Deep Cascade Residual Networks (DCRNs): Optimizing an Encoder–Decoder Convolutional Neural Network for Low-Dose CT Imaging

Zhenxing Huang, Zixiang Chen, Guotao Quan, Yanfeng Du, Yongfeng Yang, *Member, IEEE*, Xin Liu, Hairong Zheng, *Senior Member, IEEE*, Dong Liang, *Senior Member, IEEE*, and Zhanli Hu, *Senior Member, IEEE*

Abstract—To suppress noise and artifacts caused by the reduced radiation exposure in low-dose computed tomography, several deep learning (DL)-based image restoration methods have been proposed over the past few years. Many of these popular DL-based methods adopt an encoder–decoder framework, for instance, the residual encoder–decoder convolutional neural network. However, this popular framework may suffer from information loss for continual downsampling operations. In this article, deep cascaded residual networks (DCRNs) are proposed to optimize the popular encoder–decoder network. First, cross up- and downsampling operations as well as attention extraction are substitutes for the strict "downsampling and then upping" principle. What is more, four hybrid loss functions, namely, mean absolute error, edge loss, perceptual loss and adversarial loss, are engaged to achieve better visual effects and suppress noise. The experiments are conducted on three individual clinical CT datasets: dental CT data collected with a scanner manufactured by Zhongke Tianyue Company (ZTC), data from the American Association of Physicists in Medicine (AAPM) Challenge, and data collected with a commercial CT scanner from United Imaging Healthcare (UIH). The experimental results indicate the effective noise reduction and detail preservation capabilities of the proposed methods under different radiation dose-reduction strategies.

Index Terms—Adversarial learning, attention mechanism, cascaded residual network, low-dose CT (LDCT) imaging.

I. INTRODUCTION

AS AN important imaging modality, X-ray computed tomography (CT) is extensively employed in modern hospitals and clinics. Considering the health risks associated with high-dose X-ray radiation [1], research on adaptive low-dose CT (LDCT) technologies has attracted substantial interest. Traditionally, two ways exist to lower the radiation dose: 1) reducing the X-ray flux or 2) reducing the number of projection views. In the first approach, decreasing the operating current or diminishing the time under X-ray exposure helps to reduce the X-ray flux [2]. In the second approach, limited-angle and sparse-sampling methods require fewer projection views. As shown in Fig. 1, the differences between LDCT images and normal-dose CT (NDCT) images are visualized using three dose-reduction approaches. Different dose-reduction approaches will result in different types of noise distribution reflected in the ultimate image. For instance, noise appears as streak artifacts under limited angles. In general, a weaker X-ray flux compromises the diagnostic performance because of additional noise and artifacts in the resulting image [3].

To address this inherent problem, many reconstruction methods aimed at improving LDCT image quality have been developed. Generally, these methods can be divided into three categories: 1) sinogram-domain filtering methods; 2) iterative reconstruction methods; and 3) postprocessing methods. First, sinogram-domain filtering methods, which typically include structural adaptive filtering [4], bilateral filtering [5] and penalized weighted least-squares (PWLSs) algorithms [6], offer considerable convenience because of the well-known characteristics of raw data or log-transformed data before reconstruction of an image. However, it is difficult to avoid resolution loss in the spatial domain using these methods. In particular, high-frequency details, such as edges and textures, are not well preserved. The second category, iterative reconstruction methods, introduces the statistical properties of sinogram-domain data and prior information into the spatial domain to produce excellent outcomes. Several priors, such as total variation (TV) [7]–[11], PWLS [12], nonlocal means (NLM) [13]–[16], Markov random field (MRF) [17], low-rank [18], and dictionary learning priors [19], [20], have been formulated to perform iterative estimation on denoised images.
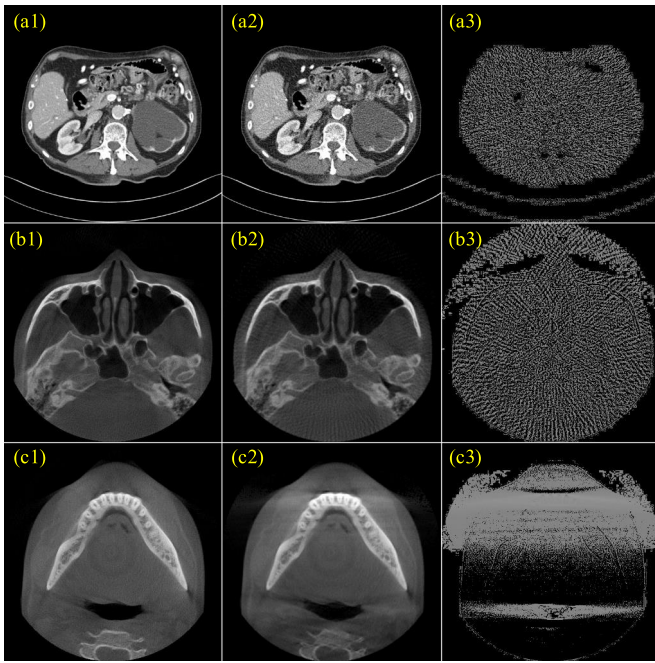
Fig. 1. Absolute differences between LDCT and NDCT using three dose-reduction approaches. The first column shows NDCT images, and the second column shows images for LDCT with current reduction, sparse view, and limited angle approaches. The third column shows the absolute differences between NDCT images and their LDCT counterparts.

Sparse transforms with compressed sensing [21] are often adopted in these iterative reconstruction methods. However, these iterative reconstruction methods still suffer from loss of detail and high computational cost. Moreover, approximating and parameterizing the priors are challenging. As the last category, postprocessing methods are directly applied to LDCT images without relying on raw data. Thus, these postprocessing methods could be conveniently integrated into any CT system. Traditional postprocessing methods can be divided into dictionary learning methods [22]–[25], NLM filtering methods [26], [27], and block-matching three-dimensional filtering (BM3D) methods [28], [29]. More recently, deep learning (DL) methods [30]–[51] have gradually become popular methods for LDCT. Because the noise in LDCT images may not obey a certain distribution, the traditional methods usually generate estimated results that contain noise and artifacts. Compared with these traditional methods, DL-based methods are more suitable for LDCT because of their excellent feature extraction capabilities based on end-to-end networks. In this article, we focus on introducing DL-based methods, especially encoder–decoder convolutional neural networks (CNNs).

In recent studies, many DL-based algorithms have achieved promising performance improvements in image restoration. Moreover, several DL-based methods, including basic CNNs [30], U-shaped networks [52], [53], wavelet networks [31], and residual encoder–decoder CNNs (RED-CNN) [32], have been employed for LDCT to obtain high-quality CT images from FBP images. There does not exist a clear rule to address the direct mapping relationship between NDCT images and LDCT images, which creates a challenging

problem for the reconstruction model. Recent studies have shown that deep neural networks can be used to efficiently solve this problem because of their excellent capacity for extracting features and representing the data distribution of the input examples. Therefore, the uncertain noise model can be disregarded when a neural network is employed. Many DL-based methods adopt an encoder–decoder structure due to its excellent performance in image denoising. However, the encoder–decoder structure may suffer from information loss because of the continual downsampling operations. Like RED-CNN, one exception is that skip connections can be used to avoid information loss [54]. In addition, revising the trained model is challenging: for example, we need to retrain a completely new encoder–decoder model instead of retraining the increased parts when introducing the encoder depth because the trained decoder may not fit the newly designed encoder.

In this article, we propose deep cascaded residual networks (DCRNs) to optimize the encoder–decoder CNNs for the estimation of a NDCT image from its low-dose counterpart. Based on the encoder–decoder structure, the continual downsampling operations have been replaced with paired up- and downsampling operations. Note that the readjusted sampling operations maintain the same parameter counts. Inspired by the attention mechanism [55], the enhanced version, named DCRN, is equipped with attention extraction in dual-sampling streams. For the sake of enhancing the high-frequency components and achieving a better visual effect, four loss functions (adversarial loss, mean absolute error, edge loss [54], [56], and perceptual loss [54], [57], [58]) are introduced for DCRN-GAN. Experiments involving the proposed method are conducted on three real clinical CT datasets, namely, dental CT data collected with a scanner manufactured by Zhongke Tianyue Company (ZTC), data from the American Association of Physicists in Medicine (AAPM) Challenge and clinical data collected with a commercial CT scanner from United Imaging Healthcare (UIH), which will be described in Section III-A. Furthermore, the quantitative and visual metrics are analyzed to evaluate performance.

This article is organized as follows: Section II introduces the optimization methods. Section III describes the experimental details, including three individual datasets. Section IV elaborates upon the experimental results with respect to visual and quantitative quality. Furthermore, the ablation studies are also conducted to validate the effectiveness of components in this work. Finally, we present a discussion and conclusion in Sections V and VI.

II. METHODS

In this section, optimization strategies conducted through adjusting the network structure and introducing multiple loss functions are described. Importantly, the upsample and downsample operations are elaborated. We then present details of the proposed cascaded residual modules and present an overall view of the network structure. Next, hybrid loss functions are introduced to improve the visual quality. Finally, parameter selection details are given. For the sake of easy distinction for the proposed models, the trained model with four hybrid loss
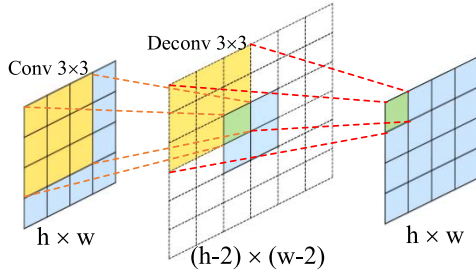
Fig. 2. Downsampling and upsampling operations on an image ($h \times w$) with convolution or deconvolution (3×3). The image size in the intermediate layer will decrease due to the zero-padding strategy, where the padding is fixed as 0.

functions is named DCRN-GAN and the training for DCRN is only constrained with mean absolute error.

A. Deep Cascaded Residual Network

Motivated by the RED-CNN, the upsampling and downsampling operations are introduced in our model. Different from an encoder–decoder structure, which includes strict rules of encoder before decoder, the sampling operations in our model can exist in any part of a network and thus offer more independence and flexibility. We treat the upsampling operation as an upsampling-based approach to enlarge an image; conversely, the downsampling operation shrinks the image size. Based on these two reciprocal operations, we can produce an output image with the same size as the input image in the local module in Fig. 2.

Let F_{in} denote the input image with dimensions of $M \times N \times C$, and let F_{out} denote the output image with dimensions of $M' \times N' \times C$. The upsampling operation is formulated as

$$F_{\text{out}} = Q_{\text{up}} * F_{\text{in}} \tag{1}$$

where "$*$" denotes the convolution operator and Q_{up} denotes the deconvolutional layers with filter size 3×3, stride 1, and zero-padding strategy, which are employed to increase the image size. Similarly, the downsampling operation is defined as

$$F_{\text{out}} = Q_{\text{down}} * F_{\text{in}} \tag{2}$$

where Q_{down} denotes the convolutional layers with the zero-padding strategy used to decrease the image size.

AEs are extensively applied in data manifold modeling [32], [40], [53]. Because they enable unsupervised feature learning for noisy inputs, AEs are suitable for image restoration. However, some high-frequency details, such as edges and textures, are lost in the estimated results due to the implementation of multiple downsampling operations [59]. To improve the preservation of details, we propose our optimized DCRN based on AEs. Inspired by the design of residual cascade blocks in ResNet [60], abundant features on multiple cascade modules [61] are extracted through paired-down and paired-up sampling operations, which mitigate the structural information loss. In Fig. 3, an overview of our proposed model structure is presented and M cascade residual modules (CRMs) are employed with several skip connections. For the generative network, we propose two

types of sampling blocks, namely, a down-to-up (DU) sampling block and an up-to-down (UD) sampling block. A CRM with attention extraction enhances the previous two blocks through a concatenation fusion technique. In terms of the sampling blocks, we realize the downsampling operation with a convolution operation and apply a deconvolutional layer for the upsampling operation. Furthermore, we set the strides of the convolutional and deconvolutional layers (filter size 3×3) to 1 with a zero-padding strategy inspired by RED-CNN. Thus, the image size in our modules is only changed twice, instead of the continual reduction in RED-CNN referenced to the depth of encoder and decoder, which alleviates the loss of structural details and reconstruction difficulty caused by the nearly continuous downsampling or upsampling operations.

For the tth CRM, the output of the DU block can be represented as

$$B_{\text{DU}}^t = P_{\text{up}} * \left(P_{\text{down}} * H^{t-1} \right) \tag{3}$$

where H^{t-1} denotes the input of the tth CRM. P_{down} denotes the downsampling operation and P_{up} denotes the upsampling operation. Inspired by the attention mechanism, we introduce the attention extraction for the DU block, which is represented as

$$A_{\text{DU}}^t = B_{\text{DU}}^t \otimes \left(A * B_{\text{DU}}^t \right) \tag{4}$$

where "$\otimes$" denotes the elementwise product. A denotes the convolution layer with the *Sigmoid* activation function used to extract spatial attention maps.

Similar to the DU block, the output of the UD block can be represented as

$$B_{\text{UD}}^t = P_{\text{down}} * \left(P_{\text{up}} * H^{t-1} \right). \tag{5}$$

Then, the attention features for the UD block can be represented as

$$A_{\text{UD}}^t = B_{\text{UD}}^t \otimes \left(B * B_{\text{UD}}^t \right) \tag{6}$$

where B denotes the convolution layer with the *Sigmoid* activation function used to extract spatial attention maps.

Last, the attention features of the DU and UD blocks as well as the original input of the CRM are combined to enhance the diversity of information, which is represented as

$$H^t = F * \left(\text{Concat}\left(A_{\text{DU}}^t, A_{\text{UD}}^t, H^{t-1} \right) \right) + H^{t-1} \tag{7}$$

where F denotes a convolutional layer and Concat denotes the concatenation function in the tth cascade module.

B. Hybrid Loss Functions

Several works [36], [50], [54], [56], [58] have achieved better performance through integration of multiple objective functions. Inspired by these works, multiple loss functions are considered for introduction into the optimized model. Similar to these methods, our proposed network adopts four objective loss functions, including mean absolute error, edge loss, perceptual loss, and adversarial loss, to enhance the high-frequency components and achieve a better visual effect.
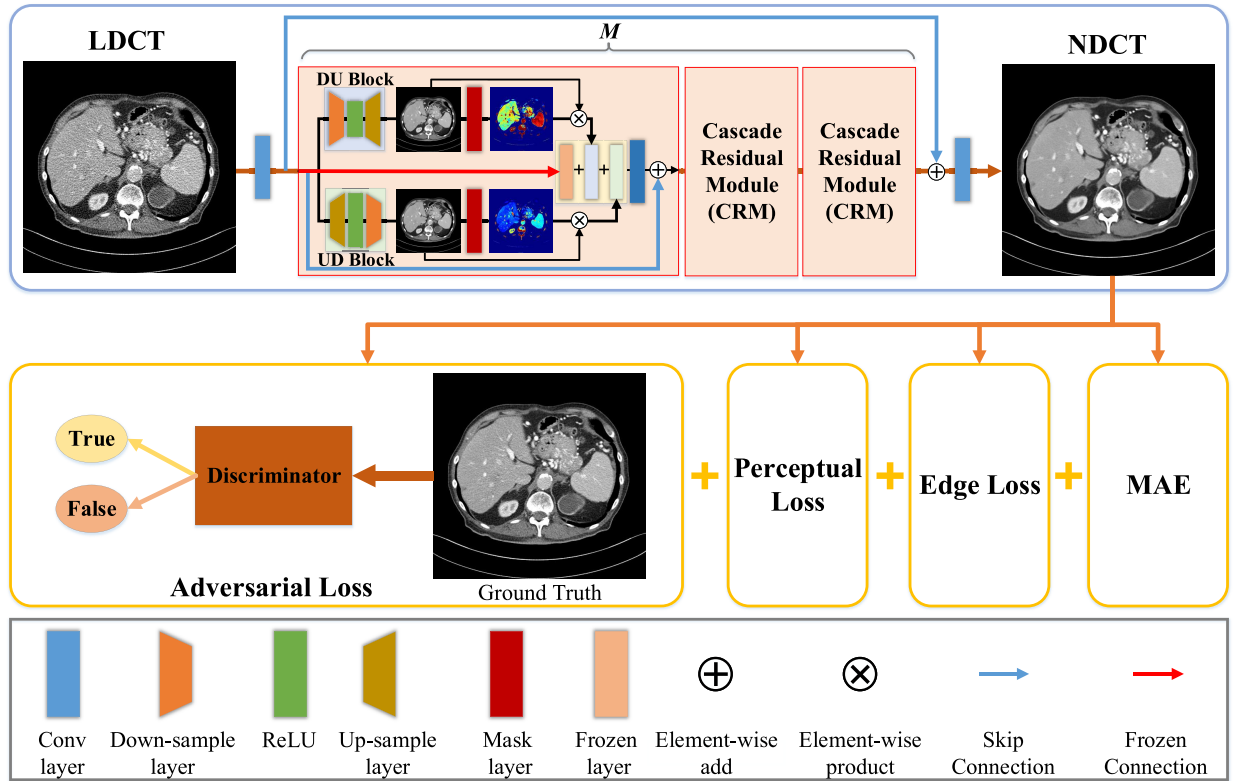
Fig. 3.　Overview of the framework for the proposed model. The mask layer denotes the attention extraction operation. The frozen connection denotes the direct concatenation for input feature maps. M denotes the total number of CRMs.

Let x denote an NDCT image, and let y denote its LDCT counterpart. The goal of the noise and artifact reduction process is to seek the inverse mapping function G between x and y. Our proposed method is based on an end-to-end network that directly maps NDCT images $x = \{x_1, x_2, \ldots, x_n\}$ from their LDCT counterparts $y = \{y_1, y_2, \ldots, y_n\}$. The mapping function is obtained by minimizing the loss between the estimated images and ground-truth images. To learn more high-frequency details, we utilize the MAE, which is also referred to as the $L1$-norm, as our loss function

$$L_{\mathrm{MAE}}(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \|G(y_i; \Theta) - x_i\|_1 \qquad (8)$$

where Θ denotes the network parameters for the end-to-end mapping function G. To enhance important high-frequency structural details, we calculate the high-frequency parts of estimated images through the *Sobel* detection template in horizontal and vertical directions shown in Fig. 4. The high-frequency parts could be formulated as

$$\Phi(x) = \sqrt{(S_h * x)^2 + (S_v * x)^2}. \qquad (9)$$

To enhance the edge structural information, we introduce the edge loss, which is formulated as

$$L_E(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \|\Phi(G(y_i; \Theta)) - \Phi(x_i)\|_2 \qquad (10)$$

where Φ denotes the edge detection function, for which we apply the *Sobel* detector template.

Fig. 4.　Sobel detection template in horizontal and vertical directions.

To improve the visual effects, perceptual loss is also introduced and formulated as

$$L_P(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \|\Psi(G(y_i; \Theta)) - \Psi(x_i)\|_2 \qquad (11)$$

where Ψ denotes the perceptual function, for which we apply VGG19 [62] to obtain the perceptual results. We extract the feature outputs of the 16th convolutional layer of VGG19 to measure the distance between the estimated CT images and referenced images.

As is shown in Fig. 5, the discriminative network consists of several convolution layers and channel-pooling operations, which yield the distinguishable answers for the predicted results or ground truth. In terms of the adversarial WGAN loss, an extra Wasserstein distance as well as a gradient penalty are engaged, which is formulated as

$$L_{\mathrm{WGAN}}(\Theta) = -E_x(D(x)) + E_y(D(G(y)))$$
$$+ \lambda E_{\hat{x}}\big[\big(\nabla_{\hat{x}} D(\hat{x}) - 1\big)\big]^2 \qquad (12)$$
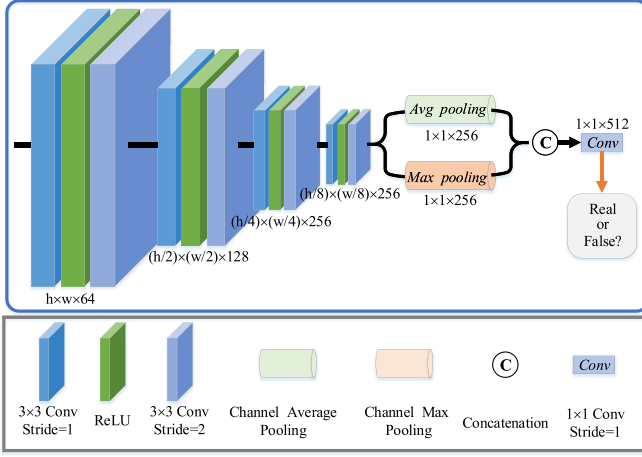
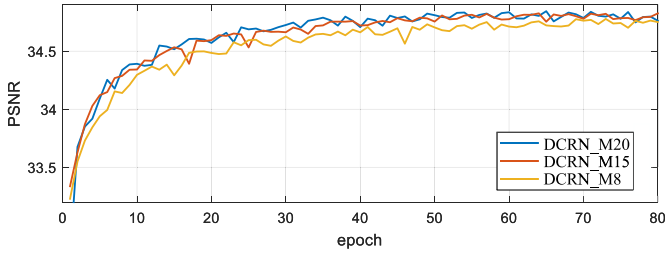Fig. 5. Discriminative network for the proposed model.

Fig. 6. Parameter selection on the number of CRMs.

where E denote the expectation operation; λ denotes the regularization parameter; and $\hat{x}$ denotes the estimated results and D denotes the discriminator networks.

Thereby, the total loss function is formulated as

$$L_T(\Theta) = L_{\text{WGAN}}(\Theta) + \alpha L_{\text{MAE}}(\Theta) + \beta L_E(\Theta)$$
$$+ \eta L_P(\Theta) \qquad (13)$$

where α, β, and η are weighting factors.

C. Parameter Selection

To achieve a tradeoff with respect to the performance, computation cost and complexity of our model, the number of cascaded residual modules is fixed as 15, when we compare the peak signal-to-noise ratio (PSNR) trend as for 8, 15 and 20 CRMs on UIH validation datasets during the training process in Fig. 6. For the convolution layers in the proposed model, filter size 3×3 and 64 channels are employed to obtain feature maps and parameter configuration of CRM is shown in Table I. In terms of multiple loss functions, the performance comparison is explored with different weight factor configurations in UIH datasets, and then the best parameter settings are chosen in Table II. For different tasks, weighting factors are fixed as shown in Table III. Similar to several works [32], [34], [35], [41], [42], paired student t-tests were used to quantify the significance of any discrepancies in terms of PSNR, structural similarity index measure (SSIM), and NMSE. If the p-value of t-test is less than 0.05, there exists a significant difference and no differences were found between the quality of the referenced and generated images.

TABLE I
PARAMETER CONFIGURATION FOR CRM

Component	Kernel size	Padding
Down-sampling layer	$3 \times 3 \times 64 \times 64$	0
Up-sampling layer	$3 \times 3 \times 64 \times 64$	0
Attention layer	$3 \times 3 \times 64 \times 64$	1
Fusion layer	$3 \times 3 \times 192 \times 64$	1

TABLE II
QUANTITATIVE RESULTS BASED ON UIH VALIDATION DATASETS FOR
THE PROPOSED MODEL WITH DIFFERENT BALANCE FACTOR
CONFIGURATIONS

Balance Factors	PSNR	SSIM
$\alpha = 0.01, \beta = 0.001, \eta = 0.01$	$39.62 \pm 1.04^*$	$0.9763 \pm 0.0057^*$
$\alpha = 0.01, \beta = 0.001, \eta = 0.1$	$39.93 \pm 1.12^*$	$0.9759 \pm 0.0058^*$
$\alpha = 0.01, \beta = 0.01, \eta = 0.1$	$39.23 \pm 0.97^*$	$0.9731 \pm 0.0061^*$
$\alpha = 0.1, \beta = 0.01, \eta = 0.1$	$39.99 \pm 1.06^*$	$0.9763 \pm 0.0057^*$
$\alpha = 0.1, \beta = 0.001, \eta = 0.01$	$39.61 \pm 1.01^*$	$0.9740 \pm 0.0058^*$
$\alpha = 0.1, \beta = 0.01, \eta = 0.01$	$39.77 \pm 1.06^*$	$0.9737 \pm 0.0060^*$
$\alpha = 0.1, \beta = 0.001, \eta = 0.1$	$\mathbf{40.24 \pm 1.05}^*$	$\mathbf{0.9776 \pm 0.0055}^*$

* denotes $P<0.05$, corresponding to a significant difference.

TABLE III
BALANCE FACTOR CONFIGURATION FOR DIFFERENT TASKS

Balance Factors	ZTC	AAPM	UIH
α	0.1	0.1	0.1
β	0.001	0.01	0.001
η	0.1	0.1	0.1

III. EXPERIMENTS

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

A. Patient Data Studies

Here, we describe the three datasets used to evaluate the effectiveness of the proposed method. The LDCT images in these datasets were obtained based on three different dose-reduction approaches: 1) decreasing the operating current; 2) sparse views; and 3) limited angles. The scanning geometry parameters of the routine dose protocols are shown in Table IV.

1) ZTC Dental Data: We employed a commercial dental CT scanner (model ZCB-100, manufactured by Zhongke Tianyue Technology Company Ltd., Shenzhen, Guangdong, China) to obtain raw sinogram data in realistic settings. We tested 42 volunteers (26 male and 18 female) aged between 22 years old and 45 years old in October 2017 and December 2017. We collected CT projections in the full-angle, normal-dose scan mode. Instead of scanning the patients twice, corresponding low-dose sinogram data were obtained by

TABLE IV
SCANNING GEOMETRY PARAMETERS FOR DIFFERENT DATASETS

Parameter	ZTC	AAPM	UIH
Source-to-detector distance	670 mm	570 mm	570 mm
Source-to-object distance	450 mm	106.23 cm	106.02 cm
Tube voltage	110 kV	120 kV	120 kV
Tube current	7.6 mA	200-500 mA	253.05 mA
Detector element size	$1.536 \times 1.920 mm^2$	$1.2856 \times 1.0947 mm^2$	$0.9150 \times 0.9150 mm^2$
Slice thickness	0.3 mm	1.0 mm	0.5 mm
Matrix size	512×512	512×512	512×512
Pixel spacing	$0.300 \times 0.300 mm^2$	$0.781 \times 0.781 mm^2$	$0.443 \times 0.443 mm^2$

choosing and retaining only certain projection data from certain angle intervals from the normal-dose sinogram data. Two sets of low-dose data were chosen to validate the performance of the proposed method. For the first set of low-dose data, we acquired 360 views from 180°, which corresponds to a limited-angle scan. For the uniformly sparsely sampled low-dose data, we acquired 180 views from 360° and apply the FDP reconstruction algorithm.

2) AAPM Challenge Data: We applied CT data from ten patients, which were provided as part of the "2016 AAPM LDCT Grand Challenge." Quarter-dose CT data and Poisson noise injected into the sinogram data were provided and the FDK reconstruction algorithm is employed. For the normal-dose data, more than 2300 projections were acquired.

3) UIH Clinical Data: We employed data collected in realistic settings using a commercial CT scanner from UIH (UIH, model uCT-760). For the normal-dose data, 1 800 projections were acquired. Low-dose data were obtained by sparse sampling of the normal-dose data to avoid scanning patients twice. In all, 150 projection views were selected to reconstruct LDCT images with the FDK algorithm, which corresponds to a ratio of 8.33% relative to the routine dose. We collected patient data corresponding to the head anatomical site from 20 patients.

B. Training Implementation Details

Data augmentation techniques were employed to obtain the training data, such as randomly rotating by 90°, 180°, and 270° and flipping in the horizontal and vertical directions, to improve the robustness of the trained model. The number of training samples for ZTC, AAPM and UIH datasets are 700, 1600, and 500, respectively. Instead of inputting the complete images into the network, we randomly cut the LDCT images into several patches with an image size of 48×48. The batch size was set to 16. We trained 1000 epochs for the end-to-end model, DCRN, on each dataset. The initial learning rate was 0.0001, which was halved every 200 epochs for the three training datasets. While the initial learning rate was fixed as 0.00001 for adversarial model DCRN-GAN after pretrained DCRN. The training time for ZTC, AAPM, and UIH datasets are about 34, 45, and 24 h, respectively. We minimized the loss function via the ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We applied the PyTorch framework to implement our model with Ubuntu 16.04 with a Titan 2080 Ti GPU.

C. Evaluation Metrics

To evaluate the performance of the proposed method, CNN [30] and RED-CNN [32] were utilized as comparison methods. For fair comparison, CNN and RED-CNN were trained with MAE and the convolution layers were as consistent as possible with the proposed model. Because the proposed method is an optimized version of the encoder–decoder network, we mainly focus on the performance comparison with RED-CNN. To quantitatively assess the image quality, we employed two popular metrics including the PSNR and SSIM.

To quantify the visual effect, we introduce the subjective evaluation from radiologists and metric perceptual index (PI) [63]. PI reflects the visual quality and has been a key evaluation index for the "2018 PIRM Challenge on Perceptual Image Super-Resolution" [64]. The lower PI corresponds with higher perceptual quality. The quality measurement for PI will be quantified as follows:

$$PI = \frac{10 - Ma}{2} + \frac{\text{NIQE}}{2} \qquad (14)$$

where *Ma* and *NIQE* denote the no-reference image quality measures of Ma [65] and NIQE [66].

IV. RESULTS

In this section, we evaluate the performance of the proposed DCRNs on three individual datasets. The statistical properties of PSNR and SSIM as evaluation metrics are employed to calculate quantitative values of different methods based on three datasets.

A. Visual Results

We trained and tested our model using the dental data from the ZTC dataset, which is subjected to limited angles and sparse sampling. As shown in Fig. 7, regions of interests (ROIs) are selected from the estimated images and ground truth for two patient cases. The observed patency of the respiratory tract is related to dental illness. For patient case 1, the proposed DCRN and DCRN-GAN can clearly preserve the curve of the respiratory tract. Conversely, the respiratory tracts are blurred in the results generated by RED-CNN and CNN. Besides, the results generated by the proposed methods appear to be closer to the ground truth for patient case 2.
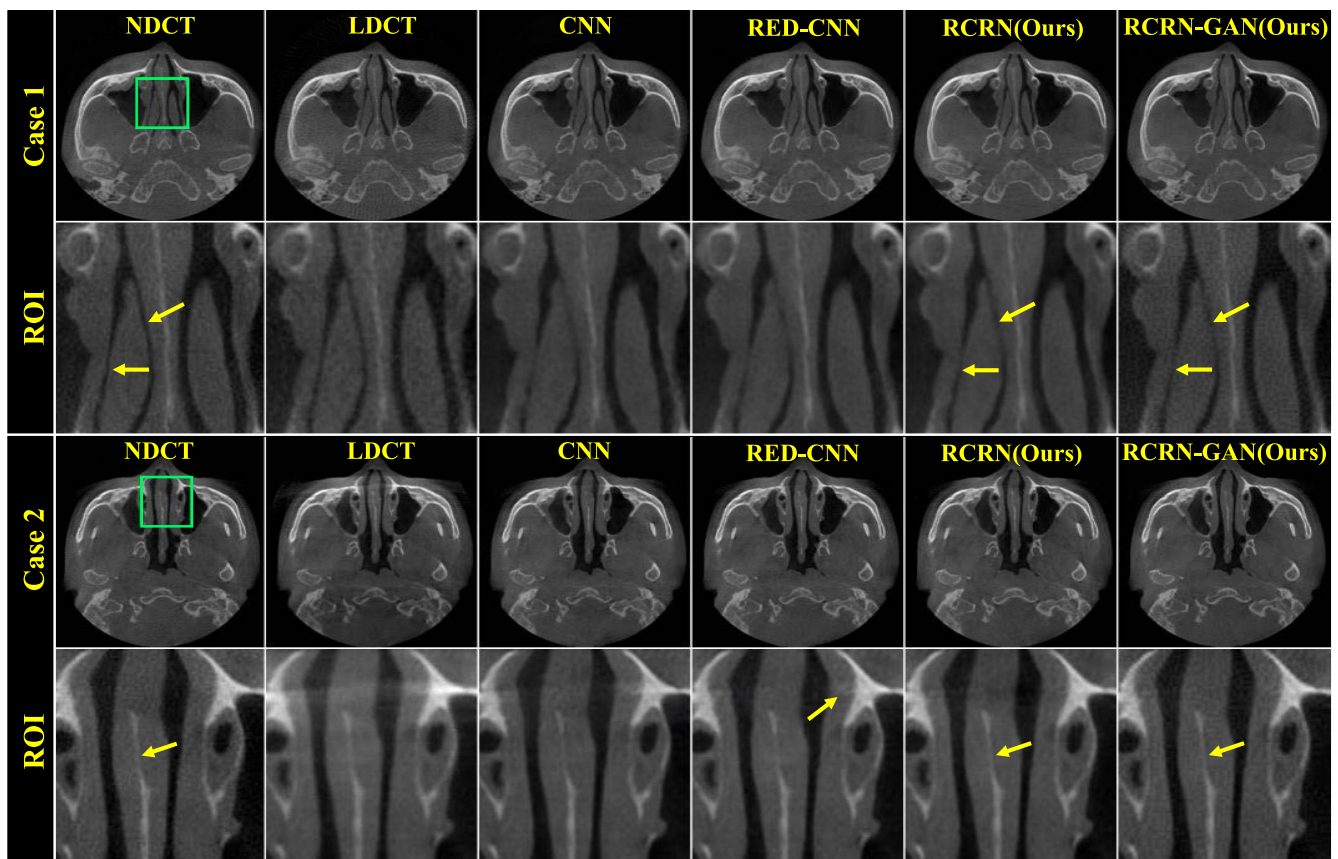
Fig. 7. Results and magnified views of dental data from the ZTC dataset for comparison. The green boxes demarcate the ROIs.
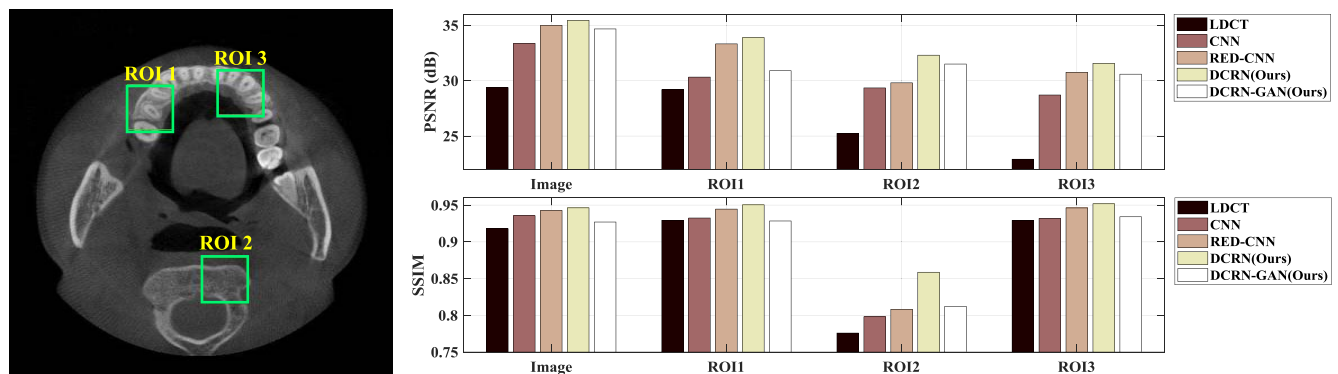

Fig. 8. Statistical results of dental data from the ZTC dataset for comparison. The green boxes demarcate the ROIs.

Quantitative results in terms of the PSNR and SSIM are calculated for the selected ROIs of the estimated images produced by different methods in Fig. 8. Compared with the other methods, our DCRN shows superior results on PSNR and SSIM. In particular, the proposed DCRN exhibits excellent performance in ROI2 in terms of PSNR and SSIM.

Overall comparison results obtained on the AAPM Challenge dataset and magnified views of the results are presented in Fig. 9. Compared with CNN and RED-CNN, the DCRN and DCRN-GAN preserve greater amounts of detail, as denoted by yellow arrows. For instance, the edge in DCRN and DCRN-GAN appears to be closer to the referenced NDCT image. DCRN-GAN generates more textural

structural details rather than smooth parts because of the use of multiple loss functions. In addition, we select three ROIs, including a bone region, a high-frequency region, and a smooth region, as shown in Fig. 10. The quantitative evaluation metrics calculated for these ROIs demonstrate the superiority of the proposed DCRN.

Visual results for the UIH data are shown in Fig. 11. The particle line is more obvious in the results obtained for the DCRNs than in results generated by RED-CNN. Furthermore, the quantitative results also show the superiority of the proposed DCRNs. Specially, DCRN-GAN achieves over 1.1-dB improvement and DCRN obtains over 1.3-dB improvement for the whole image on PSNR than
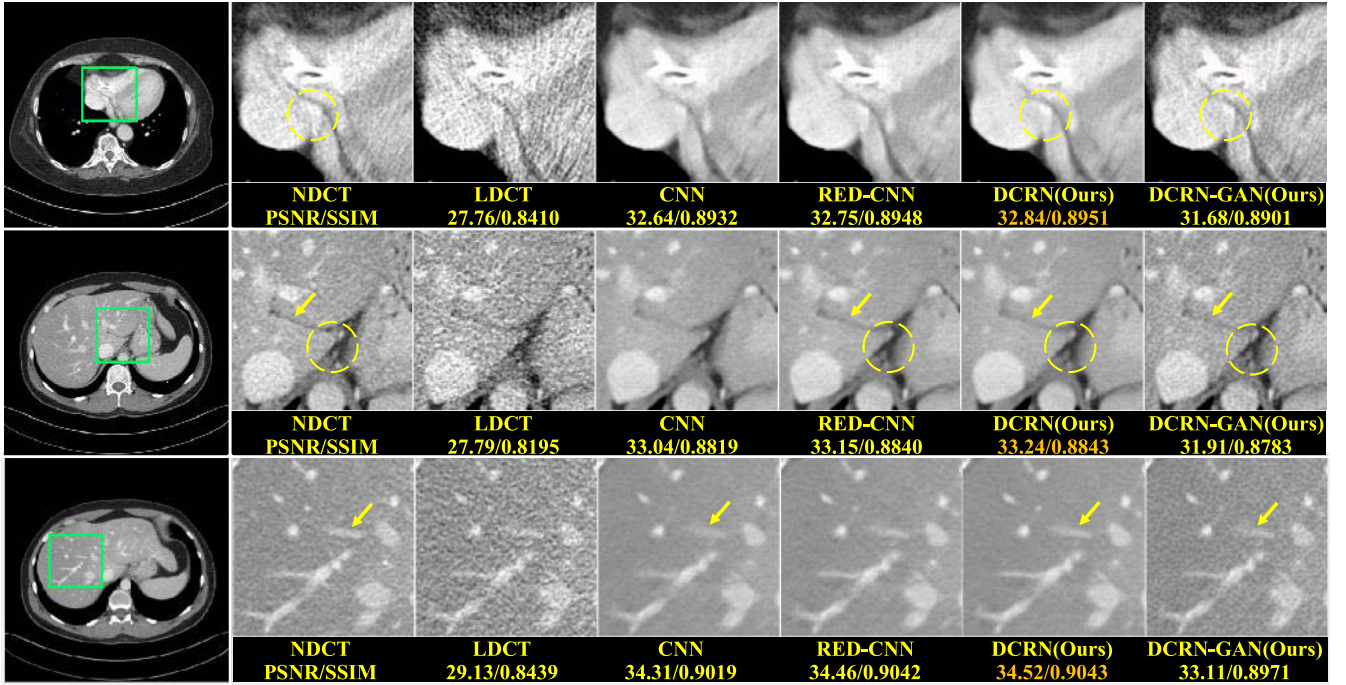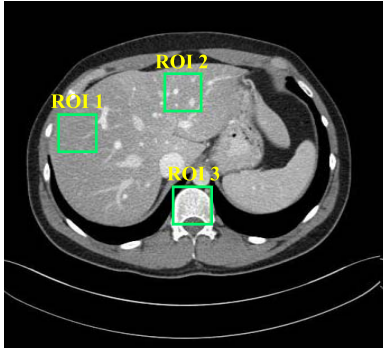
	NDCT PSNR/SSIM	LDCT 27.76/0.8410	CNN 32.64/0.8932	RED-CNN 32.75/0.8948	DCRN(Ours) 32.84/0.8951	DCRN-GAN(Ours) 31.68/0.8901

Fig. 9. Results obtained based on the AAPM Challenge dataset for comparison. The green box demarcates the ROI.

Regions	Metrics	LDCT	CNN	RED-CNN	DCRN(Ours)	DCRN-GAN(Ours)
Image	PSNR	28.27	33.49	33.61	**33.70**	32.31
	SSIM	0.8258	0.8875	0.8897	**0.8899**	0.8831
ROI1	PSNR	23.91	29.94	29.98	**29.99**	28.28
	SSIM	0.4334	0.5997	0.6013	**0.6073**	0.6000
ROI2	PSNR	24.35	27.43	27.58	**27.88**	27.10
	SSIM	0.7017	0.6970	0.7076	0.7380	**0.7646**
ROI3	PSNR	23.38	28.53	28.47	**28.79**	27.39
	SSIM	0.4716	0.6373	0.6390	**0.6486**	0.6281

Fig. 10. Results obtained on the AAPM Challenge dataset for comparison. The green box demarcates the ROI.

TABLE V
QUANTITATIVE RESULTS (MEAN ± STD) FOR THE THREE TEST DATASETS IN TERMS OF THE PSNR AND SSIM.
THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD

Datasets	Metrics	LDCT	CNN	RED-CNN	DCRN (ours)	DCRN-GAN (ours)
ZTC	PSNR	32.94 ± 3.58	$35.35 \pm 1.56^*$	$35.90 \pm 1.65^*$	$\mathbf{37.01 \pm 1.84}^*$	$35.41 \pm 1.53^*$
	SSIM	0.8914 ± 0.0258	$0.9080 \pm 0.0230^*$	$0.9122 \pm 0.0230^*$	$\mathbf{0.9164 \pm 0.0235}^*$	$0.8870 \pm 0.0311^*$
AAPM	PSNR	28.64 ± 2.12	$33.33 \pm 1.46^*$	$33.46 \pm 1.81^*$	$\mathbf{33.60 \pm 1.82}^*$	$32.32 \pm 1.81^*$
	SSIM	0.8157 ± 0.0553	$0.8894 \pm 0.0378^*$	$0.8912 \pm 0.0376^*$	$\mathbf{0.8920 \pm 0.0378}^*$	$0.8819 \pm 0.0391^*$
UIH	PSNR	27.74 ± 1.07	$39.54 \pm 1.03^*$	$39.61 \pm 1.10^*$	$\mathbf{40.51 \pm 1.11}^*$	$40.24 \pm 1.04^*$
	SSIM	0.6074 ± 0.0563	$0.9754 \pm 0.0058^*$	$0.9750 \pm 0.0061^*$	$\mathbf{0.9787 \pm 0.0054}^*$	$0.9776 \pm 0.0055^*$

* denotes $P<0.05$, corresponding to a significant difference.

RED-CNN. Meanwhile, SSIM scores of proposed methods are best.

B. Quantitative Results

In Table V, we present the quantitative results obtained for all LDCT images with three test datasets that were not included among the training datasets. The test datasets for the ZTC, AAPM Challenge, and UIH datasets contain 100, 200, and 100 LDCT images, respectively. These quantitative results illustrate the superior performance of our proposed DCRN. The DCRN gains improvement of over 1.2 dB in terms of PSNR versus RED-CNN based on the ZTC test datasets. In
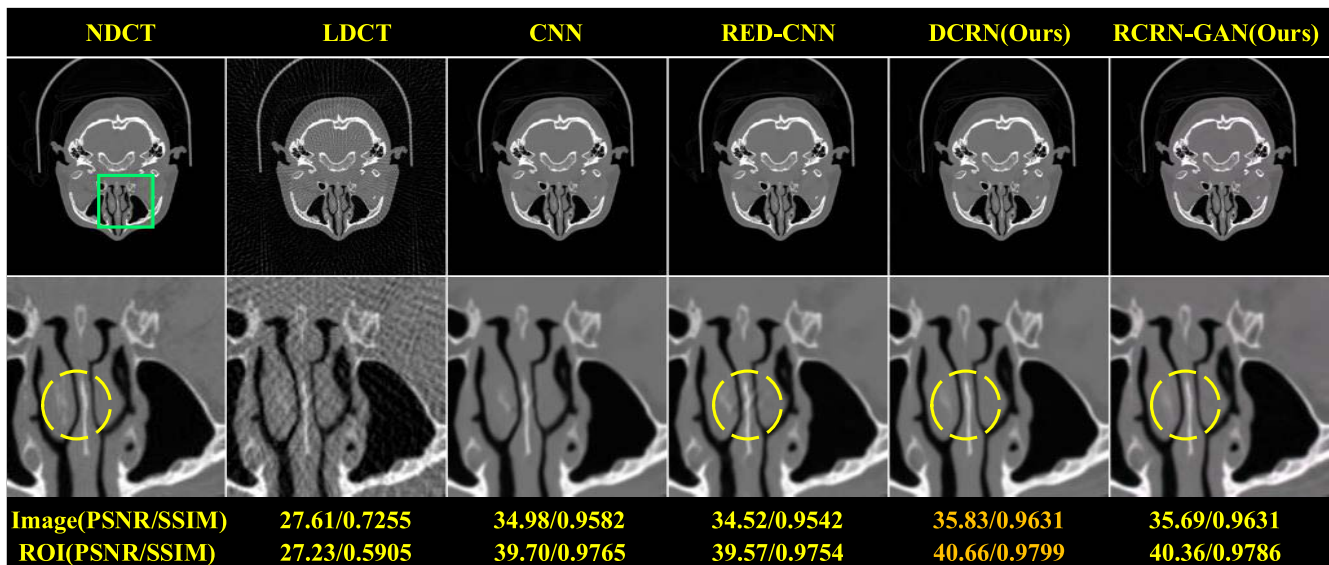
Fig. 11. Results obtained based on the UIH dataset for comparison. The green box demarcates the ROI.

TABLE VI
QUANTITATIVE PI RESULTS FOR DIFFERENT METHODS. THE LOWER PI CORRESPONDS WITH HIGHER PERCEPTUAL QUALITY, AND THE BEST SCORES ARE HIGHLIGHTED IN BOLD

Models	UIH	AAPM	ZTC
CNN	5.7407	3.7920	5.1414
RED-CNN	5.2852	3.6739	4.7399
DCRN(ours)	5.2629	3.6707	4.8678
DCRN-GAN(ours)	**4.5647**	**3.4770**	**4.0295**

TABLE VII
SUBJECTIVE RESULTS (MEAN ± STD) OF DIFFERENT METHODS ON THREE DATASETS. THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD

Methods	CNN	RED-CNN	DCRN	DCRN-GAN
ZTC	7.00 ± 0.56	7.30 ± 0.47	7.85 ± 0.49	**8.25 ± 0.44**
AAPM	6.90 ± 0.55	7.30 ± 0.47	8.00 ± 0.46	**8.25 ± 0.44**
UIH	7.05 ± 0.22	7.50 ± 0.47	8.00 ± 0.32	**8.45 ± 0.51**

TABLE VIII
QUANTITATIVE RESULTS (MEAN ± STD) FOR DCRN-DU (ONLY DU SAMPLING BLOCK), DCRN-UD (ONLY UD SAMPLING BLOCK), AND DCRN ON THE THREE TEST DATASETS IN TERMS OF THE PSNR AND SSIM. THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD

Datasets	Metrics	DCRN-DU	DCRN-UD	DCRN
ZTC	PSNR	36.43 ± 1.76*	36.75 ± 1.81*	**37.01 ± 1.84***
	SSIM	0.9138±0.0230*	0.9146±0.0232*	**0.9164±0.0235***
AAPM	PSNR	33.54 ± 1.84*	33.54 ± 1.80*	**33.60 ± 1.82***
	SSIM	0.8917±0.0378*	0.8913±0.0377*	**0.8920±0.0378***
UIH	PSNR	40.23 ± 1.12*	40.30 ± 1.13*	**40.51 ± 1.11***
	SSIM	0.9775±0.0056*	0.9776±0.0057*	**0.9787±0.0054***

* denotes $P<0.05$, corresponding to a significant difference.

addition, the DCRN achieves an over 0.9-dB improvement in PSNR versus RED-CNN based on UIH test datasets. As shown in Table VI, our DCRN-GAN with constraints from multiple loss functions achieves the lower perceptual values in terms of PI, which illustrates better visual effect.

To clearly illustrate the visual effect on medical images, subjective scores from radiologists are introduced to evaluate the visual quality. The subjective scores are provided by two radiologists with adequate clinical experience. Subjective scores ranging from 1 to 10 are assigned to 20 images selected from three test datasets. As is shown in Table VII, the subjective results also support the superiority of the proposed DCRN-GAN.

C. Ablation Analysis

In this part, ablation studies are conducted to validate the component effectiveness of the proposed model. The

optimization strategies, including the network structure design and multiple loss functions would be evaluated on PSNR and SSIM.

1) Effectiveness of Dual-Sampling Blocks: To validate the effectiveness of dual-sampling blocks, including DU and UD blocks, two version models: 1) DCRN-DU and 2) DCRN-UD are designed based on the proposed DCRN. In Table VIII, the quantitative results for DCRN achieve better PSNR and SSIM than comparison models, which demonstrates the effectiveness of dual-sampling blocks. Compared with a monotonous sampling block (DU or UD block), the integration of dual-sampling blocks provides performance improvement.

2) Effectiveness of Attention Extraction: In terms of DCRN, we validate the effectiveness of attention extraction and design the baseline model based on DCRN. In Table IX, the quantitative results for DCRN achieve better PSNR and SSIM than the baseline model, which demonstrates the effectiveness of attention introduction.

3) Effectiveness of Multiple Loss Functions: As for the proposed DCRN, we introduce four loss functions to improve the visual effect. For the sake of validation of loss functions, we compare the quantitative results for the DCRN-GAN with

TABLE IX
QUANTITATIVE RESULTS (MEAN ± STD) FOR DCRN BASELINE MODEL
(WITHOUT ATTENTION EXTRACTION) AND DCRN ON THE THREE TEST
DATASETS IN TERMS OF THE PSNR AND SSIM. THE HIGHEST SCORES
ARE HIGHLIGHTED IN BOLD

Datasets	Metrics	DCRN baseline	DCRN (Ours)
ZTC	PSNR	$36.69 \pm 1.72^*$	$\mathbf{37.01 \pm 1.84}^*$
	SSIM	$0.9148 \pm 0.0232^*$	$\mathbf{0.9164 \pm 0.0235}^*$
AAPM	PSNR	$33.53 \pm 1.82^*$	$\mathbf{33.60 \pm 1.82}^*$
	SSIM	$0.8920 \pm 0.0372^*$	$\mathbf{0.8920 \pm 0.0378}^*$
UIH	PSNR	$40.22 \pm 1.14^*$	$\mathbf{40.51 \pm 1.11}^*$
	SSIM	$0.9775 \pm 0.0058^*$	$\mathbf{0.9787 \pm 0.0054}^*$

* denotes $P<0.05$, corresponding to a significant difference.

TABLE X
QUANTITATIVE RESULTS (MEAN ± STD) FOR DCRN-GAN WITH
DIFFERENT LOSS FUNCTIONS ON THE UIH TEST DATASETS IN TERMS OF
THE PSNR AND SSIM. THE HIGHEST SCORES ARE HIGHLIGHTED IN
BOLD

Loss functions	PSNR	SSIM
WGAN	$27.90 \pm 0.52^*$	$0.9102 \pm 0.0134^*$
WGAN+M	$32.48 \pm 1.01^*$	$0.9401 \pm 0.0124^*$
WGAN+E	$39.45 \pm 1.00^*$	$0.9735 \pm 0.0058^*$
WGAN+P	$29.04 \pm 0.53^*$	$0.9224 \pm 0.0111^*$
WGAN+M+E	$39.98 \pm 1.08^*$	$0.9748 \pm 0.0061^*$
WGAN+M+P	$33.59 \pm 1.11^*$	$0.9332 \pm 0.0235^*$
WGAN+E+P	$39.57 \pm 1.01^*$	$0.9754 \pm 0.0057^*$
WGAN+M+E+P	$\mathbf{40.24 \pm 1.05}^*$	$\mathbf{0.9776 \pm 0.0055}^*$

* denotes $P<0.05$, corresponding to a significant difference.

different loss functions. In Table X, the DCRN-GAN equipped with four loss functions achieves the best quantitative results, which proves the effectiveness of total loss functions.

V. DISCUSSION

Inspired by the integration of multiple objective functions in several methods [36], [50], [54], [56], [58] to seek better performance, we consider introducing this strategy to optimized model. Multiple loss functions are employed to obtain a better visual effect. In addition to the adversarial loss function in the generative adversarial network (GAN)-based methods, mean square errors, edge loss, and perceptual loss are usually employed. Different from other methods, our proposed network adopts three objective loss functions, including mean absolute error, edge loss, and perceptual loss, to enhance the high-frequency components and achieve a better visual effect. Equipped with edge loss and perceptual loss, as well as adversarial loss, DCRN-GAN may suffer from lower PSNR or SSIM, but could improve the perceptual visual quality and reveal more details.

As for parameter counts, the proposed methods DCRN-DU and DCRN-UD have the same parameter counts compared to CNN and RED-CNN, and both DCRN-DU and DCRN-UD achieve better quantitative results in terms of PSNR and SSIM in Tables V and VIII. As the proposed model equipped with more parameters, DCRN has a DU block and a UD block for the sake of better performance, and the experimental results

prove that it could gain better quantitative values for PSNR and SSIM. The integration of dual-sampling blocks (DU and UD) provides performance improvement.

For the parameter configuration, we follow RED-CNN, set the strides of convolution and deconvolution operations (filter size 3×3) in up and down sampling units to 1 and adopt the zero-padding strategy. In addition, preserving the image size avoids information loss caused by the downsampling operations, which leads to the performance improvement. Due to our optimization strategies for an encoder–decoder network, we mainly evaluate the performance of our methods compared with RED-CNN. The proposed models exhibited superior performance compared with RED-CNN. Generally, the DCRN-DU, DCRN-UD, or enhanced DCRN are suggested to manage the task due to their convenient implementation. The DCRN-GAN model is suggested for the sake of better visual effect. What is more, achieving better visual effects necessitates further research to fit more loss functions, such as gradient difference loss [67].

VI. CONCLUSION

In this article, we optimized the popular encoder–decoder CNNs for LDCT imaging considering the information loss caused by continual downsampling operations. The optimization strategies are divided into two aspects. First, we adjusted the network structure by rearranging the down- and upsampling operations to avoid continual downsampling operations and decrease the information loss. The latter enhanced version, DCRN, maximizes the extraction of helpful information by combining the paired sampling blocks into one CRM with attention extraction. On the other hand, four loss functions (mean absolute error, edge loss, perceptual loss, and adversarial loss) were introduced to achieve superior visual perceptual quality during the training process. Through ablation studies, the components of the proposed model are validated to achieve performance improvement. The experimental results qualitatively and quantitatively showed the effectiveness of our methods.

ACKNOWLEDGMENT

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] J. Hsieh, "Adaptive streak artifact reduction in computed tomography resulting from excessive X-ray photon noise," *Med. Phys.*, vol. 25, no. 11, pp. 2139–2147, 1998.

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

[7] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.

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

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

[10] Y. Zhang, W.-H. Zhang, H. Chen, M.-L. Yang, T.-Y. Li, and J.-L. Zhou, "Few-view image reconstruction combining total variation and a high-order norm," *Int. J. Imag. Syst. Technol.*, vol. 23, no. 3, pp. 249–255, 2013.

[11] 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.

[12] 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, p. 5311, 2016.

[13] Y. Chen *et al.*, "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Comput. Med. Imag. Graph.*, vol. 33, no. 7, pp. 495–500, 2009.

[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] 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.

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

[17] 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, Jan. 2016.

[18] 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.

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

[20] 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.

[21] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[22] M. Katsura *et al.*, "Model-based iterative reconstruction technique for radiation dose reduction in chest CT: Comparison with the adaptive statistical iterative reconstruction technique," *Eur. Radiol.*, vol. 22, no. 8, pp. 1613–1623, 2012.

[23] Y. Chen *et al.*, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, no. 16, pp. 5803–5820, 2013.

[24] Y. Lu, J. Zhao, and G. Wang, "Few-view image reconstruction with dual dictionaries," *Phys. Med. Biol.*, vol. 57, no. 1, pp. 173–189, 2011.

[25] X.-Y. Cui, Z.-G. Gui, Q. Zhang, H. Shangguan, and A.-H. Wang, "Learning-based artifact removal via image decomposition for low-dose CT image processing," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 3, pp. 1860–1873, Jun. 2016.

[26] Z. Li *et al.*, "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Med. Phys.*, vol. 41, no. 1, 2014, Art. no. 11908.

[27] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Lu, and Z. Liang, "Statistical image reconstruction for low-dose CT using nonlocal means-based regularization," *Comput. Med. Imag. Graph.*, vol. 38, no. 6, pp. 423–435, 2014.

[28] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3D random noise filtering for absorption optical projection tomography," *Phys. Med. Biol.*, vol. 55, no. 18, pp. 5401–5415, 2010.

[29] D. Kang *et al.*, "Image denoising of low-radiation dose coronary ct angiography by an adaptive block-matching 3D algorithm," in *Proc. Med. Imag. Image Process.*, vol. 8669, 2013, Art. no. 86692G.

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

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

[32] 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.

[33] 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.

[34] 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.

[35] Z. Huang *et al.*, "DaNet: Dose-aware network embedded with dose-level estimation for low-dose CT imaging," *Phys. Med. Biol.*, vol. 66, no. 1, 2021, Art. no. 15005.

[36] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[37] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "Iterative low-dose CT reconstruction with priors trained by artificial neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2479–2486, Dec. 2017.

[38] Z. Huang *et al.*, "Considering anatomical prior information for low-dose ct image enhancement using attribute-augmented Wasserstein generative adversarial networks," *Neurocomputing*, vol. 428, pp. 104–115, Mar. 2021.

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

[40] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, pp. 4509–4522, 2017.

[41] Z. Huang *et al.*, "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," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 9, pp. 3416–3427, Sep. 2021.

[42] 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.

[43] F. Jiao, Z. Gui, Y. Liu, L. Yao, and P. Zhang, "Low-dose CT image denoising via frequency division and encoder-dual decoder GAN," *Signal Image Video Process.*, vol. 15, pp. 1907–1915, Jun. 2021.

[44] Z. Feng, A. Cai, Y. Wang, L. Li, L. Tong, and B. Yan, "Dual residual convolutional neural network (DRCNN) for low-dose CT imaging," *J. X-Ray Sci. Technol.*, vol. 29, no. 1, pp. 91–109, 2021.

[45] D. Zeng *et al.*, "Noise-generating-mechanism-driven unsupervised learning for low-dose CT sinogram recovery," *IEEE Trans. Radiat. Plasma Med. Sci.*, early access, May 25, 2021, doi: 10.1109/TRPMS.2021.3083361.

[46] Z. Li, S. Zhou, J. Huang, L. Yu, and M. Jin, "Investigation of low-dose CT image denoising using unpaired deep learning methods," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 2, pp. 224–234, Mar. 2021.

[47] A. M. Hasan, M. R. Mohebbian, K. A. Wahid, and P. Babyn, "Hybrid-collaborative Noise2Noise denoiser for low-dose CT images," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 2, pp. 235–244, Mar. 2021.

[48] M. Matsuura, J. Zhou, N. Akino, and Z. Yu, "Feature-aware deep-learning reconstruction for context-sensitive X-ray computed tomography," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 1, pp. 99–107, Jan. 2021.

[49] Q. Ding, Y. Nan, H. Gao, and H. Ji, "Deep learning with adaptive hyper-parameters for low-dose CT image reconstruction," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 648–660, Jun. 2021.

[50] 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.

[51] J. Liu, Y. Kang, J. Qiang, Y. Wang, D. Hu, and Y. Chen, "Low-dose CT imaging via cascaded ResUnet with spectrum loss," *Methods*, vol. 21, pp. 131–136, May 2021.

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

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

[54] H. Shan *et al.*, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nat. Mach. Intell.*, vol. 1, no. 6, pp. 269–276, 2019.

[55] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[56] H. Shan *et al.*, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[57] 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.

[58] Q. Yang, P. Yan, M. K. Kalra, and G. Wang, "CT image denoising with perceptive deep neural networks," 2017, *arXiv:1702.07019*.

[59] Y. Zhao, R.-G. Wang, W. Jia, W.-M. Wang, and W. Gao, "Iterative projection reconstruction for fast and efficient image upsampling," *Neurocomputing*, vol. 226, pp. 200–211, Feb. 2017.

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

[61] D. Wu, K. Kim, G. E. Fakhri, and Q. Li, "A cascaded convolutional neural network for X-ray low-dose CT image denoising," 2017, *arXiv:1705.04267*.

[62] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[63] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 6228–6237.

[64] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018.

[65] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Comput. Vis. Image Understand.*, vol. 158, pp. 1–16, May 2017.

[66] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[67] C.-B. Jin *et al.*, "DC^{2}Anet: Generating lumbar spine mr images from CT scan data based on semi-supervised learning," *Appl. Sci.*, vol. 9, no. 12, p. 2521, 2019.