

A Model-Based Unsupervised Deep Learning Method for Low-Dose CT Reconstruction

KAICHAO LIANG^{ID}, LI ZHANG, (Member, IEEE), HONGKAI YANG,
ZHIQIANG CHEN^{ID}, (Member, IEEE), AND
YUXIANG XING^{ID}, (Senior Member, IEEE)

Department of Engineering Physics, Tsinghua University, Beijing 100084, China

Key Laboratory of Particle and Radiation Imaging, Ministry of Education, Tsinghua University, Beijing 100084, China

Corresponding author: Yuxiang Xing (xingyx@mail.tsinghua.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771279 and Grant 61527807, and in part by the National Science Fund for Distinguished Young Scholars under Grant 11525521.

ABSTRACT Low-dose CT (LDCT) is of great significance due to the concern about the potential radiation risk. With the fast development of deep learning, neural networks have become powerful tools in LDCT enhancement. Current deep neural networks for LDCT reconstruction are often trained with paired LDCT dataset and normal-dose CT (NDCT) dataset. However, high quality NDCT dataset paired with LDCT dataset is expensive to acquire or even not available sometimes in reality. In this work, we proposed an unsupervised model-based deep learning (MBDL) for LDCT reconstruction. The network is trained based on group-wise maximum a posteriori (G-MAP) loss function with LDCT dataset only. The MBDL is a general framework. It also allows us to combine with supervised training if a small number of paired NDCT dataset accessible to help optimizing the network parameters, i.e. works in a semi-supervised mode. During inference, LDCT images are reconstructed end-to-end by the trained network. We verified the proposed method with simulated projection data from clinical CT images. The proposed method restrained noise well while restoring anatomical structures and it achieved better results than model-based iterative reconstruction (MBIR) with significantly less computational cost. The performances of MBDL were further enhanced by integrating a small paired NDCT dataset for semi-supervised training. The results suggested that MBDL is an efficient and flexible method for LDCT deep learning based reconstruction in the situations lacking of enough high quality NDCT data.

INDEX TERMS Low-dose CT, deep learning, unsupervised learning, semi-supervised learning.

I. INTRODUCTION

X-ray CT is a widely used medical imaging modality for clinical diagnosis while its radiation dose is highly concerned. High level dose received by patients raises potential radiation risk, such as an increasing possibility of cancer [1], [2]. In the past decades, lowering CT dose by adjusting tube-current-time or tube-voltage have been studied extensively [3]–[5]. However, further reducing scanning dose leads to higher photon statistical noise and the electronic noise becomes unneglectable as well. If directly reconstructed with analytical reconstruction algorithms, the noise in projection data will be magnified by the high-pass filtration (RL filter or SL filter) and degrade the image quality. Lowering dose to the lowest

level while maintaining an acceptable image quality remains a challenge in CT reconstruction area.

Previously, various researches have been conducted on low-dose CT (LDCT) optimization. These algorithms generally can be classified into three categories. The first category is analytical filtration method. Manually designed filters or filtration algorithms are adopted to restrain noise for LDCT in either projection domain or image domain denoted as pre-processing and post-processing [6]–[9]. Generally, filterers are implemented with structural-adaptive strategies thus the information of clinical structures could be preserved to a large extent [6], [7]. Normal-dose CT (NDCT) images can also be integrated into filtering process to enhance image denoising performance [9]. However, manually designed low-pass process often has limited ability in distinguishing noise and detail information so that causes image blurring in practice. The second category is model-based iterative

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

reconstruction (MBIR) methods based on Bayesian theory. They model statistical noise, image priors and system physics into reconstruction thus better reconstruction results are obtained. For example, introducing manually designed image priors as constraints to statistical iterative reconstruction (SIR) for LDCT reconstruction is commonly adopted to restrain noise and artifacts [10]–[14]. Markov random field (MRF) reduces image noise by measuring the distribution of neighboring pixels [14]. Based on compressed sensing (CS) theory, total-variation (TV) minimization makes use of the sparsity of NDCT images in variation domain and has an advantage in removing LDCT streak artifacts [12]. Instead of using manually designed priors, dictionary learning learns a data-adaptive sparsity constraint based on NDCT dataset [15], [16]. MBIR achieved significant improvements in LDCT reconstruction. However, iterative process is time consuming which hinders its usage in real-time applications. Besides, imprecise image priors might introduce new artifacts in reconstructions.

The third category is deep learning based LDCT reconstruction or LDCT image processing. Recently, deep learning methods have been widely conducted in LDCT problems and achieved significant breakthrough. Various types of convolutional neural networks (CNN) such as residual net [17]–[19], encoder-decoder [17], [18], [20], U-Net [21] were adopted in image domain for LDCT image denoising after reconstruction. The networks were trained with simulated LDCT and NDCT image pairs using L2-norm loss function. During inference, they produced images of better quality compared with MBIR both visually and quantitatively. The deep learning based denoising methods could also be implemented in projection domain [22]. Kang et. al. transformed the LDCT image denoising task to wavelet domain and developed a wavelet domain denoising network, thus the features of noise are easier to learn [23]. To overcome the over-smooth effect of networks, generative adversarial training as well as perceptual loss was integrated to train networks and the texture information was restored better [24], [25]. In recent works, deep learning was also adopted into iterative reconstruction frame to replace parts of iterative process such as the proximal operator or image priors for LDCT reconstruction [26]–[29], so that the inference results were reliable and precise with data fidelity guaranteed in inference. Despite of the powerful ability of deep learning, one common limitation in these deep learning approaches is the dependence on large high quality NDCT dataset paired with LDCT inputs. While in many real situations, large paired NDCT dataset are expensive or even impossible to acquire. Considering the training efficiency and large NDCT dataset demand, transfer learning has been investigated for LDCT deep learning in recent years [30]–[33]. Transfer learning makes use of pre-trained models in the source domain and generalizes them to new applications in the target domain considering the domain similarity. In situations lacking of large NDCT dataset, Meng et. al. trained a LDCT denoising network with 848 paired medium-dose CT image labels firstly. The pre-trained network combined with

an additional simple CNN network was further fine-tuned with 200 high-dose CT image labels [32]. Shan et. al. proposed a transfer learning framework for LDCT denoising based on basic CNN units. The CNN units pre-trained on Mayo LDCT dataset can be easily transferred to MGH LDCT dataset covering different LDCT dose-level by combining the CNN units linearly. During the transfer learning process, only one slice from each dose-level was used [33]. Transfer learning reduces the NDCT dataset demand, however most LDCT transfer learning methods still work in a supervised manner with ground-truth label needed.

In this work, we proposed a model-based deep learning (MBDL) method for LDCT reconstruction network training based on Bayesian theory. It requires no ground-truth information. The network can be directly trained with LDCT projection data only. In situations with a small paired NDCT dataset available, the MBDL can easily integrate the NDCT dataset and further improve reconstruction performance through semi-supervised training.

II. METHODS

In this section, all vectors and matrixes are denoted with bold symbols and variables are denoted with italics lower-case symbols.

A. MAXIMUM A POSTERIOR (MAP) ESTIMATION FOR LDCT

The primary concern of LDCT unsupervised training is to find an appropriate loss function without NDCT images. In MBDL, we use the idea of MAP estimation for network training. In this subsection, we briefly review the basic theory of MAP estimation for LDCT.

In monoenergetic CT imaging, the relationship between projection data and attenuation coefficient map can be modeled linearly as Eq. 1.

$$\mathbf{p} = \mathbf{H}_{\text{sys}}\boldsymbol{\mu} + \mathbf{N}, \quad \text{s.t. } \boldsymbol{\mu} \in \mathcal{U} \quad (1)$$

Here, $\mathbf{p} = \{p_1, p_2, \dots, p_M\} \in \mathbb{R}^M$ denotes the acquired projection data along all the X-ray paths, $\mathbf{H}_{\text{sys}} \in \mathbb{R}^{M \times N}$ the system matrix measuring the projection integral contribution from each pixel in the attenuation map, $\boldsymbol{\mu} \in \mathbb{R}^N$ the attenuation coefficient vector at each pixel in discrete model, \mathbf{N} the zero-mean noise, \mathcal{U} the feasible space of natural objects. X-ray photons pass through scanned objects randomly in approximate Poisson distribution and the fixed level electronic noise also influences the acquired signal [34], [35].

The essential idea of MAP estimation is to reconstruct images with the maximum posterior observing projection data \mathbf{p} [14], [36].

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu} \geq 0}{\operatorname{argmax}} \operatorname{Prob}(\boldsymbol{\mu}|\mathbf{p}) = \underset{\boldsymbol{\mu} \geq 0}{\operatorname{argmax}} \frac{\operatorname{Prob}(\mathbf{p}|\boldsymbol{\mu}) \operatorname{Prob}(\boldsymbol{\mu})}{\operatorname{Prob}(\mathbf{p})} \quad (2)$$

With negative-log-transformation, the posterior probability is transformed into an additive format loss function with the

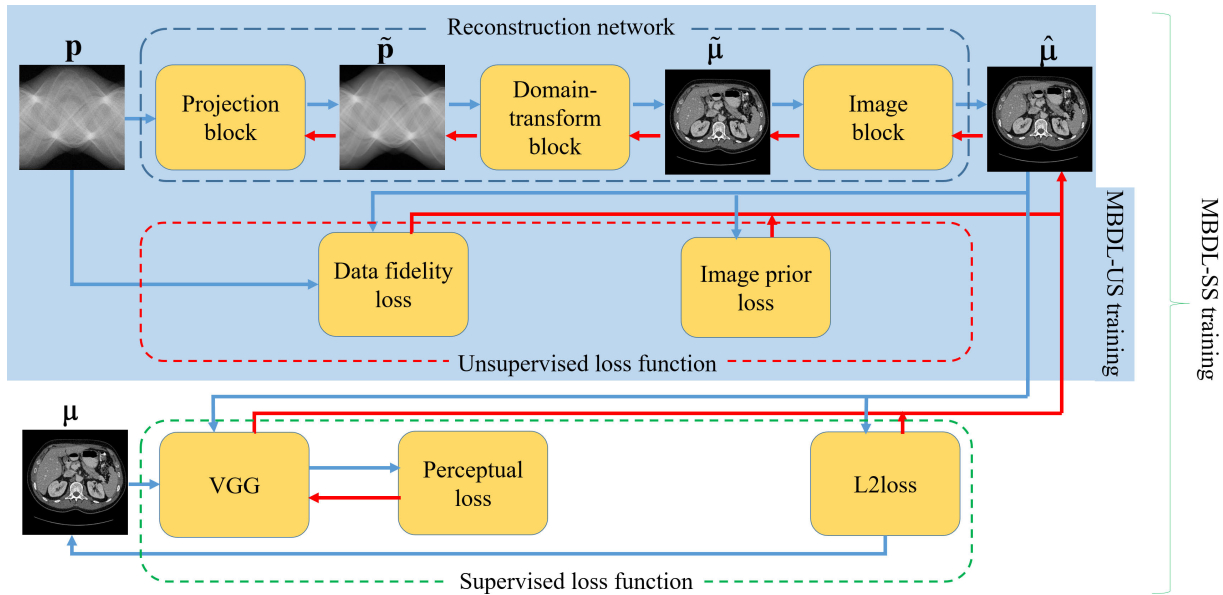


FIGURE 1. Model-based deep learning reconstruction network and its semi-supervised extension.

constant term $\text{Prob}(\mathbf{p})$ ignored:

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu} \geq 0}{\text{argmin}} \psi(\boldsymbol{\mu}, \mathbf{p}) \equiv \underset{\boldsymbol{\mu} \geq 0}{\text{argmin}} \varphi_f(\boldsymbol{\mu}, \mathbf{p}) + \lambda \varphi_p(\boldsymbol{\mu}) \quad (3)$$

Here, $\psi(\boldsymbol{\mu}, \mathbf{p}) = -\ln \text{Prob}(\boldsymbol{\mu}|\mathbf{p})$ denotes the MAP loss function. $\varphi_f(\boldsymbol{\mu}, \mathbf{p}) = -\ln \text{Prob}(\mathbf{p}|\boldsymbol{\mu})$ denotes the data fidelity term based on the distribution of \mathbf{p} conditioned on $\boldsymbol{\mu}$ after negative-log-transformation that measures the correspondence between a projection data and a reconstruction image. $\lambda \varphi_p(\boldsymbol{\mu}) = -\ln \text{Prob}(\boldsymbol{\mu})$ denotes the image prior term based on the prior distribution of $\boldsymbol{\mu}$ after negative-log-transformation that measures the rationality of a reconstruction. Here λ is a hyper-parameter adjusting the strength of prior term in implementation. Eq. 3 forms the basic optimization problem of MAP estimation.

Although the distribution of \mathbf{N} could be complicate, multiple studies have shown that it is reasonable to treat \mathbf{N} as Gaussian distributed with its variance determined by signal [8], [37]. If the noise is considered to be Gaussian, $\varphi_f(\boldsymbol{\mu}, \mathbf{p})$ is a weighted least square:

$$\varphi_f(\boldsymbol{\mu}) = (\mathbf{p} - \mathbf{H}_{\text{sys}}\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{p} - \mathbf{H}_{\text{sys}}\boldsymbol{\mu}) \quad (4)$$

Here, $\boldsymbol{\Sigma}$ is diagonal in case of independent noise in \mathbf{p} with its diagonal elements corresponding to the element-wise noise variances.

The choices for the image priors are flexible, such as TV, none-local mean (NLM), MRF, low-rank and dictionary priors [10]–[16]. The estimation results highly depend on the choices of image priors. Here, for convenience, we adopt NLM as the image prior in our experiments. The method can be extended to other priors or a combination of multiple priors.

NLM is one of the commonly used manually designed image priors for LDCT reconstruction [10], [13]. It adopts an

adaptive filtering strategy based on the assumption that two pixels are likely to come from the same type of tissues when the neighborhood patches of the two pixels are similar. NLM assigns different penalties on the difference of the two pixels according to the similarity of their neighborhood patches. The L1-norm NLM loss function is calculated as:

$$\varphi_{\text{NLM}}(\boldsymbol{\mu}) = \sum_i \sum_{i' \in \mathcal{N}(i)} w_{ii'} |\mu_i - \mu_{i'}| \quad (5)$$

Here, $\mathcal{N}(i)$ denotes the neighboring pixels of i within a searching window and $w_{ii'}$ denotes the weight of similarity computed from:

$$w_{ii'} = \frac{\exp(-\|\Pi(\mu_i) - \Pi(\mu_{i'})\|_{2,a}^2/h)}{\sum_{j \in \mathcal{N}(i)} \exp(-\|\Pi(\mu_i) - \Pi(\mu_j)\|_{2,a}^2/h)} \quad (6)$$

Here, $\Pi(\cdot)$ denotes a patch extractor, $\|\cdot\|_{2,a}$ the L2-norm distance with a Gaussian kernel a weighting the position of each pixels in a patch and h a normalization parameter. With adaptive weighting, NLM can preserve structures fairly well and reduce noise significantly.

B. LDCT RECONSTRUCTION NETWORK ARCHITECTURE AND NETWORK TRAINING

1) ARCHITECTURE OF RECONSTRUCTION NETWORK

We use a general operator Ω to denote a reconstruction process for LDCT, i.e. $\boldsymbol{\mu} \equiv \Omega(\mathbf{p})$. In theory, Ω shall be the inverse operator of the forward-projection process modelled as Eq. 1. While the function Ω is complicate, we can use a deep neural network to approximate Ω . In this work, we adopted a three-block network for LDCT reconstruction in an end-to-end mode with filtered backprojection (FBP) reconstruction integrated in the network [38], [39].

The overall network architecture is displayed in the blue dotted box in Fig. 1. A projection block reduces the noise in projection domain firstly. It takes noisy projection data as input and gives an estimation of noise reduced projection. We take the U-Net (shown in Fig. 3) as the backbone of this block because it has the advantage in catching both global features and local image details and it has achieved significant success in medical imaging processing [21], [40]. Next, a domain-transform block transfers projection domain data into image domain to form a preliminary reconstruction image. The domain-transform block is in matrix format that is equivalent to FBP reconstruction operator for 2D case. For fan-beam CT, the domain-transform block is consisted of three layers with frozen matrixes: diagonal matrix \mathbf{W} for weighting, matrix \mathbf{F} for filtration and weighted-back-projection matrix \mathbf{H}_w^T as shown in Fig. 2. The reconstruction process of the domain-transform block can be formulated as:

$$\tilde{\mu} = \mathbf{H}_w^T \mathbf{F} \mathbf{W} \tilde{\mathbf{p}} \quad (7)$$

Here, $\tilde{\mu}$ denotes the output of the domain-transform block, $\tilde{\mathbf{p}}$ the estimated projection from projection block. The three matrixes are calculated based on the FBP formula and are frozen during training. The back-propagating gradients can pass through the domain-transform block from image block to projection block in training process:

$$\frac{\partial \ell}{\partial \tilde{\mathbf{p}}} = \mathbf{W} \mathbf{F}^T \mathbf{H}_w \frac{\partial \ell}{\partial \tilde{\mu}} \quad (8)$$

Here, ℓ denotes a general network loss function.

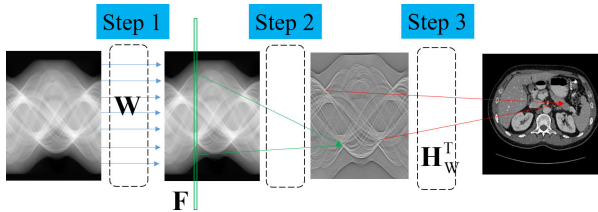


FIGURE 2. Domain-transform block. Step 1: Element-wise cosine weighting; step 2: RL filtering along detector axis; step 3: weighted back-projection.

Finally, an image block takes the preliminary reconstruction image from the second block as input and generates the final reconstruction image. It optimizes the preliminary reconstruction image to reduce artifacts and further restore structures. The image block is also consisted of a five-stage U-Net (same as the projection block) in Fig. 3 except for the different dimensions in the height and width (H and W in Fig. 3).

In all, the reconstruction can be denoted as $\hat{\mu} \equiv \Omega_{\text{Recon}}(\mathbf{p}, \theta_{\text{Recon}})$ with Ω_{Recon} denoting the network operator, θ_{Recon} the network parameter set, and $\hat{\mu}$ is the output of the whole network, i.e. reconstruction image.

2) MODEL-BASED TRAINING OF RECONSTRUCTION NETWORK

In supervised training, all the LDCT projections have corresponding NDCT image labels (or ground-truth labels). Thus, a network operator is optimized to minimize the distances such as L2-norm distance between the reconstruction images and the labels, i.e. the parameters of the network are tuned to minimize the distance:

$$\begin{aligned} \hat{\Omega} &= \underset{\Omega}{\operatorname{argmin}} \sum_{i=1:N} \mathcal{D}(\Omega(\mathbf{p}_i), \mu_i) / N \Rightarrow \hat{\theta}_{\text{Recon}} \\ &= \underset{\theta_{\text{Recon}}}{\operatorname{argmin}} \sum_{i=1:N} \mathcal{D}(\Omega_{\text{Recon}}(\mathbf{p}_i, \theta_{\text{Recon}}), \mu_i) / N \quad (9) \end{aligned}$$

Here, $\hat{\theta}_{\text{Recon}}$ denotes the optimal network parameter set, $\mathcal{D}(\bullet, \bullet)$ a distance measurement between two images, N the total number of training pairs in the training set indexed by i . However, in some situations paired NDCT images (μ_i in Eq. 9) are not available. For example, scanning a patient with both NDCT and LDCT protocols is not allowed considering the radiation dose in clinical practice. Hence, a new loss function other than Eq. 9 is needed. Fortunately, we can use the idea of MAP estimation in II. A. For each observed LDCT projection data, a network based on MAP estimation should reconstruct the corresponding image with the maximum posterior. Thus, we optimize the network parameters to reconstruct images with the maximum value of group-wise posterior for the whole LDCT training dataset (G-MAP). As different LDCT cases are mutually independent, the total posterior is calculated by multiplying the posterior of each case in the training dataset. After negative-log-transformation, the loss function further becomes additive. Hence, we get:

$$\begin{aligned} \hat{\Omega} &= \underset{\Omega}{\operatorname{argmax}} \prod_{i=1:N} \operatorname{Prob}(\Omega(\mathbf{p}_i) | \mathbf{p}_i) \Rightarrow \hat{\theta}_{\text{Recon}} \\ &= \underset{\theta_{\text{Recon}}}{\operatorname{argmax}} \prod_{i=1:N} \operatorname{Prob}(\Omega_{\text{Recon}}(\mathbf{p}_i, \theta_{\text{Recon}}) | \mathbf{p}_i) \\ &\stackrel{\text{Eq. 9}}{=} \underset{\theta_{\text{Recon}}}{\operatorname{argmin}} \sum_{i=1:N} \Psi(\Omega_{\text{Recon}}(\mathbf{p}_i, \theta_{\text{Recon}}), \mathbf{p}_i) / N \quad (10) \end{aligned}$$

Here, Ψ is the same MAP loss function as that in Eq. 3. Thus, a group of the MAP loss function is used to train the network with no label required, i.e. in an unsupervised manner. We refer this method as MBDL-US. In Fig. 1, the blue arrows denote forward-propagation (reconstruction process) while the red arrows denote the gradients back-propagation process. It should be noticed that, distinguishing from iterative reconstruction, the optimization is toward network parameters instead of reconstructed images.

One of the basic principles of deep learning is to extract knowledge from large dataset and store the knowledge in model parameters. The knowledge can be formed from multiple training routes. By introducing this strategy for constructing MBDL loss functions, we can flexibly and easily extend MBDL to integrate various data information and knowledge including paired LDCT/NDCT data and non-paired

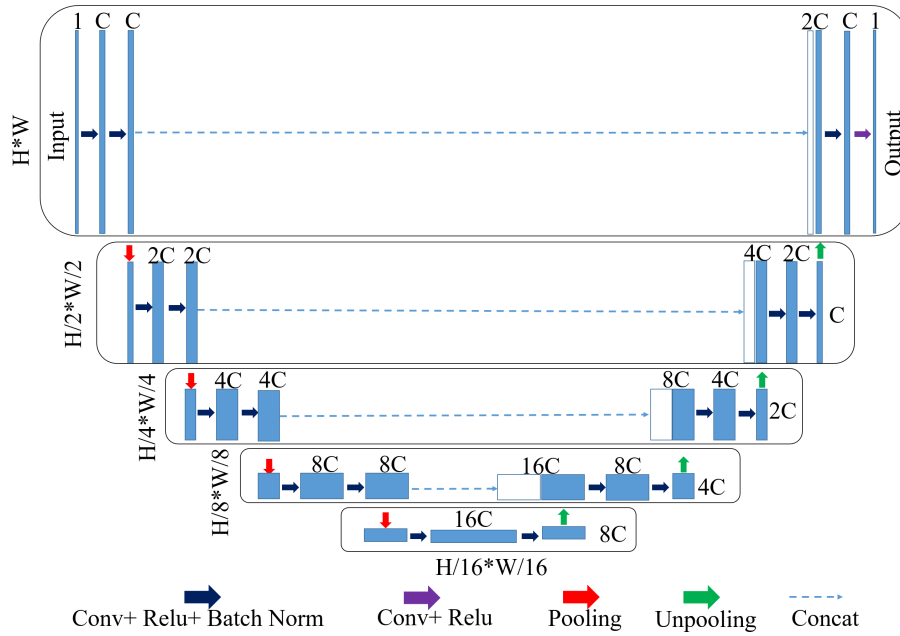


FIGURE 3. The U-Net architecture of projection block and image block.

LDCT/NDCT (often used in adversarial training). Certainly, the network performance can be further improved by these extensions. We extend MBDL to semi-supervised learning denoted as MBDL-SS as an example.

In some situations, there may be a few NDCT images available and working as labels for corresponding LDCT projection reconstruction besides a large LDCT dataset. If with enough labeled NDCT data, the network can be easily trained in a supervised route and achieve good performance. However, when labeled NDCT data is not enough, the network suffers from poor generalization during inference. By taking advantage of the large LDCT dataset, the network can gain more information in unsupervised training and become stable. Thus the network trained with both the labeled data in a supervised route and un-labeled data in an unsupervised route can achieve reliable inference performance. We adopt a combination of the perceptual loss and the L2-norm loss for the supervised training. The loss function for the supervised training is:

$$\begin{aligned} \ell_S = \min \sum_{i=1}^{N_P} [& (\|\Omega_{\text{Recon}}(\mathbf{p}_i, \theta_{\text{Recon}}) - \mu_i\|_2^2) \\ & + \beta \|\Upsilon_{\text{VGG}}(\Omega_{\text{Recon}}(\mathbf{p}_i, \theta_{\text{Recon}}), \theta_{\text{VGG}}) \\ & - \Upsilon_{\text{VGG}}(\mu_i, \theta_{\text{VGG}})\|_2^2] / N_P \end{aligned} \quad (11)$$

Here, Υ_{VGG} denotes the VGG network. θ_{VGG} denotes the VGG parameters pre-trained on ImageNet, β a hyper-parameter adjusting the importance of the perceptual loss, N_P the number of paired LDCT/NDCT data with the subscription ‘‘P’’ denoting ‘‘paired’’. The ℓ_S part of the network is illustrated with the box of green dash line in Fig. 1. Combining with the unsupervised loss function

$\ell_{\text{US}} = \sum_{i=1}^{N_{\text{UP}}} \Psi(\Omega_{\text{Recon}}(\mathbf{p}_i, \theta_{\text{Recon}}), \mathbf{p}_i) / N_{\text{UP}}$ for N_{UP} the number of unpaired LDCT data (the subscription ‘‘UP’’ meaning ‘‘unpaired’’), we acquire an overall MBDL-SS optimization problem:

$$\hat{\theta}_{\text{Recon}} = \underset{\theta_{\text{Recon}}}{\text{argmin}} (1 - \gamma) \ell_{\text{US}} + \gamma \ell_S \quad (12)$$

with $\gamma \in (0, 1)$ being a hyper-parameter adjusting the weighting of supervised training. Larger γ assigns more weighting to supervised training. In implementation, with mini-batch strategy, we can simply carry out the supervised training route and the unsupervised training route alternatively, thus the supervised training weighting γ is controlled by supervised training frequency η indirectly. For example, $\eta = 0.2$ means one supervised training step in each five training steps.

During inference, LDCT projections are fed to the network, and reconstruction images are obtained with no iteration needed.

III. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETTINGS

1) DATASET

We conducted our experiments based on thorax CT data from AAPM Low Dose CT Grand Challenge. There are 5936 CT images from 10 patients in total. Each image is of 512×512 resolution with pixel size 1 mm^2 . The LDCT projections in our experiments were simulated from NDCT images. We first simulated clean projections of fan-beam scanning from the NDCT images with forward-projection. The system-matrix was generated based on ray-tracing method [41]. In the simulation, both the source-to-center distance and detector-to-center distance were 600 mm. There were 640 detectors

with bin size 2.5 mm. We set 720 views uniformly distributed over 2π . We further generated LDCT projections by adding Poisson noise with 5×10^5 incident photons per ray. In LDCT reconstruction, we reconstructed the images of the same resolution as the original NDCT images.

2) METHODS FOR COMPARISON

To analyze the advantages and disadvantages of MBDL, we compared MBDL with three typical LDCT processing methods: non-local mean filtering after LDCT FBP reconstruction (FBP-NLM), MBIR and L2-norm supervised deep learning for LDCT reconstruction (L2-SDL).

a: FBP-NLM

FBP-NLM is a typical LDCT analytical post-processing method in 2D situations. First, a preliminary reconstruction image $\tilde{\mu}_{\text{FBP}}$ is acquired using FBP. Then the $\tilde{\mu}_{\text{FBP}}$ is further processed by NLM filtering to give the final reconstruction image $\hat{\mu}$:

$$\hat{\mu}_i = \sum_{i' \in \mathcal{N}(i)} w_{ii'} \tilde{\mu}_{\text{FBP}i'} \quad (13)$$

Here, $\hat{\mu}_i$ denotes the i^{th} pixel in $\hat{\mu}$, $\tilde{\mu}_{\text{FBP}i'}$ the i'^{th} pixel in $\tilde{\mu}_{\text{FBP}}$. The weighting $w_{ii'}$ is computed as in Eq. 6.

b: MBIR

MBIR is a well-known reconstruction method in the field which models the system geometry, noise model and image prior. For each observed LDCT projection data, MBIR estimates the reconstruction image by optimizing Eq. 3 iteratively using various optimization methods. In our experiments, to compare with MBDL, the loss function of MBIR was set the same as MBDL for each image. We optimized the loss function with the Nesterov method [42]. The number of iterations of each image was 120 to produce the best quality images on average.

c: L2-SDL

Supervised learning with L2-norm loss function is a widely studied and robust deep learning method for LDCT processing. It generally achieves good mean square error (MSE) during inference. For fair comparison, we used the exactly same end-to-end network architecture as MBDL and trained it with the loss function defined in Eq. 9, the distance function was L2-norm: $\mathcal{D}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_2^2$. In other words, the ‘‘reconstruction network’’ (marked by the dashed blue line box in Fig. 1) was adopted in L2-SDL so that L2-SDL and MBDL were of exactly the same computational complexity during inference, though they were of different network parameters because trained by different loss functions. We would like to point out that L2-SDL took the most information because exact labels were used.

3) QUANTITATIVE EVALUATION

To evaluate the reconstruction results of different methods quantitatively, we adopted root mean square error (RRMSE),

structural similarity index (SSIM) and perceptual loss for evaluation.

a: RRMSE

The RRMSE index measures the relative L2-norm error between estimated and reference images. Lower RRMSE index indicates more accurate reconstruction. RRMSE is calculated as:

$$\text{RRMSE}(\hat{\mu}, \mu) = \frac{\|\hat{\mu} - \mu\|_2}{\|\mu\|_2} \quad (14)$$

Here, $\hat{\mu}$ denotes the image reconstructed, μ the reference image.

b: SSIM

The SSIM index measures the structural similarity between estimated and reference images. It compares both the mean signal value and the distribution relevance. The SSIM index is generally calculated based on image patches instead of the whole image. Higher SSIM index indicates more precise structures. The SSIM index is calculated as:

$$\text{SSIM}(\hat{\mu}, \mu) = \frac{(2\bar{\hat{\mu}}\bar{\mu} + C_1)(2\sigma_{\hat{\mu}\mu} + C_2)}{(\bar{\hat{\mu}}^2 + \bar{\mu}^2 + C_1)(\sigma_{\hat{\mu}}^2 + \sigma_{\mu}^2 + C_2)} \quad (15)$$

Here, $\bar{\hat{\mu}}$ denotes the mean of $\hat{\mu}$, $\bar{\mu}$ the mean of μ , $\sigma_{\hat{\mu}\mu}$ the covariance of $\hat{\mu}$ and μ , $\sigma_{\hat{\mu}}^2$ the variance of $\hat{\mu}$, σ_{μ}^2 the variance of μ . C_1 and C_2 are constants avoiding zero denominator. In our experiments, the SSIM indexes were calculated based on 11×11 image patches with Gaussian distance weighting of which the standard deviation was 1.5 [43].

c: PERCEPTUAL LOSS

The Perceptual loss is from a VGG network pre-trained on ImageNet to simulate human vision and extract image features for image classification [24]. It calculates the squared L2-norm distance between the estimated image features and the reference image features as follows:

$$\text{Perceptual}(\hat{\mu}, \mu) = \|\Upsilon_{\text{VGG}}(\hat{\mu}, \theta_{\text{VGG}}) - \Upsilon_{\text{VGG}}(\mu, \theta_{\text{VGG}})\|_2^2 \quad (16)$$

Low perceptual loss also characterizes good preservation of structural information. Here, the Υ_{VGG} and θ_{VGG} are the same as those in Eq. 11. In this work, we used the features from the last convolutional layer of VGG-16 network.

In our experiments, the quantitative evaluations were conducted for the whole test set and the mean values were reported.

B. EXPERIMENTS OF MODEL-BASED UNSUPERVISED DEEP LEARNING

We trained our LDCT reconstruction network with MBDL-US as **II. B. 2**. The basic channels of the projection-domain and the image-domain U-Nets were set to be 32. We used the LDCT projection data of 5376 images from nine

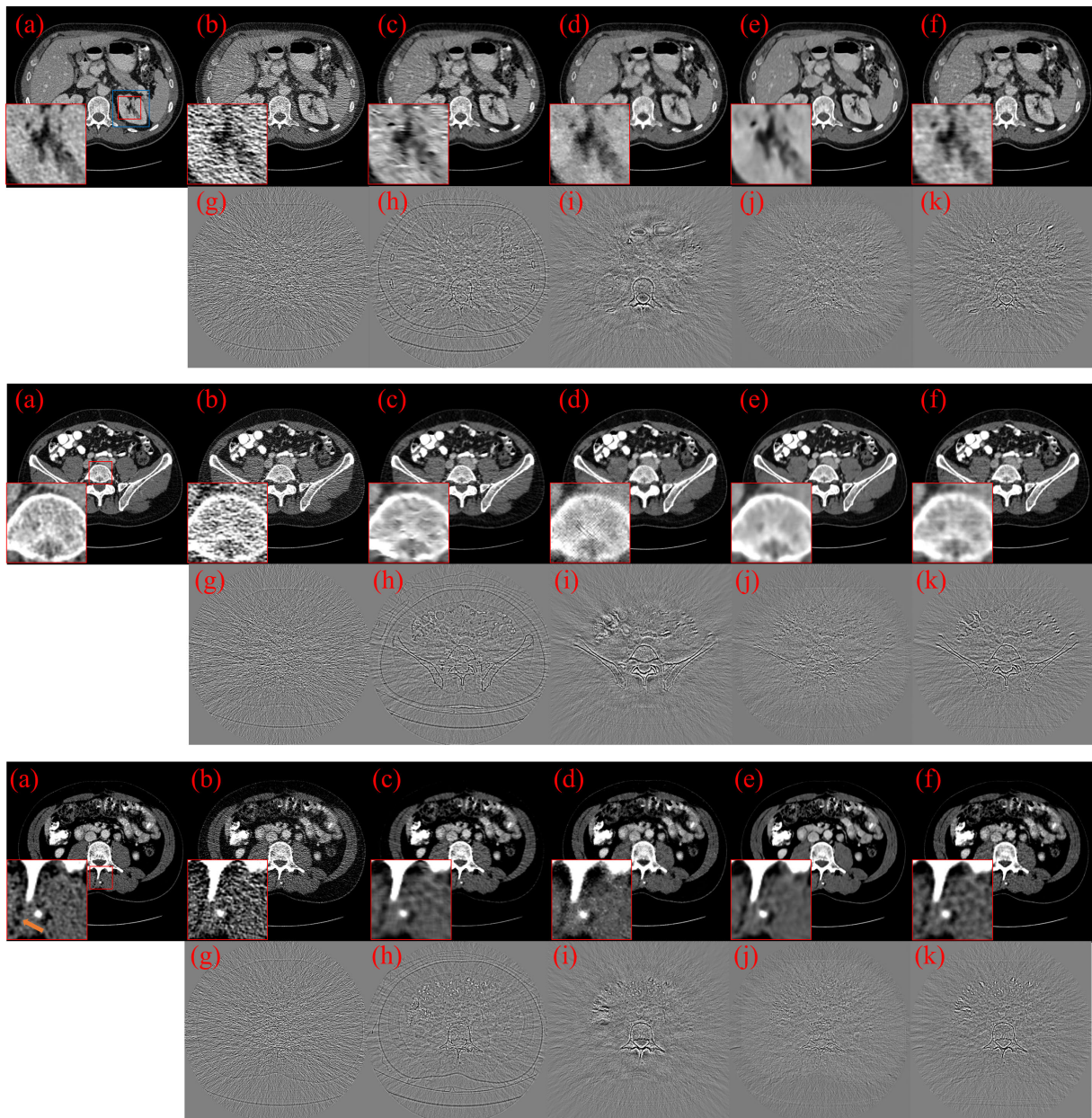


FIGURE 4. Comparison of different LDCT reconstruction methods. (a): Ground-truth, (b): LDCT FBP, (c): LDCT FBP-NLM, (d): LDCT MBIR, (e): LDCT L2-SDL, (f): LDCT MBDL-US, (g): residual error between (b) and (a), (h): residual error between (c) and (a), (i): residual error between (d) and (a), (j): residual error between (e) and (a), (k): residual error between (f) and (a). Display-window: Case 1, 2 [-160 HU, 340 HU], Case 3 [-80 HU, 250 HU].

patients to train the network while the left LDCT projection data of one patient was used to test the network. The NLM weighting λ was 0.003 which gave the best results in terms of RRMSE index and visually. The influences of λ was examined and discussed in III. C. No NDCT images were used in the unsupervised training process. NDCT images were used as the ground-truth images to validate the reconstruction results during test.

We compared the MBDL-US reconstruction results with FBP-NLM, MBIR and L2-SDL. Three general cases were displayed in Fig. 4. It's obvious that the direct FBP

reconstructions of LDCT suffered from severe streaks. FBP-NLM reduced the streaks while causing mosaic effect. MBIR, L2-SDL and MBDL-US all achieved significant improvements from FBP reconstruction with streaks removed and structures restored greatly. The L2-SDL produced results with clean structures and of least residual errors, which can be viewed as the up-limit performance because ground-truth was used in its training process. The MBDL-US reconstructed structures better compared with MBIR. The edges of MBIR were slightly blurred as we can see clearly in the corresponding residual maps. Compared with L2-SDL, MBDL-US

results were slightly noisier with streaks remained. Further observing the detail structures in the zoom-ins of case 1, the renal interstitium structures on kidney were contaminated by streak artifacts in FBP reconstruction. The MBIR, L2-SDL and MBDL-US restored most of the renal interstitium structures well. The L2-SDL results was of the lest noise and artifacts while it lost some texture information of soft tissue. The textures information of soft tissue was visually realistic in MBDL-US results. In the zoom-ins of case 2, the MBDL-US produced clear lumbar vertebrae cross section structure, while there were small streaks in MBIR reconstruction due to the high-attenuation of bone. In the zoom-ins of case 3, the three dark spots denoted by the brown arrow were severely corrupted by streak artifacts in the FBP reconstruction. FBP-NLM and MBIR restored them but they were over blurred. The MBDL-US produced the most realistic structures of the dark spots.

The LDCT reconstruction results of different methods were further evaluated quantitatively with RRMSE, SSIM and perceptual loss. The quantification results were listed in Table 1. In Table 1, the best results among all methods are in red, and the best among no-ground-truth methods are in blue. It can be seen that L2-SDL produced the best results in terms of RRMSE, SSIM and perceptual loss, which is consistent with previous study. When ground-truth images are not available, MBDL-US becomes the best choice. It achieved better results than MBIR but was ~ 100 times faster. These all suggest the great value and advantages of MBDL-US.

TABLE 1. Quantitative comparison of different LDCT reconstruction methods.

Methods	RRMSE	SSIM	Perceptual Loss	Time (s)
FBP	0.1114	0.7571	3712.2	2.98
FBP-NLM	0.0553	0.9304	1960.8	3.05
MBIR	0.0443	0.9530	898.8	330.84
L2-SDL	0.0291	0.9731	394.6	3.12
MBDL-US	0.0386	0.9638	822.8	3.12

C. THE INFLUENCE OF NLM WEIGHTING

The NLM weighting λ is a hyper parameter in MBDL-US. To study the influence of λ on reconstruction and choose a proper value, we adopted an exponential searching strategy. We conducted the MBDL-US training with $\lambda = 0$ and λ from 2.5×10^{-4} to 1.6×10^{-2} doubling each time. Results suggested the proper scope for λ was between 2×10^{-3} and 4×10^{-3} . Fig. 5 plots the RRMSE and SSIM indexes versus λ for MBDL-US results. Lowest RRMSE was achieved around $\lambda = 3 \times 10^{-3}$ (4×10^{-3} for SSIM), hence we set $\lambda = 3 \times 10^{-3}$ in all other experiments.

Fig. 6 shows the zoom-ins of the right kidney structure in the first case of Fig. 4 (denoted by a blue box) reconstructed by MBDL-US with different λ . The valuable kidney structure was the clearest with little streaks when $\lambda = 3 \times 10^{-3}$. With smaller λ , there were streaks degrading image quality. With larger λ , the edges and details were over-smoothed.

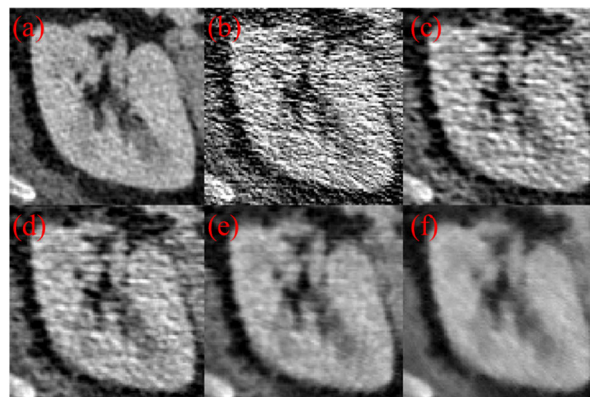


FIGURE 5. Qualitative comparison of MBDL-US results with different λ . (a): Ground-truth, (b): LDCT FBP, (c): MBDL-US $\lambda = 0$, (d): MBDL-US $\lambda = 5 \times 10^{-4}$ (e): MBDL-US $\lambda = 3 \times 10^{-3}$ (f): MBDL-US $\lambda = 8 \times 10^{-3}$. Display-window: [-80 HU, 250 HU].

Interestingly, even setting λ to 0, the network could reduce some noise. This indicated the auxiliary denoising ability of WLS learning.

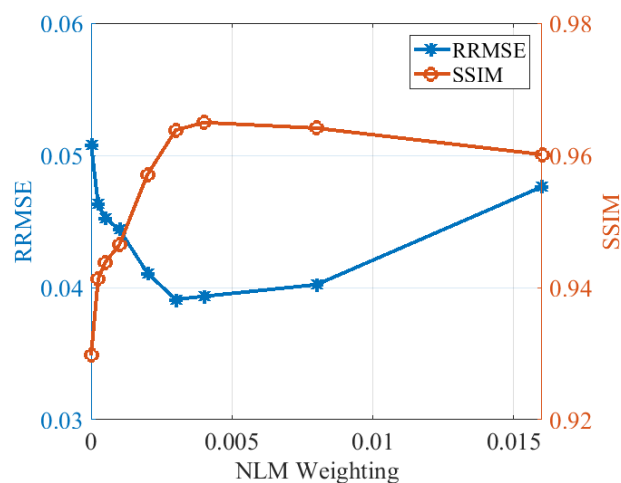


FIGURE 6. Plots of RRMSE and SSIM versus λ for MBDL-US. Best results were obtained at λ close to 0.003.

D. RESULTS OF DIFFERENT NOISE LEVELS

We further conducted MBDL-US under different noise levels by adjusting incident photon number. The incident photon number of each ray was set to be 1×10^5 , 2×10^5 , 5×10^5 , 1×10^6 , and 2×10^6 . With increasing photon numbers, projection data become less noisy. In Fig. 7, we displayed the reconstruction results of the first case in Fig. 4 under photon numbers 1×10^5 and 2×10^6 for comparison. FBP reconstruction was of very poor image quality for the high noise case with photon number being 1×10^5 . FBP-NLM reduced streaks, but detail structures were not well restored. Again, MBIR, L2-SDL and MBDL-US achieved better results than FBP-NLM. Compared with L2-SDL, the noise level of MBIR results was higher, and the low-dose streak artifacts were not completely removed in MBDL-US results. According to the residual images, the MBDL-US preserved the vertebra

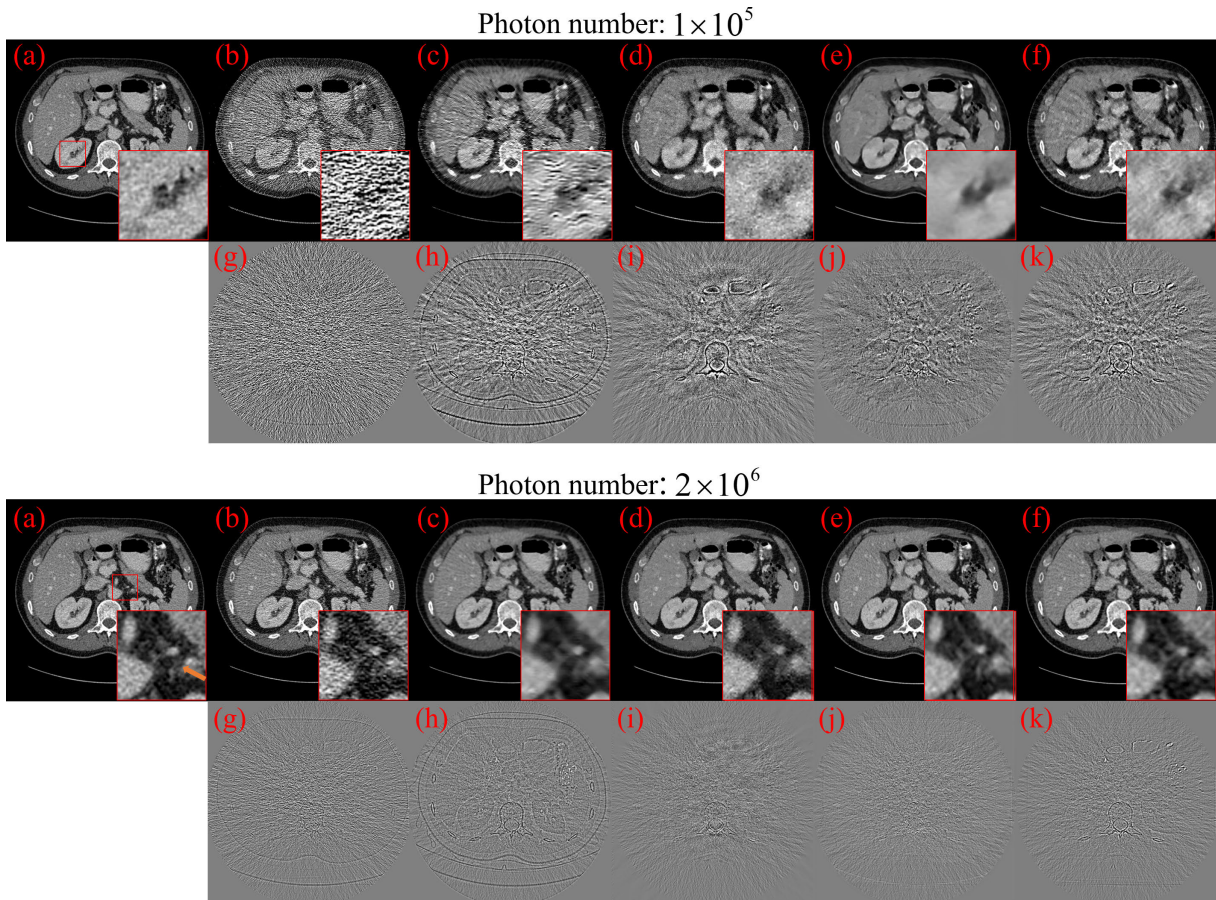


FIGURE 7. Comparison of different LDCT reconstruction methods under two typical noise levels. (a): Ground-truth, (b): LDCT FBP, (c): LDCT FBP-NLM, (d): LDCT MBIR, (e): LDCT L2-SDL, (f): LDCT MBDL-US, (g): residual error between (b) and (a), (h): residual error between (c) and (a), (i): residual error between (d) and (a), (j): residual error between (e) and (a), (k): residual error between (f) and (a). Display-window: $[-160 \text{ HU}, 330 \text{ HU}]$.

outline slightly better than MBIR. The zoom-ins of left-kidney shows that MBDL-US gave fairly good renal interstitium structures. The LDCT reconstruction task was easier for the case of 2×10^6 photons. All the methods except FBP gave images of rather good quality. FBP-NLM result was over blurred when checking the zoom-ins. The crevice structure denoted with the brown arrow was restored clearly with L2-SDL and MBDL-US, while in the MBIR reconstruction, its upper part was fused and the crevice was shortened.

We calculated the RRMSE and SSIM indexes of the compared methods for different noise levels. The results were plotted in Fig. 8. In accord with the quantitative results of III. B, the L2-SDL achieved the best results in RRMSE and SSIM while MBDL-US was the most valuable without ground-truth images. As photon number increased, the SSIM indexes of different methods become closer. The gaps between the RRMSE curves of L2-SDL and MBDL-US suggested the potential improvement space of MBDL-US.

E. ENHANCEMENT BY MODEL-BASED SEMI-SUPERVISED LEARNING

MBDL-US can be further improved with supervised training, i.e. MBDL semi-supervised learning (MBDL-SS). In this

section, we examined the performance of MBDL-SS with a small number of paired NDCT images. In this experiment, the photon number was 5×10^5 , the number of training data was still 5376, and there were 10 NDCT images from hip paired with LDCT data. In MBDL-SS, the NLM weighting λ in unsupervised training step was 3×10^{-3} the same as III. C. We set the perceptual loss weighting $\beta = 0.5$, the supervised training frequency $\eta = 0.2$ that gave the most valuable clinical structures.

To illustrate the image quality enhancement from MBDL-SS, we compared perceptual-L2 loss supervised training based on the 10 image pairs (PL2-SDL-10), MBDL-US, and MBDL-SS. Two representative cases were displayed in Fig. 9. It can be observed that PL2-SDL-10 produced visually high quality images. However, due to the small number of training data, the reconstructed detail structures were not consistent with ground-truth as denoted by the brown arrows in Case 1 and Case 2. These results suggested the insufficient generalization ability of PL2-SDL-10. In Case 2, with narrowed display-window, there were shadow artifacts left on soft-tissues in the MBDL-US results as seen in the zoom-in of Case 2. The MBDL-SS further reduced these artifacts and noise. Moreover, we can see in

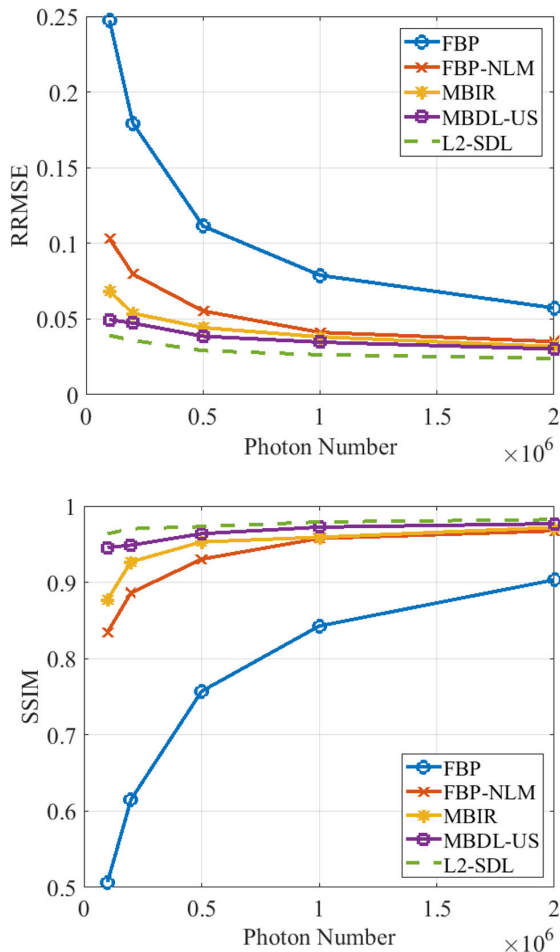


FIGURE 8. Plots of RRMSE and SSIM versus photon number. MBDL-US gives the close-to-supervised performance in all noise levels and better than all conventional methods.

the residual images that less high frequency edge loss happened in MBDL-SS than MBDL-US. These results suggested MBDL-SS can improve the reconstructions of MBDL-US with additional information from labeled data.

The different training methods were further evaluated quantitatively with RRMSE, SSIM and perceptual loss, the results were listed in Table 2. In Table 2, MBDL-SS significantly improved the RRMSE and perceptual loss from MBDL-US. Due to lacking of training data, PL2-SDL-10 results are of the worst RRMSE and SSIM, though the perceptual loss of PL2-SDL-10 was the best. This was because both PL2-SDL-10 and MBDL-SS was trained with the perceptual loss, while the PL2-SDL-10 was trained with stronger impact from the perceptual loss. The perceptual loss was further generalized to test data. Based on the qualitative and quantitative examination, MBDL-SS combined the advantage of supervised training and unsupervised training. The MBDL-US results were further enhanced by MBDL-SS, even when the number of paired data was small.

In our experiments, the supervised training frequency η influenced the MBDL-SS results. With a small η , the MBDL-SS results were similar to MBDL-US. When η was close to 1,

TABLE 2. Quantitative comparison of different training methods.

Methods	RRMSE	SSIM	Perceptual Loss
PL2-SDL-10	0.0393	0.9597	441.9
MBDL-US	0.0386	0.9638	822.8
MBDL-SS	0.0359	0.9663	485.1

the MBDL-SS results were similar to PL2-SDL-10. To determine the best frequency, we conducted grid-searching experiments for η between 0 and 1. Since paired data was of a small portion, we reduced searching intervals for η between 0 and 0.3. The plots of RRMSE and SSIM indexes versus η for MBDL-SS was shown in Fig. 10. The best η was around 0.2.

F. ABLATION STUDY FOR NETWORK ARCHITECTURE

The LDCT reconstruction network architecture plays an important role in both MBDL-US and MBDL-SS. In the previous subsections, we adopted a three block reconstruction network as presented in II. B. 1. To further demonstrate the function of each block, we conducted an ablation study for the reconstruction network architecture. We define the full network as dual-domain network (DDN), the reconstruction network with only projection block and domain-transform block as projection-domain network (PDN), the reconstruction network with only domain-transform block and image block as image-domain network (IDN). In fact, IDN is a traditional FBP + U-Net network which is widely studied in this field. The domain-transform block guaranties the reconstruction process which cannot be removed for ablation study.

We compared the reconstruction results of the three networks for both MBDL-US and MBDL-SS. Two representative cases were displayed in Fig. 11. For MBDL-US, IDN performed the worst with the smoothed streaks remained. PDN and DDN gave visually comparable results. In the zoom-ins of case 1, there were similar light streaks remained in both PDN and DDN. These indicates projection block is the core denoising block for MBDL-US. However, for MBDL-SS, the PDN performed the worst visually with little improvements from MBDL-US. The DDN again gave the best results. In the zoom-ins of case 1, least artifacts were seen in DDN results. In the zoom-ins of case 2, the circular vascular cross section denoted by the blue box was also restored the best with DDN, while both PDN and IDN deformed the circular structure. From the results in different columns, we can see that, with supervised training added, MBDL-SS with DDN significantly improved the results of MBDL-US with DDN in terms of denoising and structure restoration. However, the results of MBDL-US with PDN and MBDL-SS with PDN were similar with little improvements. This indicates the supervised training enhancement is mostly from the image-domain block.

Quantitative results of RRMSE, SSIM, and perceptual loss were listed in Table 3. They agreed with visual inspection. DDN achieved the best RRMSE, SSIM and perceptual loss

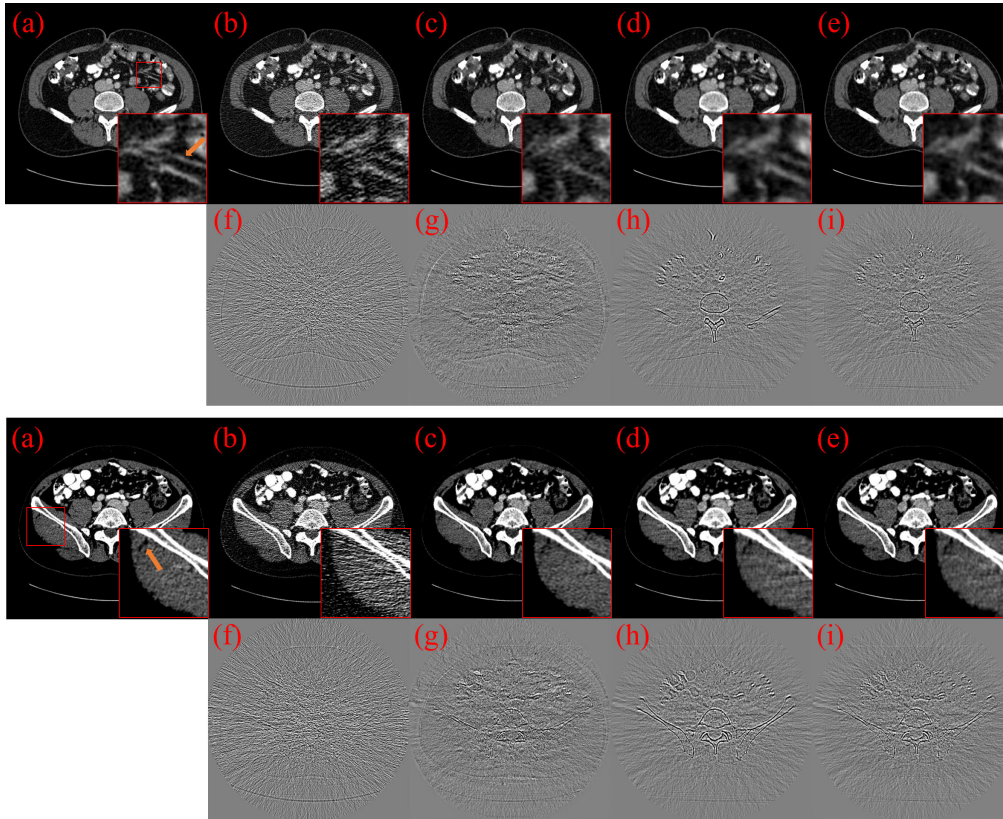


FIGURE 9. Comparison of different training methods. (a): Ground-truth, (b): LDCT FBP, (c): LDCT PL2-SDL-10, (d): LDCT MBDL-US, (e): LDCT MBDL-SS, (f): residual error between (b) and (a), (g): residual error between (c) and (a), (h): residual error between (d) and (a), (i): residual error between (e) and (a). Display window: Case 1 [-160 HU, 340 HU], Case 2 [-80 HU, 250 HU].

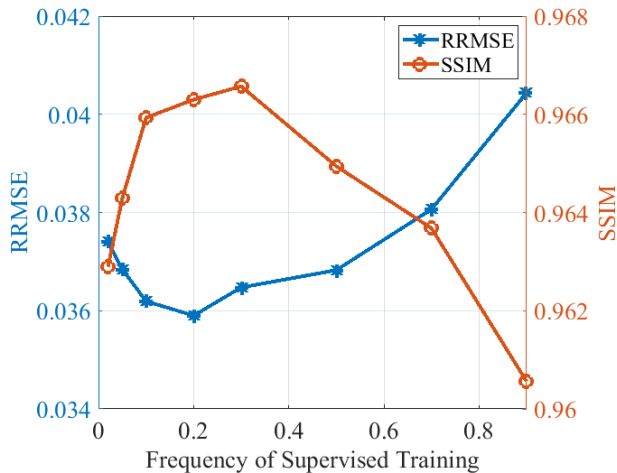


FIGURE 10. Plots of RRMSE and SSIM versus η for MBDL-SS. Best results were obtained at η close to 0.2.

in both MBDL-US and MBDL-SS which indicates the effectiveness of the DDN reconstruction network by concatenating both projection-domain processing and image-domain processing. The PDN network only achieved slightly worse quantitative results compared with DDN which indicated the projection block was the core block removing noise while the image block could further enhance the performance.

TABLE 3. Quantitative comparison of different network architectures.

Network	RRMSE	SSIM	Perceptual Loss
MBDL-US			
PDN	0.0398	0.9621	1021.2
IDN	0.0505	0.9103	1647.8
DDN	0.0386	0.9638	822.8
MBDL-SS			
PDN	0.0397	0.9617	820.1
IDN	0.0468	0.9393	877.2
DDN	0.0359	0.9663	485.1

In MBDL-SS, the enhancement of image block was more obvious than that in MBDL-US.

IV. DISCUSSION

In this work, we proposed a new method of training LDCT reconstruction network based on a G-MAP loss function. In situations with LDCT data only, MBDL-US achieved better results both qualitatively and quantitatively than MBIR with the same MAP loss function, while the reconstruction time was percentile. It's worth noting that,

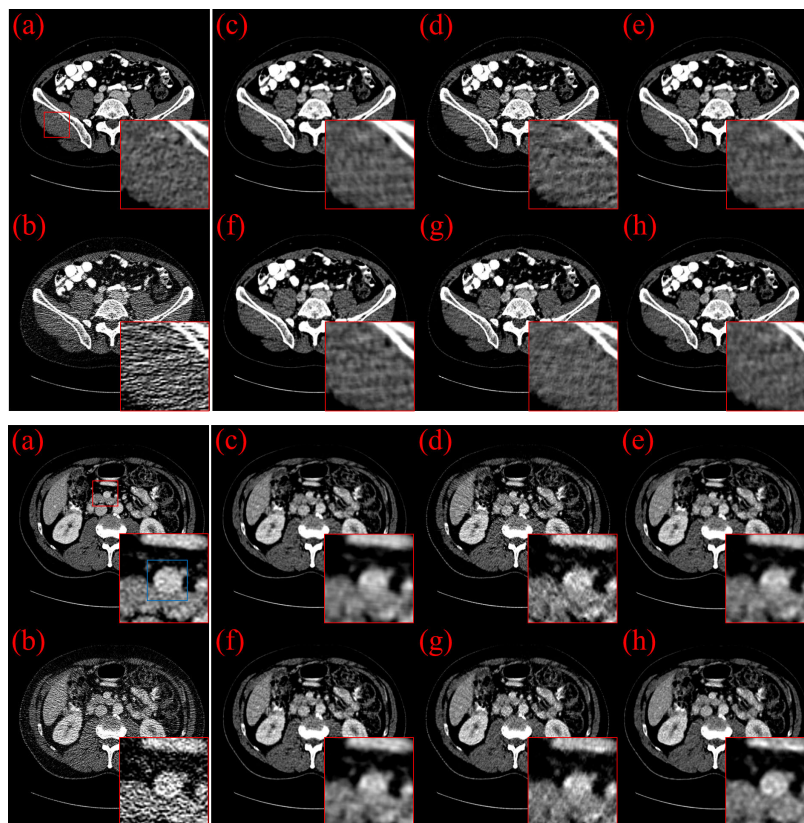


FIGURE 11. Ablation study with three different network architectures. (a): Ground-truth, (b): LDCT FBP, (c): MBDL-US with PDN, (d): MBDL-US with IDN, (e): MBDL-US with DDN, (f): MBDL-SS with PDN, (g): MBDL-SS with IDN, (h): MBDL-SS with DDN. Display window: $[-80 \text{ HU}, 150 \text{ HU}]$.

although MBDL-US and MBIR had the same loss function, MBDL-US produced slightly better results. This is because that MBDL-US benefits from the restriction of a large number of samples, while MBIR treated each sample independently.

The MBDL is a flexible frame which can integrate different training routes. In this work, by combining MBDL-US with supervised training, i.e. MBDL-SS, we achieved further improvements. It is also flexible and robust to incorporate existing deep learning strategies and other knowledge, such as additional GAN loss in case of unpaired LDCT and NDCT data available. Besides, the MBDL provides an easy way to model imaging physics, signal statistics as well as handcraft prior information, and feeds the knowledge into a deep learning framework. This gives us a great potential to combine the advantages of conventional methods and deep learning methods.

In this work, we only conducted experiments based on an NLM prior. More well-known and valuable handcraft image priors can also be integrated. MBDL can also be tailored for other reconstruction problems such as few-view and limited-view CT reconstructions. We are to further explore the potential of this MBDL frame for different ill-posed problems, different imaging modality, and test it in wide application scenarios.

In addition to the proposed MBDL method, there are also other deep learning methods applied for LDCT denoising network training without NDCT dataset. For example, Noise2Noise [44] uses noisy images as labels to train a denoising network conditioned on that labels and inputs are pairs of two independent noisy realizations. However, in clinical CT applications, two independent realizations of each patient are hard to acquire. Based on Noise2Noise, several self-supervised denoising methods have also been explored to avoid the demand for two noisy realizations of each image [45]–[47]. In Noise2Void, the central pixels are predicted from neighboring pixels taking the noisy central pixel value as label [45]. Under the assumption of pixel-independent noise, the noise in neighboring patch inputs is independent with the noise in central pixel labels, thus the network can be trained in a self-supervised manner as Noise2Noise. Similar blind-spot strategy is also used in [46]. In Noise2Self, the mapping from neighboring pixels to central pixel is extended to the concept of J-invariant functions and the implementation of self-supervised image denoising becomes more flexible assuming that the noise is pixel-independent [47]. These self-supervised methods could be useful for LDCT projection denoising but can hardly be applied to image-domain LDCT denoising currently because noise in reconstruction images is spatially

correlated. Compared with the above methods, MBDL is easily implementable in wide circumstances and can be extended to other ill-posed CT applications such as sparse-view CT and limited-angle CT with problem-specific G-MAP loss function designs.

In situations of only a small number of NDCT images paired with LDCT data available, it is also possible to fine-tune a pre-trained denoising network to obtain reasonable results, i.e. transfer learning. As most of the parameters pre-trained in source domain are frozen or constrained, the NDCT dataset demand can be reduced greatly in fine-tune. In practice, the transfer learning performance heavily relies on the available pre-trained models and the distance between source and target domains. Compared with our MBDL-SS, transfer learning acquires additional knowledge from pre-trained models, while MBDL-SS acquires additional knowledge from the MBDL-US training. The two sources of knowledge compensate for the limitation on NDCT data information. We believe these two types of knowledge can be combined together in future works, thus the LDCT denoising results can be further improved.

REFERENCES

- [1] E. J. Hall and D. J. Brenner, "Cancer risks from diagnostic radiology," *Brit. J. Radiol.*, vol. 81, no. 965, pp. 362–378, May 2008.
- [2] 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.
- [3] Y. Funama, K. Awai, Y. Nakayama, K. Kakei, N. Nagasue, M. Shimamura, N. Sato, S. Sultana, S. Morishita, and Y. Yamashita, "Radiation dose reduction without degradation of low-contrast detectability at abdominal multisection CT with a Low-Tube voltage technique: Phantom study," *Radiology*, vol. 237, no. 3, pp. 905–910, Dec. 2005.
- [4] D. P. Naidich, C. H. Marshall, C. Gribbin, R. S. Arams, and D. I. McCauley, "Low-dose CT of the lungs: Preliminary observations," *Radiology*, vol. 175, no. 3, pp. 729–731, Jun. 1990.
- [5] C. H. McCollough, A. N. Primak, N. Braun, J. Kofler, L. Yu, and J. Christner, "Strategies for reducing radiation dose in CT," *Radiol. Clinics North Amer.*, vol. 47, no. 1, pp. 27–40, Jan. 2009.
- [6] A. Borsdorf, S. Kappler, R. Raupach, F. Noo, and J. Hornegger, "Local orientation-dependent noise propagation for anisotropic denoising of CT-images," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec. (NSS/MIC)*, Orlando, FL, USA, Oct. 2009, pp. 2472–2475.
- [7] L. L. Chen, S. P. Gou, Y. Yao, J. Bai, L. Jiao, and K. Sheng, "Denoising of low dose CT image with context-based BM3D," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2016, pp. 682–685.
- [8] H. Lu, X. Li, I.-T. Hsiao, and Z. Liang, "Analytical noise treatment for low-dose CT projection data by penalized weighted least-square smoothing in the K-L domain," *Proc. SPIE*, vol. 4682, May 2002, Art. no. 465552.
- [9] Y. Zhang, J. Rong, H. Lu, Y. Xing, and J. Meng, "Low-dose lung CT image restoration using adaptive prior features from full-dose training database," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2510–2523, Dec. 2017.
- [10] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Han, L. Li, W. Moore, and Z. Liang, "Adaptive nonlocal means-based regularization for statistical image reconstruction of low-dose X-ray CT," *Proc. SPIE*, vol. 9412, Mar. 2015, Art. no. 94123K.
- [11] K. Sang Kim and J. Chul Ye, "Low-dose limited view 4D CT reconstruction using patch-based low-rank regularization," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Seoul, South Korea, Nov. 2013, pp. 1–4.
- [12] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, "Low-dose CT reconstruction via edge-preserving total variation regularization," *Phys. Med. Biol.*, vol. 56, no. 18, pp. 5949–5967, Sep. 2011.
- [13] 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, Sep. 2014.
- [14] R. Zhang, D. H. He, D. Pal, J.-B. Thibault, K. D. Sauer, and C. A. Bouman, "A Gaussian mixture MRF for model-based iterative reconstruction with applications to low-dose X-ray CT," *IEEE Trans. Comput. Imag.*, vol. 2, no. 3, pp. 359–374, Sep. 2016.
- [15] J. Liu, Y. Chen, Y. Hu, and L. Luo, "Low-dose CBCT reconstruction via 3D dictionary learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Prague, Czech Republic, Apr. 2016, pp. 735–738.
- [16] 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.
- [17] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
- [18] M. Gholizadeh-Ansari, J. Alirezaie, and P. Babyn, "Low-dose CT denoising with dilated residual network," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Honolulu, HI, USA, Jul. 2018, pp. 5117–5120.
- [19] W. Yang, H. Zhang, J. Yang, J. Wu, X. Yin, Y. Chen, H. Shu, L. Luo, G. Coatrieux, Z. Gui, and Q. Feng, "Improving low-dose CT image using residual convolutional network," *IEEE Access*, vol. 5, pp. 24698–24705, 2017.
- [20] Y. Liu and Y. Zhang, "Low-dose CT restoration via stacked sparse denoising autoencoders," *Neurocomputing*, vol. 284, pp. 80–89, Apr. 2018.
- [21] C. Zhang and Y. Xing, "CT artifact reduction via U-net CNN," *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 105741R.
- [22] L. Hoyeon, L. Jongha, K. Hyeonseo, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 2, pp. 109–119, Mar. 2019.
- [23] E. Kang, j. Min, and J. Chul Ye, "Wavelet domain residual network (WavResNet) for low-dose X-ray CT reconstruction," 2017, *arXiv:1703.01383*. [Online]. Available: <http://arxiv.org/abs/1703.01383>
- [24] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.
- [25] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.
- [26] J. Adler and O. Oktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.
- [27] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "A cascaded convolutional neural network for X-ray low-dose CT image denoising," 2017, *arXiv:1705.04267*. [Online]. Available: <http://arxiv.org/abs/1705.04267>
- [28] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, Jun. 2018.
- [29] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, "Optimizing a parameterized Plug-and-Play ADMM for iterative low-dose CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 371–382, Feb. 2019.
- [30] A. Zhong, B. Li, N. Luo, Y. Xu, L. Zhou, and X. Zhen, "Image restoration for low-dose CT via transfer learning and residual network," *IEEE Access*, vol. 8, pp. 112078–112091, 2020.
- [31] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.
- [32] M. Meng, D. Li, S. Li, M. Zhu, L. Wang, Q. Gao, Z. Bian, X. Zhang, J. Huang, D. Zeng, and J. Ma, "Progressive transfer learning strategy for low-dose CT image reconstruction with limited annotated data," *Proc. SPIE*, vol. 11312, Mar. 2020, Art. no. 1131242.
- [33] H. Shan, U. Kruger, and G. Wang, "A novel transfer learning framework for low-dose CT," in *Proc. 15th Int. Meeting Fully Three-Dimensional Image Reconstruction Radiol. Nucl. Med.*, May 2019, Art. no. 110722Y.
- [34] J. Ma, Z. Liang, and F. Yi, "Variance analysis of x-ray CT sinograms in the presence of electronic noise background," *Med. Phys.*, vol. 39, no. 7, pp. 4051–4065, 2012.
- [35] J. Xu and B. M. W. Tsui, "Electronic noise modeling in statistical iterative reconstruction," *IEEE Trans. Image Process.*, vol. 18, no. 6, pp. 1228–1238, Jun. 2009.

[36] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, Feb. 2002.

[37] J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen, and J. Manzione, "An experimental study on the noise properties of X-ray CT sinogram data in radon space," *Phys. Med. Biol.*, vol. 53, no. 12, pp. 3327–3341, Jun. 2008.

[38] K. Liang, H. Yang, and Y. Xing, "Comparison of projection domain, image domain, and comprehensive deep learning for sparse-view X-ray CT image reconstruction," 2018, *arXiv:1804.04289*. [Online]. Available: <http://arxiv.org/abs/1804.04289>

[39] H.-K. Yang, K.-C. Liang, K.-J. Kang, and Y.-X. Xing, "Slice-wise reconstruction for low-dose cone-beam CT using a deep residual convolutional neural network," *Nucl. Sci. Techn.*, vol. 30, no. 4, p. 59, Apr. 2019.

[40] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.

[41] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.*, vol. 12, no. 2, pp. 252–255, 1985.

[42] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM J. Optim.*, vol. 22, no. 2, pp. 341–362, Jan. 2012.

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

[44] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2965–2974.

[45] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void—Learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2129–2137.

[46] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Proc. Conf. Neural Inf. Process. Syst.*, 2019, pp. 6970–6980.

[47] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *Proc. Int. Conf. Mach. Learn.*, vol. 97, 2019, pp. 53–524.



LI ZHANG (Member, IEEE) received the MA degree in radiation protection and health physics from the China Institute of Atomic Energy, Beijing, China, in 1994. She is currently a Professor with the Department of Engineering Physics, Tsinghua University. Her research interests include X-ray imaging, computed tomography, non-destructive testing, and deep learning.



HONGKAI YANG received the Ph.D. degree in nuclear science and technology from Tsinghua University, Beijing, China, in 2019. He currently holds a postdoctoral position at the Department of Engineering Physics, Tsinghua University. His research interests include X-ray imaging, deep learning, and spectral CT reconstruction.



ZHIQIANG CHEN (Member, IEEE) received the Ph.D. degree in engineering physics from Tsinghua University, Beijing, China, in 1999. He is currently a Professor with the Department of Engineering Physics, Tsinghua University. His research interests include X-ray imaging, computed tomography, image processing, and reconstruction.



KAICHAO LIANG received the B.S. degree from Tsinghua University, Beijing, China, in 2018, where he is currently pursuing the Ph.D. degree. His current research interests include spectral CT reconstruction and deep learning.



YUXIANG XING (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from The State University of New York at Stony Brook, NY, USA, in 2003. She is currently a Professor with the Department of Engineering Physics, Tsinghua University. Her research interests include X-ray imaging, computed tomography, image reconstruction, and deep learning.

...