6 in Fig. 7(f)) and texture details (see ROI7 in Fig. 8(f)). CTformer and the proposed TLD-CDL have the leading effect of noise suppression, but CTformer still needs to be improved in the preservation of tissue textures (such as ROI5 in Fig. 7(h)) and lesion details (such as ROI6 in Fig. 7(h)).

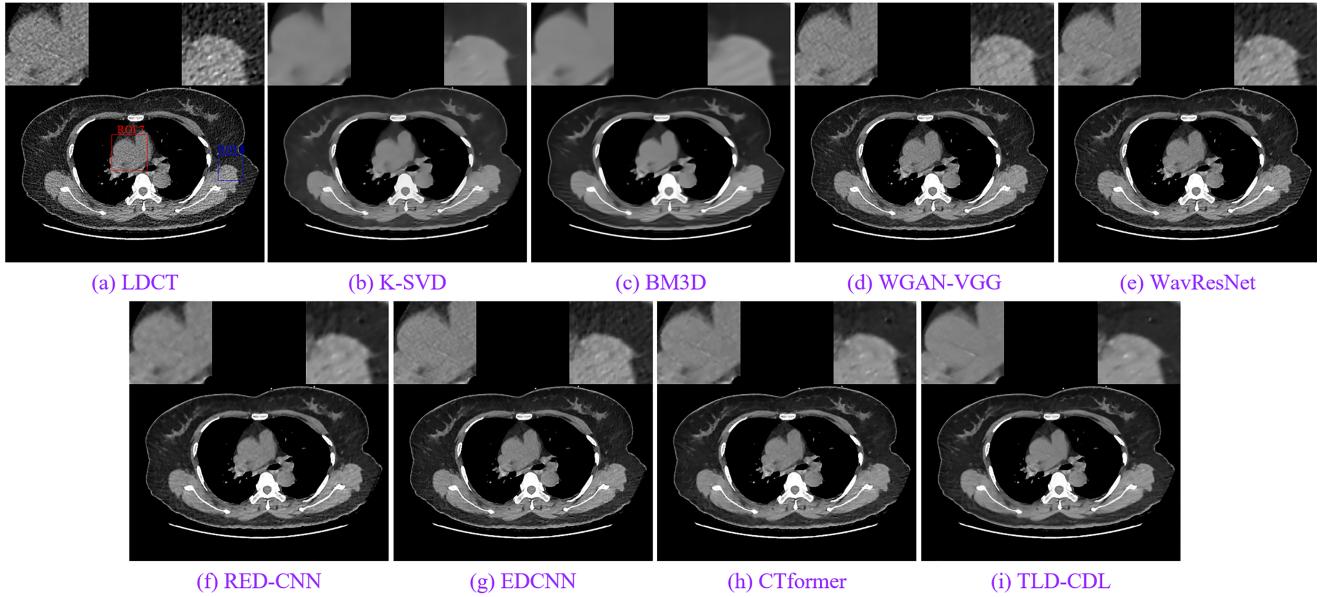


Fig. 8. The denoised results for Case4. The display window is [-160,240]HU.

TABLE III
QUANTITATIVE PERFORMANCE OF ABLATION STUDIES (MEAN VALUE)

index	Pretraining	Inception	Dense	L_1 _loss	compound_loss	PSNR	SSIM	GMSD	FSIM	VIF	VIFs
A	✓				✓	33.7242	0.9184	0.0578	0.9631	0.4380	0.5273
B	✓		✓		✓	33.8655	0.9189	0.0618	0.9612	0.4435	0.5345
C	✓	✓	✓	✓	✓	33.9325	0.9197	0.0595	0.9626	0.4454	0.5369
D	✓	✓	✓		✓	33.9225	0.9201	0.0579	0.9636	0.4467	0.5377

C. Ablation Studies

Compared with Dcdicl, our method designed a multi-scale structure at the end of Net_D and added a dense bypass connection between the convolution layers of Net_X . In addition, TLD-CDL adopted compound loss function to guide training and made the model fit the denoising task of LDCT images by means of transfer learning. In order to prove the effectiveness of the modifications, we had carried out ablation experiments, which are conducted on Mayo test data (as same as the test data in Section III-A). The average value of quantitative analysis is shown in Table III, where “Pretraining” represents pre-training process of removing Gaussian noise on the natural images, “Inception” refers to the multi-scale structure added to Net_D , and “Dense” is considered to the 6 bypass connections at Net_X . The original L1 loss function is also compared with the compound loss in the noise reduction effect. For the convenience of expression, the proposed TLD-CDL is abbreviated as D and the other variants are represented by A, B, and C, respectively. Furthermore, we selected a representative slice (denoted as Case5) and zoomed ROI regions (marked by the rectangles in Fig. 9). The values of six metrics for the representative slices are tagged in the left-upper and right-upper corner.

The comparison of A and B in Table III shows that the addition of multi-scale structure to Net_D has a significant positive contribution to the quantitative indexes, in which the increase of PSNR

is the most inspiring, and the corresponding ROI region (marked with rectangle in Fig. 9 (B1)) is not as smooth as that of A. Based on the analysis of B and C, the dense bypass connections also improved the indicators of the noise reduction result, and the noise in C is further eliminated both in the whole and ROI areas. The PSNR of C is a little higher than that of D (using compound loss function), but the other indexes are slightly lower than D. Besides, instead of using the pixel-level L1 loss to constrain the network, the overall image in D is closer to the NDCT image in terms of brightness and contrast. The contour of vessels in ROI area can hardly be seen on LDCT image, but the vessels (see the arrows in Fig. 9) and lesion in D are the clearest and complete, while the other variants show certain degrees of visual smoothness.

To compare the distribution characteristics of different models on the test dataset, we analyzed PSNR and SSIM through boxplots, as shown in Fig. 10. Boxplot summarizes a set of data through five statistics: maximum, minimum, the lower and upper quartiles, and the median values. By observing the height of the lower and upper quartiles of boxes in Fig. 10, we can conclude that models C and D acquire superb robustness. From the perspective of the gray line (median value) in the boxplots, C and D lead in PSNR and SSIM. The highest median values confirm the average quantization performance of TLD-CDL. Besides, Fig. 11 shows the changes in the PSNR and SSIM of the verification set during the transfer learning process.

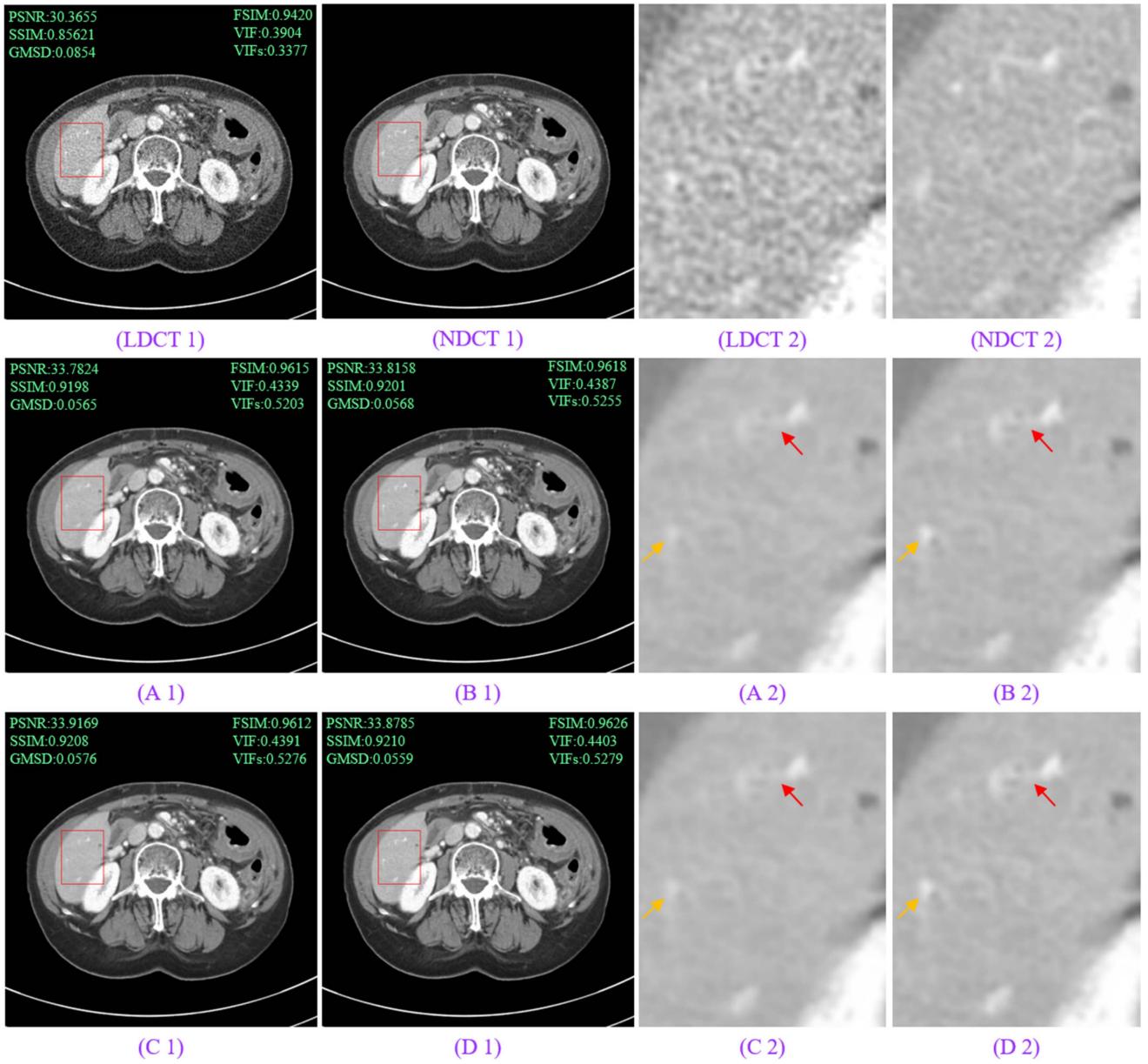


Fig. 9. The denoised results (Case 5 slice) of the ablation experiments on Mayo dataset. The display window is [-160,240]HU.

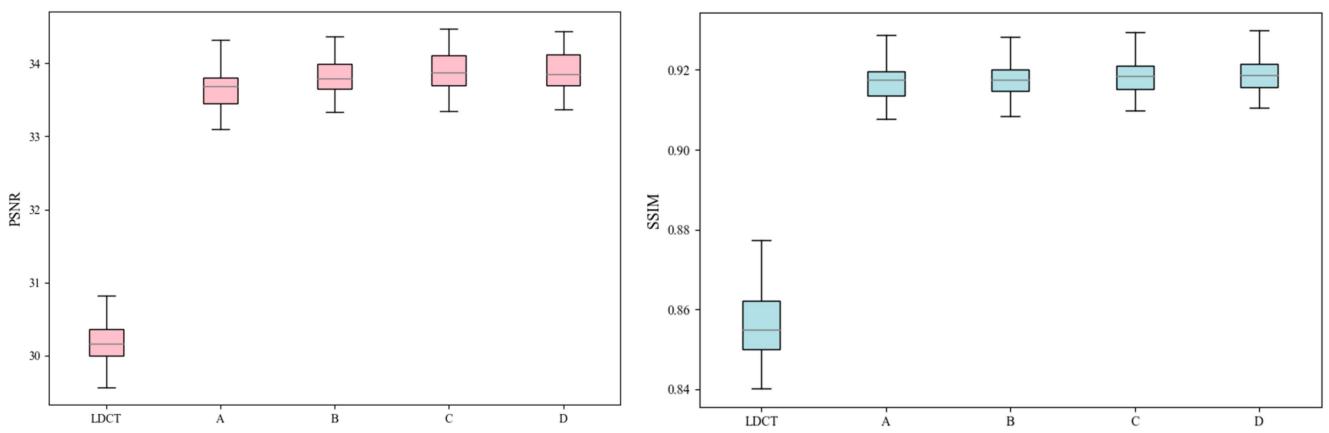


Fig. 10. Boxplot of denoised results of the ablation experiments on Mayo test set.

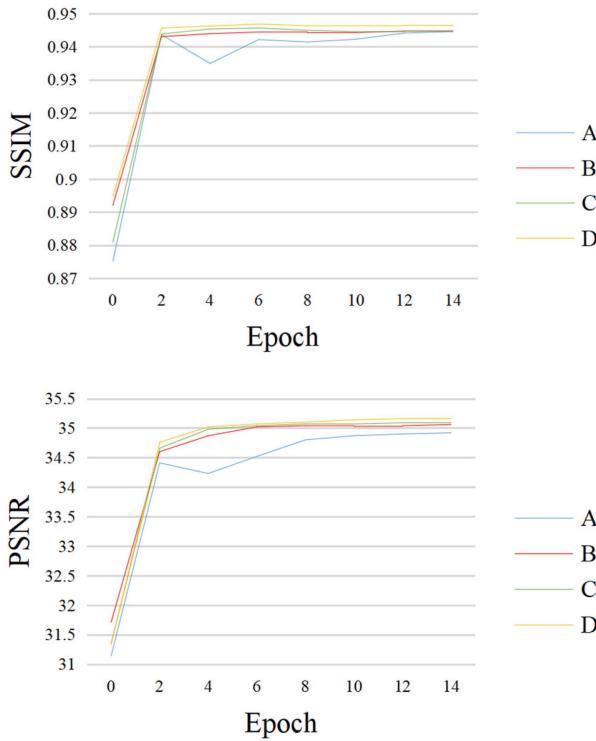


Fig. 11. Change curves in the process of transfer learning.

Unlike the pre-model A, which fluctuates greatly at the 4-th epoch, the curve of D is the most stable and has the highest index value at the end of training. A comprehensive analysis of the results of the ablation experiment shows that the changes made on the basis of Dcdic1 are effective. After processing, the noise is suppressed significantly and the small structure fidelity is perfectly maintained.

IV. CONCLUSION

For the denoising task of LDCT, it is hard to separate the noises and normal tissue texture, which causes the unsatisfactory performance of noise suppression and the serious over-smoothing phenomenon of output images. In this paper, convolutional dictionary learning and the feature extraction ability of neural network are effectively combined. The idea of transfer learning made the denoising model that performed well on natural images adapt to the LDCT image denoising. To improve the noise reduction effect in LDCT, we added dense bypass connections and multi-scale Inception structure to the sub-networks of TLD-CDL. Using the well-known mayo dataset, we utilized a compound loss function composed of L1 loss and SSIM loss to guide the training process, then conducted comparative experiments on the Mayo dataset and the real clinic dataset. Experiments show that, both objectively and subjectively, the effectiveness of TLD-CDL is ahead of all comparison algorithms, and it can reserve the details of the image while removing the noise and artifacts. Given the pairwise scarcity of medical datasets, we are working on the implementation of our approach in an unsupervised manner and plan to extend it to other medical imaging modalities.

REFERENCES

- [1] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
- [2] 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.
- [3] A. Manduca et al., "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.
- [4] M. Balda, J. Horngger, and B. Heismann, "Ray contribution masks for structure adaptive sinogram filtering," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1228–1239, Jun. 2012.
- [5] S. Bera and P. K. Biswas, "Noise conscious training of non local neural network powered by self attentive spectral normalized Markovian patch GAN for low dose CT denoising," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3663–3673, Dec. 2021.
- [6] D. Lee et al., "A feasibility study of low-dose single-scan dual-energy cone-beam CT in many-view under-sampling framework," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2578–2587, Dec. 2017.
- [7] L. Liu, X. Li, K. Xiang, J. Wang, and S. Tan, "Low-dose CBCT reconstruction using Hessian Schatten penalties," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2588–2599, Dec. 2017.
- [8] H. Shangguan, Q. Zhang, Y. Liu, X. Cui, Y. Bai, and Z. Gui, "Low-dose CT statistical iterative reconstruction via modified MRF regularization," *Comput. Methods Programs Biomed.*, vol. 123, pp. 129–141, 2016.
- [9] J. Liu et al., "3D feature constrained reconstruction for low-dose CT imaging," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 5, pp. 1232–1247, May 2018.
- [10] Y. Zhang, X. Mou, G. Wang, and H. Yu, "Tensor-based dictionary learning for spectral CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 142–154, Jan. 2017.
- [11] Z. Li et al., "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Med. Phys.*, vol. 41, no. 1, 2014, Art. no. 011908.
- [12] D. Kang et al., "Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm," *Proc. SPIE*, vol. 8669, pp. 671–676, 2013.
- [13] 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.
- [14] Y. Chen et al., "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, pp. 5803–5820, Aug. 2013.
- [15] 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.
- [16] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 366–381, Sep. 2018, doi: [10.1109/TCI.2018.2840334](https://doi.org/10.1109/TCI.2018.2840334).
- [17] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proc. Nat. Acad. Sci.*, vol. 115, no. 2, pp. 254–259, 2018.
- [18] D. Nie et al., "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.
- [19] J. Park et al., "Computed tomography super-resolution using deep convolutional neural network," *Phys. Med. Biol.*, vol. 63, no. 14, 2018, Art. no. 145011, doi: [10.1088/1361-6560/aacdd4](https://doi.org/10.1088/1361-6560/aacdd4).
- [20] M. Vardhana, N. Arunkumar, S. Lasrado, E. Abdulhay, and G. Ramirez-Gonzalez, "Convolutional neural network for bio-medical image segmentation with hardware acceleration," *Cogn. Syst. Res.*, vol. 50, pp. 10–14, 2018.
- [21] H. Chen et al., "Low-dose CT denoising with convolutional neural network," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 143–146.
- [22] 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, Oct. 2017.
- [23] E. Kang, J. Min, and J. C. Ye, "Wavelet domain residual network (WavResNet) for low-dose X-ray CT reconstruction," in *Proc. Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med. (Fully3D)*, pp. 237–241, 2017.
- [24] Q. Yang et al., "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.

- [25] T. Liang, Y. Jin, Y. Li, and T. Wang, "EDCNN: Edge enhancement-based densely connected network with compound loss for low-dose CT denoising," in *Proc. 15th IEEE Int. Conf. Signal Process.*, 2020, pp. 193–198.
- [26] F. Fan et al., "Quadratic autoencoder (Q-AE) for low-dose CT denoising," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2035–2050, Jun. 2020.
- [27] A. Zhong, B. Li, N. Luo, Y. Xu, L. Zhou, and X. Zhen, "Image restoration for low-dose CT via transfer learning and residual network," *IEEE Access*, vol. 8, pp. 112078–112091, 2020.
- [28] D. Wang, F. Fan, Z. Wu, R. Liu, F. Wang, and H. Yu, "CTformer: Convolution-free token2token dilated vision transformer for low-dose CT denoising," 2022, *arXiv:2202.13517*.
- [29] H. Sreret and R. Giryes, "Learned convolutional sparse coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 2191–2195, doi: [10.1109/ICASSP2018.8462313](https://doi.org/10.1109/ICASSP2018.8462313).
- [30] D. Simon and M. Elad, "Rethinking the CSC model for natural images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 2274–2284.
- [31] X. Fu, Z. Zha, F. Wu, X. Ding, and J. Paisley, "JPEG artifacts reduction via deep convolutional sparse coding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2501–2510.
- [32] H. Wang, Y. Li, N. He, K. Ma, D. Meng, and Y. Zheng, "DICDNet: Deep interpretable convolutional dictionary network for metal artifact reduction in CT images," *IEEE Trans. Med. Imag.*, vol. 41, no. 4, pp. 869–880, Apr. 2022.
- [33] H. Zheng, H. Yong, and L. Zhang, "Deep convolutional dictionary learning for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 630–641.
- [34] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [35] H. Huang et al., "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 1055–1059.
- [36] F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in *Proc. 5th Int. Conf. ACM*, 2017, pp. 59–62.
- [37] K. Ma et al., "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, Feb. 2017.
- [38] AAPM, "Low dose CT grand challenge, 2017. [Online]. Available: <http://www.aapm.org/GrandChallenge/LowDoseCT/>
- [39] 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.
- [40] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.
- [41] 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.
- [42] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.



Rongbiao Yan is currently working toward the M.S. degree in information and communication engineering with the North University of China, Taiyuan, China. His research interests include CT imaging and image processing.



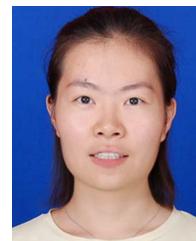
Yi Liu received the Ph.D. degree in signal and information processing from the North University of China, Taiyuan, China, in 2014. In 2015, she joined the School of Information and Communication Engineering, North University of China. Her research interests include CT imaging and image processing.



Yuhang Liu is currently working toward the M.S. degree in computer science and technology with the North University of China, Taiyuan, China. His research interests include object detection and image reconstruction.



Lei Wang is currently working toward the Ph.D. degree in signal and information processing with the North University of China, Taiyuan, China. His research interests include image processing and CT imaging reconstruction.



Rongge Zhao received the M.S. degree in information and communication engineering from the North University of China, Taiyuan, China. She is currently working toward the Ph.D. degree with Nankai University, Tianjin, China. Her research interests include image enhancement and single-molecule localization super-resolution imaging.



Yunjiao Bai received the Ph.D. degree in information and communication engineering from the North University of China, Taiyuan, China, in 2018. He is currently engaged in teaching and research with Jinzhong University, Jinzhong, China. His research interests include image denoising and enhancement.



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