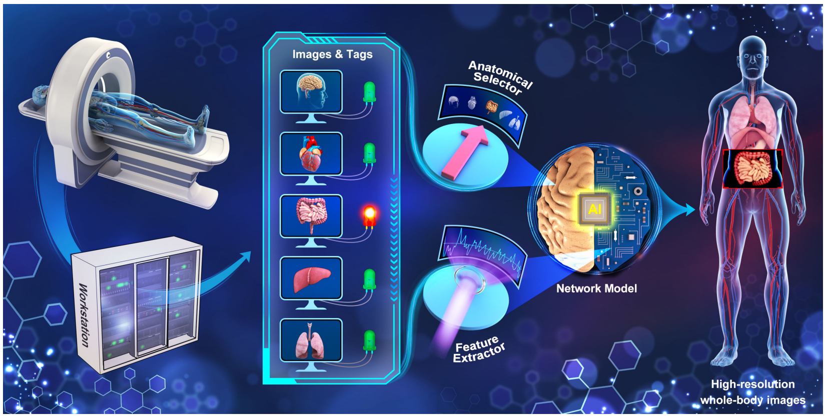
IEEE JOURNAL OF

BIOMEDICAL AND HEALTH INFORMATICS

A PUBLICATION OF
THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY
THE IEEE COMMUNICATIONS SOCIETY
THE IEEE SIGNAL PROCESSING SOCIETY

TECHNICALLY COSPONSORED BY THE IEEE COMPUTER SOCIETY

SEPTEMBER 2021 VOLUME 25 NUMBER 9 IJBHA9 (ISSN 2168-2194)



Zhenxing Huang, Xinfeng Liu, Rongpin Wang, Zixiang Chen, Yongfeng Yang, Xin Liu, Hairong Zheng, Dong Liang, Zhanli Hu, "Learning a Deep CNN Denoising Approach Using Anatomical Prior Information Implemented with an Attention Mechanism for Low-dose CT Imaging on Clinical Patient Data from Multiple Anatomical Sites."

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine

Learning a Deep CNN Denoising Approach Using Anatomical Prior Information Implemented With Attention Mechanism for Low-Dose CT Imaging on Clinical Patient Data From Multiple Anatomical Sites

Zhenxing Huang, Xinfeng Liu, Rongpin Wang, Zixiang Chen, Yongfeng Yang, Xin Liu, Hairong Zheng, *Senior Member, IEEE*, Dong Liang, *Senior Member, IEEE*, and Zhanli Hu, *Senior Member, IEEE*

***Abstract*—Dose reduction in computed tomography (CT) has gained considerable attention in clinical applications because it decreases radiation risks. However, a lower dose generates noise in low-dose computed tomography (LDCT) images. Previous deep learning (DL)-based works have investigated ways to improve diagnostic performance to address this ill-posed problem. However, most of them disregard the anatomical differences among different human body sites in constructing the mapping function between LDCT images and their high-resolution normal-dose CT (NDCT) counterparts. In this article, we propose a novel deep convolutional neural network (CNN) denoising approach by introducing information of the anatomical prior. Instead of designing multiple networks for each independent human body anatomical site, a unified network framework is employed to process anatomical information. The anatomical prior is represented as a pattern of weights of the features extracted from the corresponding LDCT image in an anatomical prior fusion module. To promote diversity in the contextual information, a spatial attention fusion mechanism is introduced to capture many local regions of interest in the attention fusion module. Although many network parameters are saved, the experimental results demonstrate that our method, which incorporates anatomical prior information, is effective in denoising LDCT images. Furthermore, the anatomical prior fusion module could be conveniently integrated into other DL-based methods and avails the performance improvement on multiple anatomical data.**

***Index Terms*—Anatomical prior information, attention mechanism, image denosing, low-dose CT.**

I. Introduction

BECAUSE of concerns about the health risk of high-dose X-ray radiation, research on low-dose computed tomography (LDCT) imaging has gained considerable attention [1]. Compared to an X-ray, for instance a chest X-ray, LDCT can substantially improve the accuracy of chest cancer diagnosis and assessments to reduce cancer deaths [2]. The common ways to lower the radiation dose are to reduce the X-ray flux and the number of projection views. However, a lower dose may lead to worse diagnostic performance, which is generally manifested as noise and artifacts in the resulting images [3].

To address this problem, many methods have been proposed to improve the image quality in LDCT. Generally, these LDCT methods are divided into three sub-methods: sinogram-domain filtering methods, iterative reconstruction methods and post-processing methods. First, several works [4], [5] based on sinogram-domain filtering have been proposed to process the raw data before image reconstruction. Although well-known noise distributions in the sinogram domain can be applied, these sinogram-domain methods may cause spatial resolution loss in the image domain. Specially, the restoration of high-frequency structural details for sinogram-domain methods, for instance, edges and textures, causes great challenges. Second, iterative

Manuscript received June 28, 2020; revised December 24, 2020 and February 19, 2021; accepted February 20, 2021. Date of publication February 24, 2021; date of current version September 3, 2021. This work was supported in part by the National Natural Science Foundation of China under Grants 32022042 and 81871441, in part by the Shenzhen International Cooperation Research Project of China under Grant GJHZ20180928115824168, in part by the Guangdong International Science and Technology Cooperation Project of China under Grant 2018A050506064, in part by the Natural Science Foundation of Guangdong Province in China under Grant 2020A1515010733, in part by the Guangdong Special Support Program of China under Grant 2017TQ04R395, and in part by the Chinese Academy of Sciences Key Laboratory of Health Informatics in China under Grant 2011DP173015. *(Zhenxing Huang and Xinfeng Liu contributed equally to this work.) (Corresponding author: Zhanli Hu.)*

Zhenxing Huang is with the Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

Xinfeng Liu and Rongpin Wang are with the Department of Radiology, Guizhou Provincial People’s Hospital, Guiyang 550002, China.

Zixiang Chen, Yongfeng Yang, Xin Liu, Hairong Zheng, Dong Liang, and Zhanli Hu are with the Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Chinese Academy of Sciences Key Laboratory of Health Informatics, Shenzhen 518055, China (e-mail: zl.hu@siat.ac.cn).

Digital Object Identifier 10.1109/JBHI.2021.3061758

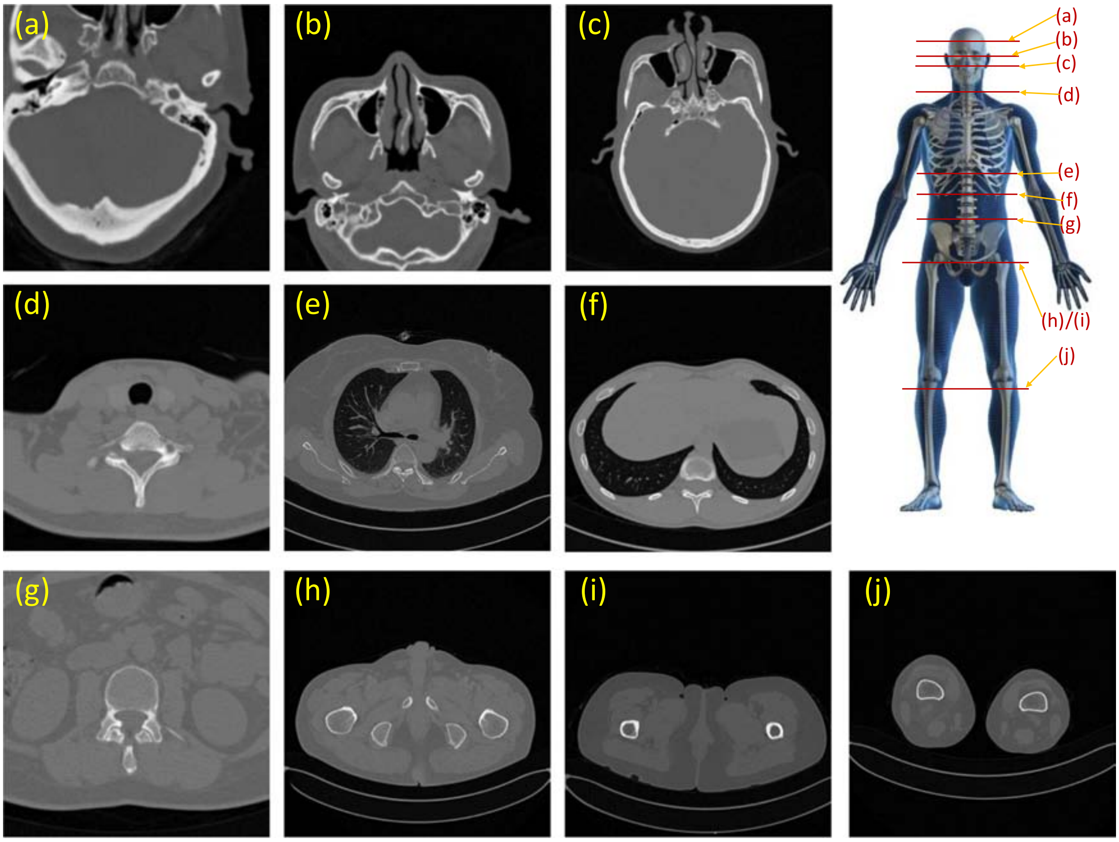


Fig. 1. Clinical patient data from different anatomical sites: (a) cranium, (b) orbit, (c) sinus, (d) neck, (e) lung, (f) abdomen, (g) waist, (h) male pelvis, (i) female pelvis and (j) knee. The data show great anatomical differences among different anatomical sites in CT imaging.

reconstruction methods [6]–[12] became popular because they allowed several priors to be formulated, such as the total variation [13] and nonlocal means [14]. Although these methods have excellent output, they are limited in detail reconstruction and expensive to compute. In recent studies, deep learning (DL) postprocessing methods [15]–[34] have been promising, popular approaches in LDCT to estimate high-dose computed tomography (CT) images via an end-to-end network. For example, Chen *et al* [16] propose a residual auto-encoder network for LDCT images, which achieves promising restoration results. Compared with traditional methods, DL-based methods are more suitable for recovering high-dose CT images because of their excellent feature extraction and representation capacity for uncertain noise models.

Although previous works have made great progress in improving image quality, most of them disregard the anatomical prior information of LDCT images. Generally, this information shows large anatomical structural differences among different human body sites, such as the sinus, neck and cranium shown in Fig. 1. On the other hand, scan parameter settings are usually different for particular human body sites in practical applications. For instance, the scan dose for the waist site is often higher than that for the cranium. A specific trained deep network model is usually applicable to the specific anatomical site. The anatomical differences and scan settings for different anatomical sites lead to the complex data distribution in the training data. Anatomical information about LDCT images could be considered extra priors to improve the denoising performance in LDCT imaging.

In this paper, we propose a deep convolutional neural network (CNN) denoising method by introducing an anatomical prior, which we refer to as the DeACNN, for noise reduction in LDCT imaging. Via an anatomical prior fusion module, the anatomical prior is fused with the features extracted from its corresponding LDCT image. Inspired by the attention mechanism [35]–[37], several spatial attention fusion modules are cascaded. To avoid information loss, we combine the original features extracted by convolution layers with attention features in these cascaded fusion modules. To reduce the network parameters, we employ several convolutions with a 1×1 kernel for the whole network framework.

We make three main contributions: 1) Considering the obvious anatomical differences among different human body sites, we introduce an anatomical prior into image denoising in LDCT imaging. 2) Instead of designing multiple independent networks for each anatomical site, we address the anatomical prior information and LDCT images in a unified framework, which means that our proposed network could process CT images from different anatomical sites. 3) Inspired by the attention mechanism, the anatomical prior fusion module and spatial attention fusion module are designed. To reduce information loss, skip connection and concatenation are also employed. The anatomical prior fusion module could be considered embedded

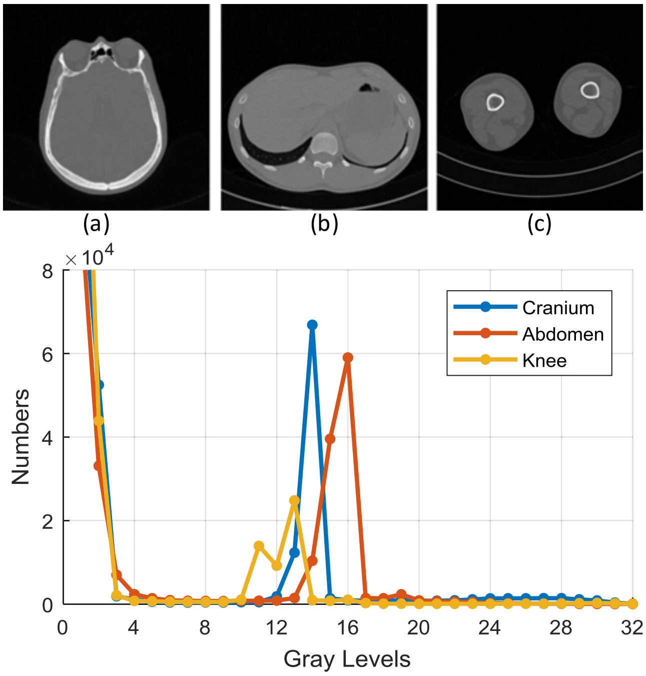


Fig. 2. Histogram distribution on 32 Gy levels for three anatomical examples: (a) cranium, (b) abdomen and (c) knee.

in other DL-based methods on multiple anatomical data to improve performance.

The remainder of this paper is organized as follows: the methods are described in Section II. In this section, the anatomical prior is explained. We then describe the network architecture, including the two fusion modules. In the next section, Section III, experiments are conducted to validate the effectiveness of our proposed method. Additionally, we provide the implementation details and show the experimental results in this section. The discussion and conclusion are given in Section IV.

II. Methods

In this section, we describe our methods. First, we introduce the anatomical prior. Second, the overview framework is elaborated. Last, two modules, the prior fusion module and the attention fusion module, are illuminated.

A. Anatomical Prior Information

Generally, large anatomical structural differences are observed among different human body sites. As shown in Fig. 2, we calculate the data distribution for three normal CT images from three anatomical sites. The distribution reflects vast differences in the data among different anatomical sites. The noise distribution in each anatomical site also seems to differ when the scanning parameters are fixed. For the clinical patient data from multiple anatomical sites, this difference avails the denoising performance when the anatomical site is given. The extra anatomical sites are considered the prior information for the LDCT images.

In our experiments, we employ clinical anatomical data for 10 anatomical sites: sinus, neck, brain, breast, abdomen, knee, orbit, waist, pelvis (male) and pelvis (female); the sites are shown in Fig. 1. First, we convert these anatomical descriptions via one-hot encoding. Thus, a low-dose image follows an anatomical vector, which is the input pattern for the proposed method, and the output is a high-resolution estimated CT image after denoising.

B. Framework Overview

Our network employs two inputs, the LDCT image and its corresponding anatomical vector, as shown in Fig. 3. We primarily employ one-hot encoding for the anatomical descriptions, and a weight prediction module follows the input anatomical vector. The anatomical prior information is fused in the prior fusion module. To take full advantage of the anatomy-fused information from the first module, we cascade *M* attention fusion modules to deepen the network. Given the LDCT images $x = \{x_1, x_2, x_3, \ldots, x_n\}$ and the corresponding anatomical vectors $a = \{a_1, a_2, a_3, \ldots, a_n\}$, we estimate the normal-dose CT (NDCT) images $y = \{y_1, y_2, y_3, \ldots, y_n\}$ after the denoising process. This restoration process, using a mean-square error (MSE) cost function, can be formulated as follows:

$$L = \frac{1}{n}\sum_{i=1}^{n} \|G(x_i; a_i; \Theta) - y_i\|_2^2, \tag{1}$$

where Θ denotes the network parameters and $G(\cdot)$ denotes the estimation function. Unlike several other denoising approaches, anatomical prior information, which is easily available for radiologists, is needed in our method.

C. Anatomical Prior Fusion

Considering the anatomical differences among human body sites, we introduce an anatomical prior during the denoising process. Instead of employing different networks for each anatomical site, the input anatomical vector is used to predict the channel weight mask in adapting to feature maps. With the weight pattern, the anatomical prior can be distinguished in the network. To avoid information loss, the original input features extracted from the original LDCT images are concatenated. For the weight prediction module, 7 convolution filters with 1×1 kernels are utilized (shown in Fig. 3). Concatenation is utililized to combine feature maps to avoid information loss. Following the concatenation operation, a convolution layer with 1×1 kernels are employed to shrink the channel to 64, which is the input channel number of this module. The $Sigmoid$ activation function helps to shrink the channel weight to $0 \sim 1$, which characterizes the influence of different anatomical sites on subsequent image features.

The prior fusion module can be formulated as follows:

$$F_o = Conv_2(Conv_1(x_i), P_c(a_i) \otimes Conv_1(x_i)), \tag{2}$$

where $P_c(\cdot)$ denotes the channel weight prediction and $Conv_1$ and $Conv_2$ denote convolution operations for the joint features via concatenation. In addition, "$\otimes$" denotes the element-wise

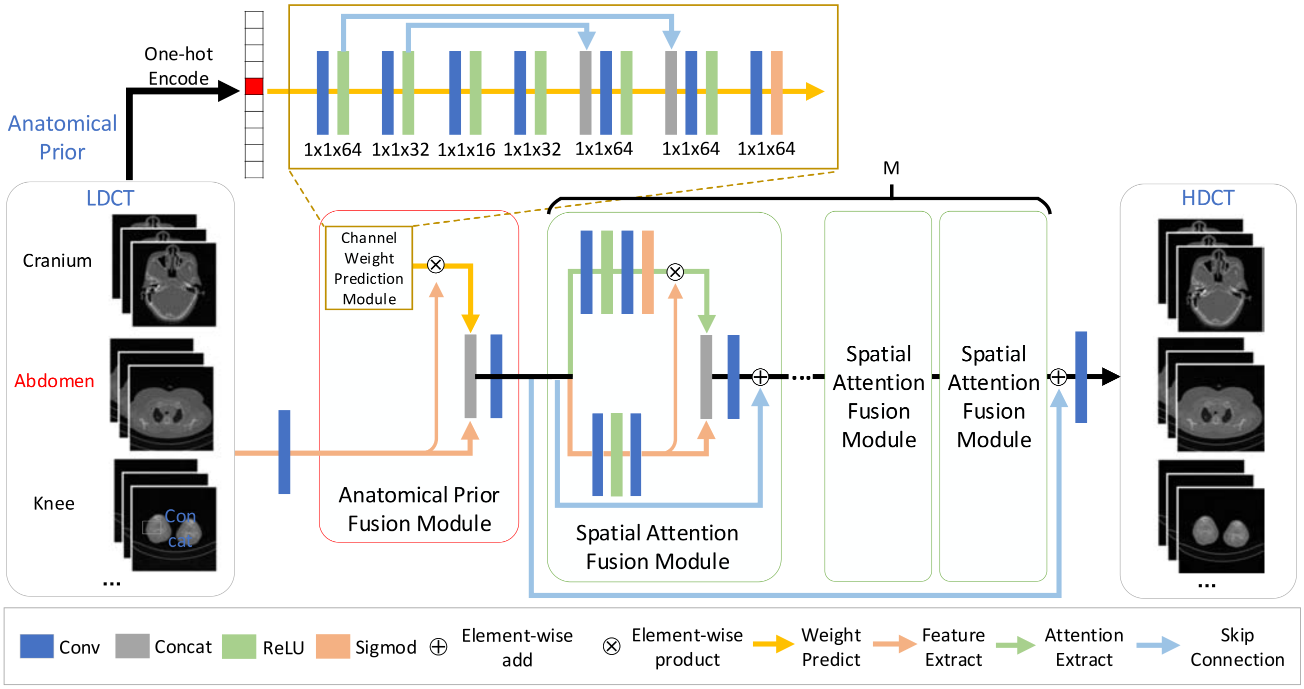


Fig. 3. Overview of the framework of the DeACNN. The overall network consists of two main parts, the prior fusion module and attention fusion module, where the weight prediction module employs an anatomical prior to obtain a weight mask for the channel of features extracted from LDCT images.

product operation, and F_o denotes the anatomy-fused outcome information. The kernel size of $Conv_1$ and $Conv_2$ are fixed to 3×3 and 1×1, respectively.

D. Spatial Attention Fusion

To take full advantage of the anatomy-fused information, we deepen the network based on the design of the cascade module architecture. Similar to ResNet [38], the local cascade module adopts down-projection and up-projection units in the cascaded modules, where the down-projection unit is implemented with a convolution operation and the up-projection unit is implemented with a deconvolution operation. Inspired by [37], we introduce a spatial attention mechanism to obtain local regions of interest (ROIs). Furthermore, the original features obtained by the convolution stream are combined with the spatial attention stream. We apply two convolution layers to extract the original features and another two convolution layers for attention extraction. The parameter details are shown in Table. I. We shrink and expand the channel number to reduce the parameter counts. Convolution layers with a filter size of 1×1 are employed during the attention extraction process. The output F_i of the i-th spatial attention fusion module can be formulated as follows:

$$F_i = Conv_3(P_{cs}(F_{i-1}), P_a(F_{i-1}) \otimes P_{cs}(F_{i-1})) \oplus F_{i-1}, \quad (3)$$

where $P_a(\cdot)$ denotes the spatial attention mask prediction and $Conv_3$ denotes the convolution operations for the joint features via concatenation. In addition, "$\otimes$" denotes the element-wise product operation, and P_{cs} denotes a feature extraction process that uses two convolution operations without changing the image size. "$\oplus$" denotes the element-wise add operation.

TABLE I
PARAMETER SETTINGS FOR THE SPATIAL ATTENTION FUSION MODULE. "CONV" DENOTES THE CONVOLUTIONAL LAYERS, AND "ACTV" DENOTES THE USED ACTIVATION FUNCTIONS

Components	Feature extraction	Attention extraction
Conv1	$3 \times 3 \times 64 \times 32$	$1 \times 1 \times 64 \times 16$
Actv1	ReLU	ReLU
Conv2	$3 \times 3 \times 32 \times 64$	$1 \times 1 \times 16 \times 16$
Actv2	-	Sigmoid
Operation 1	Element-wise product	
Operation 2	Concatenation	
Conv3	$1 \times 1 \times 128 \times 64$	
Operation 3	Element-wise add	

III. EXPERIMENTS

In this section, we conduct experiments to validate the effectiveness of our method. First, the patient data and training details are described. Second, we evaluate the performance of our method compared with that of several other DL-based methods. Last, the experimental results and ablation studies are described.

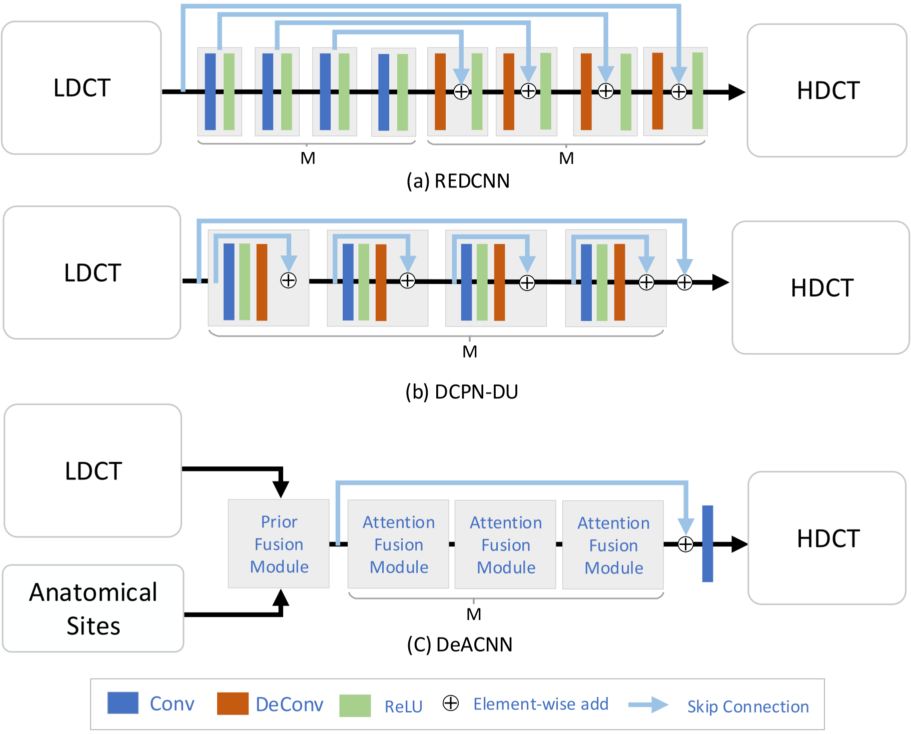


Fig. 4. Three network models for image denoising in low-dose imaging: (a) Residual encoder-decoder CNN (REDCNN), (b) Deep cascade projection network with the down-to-up projection operation (DCPN-DU), (c) Our proposed network (DeACNN).

TABLE II
SCANNING PARAMETER SETTINGS FOR ROUTINE HIGH-DOSE CT IMAGES UNDER A SEMENS CT SCANNER

Parameters	Value
Source-to-detector distance	1085.6 mm
Source-to-patient distance	595 mm
Tube voltage	120 kVp
Tube Current	200 $\sim$ 500 mA
Thickness	1.0 mm
Reconstruction Matrix	512 $\times$ 512

TABLE III
PARAMETER COUNTS AND RUNNING TIMES OF EACH TEST EXAMPLE FOR DIFFERENT METHODS

Methods	CNN	REDCNN	DCPN-DU	DeACNN(Ours)
Parameters	1.11×10^6	1.11×10^6	1.11×10^6	0.77×10^6
Run times	0.083 s	0.093 s	0.900 s	0.110 s

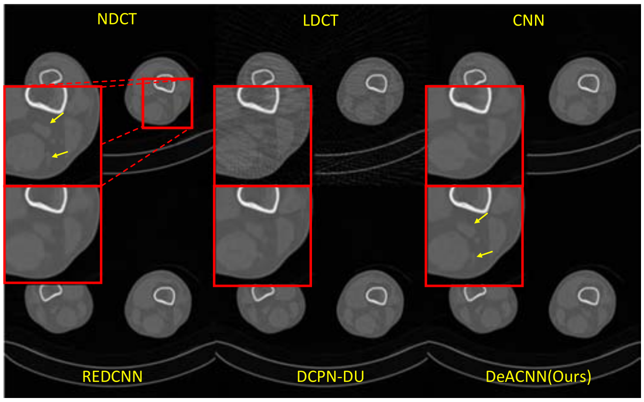


Fig. 5. Knee results of 120 sparse projection views for different methods. ROIs are marked by red boxes. Several visual differences are marked by yellow arrows.

A. Clinical Patient Data and Details of Implementation

With the research data support of Guizhou Provincial People's Hospital (Guiyang, Guizhou, China), we are able to utilize clinical data collected from more than 200 patients, with an image size of 512 $\times$ 512. The age distribution for these patients ranges from 7 to 82. Among these patients, 55% of them are male and 45% are female. The total number of CT images exceeds 80 000; 10% of the data are utilized as validation data, and 10% of the data are utilized as test data. The remainder of the data are employed for network training. The dataset contains high-resolution NDCT images and their descriptions tagged by professional radiologists for 10 human body sites: sinus, neck, brain, breast, abdomen, knee, orbit, waist, pelvis (male) and pelvis (female). Considering the continuity of the whole body, the descriptions partially overlap, which increases the robustness of the proposed method. The dataset is acquired under a Semens CT scanner(SOMATOM Definition). As shown in Table II,

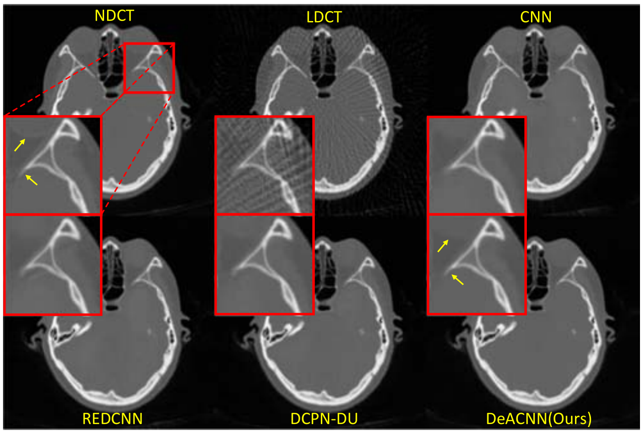


Fig. 6. Cranium results of 120 sparse projection views for different methods. ROIs are marked by red boxes. Several visual differences are marked by yellow arrows.

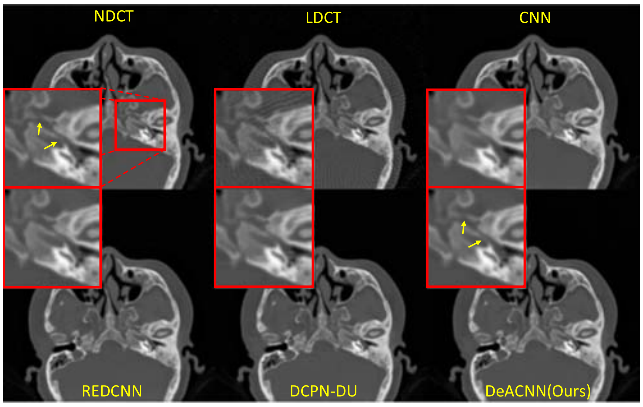


Fig. 8. Sinus results of 180 sparse projection views for different methods. ROIs are marked by red boxes. Several visual differences are marked by yellow arrows.

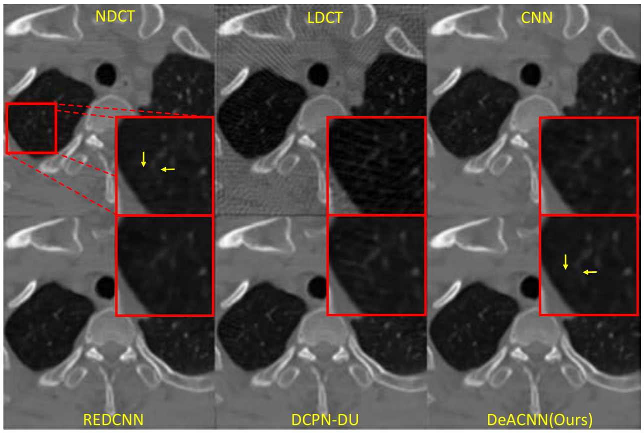


Fig. 7. Waist results of 150 sparse projection views for different methods. ROIs are marked by red boxes. Several visual differences are marked by yellow arrows.

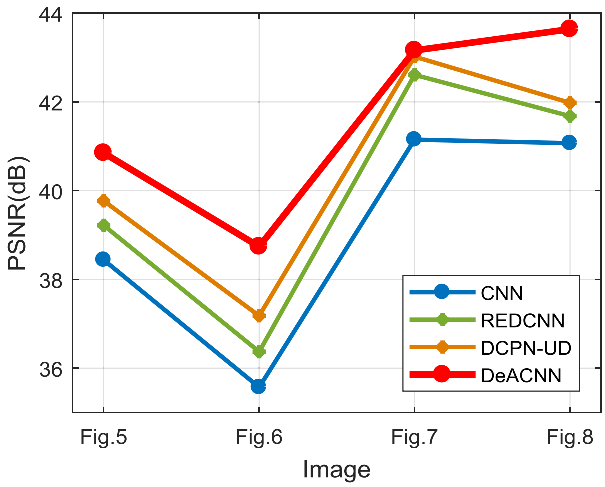


Fig. 9. Statistical results for different methods in terms of the PSNR for the whole images Fig. 5, Fig. 6, Fig. 7 and Fig. 8.

the scan tube voltage is 120 kVp, and the thickness is set to 1 mm for the routine NDCT images. We conduct the simulation process to obtain LDCT images via the MRIT toolbox[1] [39] implemented by Matlab 2017a. For the MIRT, the scanning parameters are fixed, as shown in Table II, and the system projection matrix are calculated. With the aid of the projection matrix, we obtain sinogram data under 360 projection views as the referenced NDCT images. Via uniform sparse sampling, the LDCT sinogram data under 120, 150 and 180 projection views are gained. The simulated LDCT images are reconstructed using the FBP algorithm.

For the input data, several data augmentations are adopted, such as random rotating and flipping. To reduce the training time, we use patches with an image size of 64×64. The learning rate is set to 0.0001. The ADAM optimizer [40] is applied to minimize the cost function during the network training process. We implement our model in the PyTorch framework on Ubuntu 16.04 with a Titan 1080Ti GPU during the training and test process.

We compare our method with several other methods, including the CNN [15], the residual encoder-decoder CNN (REDCNN) [16] and a baseline model named the deep cascade projection network with the down-to-up projection operation (DCPN-DU). To visually describe the network model architecture, the REDCNN and DCPN-DU are shown in Fig. 4(a) and (b). The most obvious difference is that we introduce anatomical information instead of directly applying an end-to-end network compared with another two network methods. For comparison models, we apply the kernel size 3×3 for convolution and deconvolution layers according to parameter settings following [16]. For a fair comparison, these models are retrained with the same settings as our model on our training and test datasets. Furthermore, popular metrics–the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM)–are adopted to evaluate the qualitative results. As shown

[1] The code is [Online]. Available: https://web.eecs.umich.edu/~fessler/code/

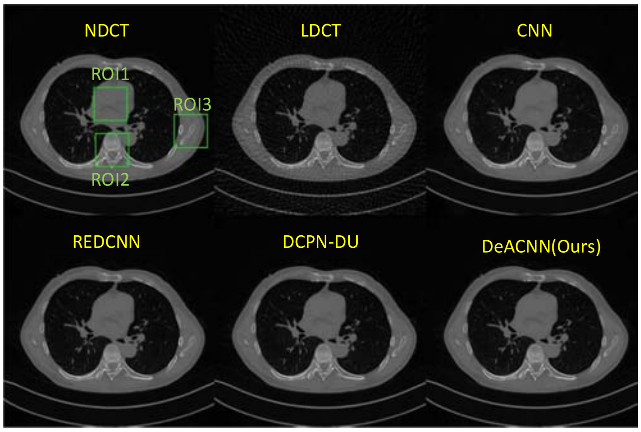


Fig. 10. Estimated results for 180-view projections. The green boxes indicate the three ROIs.

in Table III, we compare the parameter counts and running times of each test example for different methods. Due to multiple multiplication operations and multibranch data processing, the running time of our method for each test example is the slowest compared with the other three methods. Because of the reduction in the number of channels and the use of a filter size of 1×1, the parameter counts of our method are significantly less than those of other methods.

B. Quantitative and Qualitative Results on Multiple Anatomical Sites

The visual and quantitative results for different methods on multiple anatomical sites are given in this section. As shown in Fig. 5, the boundary of the soft tissue marked by the yellow arrow is distinct in Fig. 5 for the DeACNN, while the other results are blurred. In Fig. 6, the results generated by the DeACNN are similar to the reference results, and there is no blurring of the corner bones. More visual results and qualitative results for 150 and 180 projection views are shown in Fig. 7 and Fig. 8, respectively, which also illustrates the superiority of our method. In particular, the connecting lines of the bones in our method can be clearly observed in Fig. 8. We calculate the PSNR in Fig. 9 for the whole images in Fig. 5, Fig. 6, Fig. 7 and Fig. 8. Note that our method, the DeACNN, gains over 1.5 dB improvement in the PSNR in 120 and 180 projection views for the sinus and cranium anatomical sites shown in Fig. 6 and Fig. 8.

The quantitative results for local ROIs are also calculated. As shown in Fig. 10, we selected three ROIs, which are marked with green boxes, including the center, middle and edge of the abdomen site. As shown in Fig. 11, the statistical results for the PSNR and SSIM demonstrate the superiority of the DeACNN against the other methods on the ROIs. Specifically, the DeACNN performs better on ROI3 in terms of the PSNR compared with the other methods.

For the test dataset, we randomly selected 300 LDCT images, including 120, 150 and 180 projection views, to evaluate the denoising performance among different methods. The qualitative results are shown in Table IV. Our method shows improved denoising performance compared to the other methods, especially at low projection angles. In addition, we adopt the statistical test P-value, which shows the significant difference in these qualitative results.

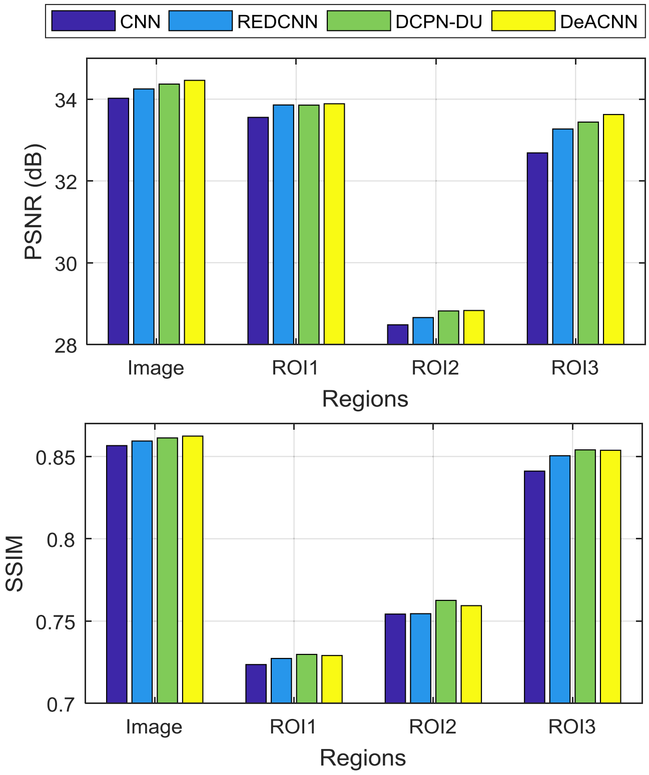


Fig. 11. Statistical results (PSNR and SSIM) for the whole image and three ROIs marked in Fig. 10.

C. Quantitative and Qualitative Results on Single Anatomical Site

Although our method is validated effective for the previously mentioned multiple anatomical sites, the denoising performance in a single anatomical site is also explored. We compare our method with several other methods, including the CNN, REDCNN and DCPN-DU. We retrain these models on two single anatomical sites, including the cranium and abdomen.

In Fig. 12, the visual results generated by different CNN-based methods are given under 120, 150 and 180 projection views. In the first row, our method appears to clearly carve the edges. The results in the CNN, DCPN-DU and REDCNN are blurry. In the second row, the results of other methods do not seem to carve the local regions marked by yellow arrows. The results of our method are closer to the ground truth. In the third row, our method can display the boundary between two organs.

For the test dataset, we randomly selected 30 LDCT images, including 120, 150 and 180 projection views, to

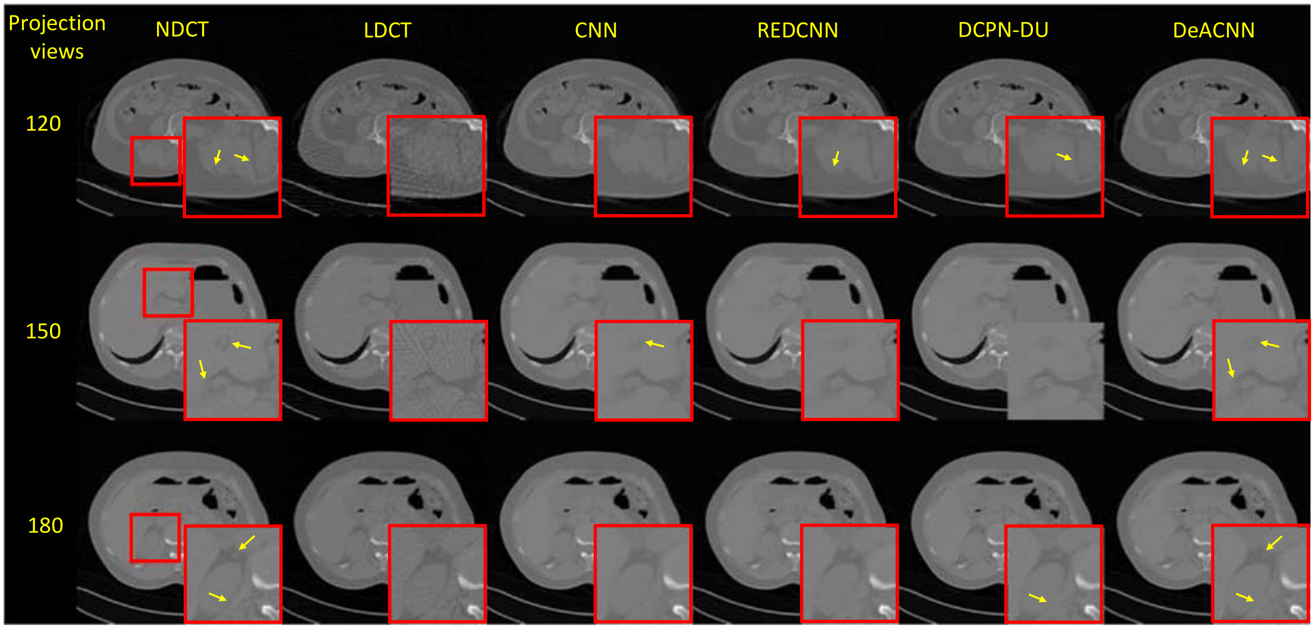


Fig. 12. Visual results of 120, 150 and180 sparse projection views for different methods on the abdomen anatomical site. ROIs are marked by red boxes. Several visual differences are marked by yellow arrows.

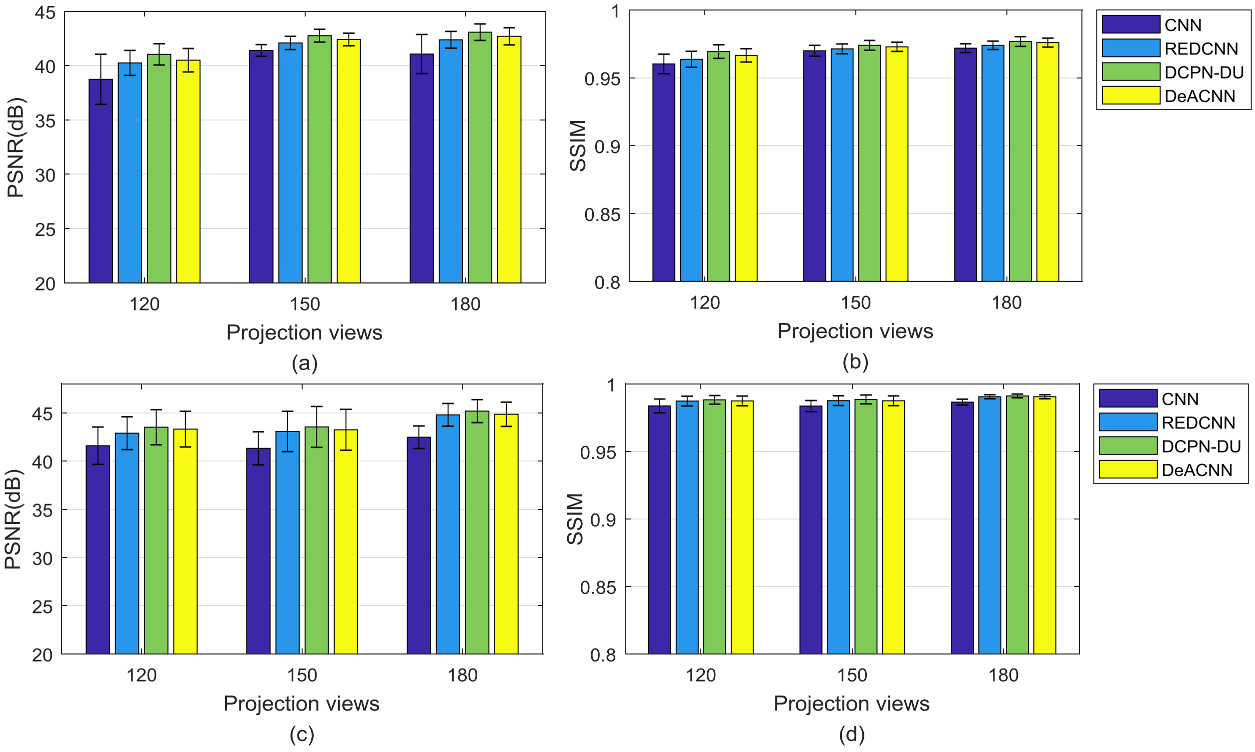


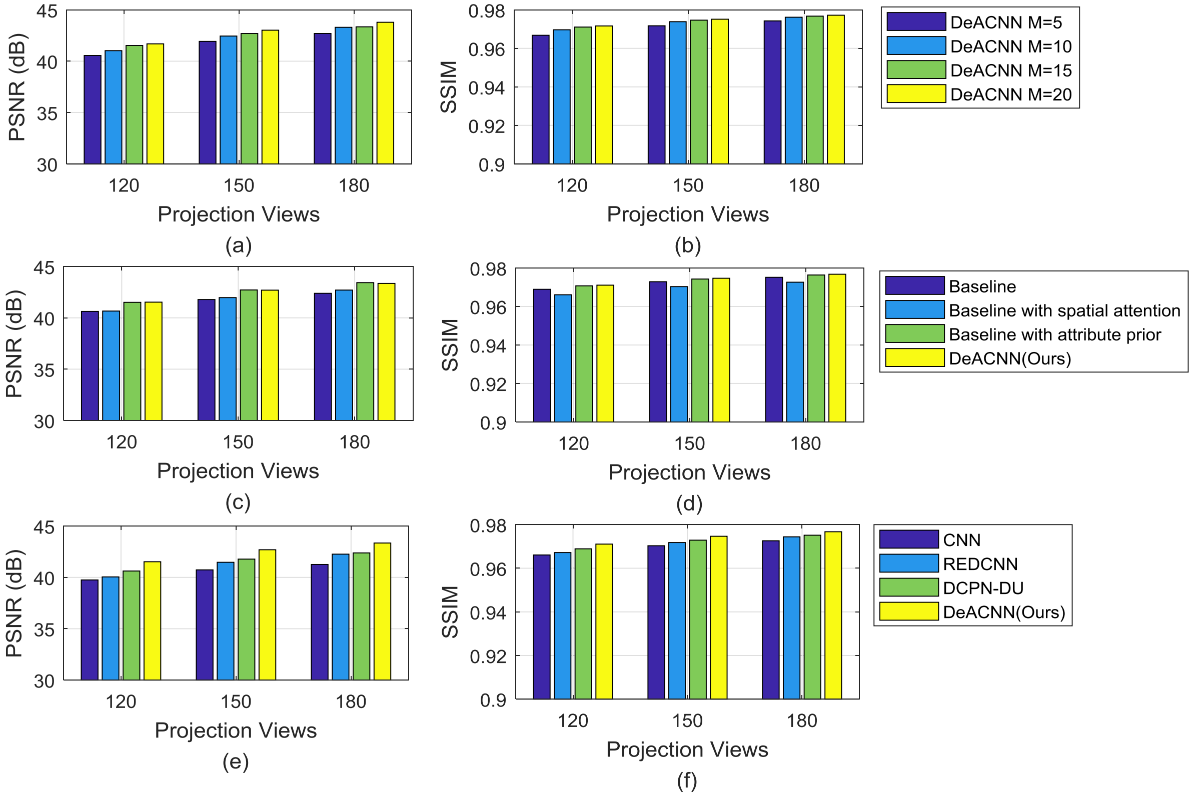
Fig. 13. Statistical quantitative results (Mean ± SDs) comparison on two single anatomical sites under 120, 150, and 180 projection views, in terms of the PSNR and SSIM. (a) and (b) are calculated on the cranium. (c) and (d) are calculated on the abdomen.

evaluate the denoising performance among different methods on two single anatomical sites. The qualitative results are shown in Fig. 13. The statistical quantitative results demonstrate that the DCPN-DU achieves the best performance while our method with less parameters gains quite excellent performance. Although the anatomical prior on a single site has minimal significance, our method can achieve considerable denoising performance.

TABLE IV
QUANTITATIVE RESULTS (MEAN ± SDS) FOR TEST DATASETS, INCLUDING 120, 150, AND 180 PROJECTION VIEWS, IN TERMS OF THE PSNR AND SSIM

Projection views	Metrics	LDCT	CNN	REDCNN	DCPN-DU	DeACNN
120	PSNR	29.88±1.89	39.43±2.48*	39.78±2.50*	40.16±2.61*	40.70±3.88*
	SSIM	0.7561±0.0616	0.9600±0.0438*	0.9609±0.0438*	0.9621±0.0435*	0.9635±0.0437*
150	PSNR	31.05±2.28	40.75±2.38*	41.03±2.36*	41.34±2.47*	41.63±2.66*
	SSIM	0.8113±0.0562	0.9654±0.0394*	0.9661±0.0395*	0.9672±0.0389*	0.9679±0.0390*
180	PSNR	32.35±2.60	41.62±2.56*	41.99±2.56*	42.26±2.64*	42.48±2.80*
	SSIM	0.8525±0.0552	0.9700±0.0342*	0.9707±0.0340*	0.9716±0.0336*	0.9722±0.0337*

* denotes P $<$ 0.05, which shows significantly different.



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 14. Quantitative results (Mean) comparison of different settings under 120, 150, and 180 projection views, in terms of the PSNR and SSIM.

D. Ablation Studies

In this section, we conduct ablation studies of our method. First, we explore the performance of our method with different numbers of spatial attention fusion modules. Second, we validate the effectiveness of anatomical prior information. Third, we validate the effectiveness of the anatomical prior and attention fusion module. Last, we compare the performance of our method during the network training process. These ablation studies are trained for 200 epochs on 10 anatomical sites; we calculate the PSNR on the same validation data.

- **The number of spatial attention fusion modules**. We explore the influence on performance of the depth of the network. We set the number of spatial attention fusion modules as 5, 10, 15 and 20. The quantitative results would improve as the depth increases shown in Fig. 14(a) and Fig. 14(b). While deeper networks could results in huge computation cost. In Fig. 15, we fix the number of spatial attention fusion modules at 15, and there is no significant

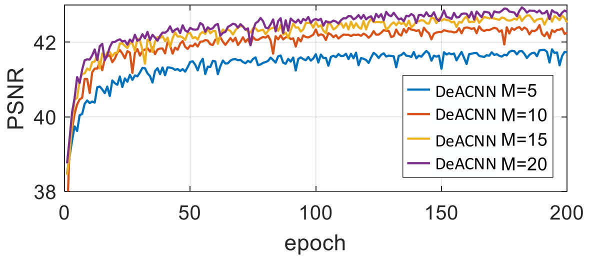


Fig. 15. PSNR with the validation datasets for different numbers of attention fusion modules.

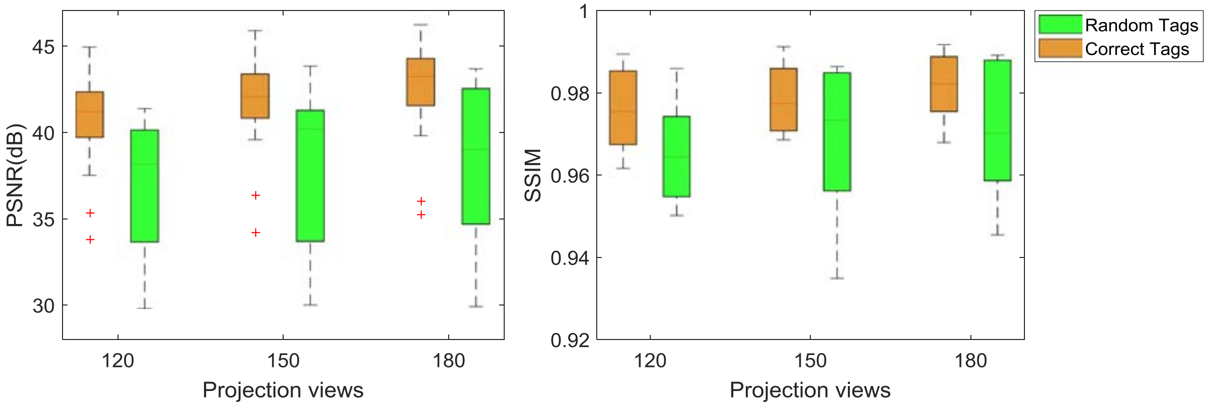


Fig. 16. Quantitative results comparison when the correct or random anatomical tags are input for test datasets, including 120, 150, and 180 projection views, in terms of the PSNR and SSIM. For each box, the central line in red denotes the median, the edges of the box represent the 25% and 75%. “+” in red indicates the outliers.

TABLE V
QUANTITATIVE RESULTS (MEAN ± SDS) FOR TEST DATASETS, INCLUDING 120, 150, AND 180 PROJECTION VIEWS, IN TERMS OF THE PSNR AND SSIM. REDCNN* DENOTES THE REDCNN WITH AN ANATOMICAL PRIOR FUSION MODULE, AND DCPN-DU* DENOTES THE DCPN-DU WITH AN ANATOMICAL PRIOR FUSION MODULE

Projection views	Metrics	REDCNN	REDCNN*	DCPN-DU	DCPN-DU*	DeACNN
120	PSNR	39.78±2.50	40.46±2.78*	40.16±2.61*	40.70±2.82*	40.70±2.77*
	SSIM	0.9609±0.0438	0.9627±0.0434*	0.9621±0.0435*	0.9633±0.0433*	0.9635±0.0437*
150	PSNR	41.03±2.36	41.49±2.62*	41.34±2.47*	41.67±2.68*	41.63±2.66*
	SSIM	0.9661±0.0395	0.9676±0.0388*	0.9672±0.0389*	0.9682±0.0386*	0.9679±0.0390*
180	PSNR	41.99±2.56	42.36±2.77*	42.26±2.64*	42.56±2.83*	42.48±2.80*
	SSIM	0.9707±0.0340	0.9719±0.0335*	0.9716±0.0336*	0.9725±0.0334*	0.9722±0.0337*

[1] * denotes P $<$ 0.05, which shows significantly different.

improvement when it is increased to 20. Thus, we set the number of attention fusion modules to 15.

- **The effectiveness of the anatomical prior information**. We validate the effectiveness of anatomical prior information by input the correct and random anatomical tags into the trained model. As is shown in Fig. 16, it would achieve better performance in PSNR and SSIM in correct anatomical tags than random ones. Besides, the value distribution on correct tags are more centralized than random tags, which proves that anatomical prior information avails the proposed method.
- **The effect of the anatomical prior and attention fusion module**. We validate the effectiveness of the channel weight prediction and the spatial attention stream for the baseline model (shown in Fig. 17). The performance gains on PSNR show the effectiveness of these two streams. As is shown in Fig. 14(c) and Fig. 14(d), the quantitative results also proves that anatomical prior information and spatial attention avails PSNR improvements. In particular, the anatomical prior stream plays a vital role compared with the spatial attention.
- **The effectiveness of the whole network framework**. Finally, we compare the qualitative results for PSNR during the training process (shown in Fig. 18); the results

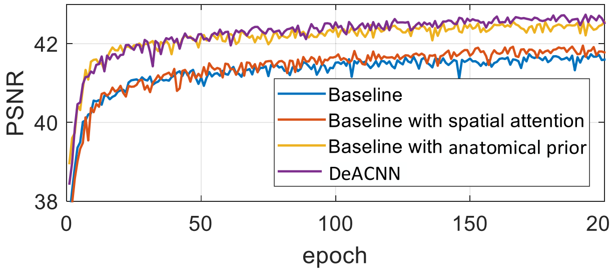


Fig. 17. PSNR with the validation datasets for different settings of the network framework.

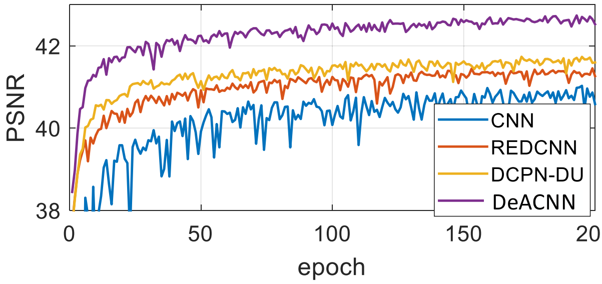


Fig. 18. PSNR with the validation datasets for different DL-based methods.

TABLE VI
PARAMETER COUNTS FOR DIFFERENT METHODS WITH AN ANATOMICAL PRIOR FUSION MODULE. REDCNN* DENOTES THE ENHANCED REDCNN WITH AN ANATOMICAL PRIOR FUSION MODULE, AND DCPN-DU* DENOTES THE ENHANCED DCPN-DU

Methods	REDCNN*	DCPN-DU*	DeACNN(Ours)
Parameters	1.14×10^6	1.14×10^6	0.77×10^6

in Fig. 14(e) and Fig. 14(f) also show that our method outperforms other methods during the training process.

IV. DISCUSSION AND CONCLUSION

From the ablation study to validate the effectiveness of the anatomical prior and attention fusion module, we discover that the anatomical prior is a key factor in noticeably improving performance. The experimental results for multiple anatomical sites can better reflect our advantages than those for the single anatomical site. The results for the single anatomical site could not reflect the difference in the learned weight caused by different parts, while the DeACNN still obtained quite excellent performance. We tend to migrate the anatomical prior fusion module to other DL-based methods, such as the REDCNN and DCPN-DU. We implement the enhanced models REDCNN* and DCPN-DU*. The quantitative results are compared in Table V, and their parameter counts are compared in Table VI. Although the parameter counts of our method are significantly less than those of other methods, which could obtain excellent denoising results. The enhanced models embedded with anatomical priors would improve performance better than the original versions of these models. The spatial attention may lose structural information due to the local attention regions and produce a reasonable SSIM. Our spatial attention mechanism works on local modules and the primary latter attention module can be corrected by the latter primary attention operation, which may lose the part that causes global concern. In the future, we will seek a better fusion method to improve performance. In addition, we applied clinical patient data for only 10 human body sites. More than 10 sites should be considered for real clinical applications. Further exploration of additional sites as well as different scanning parameter settings are needed.

In this paper, we introduce the anatomical prior for image denoising in low-dose imaging while considering the evident anatomical differences among human body sites. By using the channel weight prediction module, we fuse the anatomical prior with the coarse features extracted from the original LDCT images. Because the weight is different for each anatomical description, the united framework of our proposed method processes different anatomical prior inputs. In addition, the spatial attention fusion module is employed to explore the improved performance of our model while deepening the network. By this means, we incorporate anatomical descriptions into CT image denoising, which can be applied to real-world clinical data, especially in scanning certain human body sites. In ablation studies, the anatomical prior information is validated to be beneficial for proposed networks similar to the prior fusion and attention fusion module. Based on the qualitative results, the anatomical prior is proven to be a vital factor in noticeably improving performance. The experimental results demonstrate the effectiveness of our method both in terms of the visual effect and the qualitative measurements. Furthermore, our method can take advantage of clinical data from multiple anatomical sites rather than a single anatomical site.

REFERENCES

[1] D. J. Brenner and E. J. Hall, “Computed tomography-an increasing source of radiation exposure,” *New England J. Med.*, vol. 357, no. 22, pp. 2277–2284, 2007.

[2] H. Guo, U. Kruger, G. Wang, M. K. Kalra, and P. Yan, “Knowledge-based analysis for mortality prediction from CT images,” *IEEE J. Biomed. Health Inf.*, vol. 24, no. 2, pp. 457–464, Feb. 2020.

[3] C. Jiang, N. Zhang, J. Gao, and Z. Hu, “Geometric calibration of a stationary digital breast tomosynthesis system based on distributed carbon nanotube x-ray source arrays,” *PLoS One*, vol. 12, no. 11, 2017, Art. no. e0188367.

[4] 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.

[5] 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, 2009.

[6] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, and G. Wang, “Spectral CT reconstruction with image sparsity and spectral mean,” *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 510–523, Dec. 2016.

[7] 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 *Int. Conf. Med. Image Comput. Comput.-Assist. Interv*., 2016, pp. 423–431.

[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] Y. Chen *et al.*, “Artifact suppressed dictionary learning for low-dose CT image processing,” *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2271–2292, Dec. 2014.

[10] 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.

[11] Z. Hu, Y. Zhang, J. Liu, J. Ma, H. Zheng, and D. Liang, “A feature refinement approach for statistical interior CT reconstruction,” *Phys. Med. Biol.*, vol. 61, no. 14, pp. 5311–5334, 2016.

[12] Z. Hu *et al.*, “Image reconstruction from few-view ct data by gradient-domain dictionary learning,” *J. X-Ray Sci. Technol.*, vol. 24, no. 4, pp. 627–638, 2016.

[13] J. Huang *et al.*, “Iterative image reconstruction for sparse-view CT using normal-dose image induced total variation prior,” *PLoS One*, vol. 8, no. 11, 2013, Art. no. e79709.

[14] J. Ma *et al.*, “Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior,” *Phys. Med. Biol.*, vol. 57, no. 22, pp. 7519–7542, 2012.

[15] H. Chen *et al.*, “Low-dose CT via convolutional neural network,” *Biomed. Opt. Exp.*, vol. 8, no. 2, pp. 679–694, 2017.

[16] 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.

[17] H. Li and K. Mueller, “Low-dose CT streak artifacts removal using deep residual neural network,” in *Proc. Fully 3D Conf.*, 2017, pp. 191–194.

[18] Z. Hu *et al.*, “Artifact correction in low-dose dental CT imaging using wasserstein generative adversarial networks,” *Med. Phys.*, vol. 46, no. 4, pp. 1686–1696, 2019.

[19] Y. Wang *et al.*, “Iterative quality enhancement via residual-artifact learning networks for low-dose CT,” *Phys. Med. Biol.*, vol. 63, no. 21, 2018, Art. no. 215004.

[20] 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, Apr. 2018.

[21] X. Yin *et al.*, "Domain progressive 3 d residual convolution network to improve low dose CT imaging," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2903–2913, Dec. 2019.

[22] Y. Liu and Y. Zhang, "Low-dose CT restoration via stacked sparse denoising autoencoders," *Neurocomputing*, vol. 284, pp. 80–89, 2018.

[23] E. Kang, J. Min, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Trans. Med. Imag.*, vol. 36, no. 6, pp. 1358–1369, 2018.

[24] W. Du, H. Chen, P. Liao, H. Yang, G. Wang, and Y. Zhang, "Visual attention network for low-dose CT," *IEEE Signal Process. Lett.*, vol. 26, no. 8, pp. 1152–1156, Aug. 2019.

[25] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[26] Z. Hu *et al.*, "Artifact correction in low-dose dental CT imaging using wasserstein generative adversarial networks," *Med. Phys.*, vol. 46, no. 4, pp. 1686–1696, 2019.

[27] 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.

[28] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Deep learning for low-dose CT denoising using perceptual loss and edge detection layer," *J. Digit. Imag.*, vol. 33, no. 2, pp. 504–515, 2020.

[29] F. Yang, D. Zhang, H. Zhang, K. Huang, Y. Du, and M. Teng, "Streaking artifacts suppression for cone-beam computed tomography with the residual learning in neural network," *Neurocomputing*, vol. 378, pp. 65–78, 2020.

[30] L. Huang, H. Jiang, S. Li, Z. Bai, and J. Zhang, "Two stage residual CNN for texture denoising and structure enhancement on low dose CT image," *Comput. Methods Programs Biomed.*, vol. 184, 2020, Art. no. 105115.

[31] M. Meng *et al.*, "Semi-supervised learned sinogram restoration network for low-dose CT image reconstruction," in *Med. Imag.: Phys. Med. Imag.*, 2020, Art. no. 113120B.

[32] D. Li *et al.*, "Unsupervised data fidelity enhancement network for spectral CT reconstruction," in *Med. Imag.: Phys. Med. Imag.*, 2020, Art. no. 113124D.

[33] Z. Huang *et al.*, "Cagan: A cycle-consistent generative adversarial network with attention for low-dose CT imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1203–1218, 2020.

[34] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "Sacnn: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2289–2301, Jul. 2020.

[35] J. Lu, C. Xiong, D. Parikh, and R. Socher, "Knowing when to look: Adaptive attention via a visual sentinel for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 375–383.

[36] J. Lee, J. H. Shin, and J. S. Kim, "Interactive visualization and manipulation of attention-based neural machine translation," *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2017, pp. 121–126.

[37] L. Chen *et al.*, "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5659–5667.

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

[39] L. Shi *et al.*, "Review of CT image reconstruction open source toolkits," *J. X-Ray Sci. Technol.*, vol. 28, no. 4, pp. 619–639, 2020.

[40] A. Barakat and P. Bianchi, "Convergence and dynamical behavior of the ADAM algorithm for nonconvex stochastic optimization," *SIAM J. Optimiz.*, vol. 31, no. 1, pp. 244–274, 2021.