Considering anatomical prior information for low-dose CT image enhancement using attribute-augmented Wasserstein generative adversarial networks

Zhenxing Huang [a,b,c,d,e,1], Xinfeng Liu [f,1], Rongpin Wang [f], Jincai Chen [a,d,e,*], Ping Lu [a,d,e], Qiyang Zhang [b], Changhui Jiang [b], Yongfeng Yang [b,c], Xin Liu [b,c], Hairong Zheng [b,c], Dong Liang [b,c], Zhanli Hu [b,c,*]

[a] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science & Technology, Wuhan 430074, China
[b] Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] Chinese Academy of Sciences Key Laboratory of Health Informatics, Shenzhen 518055, China
[d] School of Computer Science & Technology, Huazhong University of Science & Technolog, Wuhan 430074, China
[e] Key Laboratory of Information Storage System, Engineering Research Center of Data Storage Systems and Technology, Ministry of Education of China, Wuhan 430074, China
[f] Department of Radiology, Guizhou Provincial People's Hospital, Guiyang 550002, China

ARTICLE INFO

ABSTRACT

Currently, many deep learning (DL)-based low-dose CT image postprocessing technologies fail to consider the anatomical differences in training data among different human body sites, such as the cranium, lung and pelvis. In addition, we can observe evident anatomical similarities at the same site among individuals. However, these anatomical differences and similarities are ignored in the current DL-based methods during the network training process. In this paper, we propose a deep network trained by introducing anatomical site labels, termed attributes for training data. Then, the network can adaptively learn to obtain the optimal weight for each anatomical site. By doing so, the proposed network can take full advantage of anatomical prior information to estimate high-resolution CT images. Furthermore, we employ a Wasserstein generative adversarial network (WGAN) augmented with attributes to preserve more structural details. Compared with the traditional networks that do not consider the anatomical prior and whose weights are consequently the same for each anatomical site, the proposed network achieves better performance by adaptively adjusting to the anatomical prior information.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

X-ray computed tomography (CT) is a highly valuable imaging modality used to visualize anatomical structures. Generally, X-ray CT is applied to whole-body diagnostic imaging; however, in real-world clinical diagnoses, it is also applied at specific human body sites, such the cranium, abdomen or waist. It is well known that radiation is harmful to patients due to its increased risk of cancer [3,33]. To reduce this health risk, low-dose techniques, whose goal is to build estimated high-resolution normal-dose CT (NDCT) images from corresponding low-dose CT (LDCT) images, have gradually gained considerable attention in the CT imaging domain [41,5,39,43,2,19]. To achieve low radiation doses, common strategies include directly lowering the current and using sparse-view or limited-view projections. However, lowering the radiation exposure can significantly degrade the imaging quality, which may lead to less accurate diagnosis, and the limited-view projection strategy may introduce streaking artifacts because of insufficient projection data. In recent decades, many approaches have been proposed to address the ill-posed problem of improving the quality of LDCT images. These approaches can be broadly categorized into three types: sinogram restoration techniques, iterative reconstruction methods and postprocessing methods.

Before the reconstruction of CT images, sinogram restoration methods are utilized to suppress noise in the raw or log-transformed data in the sinogram domain. Over the past few decades, several typical sinogram restoration methods, including bilateral filtering [32] and penalized weighted least-squares (PWLS) [36] algorithms, have been proposed. Although several well-known noise characteristics in the sinogram domain can be

* Corresponding authors at: Wuhan National Laboratory for Optoelectronics, Huazhong University of Science & Technology, Wuhan 430074, China (J. Chen). Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China (Z. Hu).
E-mail addresses: jcchen@hust.edu.cn (J. Chen), zl.hu@siat.ac.cn (Z. Hu).
[1] These authors contributed equally.

effectively exploited for this purpose, sinogram restoration methods often suffer from resolution loss in the image domain.

Substantial progress has been achieved using iterative reconstruction methods for noise suppression and the preservation of detailed information due to good phantom statistics and prior information from the input LDCT images. These methods have achieved promising results under limited data through the use of several types of priors, such as total variation (TV) [46,47,18,17], nonlocal means (NLM) [31,45,13], low-rank [4] and dictionary learning [40,38,16] priors. Moreover, in iterative reconstruction methods, compressive sensing (CS) techniques can also be used to improve the reconstructed image quality [10]. However, the final estimated CT images are highly dependent on the regularization terms or parameter settings used.

Postprocessing methods can be directly applied to CT images in DICOM format, thus allowing them to be easily integrated into the CT workflow without relying on projected data. In recent decades, dictionary-learning-based methods [25,8,9], NLM [29,44] and three-dimensional filtering methods [12,22] have been proposed. Recently, deep learning (DL)-based methods have demonstrated the potential to improve image quality for LDCT images. Due to the excellent feature extraction capabilities of neural networks, DL-based methods offer a convenient way to solve this ill-posed problem. Chen et al. [7] proposed a basic convolutional neural network (CNN) architecture for LDCT images. Inspired by autoencoders (AEs), Chen et al. [6] built a residual encoder-decoder CNN to suppress noise. Additionally, several U-shaped networks [34,28,30] have been introduced to perform this task. Kang et al. [23,24] applied a CNN with the wavelet transform. More recently, generative adversarial networks (GANs) [37,11,15,20] have been introduced to enable the estimation of CT images that retain more structural details.

Although several previous works have achieved improvements in image quality, most of them ignored the available anatomical structural prior information in the training data, which reflects the obvious structural differences among various human body sites, as can be observed from the differences in content in the CT image domain seen in Fig. 1. In addition, the same site shows similar anatomical structures among individuals. In this paper, we consider both the anatomical differences and similarities by introducing an anatomical description, called an attribute, as prior information. Through this additional attribute, the prior information is introduced into an end-to-end network in the form of an adaptive weight pattern such that input samples with the same attribute will share the same weight pattern. In addition, input samples with different attributes can be used to estimate the response of high-resolution CT images without the need to design a different model for each individual body site. Furthermore, we combine our supervised network with a Wasserstein GAN (WGAN) [1]. Because an input LDCT image and the corresponding estimated CT image should share the same attribute, we introduce an attribute loss to measure the distance between each given attribute and the corresponding attribute estimated by the discriminator network. We refer to our method as image enhancement with attribute augmentation (IEAA), and describe several related models and their loss functions in Table 1. Four loss functions, including L_{MSE}, L_{WGAN}, $L_{Attribute}$ and $L_{Perceptual}$ are elaborated in Section 2.3.

We summarize our contributions as follows:

(1) We introduce anatomical prior information to estimate high-resolution CT images. Anatomical structures exhibit evident differences among different human body sites; however, the input samples at the same site are quite similar among individuals. Based on these observations, we introduce anatomical prior information for low-dose CT. The

prior information, which is in the form of an anatomical description for each input LDCT image, is considered as an attribute that can be used to enhance the image restoration performance. To our knowledge, this is the first work to combine anatomical prior information with input CT images to enhance image restoration in the low-dose CT domain.

(2) We apply a united framework instead of designing individual models to address multiple human body sites. In our framework, the adaptive weights learned from the input attributes are adjusted for each human body site by a weight prediction module through channel attention. Different human body sites can share different weights, thus, it is possible to integrate multiple models for different human body sites into a single united framework. In this way, the training effort required for each body site is greatly reduced because all the data from each body site needs to be used only once for training.

(3) We achieve better performance with attribute-augmented WGANs. We introduce adversarial learning for our supervised IEAA model. During the training process, we employ four loss functions, namely, a mean square error (MSE) loss, an adversarial loss, a perceptual loss [21] and an attribute loss. To achieve both a good visual effect and high performance in terms of quantitative metrics, we recommend training the model with all four loss functions to obtain more precisely estimated CT images.

The remainder of this paper is organized as follows. Section 2 describes the proposed method in detail, including the overall framework, the attribute augmentation modules and the four loss functions. Experiments are reported in Section 3, which describes the patient data, training details, quantitative metrics, experimental results and ablation studies. Then, we elaborate a discussion of the findings and plans for future work in Section 4, and finally, we present our conclusion in Section 5.

2. Methods

This section presents the details of our method for solving the low-dose CT problem. First, we describe the overall framework, including the generator network and the discriminator network. Next, the attribute augmentation strategies used to incorporate anatomical prior information into deep networks are presented. Finally, we describe the various loss functions used during the training process.

2.1. Overall framework

Our overall framework consists of two parts: one that generates estimated high-dose CT images from their low-dose counterparts and one that discriminates estimated images from ground-truth images. In the generator network, we introduce the additional attribute inspired by considering the anatomical differences among different human body sites. Generally, considerable differences in anatomical structure are evident among different human body sites, such as the cranium, neck and knee. Therefore, we introduce human body site descriptions as prior information attributes to enhance the reconstruction quality. The discriminator network tries to distinguish estimated high-dose CT images from real ground-truth images. In particular, we introduce an attribute loss function, which measures the distance between the output attribute index and the input one.

We illustrate our overall framework in Fig. 2. For the generator network, we first introduce the weight prediction module into the attribute stream to obtain a weight mask. In each cascaded attri-
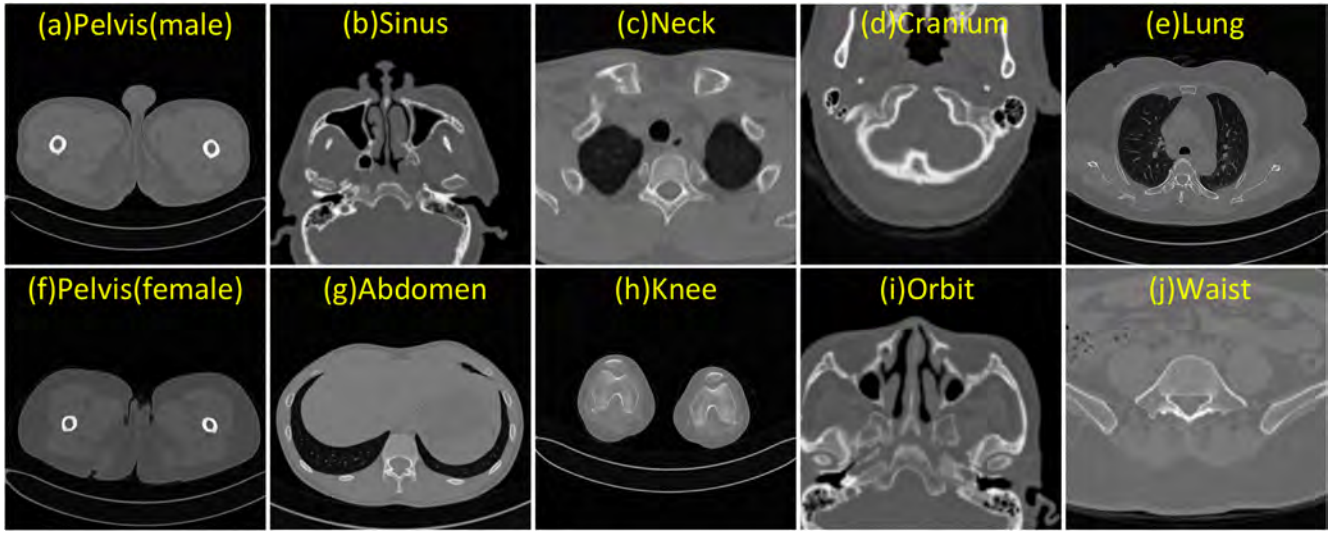
Fig. 1. Anatomical images from 10 different human body sites collected by a CT scanner. (a)–(j) Anatomical example for each human body site.

Table 1
Network models and loss functions. α, β, γ and θ denote the balance factors among the different loss functions.

Model	Loss Function(s)
IEAA	L_{MSE}
IEAA-W	αL_{WGAN}
IEAA-WP	$\alpha L_{WGAN} + \theta L_{Perceptual}$
IEAA-WM	$\alpha L_{WGAN} + \beta L_{MSE}$
IEAA-WA	$\alpha L_{WGAN} + \gamma L_{Attribute}$
IEAA-WMP	$\alpha L_{WGAN} + \beta L_{MSE} + \theta L_{Perceptual}$
IEAA-WAM	$\alpha L_{WGAN} + \gamma L_{Attribute} + \beta L_{MSE}$
IEAA-WAP	$\alpha L_{WGAN} + \gamma L_{Attribute} + \theta L_{Perceptual}$
IEAA-WAMP	$\alpha L_{WGAN} + \gamma L_{Attribute} + \beta L_{MSE} + \theta L_{Perceptual}$

bute augmentation module, the shared weight mask is utilized to redistribute the feature maps through a down-to-up projection unit. To maximize the helpful contextual information, we employ concatenation fusion with a convolutional layer. In addition, long skip connections are used to ensure that sufficient contextual information is preserved. In the discriminator network, we use a large stride instead of pooling operations to capitalize on the convolution operations. The outcomes of the discriminator network include the discrimination results and the predicted attribute indices. From each predicted attribute, we can naturally measure the distance with respect to the corresponding real attribute, which we call the attribute loss.

2.2. Attribute augmentation

In previous works, estimated CT images have generally been produced on the basis of their original low-dose counterparts without considering the anatomical prior information associated with the LDCT images. Moreover, several methods have been designed that are applicable to only a single specific human body site, such as the cranium or abdomen, which limits the robustness and generalizability of these methods. In this study, we propose a united framework for CT images acquired from multiple body sites by introducing an attribute prior based on the anatomical structures of the human body. Using this corresponding auxiliary attribute stream, the network model can incorporate anatomical information to achieve enhanced image quality. Furthermore, the anatomical information serves as a constraint on the output of the discriminator network to strengthen convergence during training.

We represent the attribute prior in the form of channel weights that reflect anatomical information. Because our dataset includes data from 10 human body sites, we primarily encode each attribute in the form of a fixed-length vector using a one-hot code. Next, we apply the weight prediction module to obtain a shared channel mask, as shown in Fig. 3. Similar to the approach in an encoder-decoder structure, we perform scale expansion and compression of the feature maps with respect to the channel dimensionality. In addition, we use skip connections to help preserve useful information. Finally, we produce the weight mask by means of the *sigmoid* activation function. The entire weight prediction process can be formulated as follows:

$$F_{mask} = Sigmoid(C_E(C_S(F_{att}))), \tag{1}$$

where F_{att} and F_{mask} denote the input attribute data and the output weight mask, respectively, and C_S and C_E denote the channel compression and expansion operations, respectively. We achieve channel scaling through convolution with a 1×1 kernel size.

Through the weight prediction module, we obtain the weight mask; then, we apply this shared mask in each cascaded module, called an attribute augmentation module, following the coarse feature extraction layer. In each attribute augmentation module, we use a down-to-up projection unit to obtain the initial features. After the weight mask is applied, the features are redistributed based on the attribute prior. To maximize the useful contextual information, we perform concatenation fusion. Finally, a short skip connection is also used here. The detailed parameters regarding the attribute augmentation modules are shown in Table 2. The process of applying the i-th attribute augmentation module can be formulated as follows:

$$F_{DU} = P_{up}(P_{down}(F_i)) \tag{2}$$
$$F_{i+1} = Conv(\{F_{DU}, F_{mask} * F_{DU}\}) + F_i, \tag{3}$$

where P_{up} and P_{down} denote the up- and down-projection units, respectively, which are based on convolution and deconvolution operations and do not change the image size, and $*$ denotes the elementwise product operation.

2.3. Loss functions

In contrast to other low-dose methods, our model requires a specific attribute as well as an input LDCT image. Given the NDCT images $x = \{x_1, x_2, \ldots, x_n\}$, we obtain the corresponding degraded
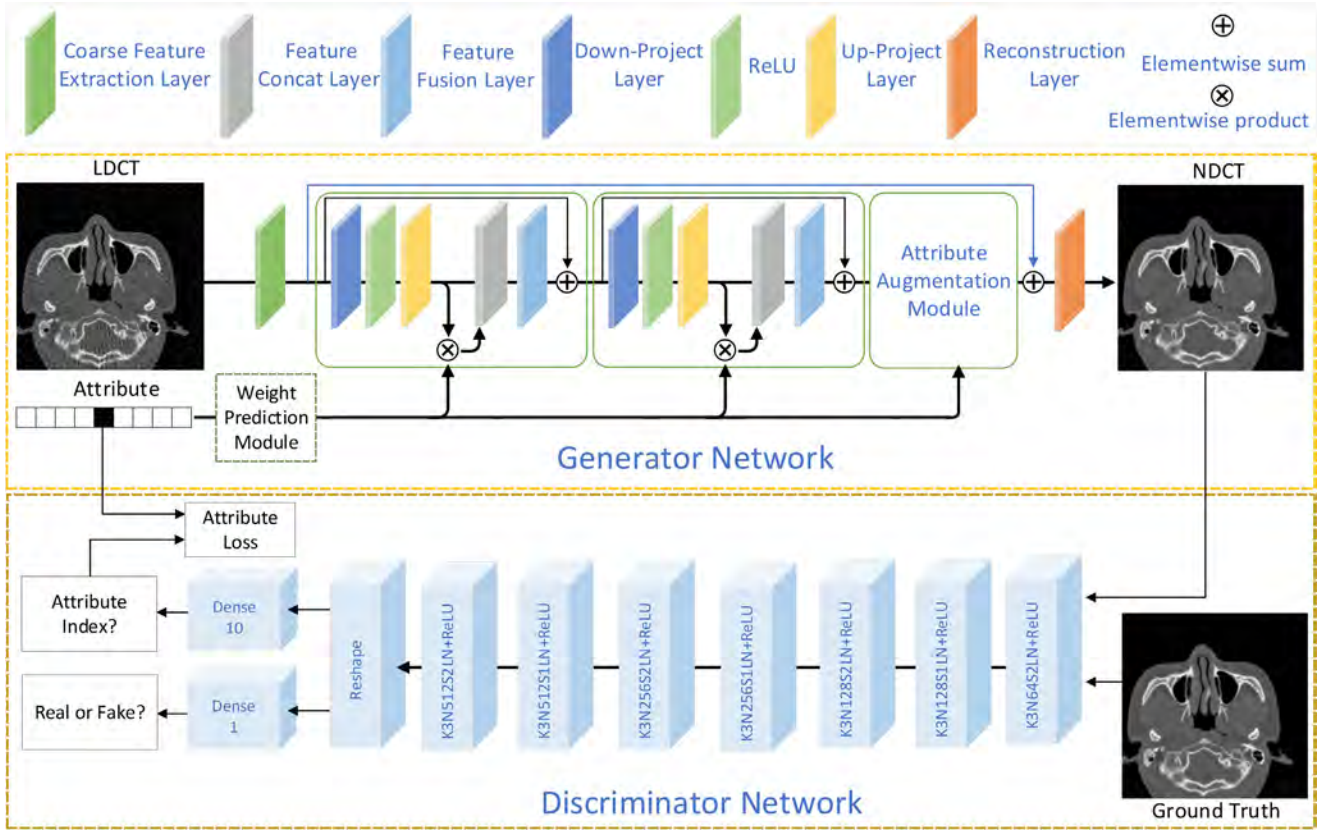
Fig. 2. Overall framework of our network model. The weight prediction module is illustrated in Fig. 3. For the discriminator network, "K" denotes the kernel size; "N" denotes the number of convolution output channels; "S" denotes the stride of a convolution operation; "LN" denotes the layer normalization suggested by [14]; and "ReLU" denotes the rectified linear unit activation function.
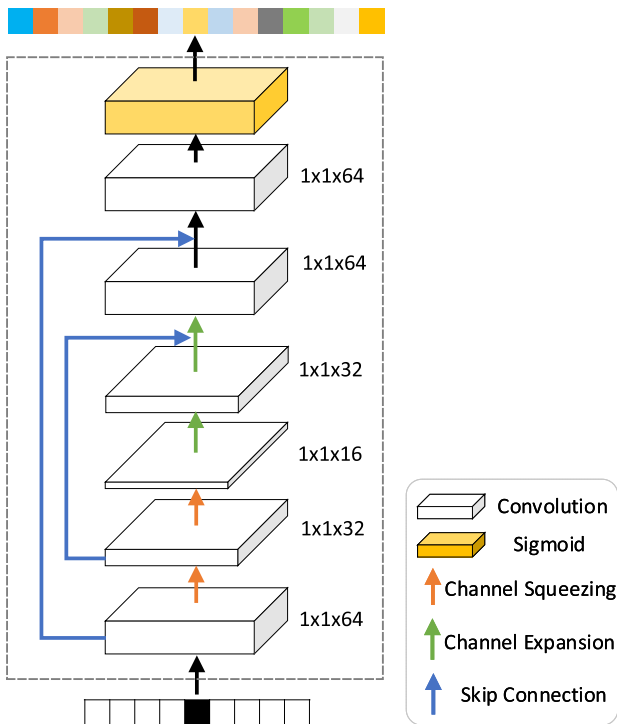
Fig. 3. Weight prediction module.

Table 2
Parameters of an attribute augmentation module.

Component	Kernel Size	Input/Output Channels
Down-projection layer	3×3	64/64
Up-projection layer	3×3	64/64
Feature fusion layer	3×3	128/64

LDCT images $y = \{y_1, y_2, \ldots, y_n\}$ and the attributes $a = \{a_1, a_2, \ldots, a_n\}$. Before training the adversarial model, we obtain a pretrained model through supervised learning with the MSE loss function.

2.3.1. Supervised loss

During the supervised training process, we apply the popular MSE loss function to ensure convergence. This training process can be formulated as

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^{n} \|G(y_i; a_i; \Theta) - x_i\|_2^2, \tag{4}$$

where Θ denotes the network parameters and G denotes the generator network.

2.3.2. WGAN loss

For the adversarial training process, we apply the Wasserstein distance with a gradient penalty instead of the negative log-likelihood to improve the training quality. This adversarial training process can be formulated as follows:

$$L_{WGAN} = -E_x[D_d(x)] + E_y[D_d(G(y; x; \Theta))] \tag{5}$$

$$+ \beta E_{\hat{x}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2], \tag{6}$$

where E denotes the expectation operator; β denotes the regularization parameter used to balance the Wasserstein estimation and the gradient penalty term; $\hat{x}$ denotes the uniform sampling of pairs of estimated and reference images; and D_d denotes the operation to distinguish an estimated CT image from a ground-truth image.

2.3.3. Attribute loss

Because the corresponding input and output CT images should share the same anatomical description of the human body site, to strengthen the discriminative capability of the discriminator network, we output the predicted attribute along with the conventional discrimination outcome. The attribute loss can then be formulated as shown below:

$$L_{Attribute} = E_x[D_a(x) - a] + E_y[D_a(G(y; a; \Theta)) - a], \tag{7}$$

where D_a denotes the operation of predicting the attribute index for the input CT image, and E_y denotes the expectation operator for the input low-dose CT image. Similarly, E_x denotes the expectation operator for the input ground truth. These operators measure the distance between the correct attribute and the estimated attribute.

2.3.4. Perceptual loss

Similar to several previous works [21,27,42] on enhancing perceptual effects, we also introduce a perceptual loss:

$$L_{Perceptual} = E_{(y,x)}[\|\Phi(G(y; a; \Theta)) - \Phi(x)\|_2^2], \tag{8}$$

where Φ denotes the operation to extract perceptual results using a pretrained VGG [35] model.

3. Experiments

In this section, we first describe the patient data used in our experiments and then report the training details. Then, we visually and quantitatively compare the results of our method with the results of several state-of-the-art methods and present ablation studies.

3.1. Patient data

Thanks to the support provided by Guizhou General People's Hospital, we were able to obtain high-resolution CT images acquired from 10 human body sites using a Siemens CT scanner. The tube voltage was fixed at 120 kVp, and the thickness was 1 mm. The reconstruction matrix size was 512×512. For consistency with real clinical applications, the anatomical descriptions used to define the attributes were determined based on the sampling position of the CT scanner for each known human body site. Next, we generated simulated 120, 150 and 180 sparse-projection LDCT images from the original high-resolution data using MATLAB 2017 with the aid of the Michigan Image Reconstruction Toolbox (MIRT).[2] We introduced attribute priors corresponding to 10 human body sites: pelvis (male), sinus, neck, cranium, lung, pelvis (female), abdomen, knee, orbit and waist. For each human body site, we collected data from 10 patients. In total, 100 sets of patient data comprising over 80,000 CT image files were utilized because data from different patients were used for different body sites. We divided the data such that 80 sets of patient data were used for training, 10 were used for validation and 10 were used for testing. During the training process, horizontal and vertical flipping and rotation were applied to increase the number of samples.

3.2. Training details

Regarding the attributes, we converted the anatomical descriptions into one-hot codes. To reduce the computational cost, we used LDCT image patches with an image size of 64×64 as the input to the network model along with the corresponding attributes. For both the coarse feature extraction layer and the reconstruction layer, we used a convolutional layer with a kernel size of 3×3. The learning rate was fixed to $1e^{-4}$. As shown in Table 1, the parameter settings for these balance factors were fixed to $\alpha = 0.001$, $\beta = 1.0$, $\gamma = 0.001$ and $\theta = 0.001$. Before training the adversarial network, we pretrained the model for 200 epochs. This initial training required approximately 8 h. The ADAM optimizer [26] was used to minimize the loss function during the supervised training process. Then, the SGD optimizer was applied for the unsupervised training process, with the learning rate fixed to $1e^{-5}$. We implemented our model in PyTorch on a TITAN 1080 Ti GPU.

3.3. Results

We compared our method with several other DL-based methods, namely, the CNN, REDCNN and DCPN-DU methods. To ensure fair comparisons, we retrained these methods on the same training and test datasets. Moreover, we maintained the same number of network parameters among all methods. In DCPN-DU, the main framework consisted of cascaded residual down- and up-projection units. We based our model on the DCPN-DU architecture but introduced additional branches, such as the weight prediction module. To reduce the number of parameters, we applied several convolution operations with a 1×1 kernel size. Furthermore, we compared a number of GAN-based methods with the various loss functions shown in Table 1. For the GAN-based methods that include the attribute loss, the network structure includes the additional discriminant attribute stream. Two common quantitative metrics, the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM), were used to evaluate the performances of the different methods.

3.3.1. Comparison with CNN-based methods

As shown in Fig. 4, our IEAA and IEAA-WAMP methods yield clear results, while the others produce blurred outputs at sites with bone links. Fig. 5 shows that the output of our methods also correctly reflects the normal structures of the soft tissues at the abdomen and cranium sites. In addition, the results of our methods appear to be clearer than the results produced by the other three methods in Fig. 6.

3.3.2. Comparison with GAN-based methods

Fig. 7 illustrates that our IEAA-WAMP method appears to yield results that are more similar to the reference NDCT image, while several noise stripes exist in the estimated results of the other methods. As shown in Fig. 8, unlike the other GAN-based methods, our IEAA-WAMP method successfully captures the normal structure of the tendon at the knee site.

3.3.3. Quantitative evaluation

We calculated the PSNR and SSIM for 60 selected test CT images to estimate the quantitative results. The results are listed in Table 3, showing that our IEAA method achieves the best quantitative results in terms of both the PSNR and SSIM among all the supervised methods. However, the unsupervised methods with MSE loss achieve better quantitative scores. Based on this observation, our IEAA-WAMP method with all four loss functions achieves better PSNR and SSIM performances.

---

[2] The code is available at https://web.eecs.umich.edu/ fessler/code/.
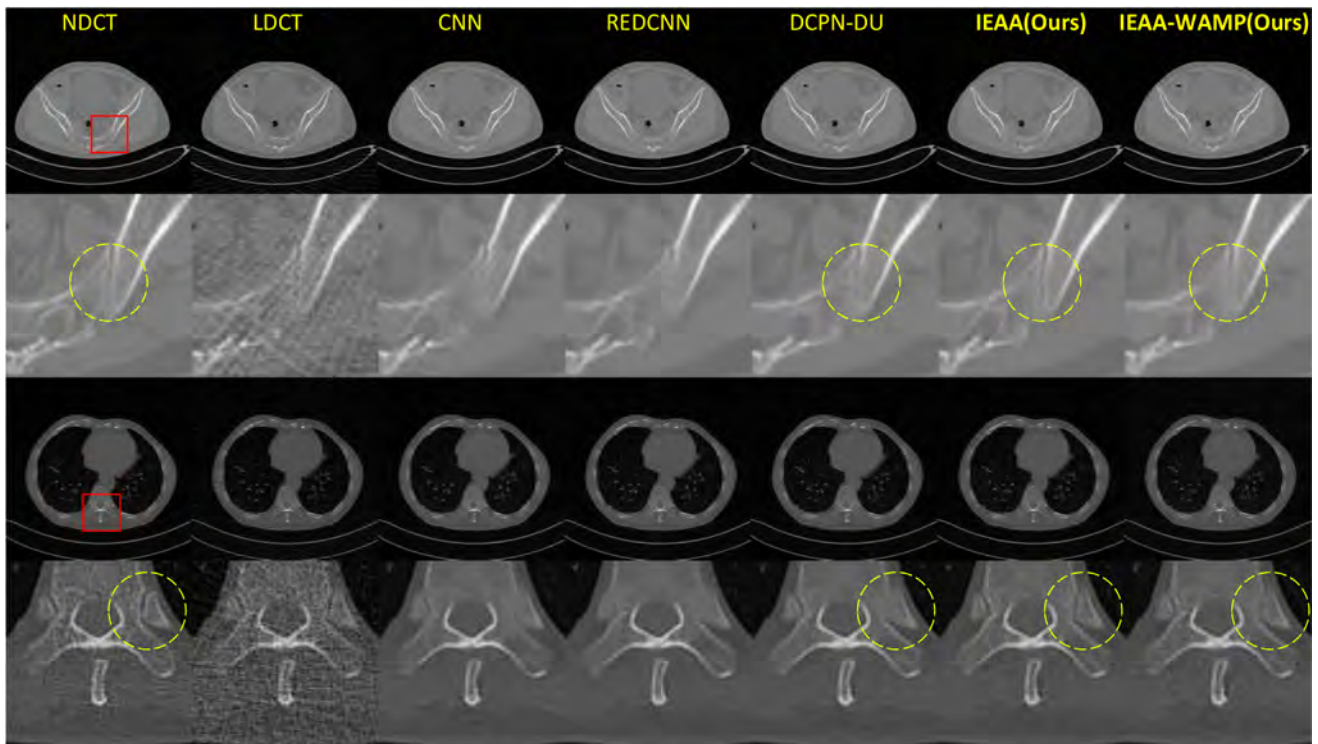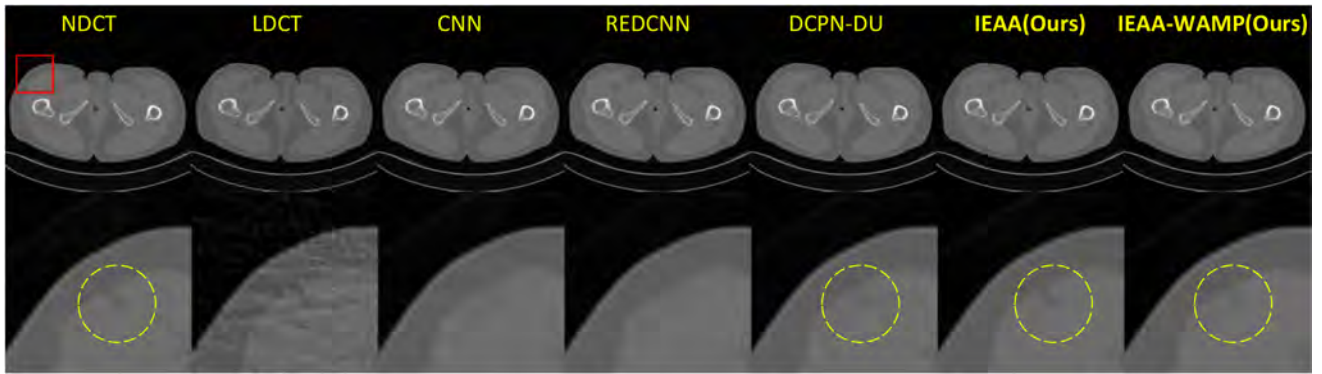
Fig. 4. Estimated results for 120-view projections. The red boxes demarcate the regions of interest (ROIs), and the yellow circles highlight several differences among the estimated results.
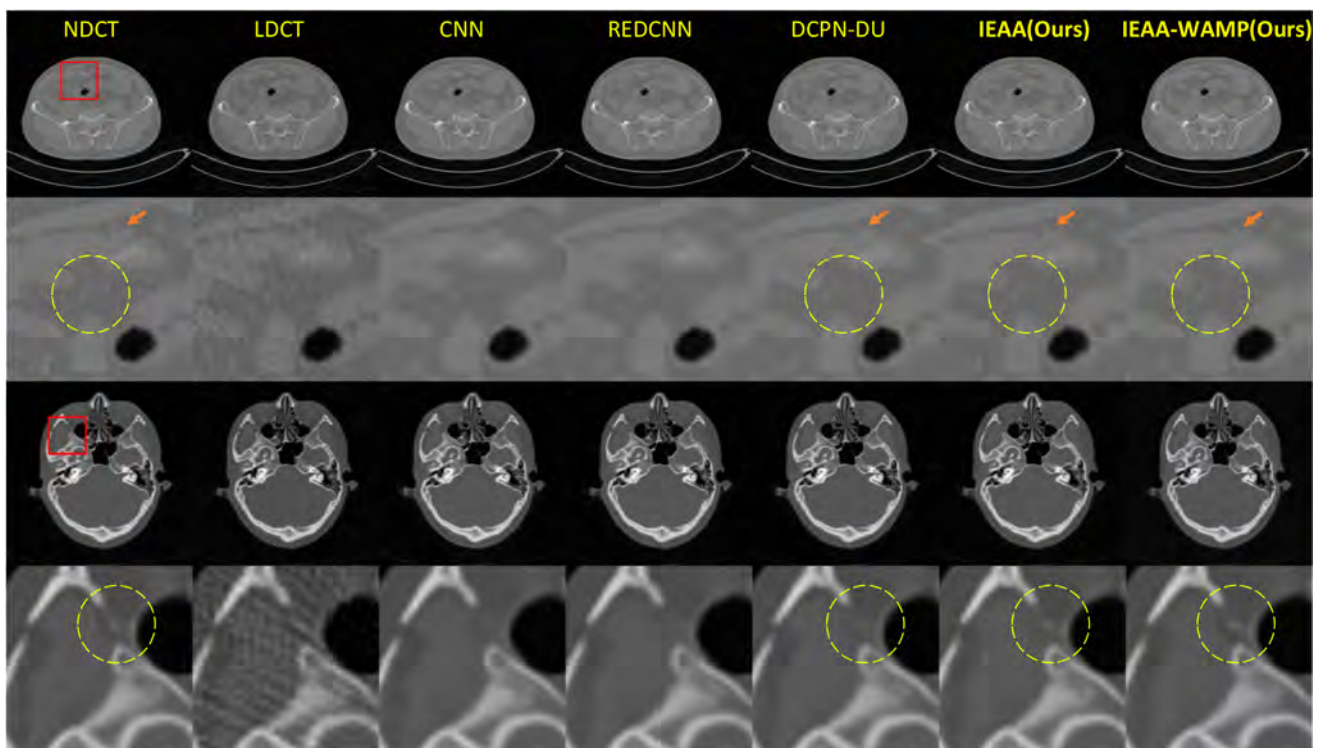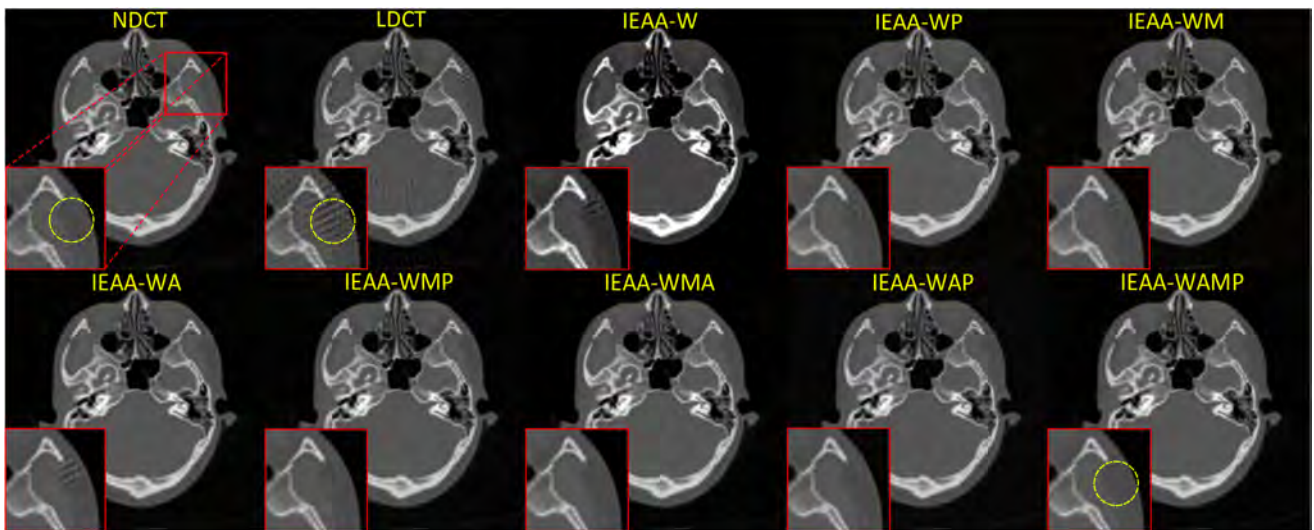

Fig. 5. Estimated results for 150-view projections. The red boxes demarcate the ROIs, and the yellow circles and orange arrows highlight several differences among the estimated results.
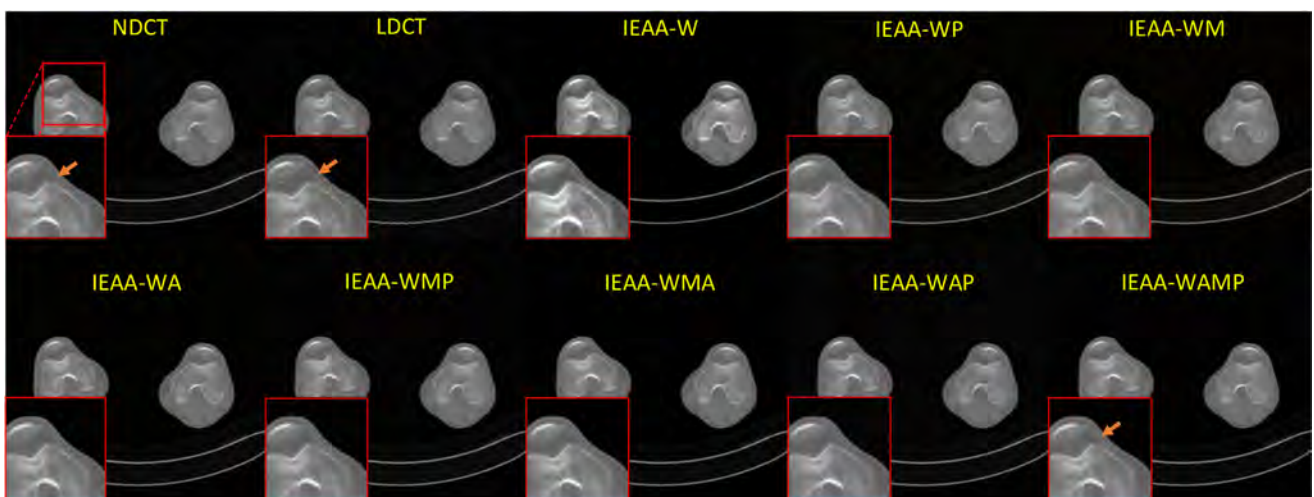
Fig. 6. Estimated results for 180-view projections, the red box demarcates the ROI. The yellow circles highlight several differences among the estimated results.

Fig. 7. Estimated results of different GAN-based methods. The red boxes demarcate the ROIs, and the yellow circles highlight several differences among the estimated results.


Fig. 8. Estimated results of different GAN-based methods. The red boxes demarcate the ROIs, and the orange arrows highlight several differences among the estimated results.

Table 3
Quantitative results of different methods in terms of the PSNR and SSIM. The best values are shown in red, and the second best values are presented in blue.

Methods		120 projection views		150 projection views		180 projection views	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CNN-based	LDCT	30.09	0.7660	31.25	0.8132	32.37	0.8531
	CNN	39.91	0.9640	40.83	0.9681	41.69	0.9714
	REDCNN	40.24	0.9647	41.13	0.9688	42.05	0.9722
	DCPN-DU	40.65	0.9659	41.39	0.9699	42.32	0.9731
	IEAA	41.27	0.9674	41.97	0.9713	42.74	0.9741
GAN-based	IEAA-W	28.36	0.9178	28.03	0.9203	28.38	0.9250
	IEAA-WP	40.16	0.9664	40.91	0.9700	41.56	0.9729
	IEAA-WM	40.08	0.9677	41.86	0.9713	42.57	0.9741
	IEAA-WA	40.46	0.9627	41.21	0.9674	41.95	0.9712
	IEAA-WMP	41.08	0.9679	41.87	0.9715	42.66	0.9743
	IEAA-WAM	40.79	0.9677	41.33	0.9711	42.06	0.9738
	IEAA-WAP	38.11	0.9540	38.97	0.9613	39.63	0.9665
	IEAA-WAMP	41.39	0.9681	42.07	0.9716	42.86	0.9743
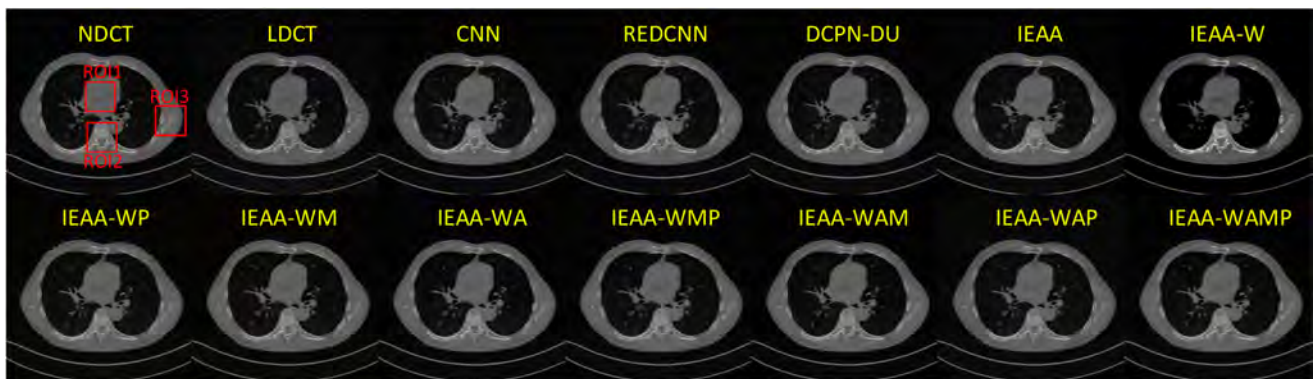
Fig. 9. Estimated results for 180-view projections. The red boxes demarcate the ROIs.
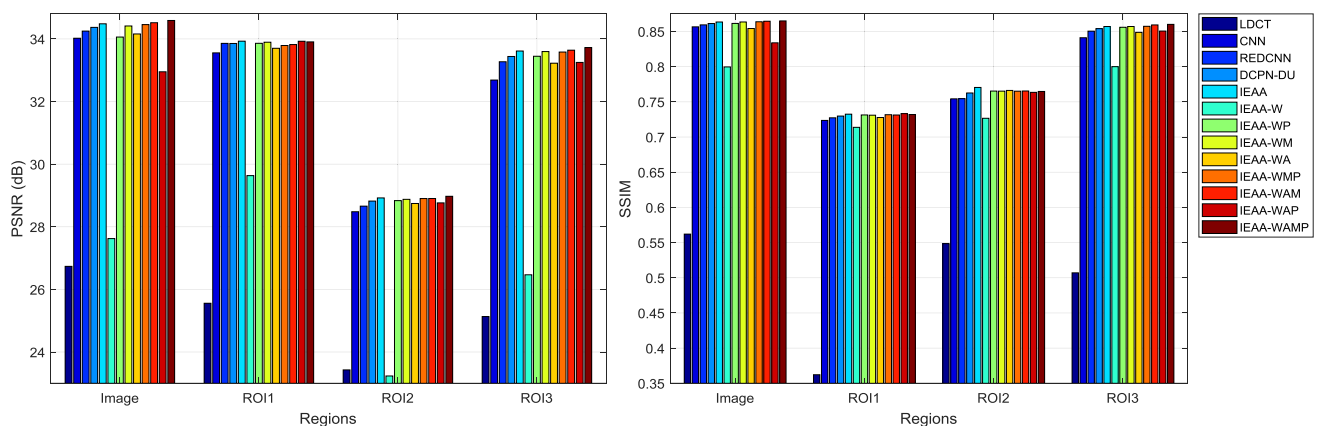
Fig. 10. Statistical results on the whole image and the three ROIs marked in Fig. 9.

We also calculated the PSNR and SSIM values for several local regions of interest (ROIs). As shown in Fig. 9, we selected three ROIs in the center and edge regions of the estimated images. The statistical results presented in Fig. 10 further demonstrate the superiority of IEAA-WAMP in most cases.

3.4. Ablation studies

3.4.1. Number of attribute augmentation modules
We explored the trade-off between performance and the number of network parameters. As shown in Fig. 11(a), we varied the
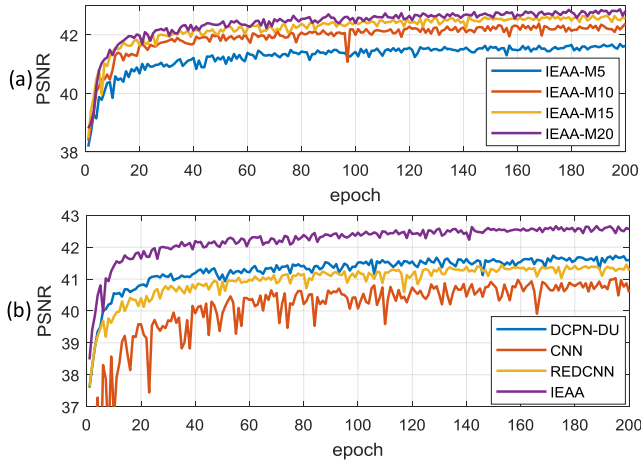
Fig. 11. PSNR results on the validation dataset.

number of attribute augmentation modules, M, between 5, 10, 15 and 20 and found no evident improvement when the number of modules increased from $M = 15$ to $M = 20$. Thus, we fixed the number of attribute augmentation modules to 15.

3.4.2. Effectiveness of attribute augmentation

Based on DCPN-DU, we incorporated the attribute prior into the cascaded modules. As shown in Fig. 11(b), our method outperforms both the baseline DCPN-DU model and the other models. Although our method requires a few additional parameters, the performance is significantly improved. In Table 4, the impact of attribute loss on performance is explored. For the trained models, we input the random attributes and correct attributes into the network framework. Our methods(IEAA-WA, IEAA-WAM, IEAA-WAP and IEAA-WAMP) with attribute loss are chosen to conduct the performance comparison. Note that the quantitative results on PSNR and SSIM would degrade by random attributes, which demonstrates that attribute information plays a role in improving performance. Specially, the quantitative results of IEAA-WA with correct attributes would gain over 4.0 dB improvement on PSNR under 120 projection views.

3.4.3. Effectiveness of the attribute loss

Compared with IEAA-W, IEAA-WA achieves higher PSNR values, as shown in Figs. 7 and 8. In Fig. 7, the results generated by IEAA-WAMP appear to be closer to the reference image than are the results generated by IEAA-WMP. Moreover, IEAA-WAP seems to be less susceptible to stripe noise than IEAA-WP, as shown in Fig. 7. Thus, we conclude that the attribute loss is beneficial for suppressing noise and improving the quantitative results.

4. Discussion and future work

In this section, we first discuss the influence of the number of parameters on improving performance. Then, we discuss the trade-off between the visual effect and the PSNR. Finally, we describe our future work.

The trade-off between the increase in the number of parameters and the performance improvement should be considered. To ensure fair comparisons, we rebuilt several models, including the CNN, REDCNN and DCPN-DU models. The parameter count of these models was 1.1M, while for the IEAA, the parameter count was 1.2M. However, this 0.1M increase in the number of parameters resulted in an evident performance improvement.

Multiple loss functions should be considered to ensure both a high PSNR and a good visual effect. In general, MSE loss will lead to a higher PSNR, and adversarial loss will ensure the preservation of more textural details, thus facilitating a better visual effect. However, adversarial loss may also introduce noise or artifacts. Consequently, a trade-off is to expected between obtaining a better visual effect or a higher PSNR. Fig. 7 shows that GAN-based models with adversarial loss, attribute loss and MSE loss are recommended for optimizing this trade-off.

In future work, we plan to explore the possibility of adaptive discrimination among human body sites rather than relying on the input prior in the generator network for whole-body data. In addition, we will consider multiple scan parameters for the same sites to adapt to different clinical applications.

5. Conclusion

In this paper, we proposed a novel low-dose image enhancement method that considers anatomical prior information in the low-dose CT domain for the first time by introducing an attribute in the form of an anatomical description prior for each specific human body site that guides the reconstruction process by means of channel weights incorporated into the main framework. Low-dose CT images with the same attribute share the same weight pattern, while different body sites have other weights. This approach makes it easy to combine multiple sets of data acquired from different body sites into a single united framework. In addition, we introduce adversarial learning to seek a better visual effect. We built the GAN-based models based on the WGAN architecture. The experiments demonstrate that the additional attribute considered here is beneficial to the performance of the supervised model and that the attribute loss is effective for improving the results of GAN-based models. In quantitative terms, our supervised IEAA method outperforms other CNN-based methods. Four loss functions, namely, adversarial loss, attribute loss, MSE loss and perceptual loss, are all recommended for use during the training process to achieve more accurate outcomes.

Table 4
Quantitative results of different methods on PSNR and SSIM with random or correct attributes.

Methods		120 projection views		150 projection views		180 projection views	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Random Attributes	IEAA-WA	36.23	0.9425	37.85	0.9566	39.06	0.9638
	IEAA-WAM	37.72	0.9543	37.45	0.9557	38.88	0.9606
	IEAA-WAP	37.27	0.9520	37.61	0.9567	39.33	0.9627
	IEAA-WAMP	37.92	0.9591	38.20	0.9612	39.53	0.9673
Correct Attributes	IEAA-WA	40.46	0.9627	41.21	0.9674	41.95	0.9712
	IEAA-WAM	40.79	0.9677	41.33	0.9711	42.06	0.9738
	IEAA-WAP	38.11	0.9540	38.97	0.9613	39.63	0.9665
	IEAA-WAMP	41.39	0.9681	42.07	0.9716	42.86	0.9743

CRediT authorship contribution statement

Zhenxing Huang: Conceptualization, Methodology. **Xinfeng Liu:** Data curation, Investigation, Methodology. **Rongpin Wang:** Data curation, Investigation. **Jincai Chen:** Methodology, Writing - original draft. **Ping Lu:** Methodology. **Qiyang Zhang:** Software, Validation. **Changhui Jiang:** Software, Validation. **Yongfeng Yang:** Supervision, Writing - review & editing. **Xin Liu:** Supervision, Writing - review & editing. **Hairong Zheng:** Supervision, Writing - review & editing. **Dong Liang:** Supervision, Writing - review & editing. **Zhanli Hu:** Methodology, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: Proceedings of the 34th International Conference on Machine Learning, 2017, pp. 214–223.

[2] P. Bao, W. Xia, K. Yang, W. Chen, M. Chen, Y. Xi, S. Niu, J. Zhou, H. Zhang, H. Sun, et al., Convolutional sparse coding for compressed sensing ct reconstruction, IEEE Transactions on Medical Imaging 38 (2019) 2607–2619.

[3] D.J. Brenner, E.J. Hall, Computed tomography–an increasing source of radiation exposure, New England Journal of Medicine 357 (2007) 2277–2284.

[4] J.F. Cai, X. Jia, H. Gao, S.B. Jiang, Z. Shen, H. Zhao, Cine cone beam ct reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study, IEEE Transactions on Medical Imaging 33 (2014) 1581–1591.

[5] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, G. Wang, Learn: Learned experts– assessment-based reconstruction network for sparse-data ct, IEEE Transactions on Medical Imaging 37 (2018) 1333–1347.

[6] H. Chen, Y. Zhang, M.K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, G. Wang, Low-dose ct with a residual encoder-decoder convolutional neural network, IEEE Transactions on Medical Imaging 36 (2017) 2524–2535.

[7] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, G. Wang, Low-dose ct via convolutional neural network, Biomedical Optics Express 8 (2017) 679–694.

[8] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.L. Coatrieux, C. Toumoulin, Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing, Physics in Medicine & Biology 58 (2013) 5803.

[9] X.Y. Cui, Z.G. Gui, Q. Zhang, H. Shangguan, A.H. Wang, Learning-based artifact removal via image decomposition for low-dose ct image processing, IEEE Transactions on Nuclear Science 63 (2016) 1860–1873.

[10] D.L. Donoho et al., Compressed sensing, IEEE Transactions on Information Theory 52 (2006) 1289–1306.

[11] W. Du, H. Chen, P. Liao, H. Yang, G. Wang, Y. Zhang, Visual attention network for low-dose ct, IEEE Signal Processing Letters 26 (2019) 1152–1156.

[12] P.F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, R. Weissleder, Block matching 3d random noise filtering for absorption optical projection tomography, Physics in Medicine & Biology 55 (2010) 5401.

[13] M. Green, E.M. Marom, N. Kiryati, E. Konen, A. Mayer, Efficient low-dose ct denoising by locally-consistent non-local means (lc-nlm), in: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, 2016, pp. 423–431.

[14] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans, 2017, arXiv preprint arXiv:1704.00028..

[15] Z. Hu, C. Jiang, F. Sun, Q. Zhang, Y. Ge, Y. Yang, X. Liu, H. Zheng, D. Liang, Artifact correction in low-dose dental ct imaging using wasserstein generative adversarial networks, Medical Physics 46 (2019) 1686–1696.

[16] Z. Hu, Q. Liu, N. Zhang, Y. Zhang, X. Peng, P.Z. Wu, H. Zheng, D. Liang, Image reconstruction from few-view ct data by gradient-domain dictionary learning, Journal of X-ray Science and Technology 24 (2016) 627–638.

[17] Z. Hu, Y. Zhang, J. Liu, J. Ma, H. Zheng, D. Liang, A feature refinement approach for statistical interior ct reconstruction, Physics in Medicine & Biology 61 (2016) 5311.

[18] J. Huang, Y. Zhang, J. Ma, D. Zeng, Z. Bian, S. Niu, Q. Feng, Z. Liang, W. Chen, Iterative image reconstruction for sparse-view ct using normal-dose image induced total variation prior, PloS One 8 (2013), e79709.

[19] Y. Huang, Z. Lu, Z. Shao, M. Ran, J. Zhou, L. Fang, Y. Zhang, Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network, Optics Express 27 (2019) 12289–12307.

[20] Z. Huang, Z. Chen, Q. Zhang, G. Quan, M. Ji, C. Zhang, Y. Yang, X. Liu, D. Liang, H. Zheng, et al., Cagan: a cycle-consistent generative adversarial network with attention for low-dose ct imaging, IEEE Transactions on Computational Imaging 6 (2020) 1203–1218.

[21] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: ECCV 2016, 2016, pp. 694–711.

[22] D. Kang, P. Slomka, R. Nakazato, J. Woo, D.S. Berman, C.C.J. Kuo, D. Dey, Image denoising of low-radiation dose coronary ct angiography by an adaptive block-matching 3d algorithm, in: Medical Imaging 2013: Image Processing, International Society for Optics and Photonics, 2013, p. 86692G.

[23] E. Kang, J. Min, J.C. Ye, A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction, Medical Physics 44 (2017) e360–e375.

[24] E. Kang, J. Min, J.C. Ye, Wavelet domain residual network (wavresnet) for low-dose x-ray CT reconstruction, CoRR abs/1703.01383, 2017..

[25] M. Katsura, I. Matsuda, M. Akahane, J. Sato, H. Akai, K. Yasaka, A. Kunimatsu, K. Ohtomo, Model-based iterative reconstruction technique for radiation dose reduction in chest ct: comparison with the adaptive statistical iterative reconstruction technique, European Radiology 22 (2012) 1613–1623.

[26] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980..

[27] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017..

[28] H. Li, K. Mueller, Low-dose ct streak artifacts removal using deep residual neural network, in: Proceedings of Fully 3D Conference, 2017, pp. 191–194.

[29] Z. Li, L. Yu, J.D. Trzasko, D.S. Lake, D.J. Blezek, J.G. Fletcher, C.H. McCollough, A. Manduca, Adaptive nonlocal means filtering based on local noise level for ct denoising, Medical Physics 41 (2014).

[30] Y. Liu, Y. Zhang, Low-dose ct restoration via stacked sparse denoising autoencoders, Neurocomputing 284 (2018) 80–89.

[31] J. Ma, H. Zhang, Y. Gao, J. Liang, Z. Liang, Q. Feng, W. Chen, Iterative image reconstruction for cerebral perfusion ct using a pre-contrast scan induced edge-preserving prior, Physics in Medicine & Biology 57 (2012) 7519.

[32] A. Manduca, L. Yu, J.D. Trzasko, N. Khaylova, J.M. Kofler, C.M. McCollough, J.G. Fletcher, Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct, Medical Physics 36 (2009) 4911–4919.

[33] A.E. Othman, S. Afat, M.A. Brockmann, O. Nikoubashman, C. Brockmann, K. Nikolaou, M. Wiesmann, Radiation dose reduction in perfusion ct imaging of the brain: a review of the literature, Journal of Neuroradiology 43 (2016) 1–5.

[34] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-assisted Intervention, Springer, 2015, pp. 234–241.

[35] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556..

[36] J. Wang, T. Li, H. Lu, Z. Liang, Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography, IEEE Transactions on Medical Imaging 25 (2006) 1272–1283.

[37] J.M. Wolterink, T. Leiner, M.A. Viergever, I. Išgum, Generative adversarial networks for noise reduction in low-dose ct, IEEE Transactions on Medical Imaging 36 (2017) 2536–2545.

[38] D. Wu, L. Li, L. Zhang, Feature constrained compressed sensing CT image reconstruction from incomplete data via robust principal component analysis of the database, Physics in Medicine and Biology 58 (2013) 4047–4070.

[39] W. Xia, W. Wu, S. Niu, F. Liu, J. Zhou, H. Yu, G. Wang, Y. Zhang, Spectral ct reconstruction–assist: Aided by self-similarity in image-spectral tensors, IEEE Transactions on Computational Imaging 5 (2019) 420–436.

[40] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, G. Wang, Low-dose x-ray ct reconstruction via dictionary learning, IEEE Transactions on Medical Imaging 31 (2012) 1682–1697.

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

[42] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M.K. Kalra, Y. Zhang, L. Sun, G. Wang, Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss, IEEE Transactions on Medical Imaging 37 (2018) 1348–1357.

[43] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, et al., Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle), IEEE Transactions on Medical Imaging 39 (2019) 188–203.

[44] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Han, H. Lu, W. Moore, Z. Liang, Statistical image reconstruction for low-dose ct using nonlocal means-based regularization. Part ii: An adaptive approach, Computerized Medical Imaging and Graphics 43 (2015) 26–35.

[45] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, G. Wang, Spectral ct reconstruction with image sparsity and spectral mean, IEEE Transactions on Computational Imaging 2 (2016) 510–523.

[46] Y. Zhang, W. Zhang, Y. Lei, J. Zhou, Few-view image reconstruction with fractional-order total variation, JOSA A 31 (2014) 981–995.

[47] Y. Zhang, W.H. Zhang, H. Chen, M.L. Yang, T.Y. Li, J.L. Zhou, Few-view image reconstruction combining total variation and a high-order norm, International Journal of Imaging Systems and Technology 23 (2013) 249–255.

Zhenxing Huang received his B.S. degree in network engineering from the South-Central University for Nationalities in 2016. He is currently pursuing a Ph.D. degree at Huazhong University of Science and Technology (HUST). His research interests include CT imaging and image processing.

Xinfeng Liu received the M.S. degree in medical imaging from Zunyi Medical University, China, in 2012. He is currently an attending physician at the Guizhou Provincial People's Hospital of China. His research interests include cardiovascular disease diagnosis, medical image processing and segmentation.

Rongpin Wang received a Ph.D. degree in medical imaging from Southern Medical University, China, in 2010. He is currently a chief physician at the Guizhou Provincial People's Hospital of China. His research interests include8 diffusion tensor imaging of fetal brain development and developmental disorders.

Jincai Chen is a professor at the Wuhan National Laboratory for Optoelectronics (WNLO), Huazhong University of Science and Technology (HUST). He received his doctoral degree in computer architecture from Xi'an Jiaotong University in 2000. From 2001 to 2003, he was engaged in postdoctoral research at the National Laboratory of Storage Systems, HUST. From 2011–2012, he was a visiting researcher at the University of California, Santa Cruz, USA. His main research directions include storage theory and technology, big data, machine learning, image processing and affective computing.

Ping Lu received her master's degree in computer architecture from Huazhong University of Science and Technology, Wuhan, Hubei, China, in 1995. She is currently an associate professor of the School of Computer Science and Technology, Huazhong University of Science and Technology. Her research fields are information storage technology and big data technology.

Qiyang Zhang received the M.S. degree from Liaoning Technical University, China, in 2010 and is currently pursuing a Ph.D. degree at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include computed tomography, PET, image processing and machine learning.

Changhui Jiang received an M.S. degree in mechatronic engineering from the Harbin Institute of Technology in 2011. Then, he joined The Shenzhen Institutes of Advanced Technology, Chinese Academy of Science. His research interests include CT image processing and the application of artificial intelligence in medical images.

Yongfeng Yang received his Ph.D. degree in nuclear physics from Kyushu University, Japan, in 2001. He is currently a professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include PET molecular imaging instruments.

Xin Liu received his Ph.D. degree in imaging and nuclear medicine from PLA General Hospital, China, in 2006. He is currently a professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include magnetic resonance imaging technology.
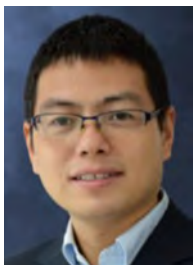
Hairong Zheng received his Ph.D. degree in biomedical engineering from the University of Colorado, United States, in 2006. He is currently a professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include medical imaging technology and instruments.

Zhanli Hu received his Ph.D. degree in computer application technology from the Chinese Academy of Sciences University, China, in 2014. He is currently an Associate Professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include CT and PET imaging and medical artificial intelligence.

Dong Liang received his Ph.D. degree in pattern recognition and intelligent systems from Shanghai Jiaotong University, China, in 2006. He is currently a professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include signal processing, machine learning and biomedical imaging.