Unsharp Structure Guided Filtering for Self-Supervised Low-dose CT Imaging

Qianyu Wu, Xu Ji, Yunbo Gu, Jun Xiang, Guotao Quan, Baosheng Li, Jian Zhu, Gouenou Coatrieux, Jean-Louis Coatrieux, Yang Chen

Unsharp Structure Guided Filtering for Self-Supervised Low-dose CT Imaging

Qianyu Wu, Xu Ji, Yunbo Gu, Jun Xiang, Guotao Quan, Baosheng Li, Jian Zhu, Gouenou Coatrieux, *Senior Member, IEEE*, Jean-Louis Coatrieux, *Life Fellow, IEEE*, Yang Chen, *Senior Member, IEEE*

***Abstract*— Low-dose computed tomography (LDCT) imaging faces great challenges. Although supervised learning has revealed great potential, it requires sufficient and high-quality references for network training. Therefore, existing deep learning methods have been sparingly applied in clinical practice. To this end, this paper presents a novel Unsharp Structure Guided Filtering (USGF) method, which can reconstruct high-quality CT images directly from low-dose projections without clean references. Specifically, we first employ low-pass filters to estimate the structure priors from the input LDCT images. Then, inspired by classical structure transfer techniques, deep convolutional networks are adopted to implement our imaging method which combines guided filtering and structure transfer. Finally, the structure priors serve as the guidance images to alleviate over-smoothing, as they can transfer specific structural characteristics to the generated images. Furthermore, we incorporate traditional FBP algorithms into self-supervised training to enable the transformation of projection domain data to the image domain. Extensive comparisons and analyses on three datasets demonstrate that the proposed USGF has achieved superior performance in terms of noise suppression and edge preservation, and could have a significant impact on LDCT imaging in the future.**

***Index Terms*— Guided filtering, computed tomography imaging, deep learning, structure transfer.**

This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0116700, in part by the State Key Project of Research and Development Plan under Grant 2022YFC2401600, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and development Programs in Jiangsu Province of China under Grant BE2021703 and BE2022768. Yang Chen is with the Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China, and also with the Laboratory of Image Science and Technology, the School of Computer Science and Engineering, Southeast University, Nanjing 210096, China (Corresponding author: Yang Chen).

Qianyu Wu, Xu Ji, Yunbo Gu, and Yang Chen are with the Laboratory of Image Science and Technology, Southeast University, Nanjing 210096, China, and also with the Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 210096, China (e-mail: wqy021434@163.com; xuji@seu.edu.cn; 230189874@seu.edu.cn; chenyang.list@seu.edu.cn;).

Guotao Quan is with the CT RPA Department, United Imaging Healthcare Co., Ltd., Shanghai 201807, China (e-mail: guotao.quan@united-imaging.com).

Jun Xiang is with the X-Ray Department, United Imaging Healthcare Co., Ltd., Shanghai 201807, China (e-mail: Jun.xiang@united-imaging.com).

Baosheng Li and Jian Zhu are with the Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan250117, China (e-mail: baoshli1963@163.com; zhujian.cn@163.com).

G. Coatrieux is with the IMT Atlantique, Inserm, LaTIM UMR1101,Brest 29000, France (e-mail: gouenou.coatrieux@imt-atlantique.fr).

J.-L. Coatrieux is with the Laboratoire Traitement du Signal et de l'Image, Universit de Rennes 1, F-35000 Rennes, France, with the Centre de Recherche en Information Biomedicale Sino-francais, 35042 Rennes, France, and also with the National Institute for Health and Medical Research, F-35000 Rennes, France (e-mail: jeanlouis.coatrieux@univ-rennes1.fr).

I. Introduction

X-ray computed tomography (CT) is a common imaging modality, which has been widely used in clinical diagnosis, disease evaluation, and treatment planning. For example, it can be applied to visualize and target tumors, allowing oncologists to develop appropriate treatment plans. Despite such benefits, the inherent radiation exposure of CT may increase the risk of radio-induced cancers [1]. However, reducing the radiation dose usually increases the noise level in x-ray measurements, which may lead to severe degradation of reconstructed images, affecting diagnostic accuracy [2], [3].

Over the past few decades, various works have been extensively investigated for low-dose CT (LDCT) imaging, such as iterative reconstruction [4]–[6] and image post-processing [7], [8]. Although these methods are capable of reconstructing high-quality images from low-dose projections, their performance is heavily dependent on hand-crafted regularization. The computing consumption also prevents them from being deployed to the clinic. In recent years, deep learning (DL) has shown great potential in medical imaging tasks [9], [10] and has become the most popular course in the field of CT imaging. DL-based CT reconstruction is a data-driven method, which can find the optimal solution without completely relying on the exact mathematical model. Compared with traditional algorithms, it is more convenient in algorithm development and more suitable for clinical practice. However, most existing DL-based CT reconstruction methods are supervised learning [11]–[13], which means that the network training requires accurate and sufficient data pairs, such as normal-dose CT (NDCT) and LDCT pairs. Clinically, it is impractical to perform multiple NDCT and LDCT scans on the same patient. Moreover, clinical datasets are typically limited or inaccurate, and may not be adequate as the ground truth for supervised learning. Although some studies have attempted to restore degraded images without supervision, they highly rely on supervised pre-training models [14], [15] or unrealistic

assumptions, e.g., the underlying noise distribution [16], [17]. This is why DL-based methods, while achieving promising results, are rarely deployed on commercial CT scanners.

To address the above-mentioned problems in clinical practice, we propose a novel self-supervised learning network that can reconstruct high-quality CT images directly from low-dose projections. Considering that traditional DL-based methods often suffer from content blindness without clean references, we take inspiration from guided filtering [18]–[22], which can transfer the structure features of the guidance image to the target image. As shown in Fig. 1, existing self-supervised methods deal with noise and structure identically, resulting in the blurring of salient details.

The core idea of this paper is to implement guided filtering with convolutional neural networks (CNNs) and introduce unsharp structures for edge enhancement. Starting from the classical guided filtering principle, combined with the previous structure transfer methods, such as unsharp masking [23], [24], we finally derive a guided filtering formula based on deep neural networks. Unlike conventional structure transfer methods, our guidance images are learned from the network instead of just using a fixed Gaussian filter or box mean filter. Thus, it allows the network to select appropriate structural features during inference. In addition, by utilizing the conventional FBP (filtered back projection) algorithms [25], [26], we can directly perform image post-processing on the reconstructed CT images. To evaluate the proposed method, qualitative and quantitative analyses have been performed based on the 2016 AAPM Low Dose CT Grand Challenge data [27], real mice data and Siemens head data.

The main contributions of our work can be summarized as follows: **(i)** We propose a novel self-supervised learning framework for low-dose CT imaging without any clean reference and noise assumption. It is suitable for LDCT imaging under different noise levels in clinical practice. **(ii)** The proposed method employs the unsharp structure priors as the guidance image, and implements the function of guided filtering with deep neural networks. By doing so, the specific functions of this framework, i.e., image restoration and structure enhancement, can be intuitively understood. This method has a solid theoretical rationale. **(iii)** Experimental results demonstrate that it is feasible to reconstruct high-quality CT images directly from low-dose projections using self-supervised learning. **(iv)** The proposed method achieves competitive results in terms of noise suppression, structural fidelity, and visual perception improvement.

II. Related Work

A. Conventional CT Imaging Methods

Among the various conventional CT imaging methods, analytical reconstruction has achieved much attention due to its short reconstruction time, e.g., the filtered back-projection (FBP) methods [26], [30]. However, low-dose and sparse-view sampling increase the noise level in projection data, which may lead to the severe degradation of reconstructed images [31]. As a more sophisticated method, iterative reconstruction (IR) formulates the statistical model of measurements and prior knowledge of the unknown object into a cost function, and then performs iterative optimization. Conventional IR methods [32], [33] employ a data-fidelity term to represent the forward imaging model and statistical model of measurements, and a regularization term to capture prior knowledge. In particular, extensive efforts have been explored to adopt suitable regularizers to model sparse priors, such as dictionary learning [34] and sparse transform [35]. Although these IR methods have demonstrated impressive performance, the high computational complexity severely hinders their clinical practice.

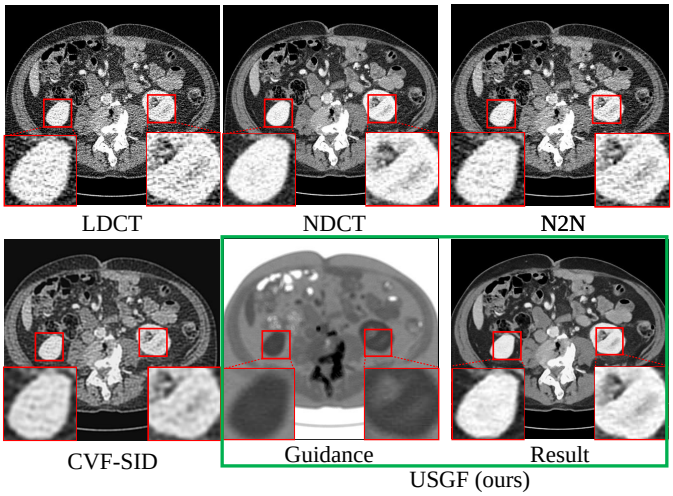


Fig. 1: Motivation of this paper. DL-based unsupervised methods suffer from content blindness. They deal with noise and structure identically, resulting in blurred details (e.g., CVF-SID [28]) or inability to remove noise (e.g., N2N [29]).

B. Supervised CT Imaging Methods

The tremendous progress of deep learning has made it increasingly popular in CT imaging. Supervised CT imaging methods rely on paired data, i.e., LDCT and corresponding NDCT. A typical approach is to employ CNN to learn the mapping from LDCT inputs to suitable NDCT outputs in the image domain. For example, Chen et al. [36] utilized CNN to suppress the noise of the reconstructed CT images. On this basis, residual learning [11] was introduced to preserve subtle details in LDCT images. Some studies have also demonstrated that adding attention mechanism [37] and adjusting the loss function [38] significantly improve the reconstruction quality. Furthermore, projection domain pre-processing methods are suggested to deal with the problems of limited-angle and sparse-view CT imaging. Lee et al. [39] employed U-net to remove artifacts by predicting the missing projections. In general, the above-mentioned methods based on supervised learning have achieved superior performance over conventional CT imaging methods, and are expected to further improve the reconstruction accuracy. However, sufficient and accurate data pairs for network training are not readily available in the clinical scenarios.

C. Unsupervised CT Imaging Methods

Recently, attempts have been made to overcome the limitation of supervised learning. For instance, Liao et al. [40] proposed an artifact disentanglement network (ADN) that does not require paired clean and degraded images for training. Lee et al. [41] simplified the structure of the ADN method and employed the attention mechanism to improve the efficiency of reconstruction. In [42], self-supervised learning was combined with cyclic adversarial learning for unpaired CT image denoising. Moreover, some works turned their attention to reconstructing CT images without clean references [43], [44]. For example, Liang et al. [44] integrated a model-based reconstruction method into self-supervised learning networks to reduce noise without ground-truth information. However, they also suffered from inefficiency in reconstruction due to highly complex networks. Furthermore, most existing unsupervised methods focus on processing low-dose CT data in the image domain. Ignoring the projection information may cause the network unable to recover the corrupted information in the projection domain, resulting in the loss of subtle details of the reconstructed CT image [45], [46].

III. METHODOLOGY

The proposed unsharp structure guided filtering is formed by an unsharp structure generator and a deep guided filtering module. As an initial motivation, we expect structural features in the form of images to guide self-supervised networks for edge enhancement. In this section, we first outline the filtering formulation to illustrate how to use the unsharp structure for guided filtering and structure transfer. Then, we introduce the proposed self-supervised LDCT imaging network using guided filtering. Finally, we present our loss function for self-supervised learning.

A. Unsharp Structure Guided Filtering

Classical guided filtering [18] assumes that there is a local linear relation between the input image I and the guidance image G. In this case, the information of G is transferred to I. Such correlation is defined by the statistics of the inputs. Supposing that w_k is a local window centered at pixel k, the statistics can be calculated by:

$$a_k = \frac{\frac{1}{|w|}\sum_{i\in w_k} I_i G_i - \overline{I}_k \overline{G}_k}{\sigma_k^2 + \epsilon}, \tag{1}$$

$$b_k = \overline{I}_k - a_k \overline{G}_k, \tag{2}$$

where $\overline{G}_k$ and $\overline{I}_k$ represent the mean of G and I in w_k, σ_k^2 is the variance of G in w_k, $|w|$ represents the number of pixels in w_k, and ϵ is a regularization term. According to the linear coefficients a_k and b_k, the predicted output P can be expressed as:

$$P_i = a_k G_i + b_k, \forall i \in w_k. \tag{3}$$

Since the entire image contains multiple windows w_k convering pixel i, the filtered results for that point are different. The final result can be obtained by taking the average of all values of P_i:

$$P_i = \frac{1}{|w|}\sum_{k\in w_i} \left(a_k G_i + b_k\right). \tag{4}$$

However, traditional guided filtering requires empirical adjustment of parameters to obtain the optimal a_k and b_k, which greatly limits the denoising performance. Moreover, estimating two coefficients simultaneously through a self-supervised learning network may exacerbate the instability of training, and lead to structure inconsistency in the predicted images. Therefore, the coefficient b_k can be eliminated by putting Eq. 2 into Eq. 4:

$$P_i = \frac{1}{|w|}\sum_{k\in w_i} a_k G_i + \frac{1}{|w|}\sum_{k\in w_i} \left(\overline{I}_k - a_k \overline{G}_k\right). \tag{5}$$

Then, we can obtain the following formulation:

$$P_i = \frac{1}{|w|}\sum_{k\in w_i} a_k \left(G_i - \overline{G}_k\right) + \frac{1}{|w|}\sum_{k\in w_i} \overline{I}_k. \tag{6}$$

In this paper, a box mean filter is employed to obtain the value of G_i^*. Thus, G_i^* is very close to its mean $\overline{G}_k$ in the window w_i. Then, Eq. 6 can be rewritten as

$$P_i = a_i^* \left(G_i - G_i^*\right) + I_i^*, \tag{7}$$

where $a_i^* = \frac{1}{|w|}\sum_{k\in w_i} a_k$, $I_i^* = \frac{1}{|w|}\sum_{k\in w_i} \overline{I}_k$. It can be observed that $(G - G^*)$ denotes the unsharp structures of the guidance images. The coefficient a^* controls the intensity of structures. Specifically, the term $a^*(G - G^*)$ allows the structural information to be transferred to the filtered image I^*.

Previous works have demonstrated the advantages of deep learning-based guided filtering in image processing. In [47], Pan et al. employed two trainable deep neural networks to solve the coefficients in Eq. 3, which is expressed as:

$$P = \mathcal{F}_a\left(I, G\right) * G + \mathcal{F}_b\left(I, G\right), \tag{8}$$

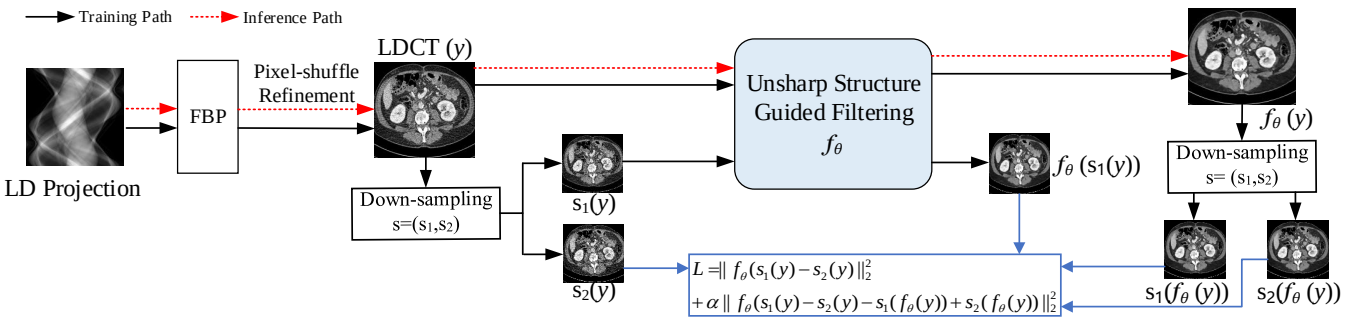
where $\mathcal{F}_a$ and $\mathcal{F}_b$ are two CNNs, $*$ represents element-wise multiplication. On this basis, Eq. 7 can be rewritten as:

$$P = \mathcal{F}_a\left(I_{us}, G_{us}\right) * G_{us} + f\left(I\right), \tag{9}$$

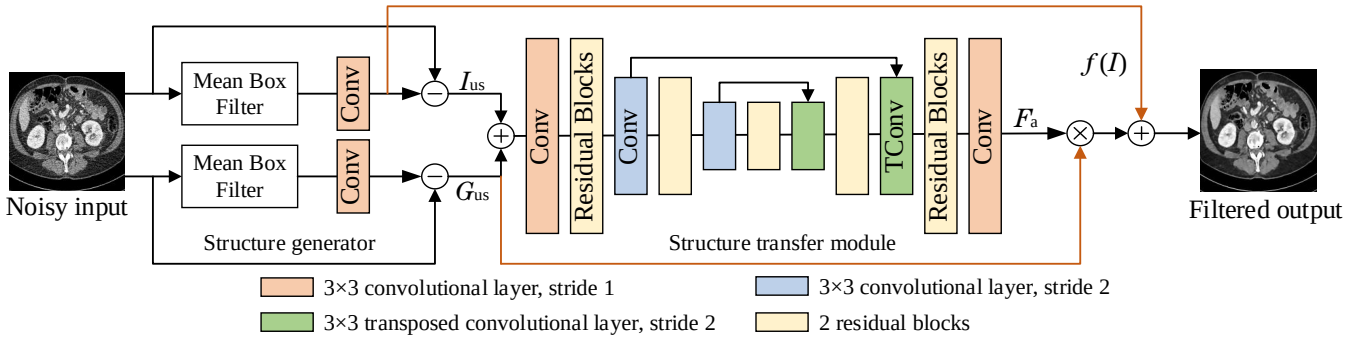
where the function $f(\cdot)$ represents a classical mean box filter, followed by a 3×3 convolutional layer. $I_{us} = I - f(I)$, $G_{us} = G - f(G)$ are the unsharp structures of the input image I and guidance image G. From this formula we can clearly know the rationale of the proposed method.

B. Network Overview

In this section, we introduce how to implement unsharp structure guided filtering for self-supervised LDCT imaging. As shown in Fig. 2(a), conventional FBP algorithms [25], [26] are employed to reconstruct CT from the original low-dose projection. The reason is two-fold. First, this paper aims to explore a reconstruction algorithm for clinical application. Conventional analytic reconstruction is very suitable to be combined with deep learning due to its simplicity and fast reconstruction speed. Second, the core idea of the proposed



(a) Overview of our method for LDCT imaging



(b) Implementation of unsharp structure guided filtering on CNNs

Fig. 2: The proposed unsharp structure guided filtering for LDCT imaging. (a) is the overview of the method. During training, the down-sampling module $s = (s_1, s_2)$ generates two similar but independent images that are half the scale of the original image. The inference process only includes the FBP operation and the proposed filtering method. (b) is the details of the proposed unsharp structure guided filtering. The structure generator adopts two box mean filters and convolutional layers to obtain the unsharp structure G_{us} and I_{us} in Eq. 9.

algorithm is to utilize guided filtering for image restoration. Therefore, the noise and artifacts that may be introduced by the FBP algorithm are acceptable, which can effectively demonstrate the reconstruction performance of the proposed method.

According to [16], [29], we suppose (x, y) is a clean-noisy image pair, if there exists an image z that is very similar to y and independent with each other, the network trained by paired images (y, z) is a reasonable approximate solution of the supervised training network using (x, y). Detailed theoretical proof can be found in [29]. Since CT reconstruction involves the filtering and back-projection operations, there is a non-negligible correlation between adjacent pixels in the reconstructed CT image. To this end, pixel-shuffle refinement strategies [48], [49] are introduced to suppress the correlation between pixels. Specifically, it first divides the original image into several regions. Second, each sub-image is refilled with noise blocks and pixel shuffled separately. Then, they are up-sampled using convolutional networks to recover the missing details. Finally, the refilled sub-images are weighted to get the final result.

In this paper, we adopt the method proposed in [29] to generate noisy training pairs. The down-sampling module $s = (s_1, s_2)$ in Fig. 2(a) generates two sub-sampled images that are half the scale of the original image. For example, in each 2×2 pixel unit of the input LDCT (y), two adjacent pixels are randomly selected as the pixels of $s_1(y)$ and $s_2(y)$, respectively. Fig. 2(b) presents the implementation of unsharp structure guided filtering using deep learning. It can be observed that the mean box filter and subsequent convolutional layer represent the function $f(\cdot)$ in Eq. 9. Specifically, we use two box mean filters with the same radius r to obtain unsharp structure G_{us} and I_{us}. Then, they are concatenated with each other and serve as inputs of the proposed structure transfer module. The backbone of this module is U-net [50] and it contains multiple 3×3 convolutional layers, transposed convolutional layers and residual blocks [51]. Each convolutional or transposed convolutional layer is followed by a ReLU activation function.

C. Reconstruction loss

In this section, we explain the theoretical foundation and optimization method in detail. Fully-supervised methods try to optimize the following term:

$$\arg\min_{\theta} \mathbb{E}_{x,y} \| f_\theta(y) - x \|_2^2 \tag{10}$$

where $y_\theta(z)$ is a neural network f_θ with a noisy input y, x is the clean target (ground-truth). We assume that y and z are independent noisy images conditioned on x, and there exists an $\varepsilon \neq 0$ such that $\mathbb{E}_{y|x}(y) = x$, $\mathbb{E}_{z|x}(z) = x + \varepsilon$. If the variance

of $z=\sigma_z^2$, we can obtain

$$\begin{aligned}
&\mathbb{E}_{y|x}\|f_\theta(y)-x\|_2^2\\
=&\mathbb{E}_{y,z|x}\|f_\theta(y)-z+z-x\|_2^2\\
=&\mathbb{E}_{y,z|x}\|f_\theta(y)-z\|_2^2+\mathbb{E}_{z|x}\|f_\theta(y)\,z-x\|_2^2\\
&+2\mathbb{E}_{y,z|x}\left(f_\theta(y)-z\right)^\top(z-x)\\
=&\mathbb{E}_{y,z|x}\|f_\theta(y)-z\|_2^2+\sigma_z^2\\
&+2\mathbb{E}_{y,z|x}\left(f_\theta(y)-x+x-z\right)^\top(z-x)\\
=&\mathbb{E}_{y,z|x}\|f_\theta(y)-z\|_2^2+\sigma_z^2\\
&+2\mathbb{E}_{y,z|x}\left(f_\theta(y)-x\right)^\top(z-x)+2\mathbb{E}_{z|x}\left(x-z\right)^\top(z-x)\\
=&\mathbb{E}_{y,z|x}\|f_\theta(y)-z\|_2^2-\sigma_z^2+2\mathbb{E}_{y,z|x}\left(f_\theta(y)-x\right)^\top(z-x)\\
=&\mathbb{E}_{y,z|x}\|f_\theta(y)-z\|_2^2-\sigma_z^2+2\varepsilon\mathbb{E}_{y|x}\left(f_\theta(y)-x\right).
\end{aligned}\tag{11}$$

When $\varepsilon\to 0$, which means the difference of the noisy inputs y, z are small enough. Then, $2\varepsilon\mathbb{E}_{y|x}\left(f_\theta(y)-x\right)\to 0$. In this case, a network trained with noisy image pairs (y,z) can be used as a reasonable approximation of a supervised training network.

In this paper, a specific down-sampling operation $s=(s_1,s_2)$ is used to generate similar and independent training pairs $(s_1(y),s_2(y))$. The proposed self-supervised optimization problem has become:

$$\arg\min_\theta \mathbb{E}_{x,y}\|f_\theta\left(s_1(y)\right)-s_2(y)\|_2^2,\tag{12}$$

where f_θ is the guided filtering based on deep neural network. Assuming that f_θ is trained with clean targets, it should be an optimal denoising network. And we have $f_\theta(y)=x$, $f_\theta\left(s(y)\right)=s(x)$. Thus, we can exploit the following ideal constraint:

$$\begin{aligned}
&\mathbb{E}_{y|x}\left\{f_\theta\left(s_1(y)\right)-s_2(y)-\left[s_1\left(f_\theta(y)\right)-s_2\left(f_\theta(y)\right)\right]\right\}\\
=&s_1(x)-\mathbb{E}_{y|x}\left\{s_2(y)\right\}-s_1(x)+s_2(x)
\end{aligned}\tag{13}$$

The last two subtraction terms are the corrections of the difference the first two terms. Finally, the optimization problem is expressed as:

$$\begin{aligned}
&\min_\theta \mathbb{E}_{x,y}\|f_\theta\left(s_1(y)\right)-s_2(y)\|_2^2\\
&+\alpha\mathbb{E}_{y|x}\|f_\theta\left(s_1(y)\right)-s_2(y)-s_1\left(f_\theta(y)\right)+s_2\left(f_\theta(y)\right)\|_2^2.
\end{aligned}\tag{14}$$

As shown in Fig. 2(a), we denote two images down-sampled by the LDCT (y) as $s_1(y)$ and $s_2(y)$, two images down-sampled by the reconstructed CT as $s_1\left(f_\theta(y)\right)$ and $s_2\left(f_\theta(y)\right)$, and the filtered $s_1(y)$ as $f_\theta\left(s_1(y)\right)$. The proposed self-supervised loss function can be expressed as:

$$\begin{aligned}
&L=\|f_\theta\left(s_1(y)\right)-s_2(y)\|_2^2\\
&+\alpha\|f_\theta\left(s_1(y)\right)-s_2(y)-s_1\left(f_\theta(y)\right)+s_2\left(f_\theta(y)\right)\|_2^2.
\end{aligned}\tag{15}$$

where α is used to balance the weight of the regularization term.

IV. EXPERIMENT AND RESULTS

A. Experimental Datasets

1) AAPM Challenge Data: The AAPM Challenge data [27], produced by Mayo Clinics, is a low-dose CT dataset containing 6687 LDCT and corresponding NDCT image pairs from 10 patients. Each CT slice is of size 512×512 pixels. The proposed network is trained using NDCT images from 8 of the patients, and the rest for testing. In this paper, a cone-beam scanning geometry is adopted to acquire the low-dose projections. The distances from the source to detector and object are set to 100cm and 50cm, respectively. The detector has 896×400 elements, and each element is 1.5×1.5 mm^2. The pixel dimension of all the CT images is 0.9×0.9 mm^2.

2) Real Mice Data: The real mice data is used to evaluate the robustness of the proposed method to different noise levels. The tube is Hamamatsu L9421-02. The tube current and voltage during scanning are 130μA and 60kVp, respectively. The detector is Dexela1512 which has 944×768 elements, and each element is 0.072×0.072 mm^2. The distances from the source to detector and object are set to 44 and 22, respectively. Projections are collected with cone-beam geometry from three mice (two for training), each containing 1000 projections. The reconstructed volume for each mouse is 872×872×600 voxels. The pixel dimension of all the CT images is 0.072×0.072 mm^2. Different levels of Poisson noise are added to the real projections to generate the low-dose projections.

3) Siemens Head Data: The Siemens head data are performed in Nanjing PLA General Hospital, China, with the approval of the institutional review board and patient consent forms. The CT images are acquired using a Siemens SOMATOM Definition Flash DECT scanner (1569 slices in total). The high-energy and low-energy of DECT scans were 140kV and 100kV. In this study, we only use the high-energy data for validation, and the original helical geometry is converted to the fan-beam geometry. The distances from the source to detector and patient are 1085.6 and 595 respectively. The resolution of the CT images is 512×512 pixels, and the pixel size is 0.4451×0.4451 mm^2.

B. Implementation Details

The proposed network was performed on the Pytorch platform with one NVIDIA RTX 3090 GPU. Adam algorithm was employed to optimize the parameters of our network and all comparison methods. Two exponential decay rates β_1 and β_2 for Adam were 0.9, 0.999, respectively. The learning rate was 1×10^{-4}, and it was reduced to 50 percent every 10 epochs. The batch size was set to 32. We performed K-fold cross validation, which divided the whole dataset into K folds equally. One of the folds is used for testing, and the remaining $K-1$ folds are used for training. We repeated K times until all folds are used as the testing set. The final result of the model is the average of K predictions. According to the number of cases contained in the AAPM challenge data (10 cases), mice data (3 cases) and Siemens head data (6 cases), we set the value of K as 5, 1 and 3 respectively.

Note that LDCT images were first reconstructed from the raw projections using the FBP algorithm (Ramp-filter), then

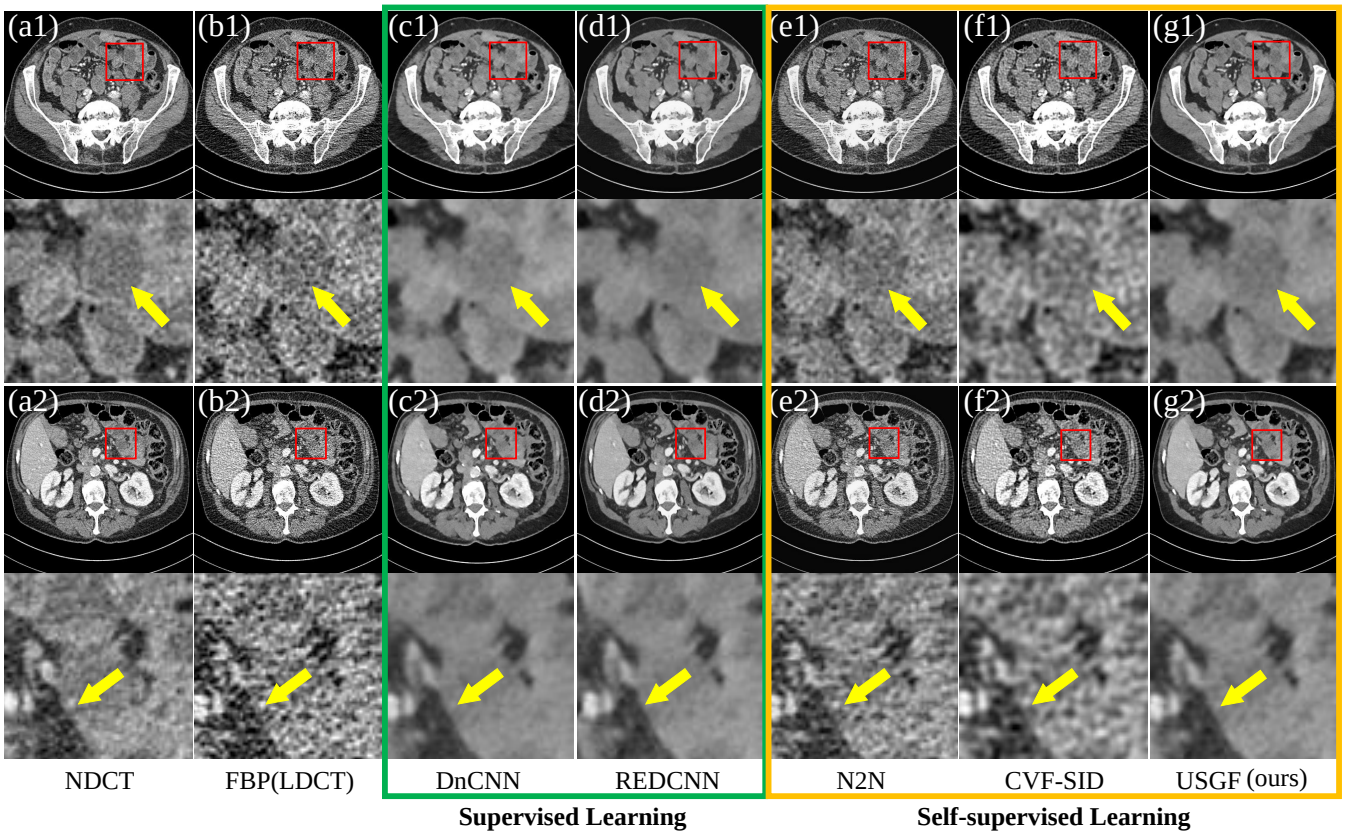


Fig. 3: Visual comparisons of different methods on the AAPM challenge data. The reconstruction results are normalized under [-160,240] Hounsfield Unit (HU). The zoomed regions marked by the red rectangles are located below the corresponding images.

noise suppression, artifact reduction, and edge enhancement were performed on the reconstructed LDCT using the proposed unsharp structure guided filtering. In this case, all the reconstructed LDCT were randomly cropped to the size of 128×128 to reduce computational cost. The radius of two box mean filters for G_{us} and I_{us} in Eq. 9 was set to $r = 3$. The hyper-parameter α in Eq. 15 was set to 0.1. We used a Poisson noise model [52] to simulate the physical effects of a monochromatic X-ray source, and its forward projection process can be expressed as:

$$I_i \sim Poisson\left\{I_{0i} \cdot \exp\left(-P_i\right) + E_i\right\}, i = 1, 2, \ldots, N, \quad (16)$$

where I_i and I_{0i} are the number and intensity of X-ray photons transmitted along the i^{th} path, respectively. P_i denotes the attenuation coefficient of X-ray beams and E_i denotes the inherent electronic noise. In this paper, different values of I_{0i} were set to simulate low-dose projections. A small value of I_{0i} results in more severe noise.

C. Experimental Results

In this section, we compared our USGF with five methods, including FBP algorithm, DnCNN [53], RED-CNN [54], CVF-SID [28] and N2N [29]. DnCNN and RED-CNN are two well-known image processing methods that adopt supervised learning. CVF-SID and N2N are two self-supervised learning methods that can be trained without clean references. Specifically, CVF-SID directly disentangles clean and noise maps from the input by leveraging various CNNs and self-supervised loss functions. N2N samples the input and target from the same noisy image to train the U-net [50] denoiser. In our experiment, all self-supervised methods do not involve any references during imaging, and these references are only used for metric calculation. We used MAE (Mean Absolute Error), PSNR (Peak Signal to Noise Ratio), and SSIM (Structural Similarity) to objectively evaluate the proposed method.

TABLE I: Comparison results of different methods on the AAPM challenge data. P.S.: Paired Supervision. S.S.: Self-Supervision. The best results for case S.S. are highlighted in **bold**.

Type	Method	PSNR (dB)	SSIM (%)	MAE (HU)
P.S.	DnCNN	44.75 ± 1.55	97.02 ± 0.72	17.27 ± 2.79
	REDCNN	45.10 ± 1.42	97.58 ± 0.81	16.42 ± 2.85
S.S.	FBP	39.99 ± 1.79	91.79 ± 2.38	30.19 ± 5.59
	N2N	42.19 ± 1.57	94.62 ± 1.66	23.52 ± 4.39
	CVF-SID	42.32 ± 1.45	95.02 ± 1.51	23.24 ± 4.01
	USGF (ours)	$\mathbf{44.82 \pm 1.38}$	$\mathbf{97.51 \pm 1.08}$	$\mathbf{16.88 \pm 3.66}$

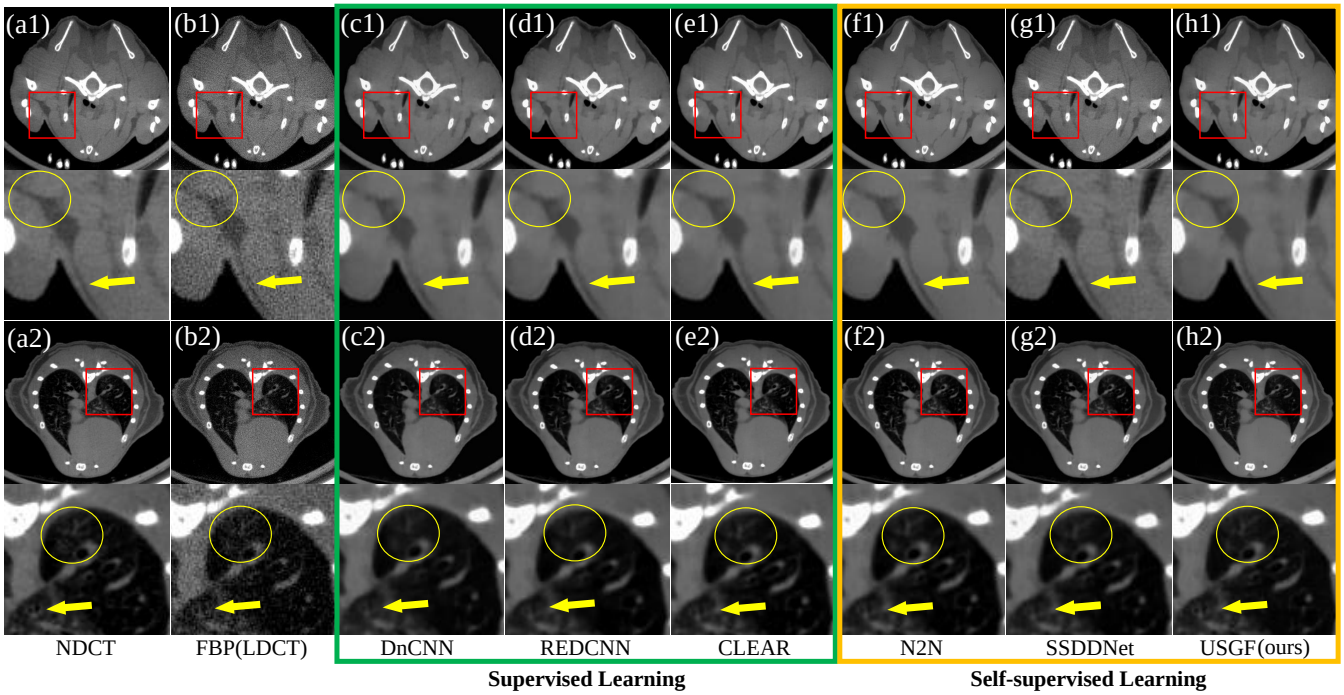


Fig. 4: Visual comparisons of different methods on the real mice data with noise level $I_0 = 5 \times 10^4$. The reconstruction results are normalized under [-600,600] Hounsfield Unit (HU). The zoomed regions marked by the red rectangles are located below the corresponding images.

TABLE II: Comparison results of different methods on the real mice data. The Poisson noise is added to the real normal-dose projections to generate the low-dose projections. The best results for case *Self-Supervision* are highlighted in **bold**.

Dose Level	Method / Index	Paired Supervision			Self-Supervision			
		DnCNN	REDCNN	CLEAR	FBP	N2N	SSDDNet	USGF (ours)
5×10^4	PSNR(dB)	38.65 ± 1.07	38.92 ± 1.11	39.33 ± 1.13	28.74 ± 1.08	38.16 ± 1.03	38.48 ± 1.09	$\mathbf{38.99 \pm 1.04}$
	SSIM(%)	94.47 ± 1.51	94.83 ± 1.53	95.05 ± 1.29	74.33 ± 5.04	92.98 ± 1.74	93.90 ± 1.63	$\mathbf{94.95 \pm 1.55}$
	MAE(HU)	6.59 ± 1.10	5.97 ± 1.14	5.69 ± 1.13	20.01 ± 3.82	6.96 ± 1.13	6.84 ± 1.14	$\mathbf{6.03 \pm 1.14}$
7.5×10^4	PSNR(dB)	39.26 ± 1.07	39.63 ± 1.15	39.87 ± 1.17	30.38 ± 1.08	39.14 ± 1.04	39.12 ± 1.15	$\mathbf{39.42 \pm 1.07}$
	SSIM(%)	94.97 ± 1.43	95.31 ± 1.44	95.40 ± 1.24	78.65 ± 4.58	95.15 ± 1.66	94.33 ± 1.51	$\mathbf{95.30 \pm 1.45}$
	MAE(HU)	6.12 ± 1.06	5.55 ± 1.12	5.39 ± 1.09	16.51 ± 3.18	5.99 ± 1.07	5.78 ± 1.10	$\mathbf{5.62 \pm 1.09}$
1×10^5	PSNR(dB)	39.24 ± 1.10	39.79 ± 1.16	40.14 ± 1.18	31.52 ± 1.09	39.29 ± 1.04	$\mathbf{39.74 \pm 1.18}$	39.55 ± 1.07
	SSIM(%)	95.28 ± 1.36	95.45 ± 1.39	95.64 ± 1.19	81.41 ± 4.18	94.33 ± 1.44	94.62 ± 1.37	95.43 ± 1.29
	MAE(HU)	6.33 ± 1.06	5.47 ± 1.08	5.27 ± 1.05	14.47 ± 2.81	6.22 ± 1.07	$\mathbf{5.51 \pm 1.06}$	5.92 ± 1.09
2.5×10^5	PSNR(dB)	41.03 ± 1.15	41.09 ± 1.20	41.28 ± 1.21	34.84 ± 1.08	39.51 ± 1.14	40.69 ± 1.24	$\mathbf{40.88 \pm 1.10}$
	SSIM(%)	96.39 ± 1.10	96.38 ± 1.12	96.94 ± 1.02	88.46 ± 2.81	96.02 ± 1.11	96.19 ± 1.12	$\mathbf{96.43 \pm 1.13}$
	MAE(HU)	5.01 ± 0.93	4.75 ± 0.97	4.50 ± 1.05	9.83 ± 1.91	5.83 ± 0.94	5.09 ± 0.95	$\mathbf{4.80 \pm 0.98}$
5×10^5	PSNR(dB)	41.50 ± 1.05	42.32 ± 1.24	42.26 ± 1.25	36.93 ± 1.07	41.30 ± 1.17	41.45 ± 1.26	$\mathbf{41.70 \pm 1.18}$
	SSIM(%)	95.81 ± 0.87	97.09 ± 0.81	97.06 ± 0.93	91.90 ± 2.03	96.76 ± 0.93	96.07 ± 0.93	$\mathbf{96.91 \pm 0.95}$
	MAE(HU)	4.77 ± 0.79	4.14 ± 0.88	4.18 ± 0.90	7.72 ± 1.49	5.58 ± 0.85	4.84 ± 0.87	$\mathbf{4.50 \pm 0.82}$

1) AAPM Challenge Data Results: The self-supervised learning method is proposed to deal with clinical CT imaging tasks. Note that, the proposed method adopts a conventional FBP algorithm to reconstruct CT from the original low-dose projection. Therefore, we denote the results obtained by the FBP algorithm as LDCT. The quantitative comparisons of different methods on the AAPM data are listed in Table I. In general, all methods exhibit different levels of edge-preserving and denoising performance. Based on the FBP algorithm, the proposed USGF method improves PSNR and SSIM by 4.84dB and 5.72%, and MAE is reduced by 13.31. It can be seen that the performance of our USGF is superior to all self-supervised methods, as well as the classical supervised learning method DnCNN. Specifically, the average PSNR and SSIM of USGF were 2.50dB and 2.49% higher than CVF-SID, respectively, and the average MAE was reduced by 6.36.

To further analyze the effectiveness of our method, the visual comparisons of different methods are illustrated in Fig. 3. We present the zoomed regions of interest marked by red rectangles, as shown below each image. The reconstruction

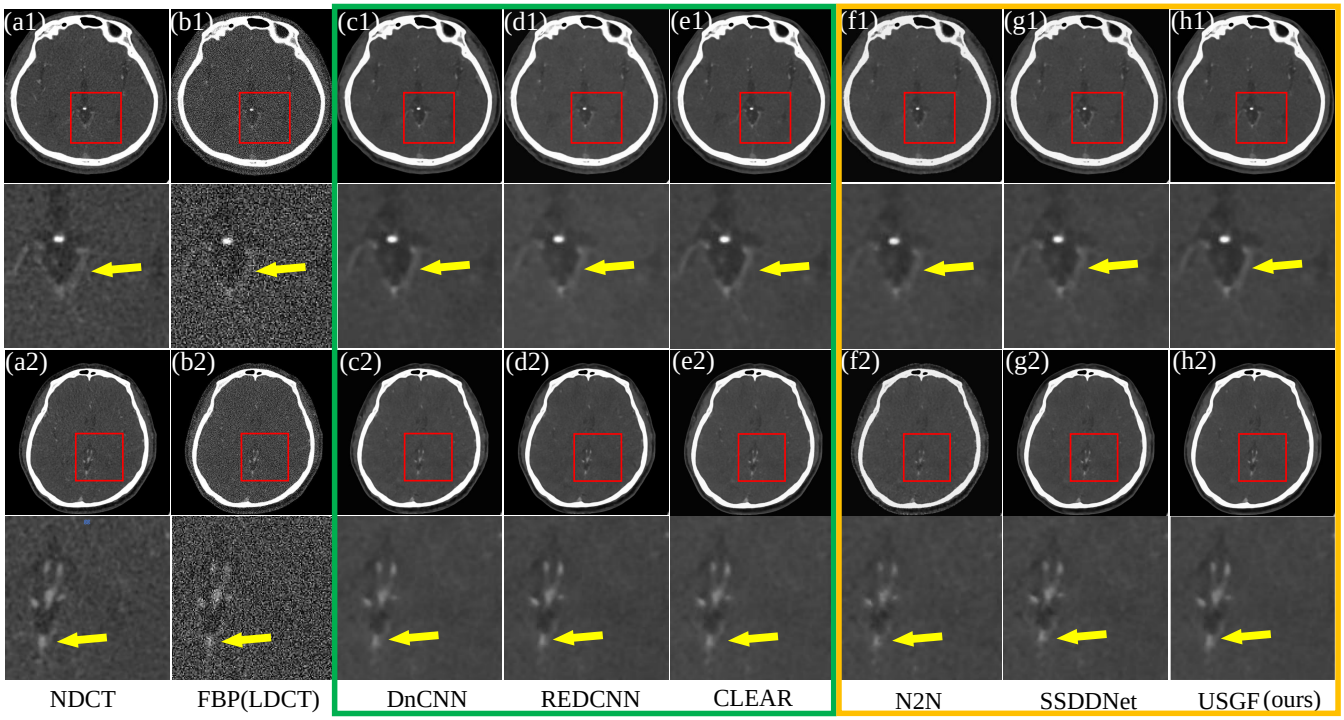


Fig. 5: Visual comparisons of different methods on the Siemens head data with noise level $I_0 = 5 \times 10^4$. The reconstruction results are normalized under [-100,400] Hounsfield Unit (HU). The zoomed regions marked by the red rectangles are located below the corresponding images.

TABLE III: Comparison results of different methods on the Siemens head data. The Poisson noise is added to the real normal-dose projections to generate the low-dose projections. The best results for case *Self-Supervision* are highlighted in **bold**.

Dose Level	Method / Index	Paired Supervision			Self-Supervision			
		DnCNN	REDCNN	CLEAR	FBP	N2N	SSDDNet	USGF (ours)
5×10^4	PSNR(dB)	36.91 ± 3.17	37.20 ± 3.30	37.56 ± 3.33	22.37 ± 5.19	36.80 ± 3.10	37.01 ± 3.25	$\mathbf{37.19 \pm 3.19}$
	SSIM(%)	94.01 ± 3.77	94.31 ± 3.72	94.48 ± 3.62	69.53 ± 8.50	93.25 ± 3.71	93.34 ± 3.67	$\mathbf{93.78 \pm 3.69}$
	MAE(HU)	3.61 ± 1.85	3.31 ± 1.82	3.15 ± 1.73	21.86 ± 10.85	3.37 ± 1.79	3.28 ± 1.79	$\mathbf{3.23 \pm 1.77}$
7.5×10^4	PSNR(dB)	37.51 ± 3.23	38.03 ± 3.36	38.31 ± 3.46	23.88 ± 5.30	37.41 ± 3.26	37.69 ± 3.27	$\mathbf{37.91 \pm 3.36}$
	SSIM(%)	94.09 ± 3.30	94.85 ± 3.45	94.98 ± 3.31	71.62 ± 7.47	93.82 ± 3.39	94.47 ± 3.37	$\mathbf{94.78 \pm 3.36}$
	MAE(HU)	3.62 ± 1.67	3.08 ± 1.71	2.91 ± 1.63	18.46 ± 9.83	3.07 ± 1.66	$\mathbf{2.99 \pm 1.64}$	3.01 ± 1.66
1×10^5	PSNR(dB)	38.25 ± 3.39	38.43 ± 3.43	38.85 ± 3.52	24.94 ± 5.36	38.33 ± 3.32	38.45 ± 3.29	$\mathbf{38.60 \pm 3.44}$
	SSIM(%)	95.07 ± 3.24	95.13 ± 3.22	95.29 ± 3.13	73.07 ± 6.92	95.16 ± 3.19	95.03 ± 3.15	$\mathbf{95.20 \pm 3.11}$
	MAE(HU)	3.03 ± 1.62	2.91 ± 1.62	2.79 ± 1.57	16.36 ± 8.55	2.96 ± 1.59	2.85 ± 1.57	$\mathbf{2.83 \pm 1.58}$
2.5×10^5	PSNR(dB)	39.93 ± 3.55	40.28 ± 3.65	40.52 ± 3.74	28.14 ± 5.44	39.99 ± 3.49	39.71 ± 3.37	$\mathbf{40.31 \pm 3.65}$
	SSIM(%)	96.17 ± 2.56	96.28 ± 2.51	96.35 ± 2.47	77.45 ± 6.14	96.24 ± 2.49	96.29 ± 2.50	$\mathbf{96.32 \pm 2.48}$
	MAE(HU)	2.57 ± 1.39	2.39 ± 1.37	2.35 ± 1.28	11.27 ± 7.33	2.53 ± 1.35	2.52 ± 1.38	$\mathbf{2.37 \pm 1.36}$
5×10^5	PSNR(dB)	41.05 ± 3.63	41.39 ± 3.72	41.71 ± 3.86	30.24 ± 5.40	41.13 ± 3.58	41.28 ± 3.75	$\mathbf{41.22 \pm 3.25}$
	SSIM(%)	96.90 ± 2.08	96.96 ± 2.06	97.04 ± 2.02	80.55 ± 5.79	95.21 ± 2.62	96.50 ± 2.04	$\mathbf{96.77 \pm 1.66}$
	MAE(HU)	2.33 ± 1.26	2.13 ± 1.22	2.07 ± 1.20	8.80 ± 5.72	2.24 ± 1.21	$\mathbf{2.12 \pm 1.05}$	2.22 ± 1.19

results are normalized under [-160,240] Hounsfield Unit (HU) for a better visual effect. It can be noticed that the images produced by N2N and CVF-SID are severely degraded, as pointed by the yellow arrows in Fig. 3(e2) and Fig. 3(f2). The supervised-learning methods DnCNN and REDCNN suffer from the blurring effect, as pointed by the yellow arrows in Fig. 3(c2) and Fig. 3(d2). Moreover, by integrating the unsharp structure priors to the self-supervised framework, USGF demonstrates promising performance in structural fidelity and has better visual perception compared to DnCNN and REDCNN. For instance, the zoomed region of Fig. 3(g2) has richer texture details compared to other methods.

2) Real Mice Data Results: In real data experiments, the Poisson noises with different intensities were added to the real normal-dose projections to generate the low-dose projections. Specifically, the I_{0i} in Eq. 16 was set to 5.0×10^4, 1.0×10^5

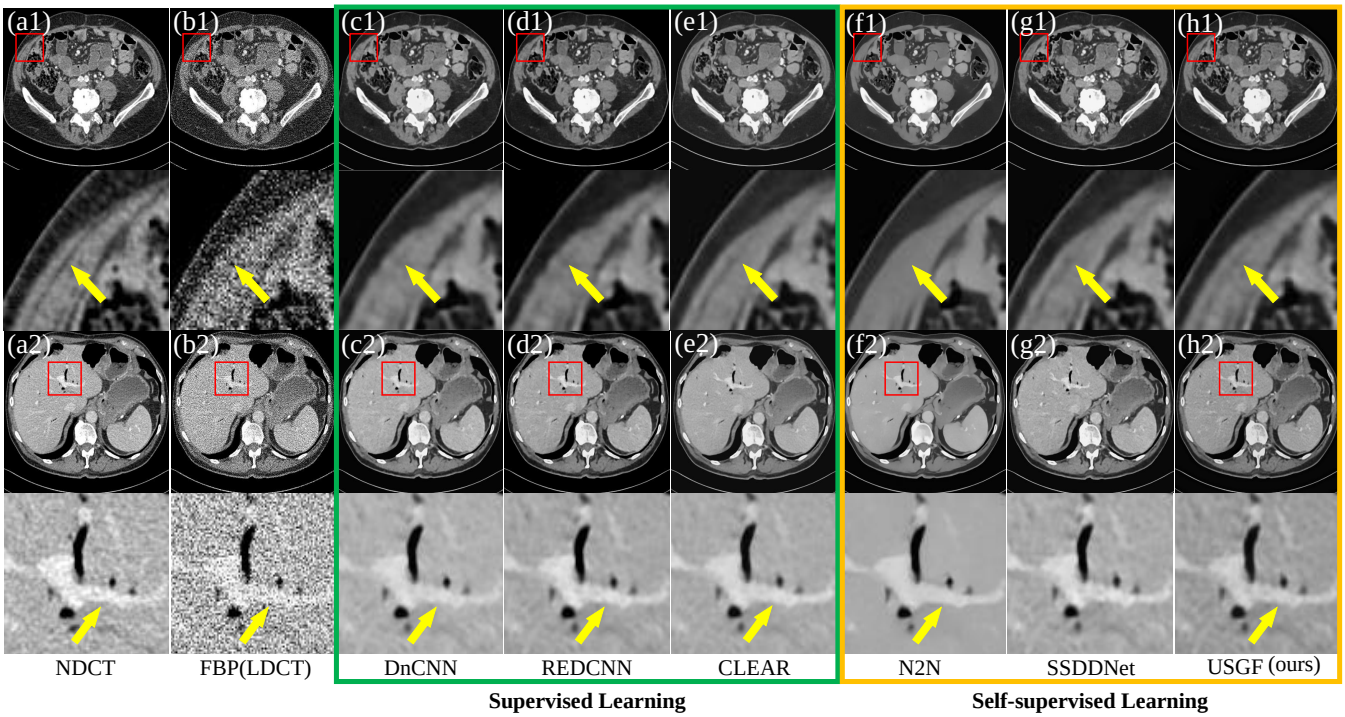


Fig. 6: Visual comparisons of different methods on the AAPM challenge data with noise level $I_{0i} = 5 \times 10^4$. The reconstruction results are normalized under [-160,240] Hounsfield Unit (HU). The zoomed regions marked by the red rectangles are located below the corresponding images.

TABLE IV: Comparison results of different methods on the AAPM challenge data with different noise levels. The Poisson noise is added to the simulated AAPM projections to generate the low-dose projections. The best results for case *Self-Supervision* are highlighted in **bold**

Dose Level	Method / Index	Paired Supervision			Self-Supervision			
		DnCNN	REDCNN	CLEAR	FBP	N2N	SSDDNet	USGF (ours)
5×10^4	PSNR(dB)	29.57 ± 1.32	29.87 ± 1.31	29.96 ± 1.26	18.24 ± 1.32	28.03 ± 1.47	29.63 ± 1.53	$\mathbf{29.96 \pm 1.59}$
	SSIM(%)	86.81 ± 2.55	87.01 ± 2.36	88.18 ± 1.58	66.86 ± 4.38	81.42 ± 3.01	84.10 ± 2.54	$\mathbf{86.92 \pm 2.63}$
	MAE(HU)	6.94 ± 1.27	6.78 ± 1.25	6.67 ± 1.14	26.21 ± 4.85	8.83 ± 1.17	7.53 ± 1.21	$\mathbf{6.48 \pm 1.29}$
7.5×10^4	PSNR(dB)	29.52 ± 1.09	30.06 ± 1.25	30.71 ± 1.39	19.70 ± 1.34	29.61 ± 1.36	$\mathbf{30.32 \pm 1.33}$	30.29 ± 1.37
	SSIM(%)	85.93 ± 1.89	87.35 ± 2.20	88.54 ± 2.31	69.85 ± 4.09	86.55 ± 3.13	87.81 ± 3.21	$\mathbf{87.91 \pm 3.07}$
	MAE(HU)	7.60 ± 1.13	7.04 ± 1.18	6.28 ± 1.20	22.11 ± 4.13	7.30 ± 1.22	$\mathbf{6.65 \pm 1.18}$	7.02 ± 1.23
1×10^5	PSNR(dB)	30.70 ± 1.39	31.05 ± 1.43	31.44 ± 1.32	20.75 ± 1.36	29.64 ± 1.40	30.79 ± 1.81	$\mathbf{30.95 \pm 1.27}$
	SSIM(%)	88.81 ± 1.97	89.44 ± 2.08	90.44 ± 1.92	72.06 ± 3.86	85.11 ± 2.76	89.19 ± 2.07	$\mathbf{88.93 \pm 2.14}$
	MAE(HU)	6.33 ± 1.12	5.95 ± 1.15	5.72 ± 1.18	19.59 ± 3.68	7.72 ± 1.20	6.71 ± 1.28	$\mathbf{6.09 \pm 1.14}$
2.5×10^5	PSNR(dB)	32.02 ± 1.39	32.27 ± 1.24	32.78 ± 1.10	23.94 ± 1.35	30.86 ± 1.30	31.07 ± 1.48	$\mathbf{31.20 \pm 1.28}$
	SSIM(%)	90.74 ± 1.54	92.84 ± 1.34	93.01 ± 1.41	79.02 ± 2.99	87.07 ± 2.62	90.01 ± 1.55	$\mathbf{91.45 \pm 1.52}$
	MAE(HU)	5.98 ± 0.82	5.33 ± 0.94	5.20 ± 0.87	13.53 ± 2.53	6.63 ± 1.15	6.01 ± 1.17	$\mathbf{5.18 \pm 1.09}$
5×10^5	PSNR(dB)	33.10 ± 1.45	33.32 ± 1.35	33.40 ± 1.38	26.06 ± 0.06	32.02 ± 1.23	32.36 ± 1.50	$\mathbf{32.88 \pm 1.45}$
	SSIM(%)	91.46 ± 1.30	93.39 ± 1.16	93.64 ± 1.19	83.52 ± 0.24	90.82 ± 2.24	92.55 ± 1.25	$\mathbf{92.65 \pm 1.21}$
	MAE(HU)	5.17 ± 0.90	4.66 ± 0.85	4.56 ± 0.86	10.58 ± 0.12	5.58 ± 1.03	5.89 ± 0.87	$\mathbf{5.42 \pm 0.85}$

and 5.0×10^5, respectively. The quantitative results are listed in Table II. It can be seen that the proposed USGF method outperforms other self-supervised learning methods at five noise levels. For example, USGF surpasses N2N by an average PSNR of 0.83dB, and has an average SSIM gain of about 1.97% over N2N when the noise of $I_{0i} = 5 \times 10^4$. Moreover, as the noise level increases, the performance of our method gradually exceeds that of the supervised learning method REDCNN. The reason is two-fold. First, supervised learning methods usually minimize the Euclidean distance between LDCT and NDCT images, which can lead to the blurring effect and subtle structural distortion when there is severe noise near tissue edges. Second, the proposed method employs the unsharp structure priors as the guidance, which greatly preserves details such as edges and textures in the original image.

TABLE V: Influence of unsharp structure guidance on the AAPM challenge data with different noise levels. (w/o: without)

Dose Level	Index	Method		
		FBP	USGF (w/o guidance)	USGF
5×10^4	PSNR(dB)	18.24 ± 1.32	28.94 ± 1.49	**29.96 ± 1.59**
	SSIM(%)	66.86 ± 4.38	80.36 ± 2.97	**86.92 ± 2.63**
	MAE(HU)	26.21 ± 4.85	8.16 ± 1.33	**6.48 ± 1.29**
1×10^5	PSNR(dB)	20.75 ± 1.36	30.12 ± 1.30	**30.95 ± 1.27**
	SSIM(%)	72.06 ± 3.86	86.53 ± 2.25	**88.93 ± 2.14**
	MAE(HU)	19.59 ± 3.68	6.64 ± 1.20	**6.09 ± 1.14**
5×10^5	PSNR(dB)	26.06 ± 0.06	31.81 ± 1.52	**32.88 ± 1.45**
	SSIM(%)	83.52 ± 0.24	91.48 ± 1.28	**92.65 ± 1.21**
	MAE(HU)	10.58 ± 0.12	5.74 ± 0.92	**5.42 ± 0.85**

The reconstruction results of real mice data with noise level $I_{0i} = 5 \times 10^4$ are present in Fig. 4. It can be observed that the images produced by FBP were seriously polluted by the noise, especially in Fig. 4 (b2). Although the REDCNN method can suppress the noise well, the edges pointed by yellow arrows cannot be well preserved, as illustrated in Fig. 4 (d2). The N2N method prevents edge degradation but introduces slight artifacts, as marked in yellow circles in Fig. 4 (e2). Compared with these methods, the proposed USGF can effectively suppress noise and preserve sharp edges.

3) Siemens Head Data Results: The other was discussed to evaluate the proposed method. In this study, the original geometry helical geometry was converted to the fan-beam geometry. The different levels of Poisson noises were also added to the simulated projections to generate the low-dose projections. Visual comparisons of different methods are show in Fig. 5. It can be seen that USGF and CLEAR can effectively preserve the edge details of the tissue. In contrast, the details of the images generated by DNCNN, N2N, and SSDDNet methods are severely damaged, as marked by the red rectangles in Fig. 5 (c2), Fig. 5 (f2) and Fig. 5 (g2). The quantitative results are listed in Table III. Among all the self-supervised methods, the effect of our method is closest to the new emerging supervised method. However, due to the lack of clean images, although the quantitative results of our USGF exceeds DNCNN, it still falls short of the latest supervised learning method CLEAR.

4) Robustness to Noise: In this section, we evaluate the robustness of our method to different noise levels on the AAPM data. The Poisson noises with different intensities were added to the simulated projections acquired by a cone-beam imaging scanning geometry. As listed in Table IV, the proposed method outperforms other self-supervised learning methods on five noise levels, significantly improving the reconstruction quality without clean references. Moreover, our method can not only preserve richer subtle textures in the case of severe noise, but also mitigate the blurring effect and maintain a better visual perception. Fig. 6 presents several examples that demonstrate the edge-preserving capability of the proposed method. For example, the edge indicated by the yellow arrow in Fig. 6 (b1) is degraded by noise. Compared with REDCNN and DnCNN, the image reconstructed by the proposed USGF contains finer textures, as shown in Fig. 6 (g1).

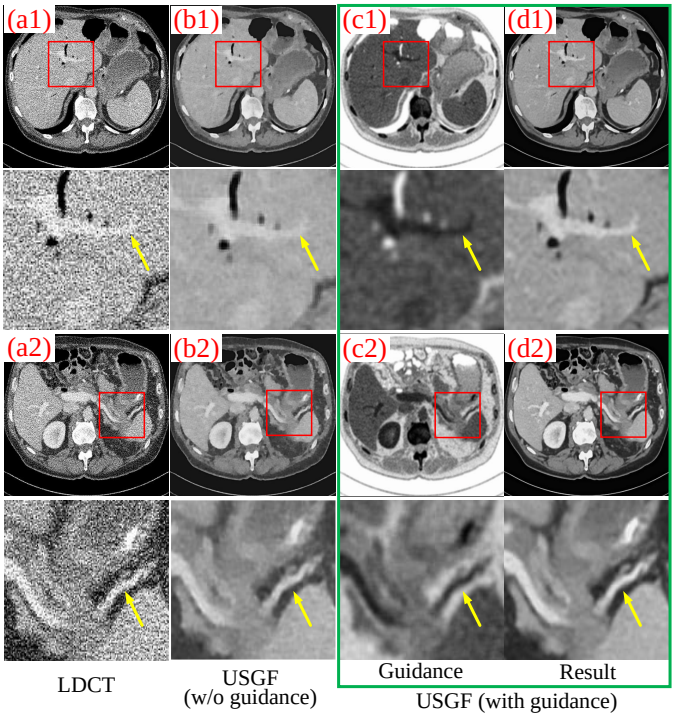


Fig. 7: Visual results of the AAPM challenge data (noise level $I_{0i} = 5 \times 10^4$) with and without guidance images. The reconstruction results are normalized under [-160,240] Hounsfield Unit (HU). The zoomed regions marked by the red rectangles are located below the corresponding images. (w/o: without)

D. Ablation Study

The inherent property of self-supervised learning makes them suffer from content blindness, which means that they treat noise and structures identically, resulting in the blurring of edge details. In this section, we analyze the impact of unsharp structure priors on the final LDCT imaging and edge preservation. For convenience, we removed the structure

generator in Fig. 2(b). That is, the unsharp structure images were replaced by the original noisy images. It can be observed that the reconstruction results were improved after introducing the guidance images. Furthermore, when the noise becomes heavier, the unsharp structure priors can improve the reconstruction results more significantly. Specifically, The SSIM metric improved by only 1.17% at $I_{0i} = 5 \times 10^5$, and by 6.56% at $I_{0i} = 5 \times 10^4$.

Fig. 7 presents the visual results of the AAPM challenge data with and without the guidance of the unsharp sturcture information. The zoomed regions in the red rectangle are present below the corresponding images. It can be seen that the guidance images produced by our method have sharp tissue edges, as pointed by the yellow arrow in Fig. 7(b2). However, the edges of the same region in LDCT are difficult to distinguish, e.g., Fig. 7(a2). Therefore, the reconstructed image without the guidance of unsharp structure priors may lose subtle details, especially in Fig. 7(c2).

E. Sensitivity Analysis of the Reconstruction Loss

The parameter α in Eq. 15 is used to balance the weight of the regularization term. The sensitivity analysis on the AAPM challenge data with noise level $I_{0i} = 5 \times 10^4$ is shown in Fig. 8. It can be seen that when α=0.1, the model can be trained stably. As α increases, the network training is more likely to suffer mode collapse. Therefore, we select a smaller value, i.g., α=0.1.

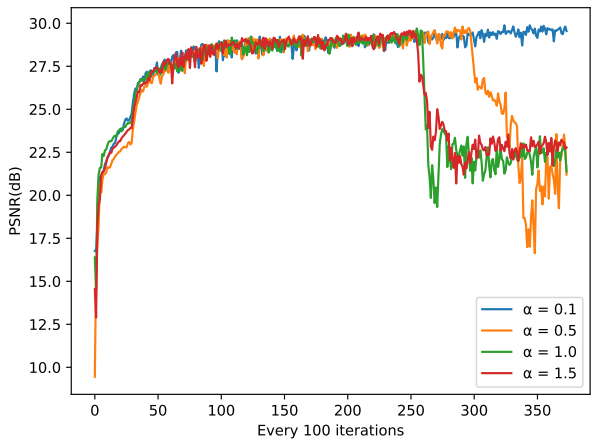


Fig. 8: Sensitivity analysis of the reconstruction loss on the AAPM challenge data with noise level $I_{0i} = 5 \times 10^4$.

V. Discussions and Conclusion

In the field of low-dose CT imaging, existing deep learning methods have achieved satisfactory results. However, limited by training data, most of them are rarely deployed in the clinic. In this paper, we have shown the proposed self-supervised learning method capable of reconstructing CT images from low-dose projections without clean references. Inspired by traditional guided filtering, we employ deep neural networks to implement unsharp structure guided filtering (USGF). This method has a clear theoretical basis. By introducing structural priors as guidance, tissue edges and other subtle details in the guidance image can be transferred into the reconstruction results for edge enhancement.

As an extension of N2N [29], the proposed USGF method significantly improves the reconstruction performance and restores more realistic anatomical details. There are two main reasons. First, the proposed method introduces additional unsharp structure priors to recover accurate edge features. Second, N2N requires pairs of independent noisy images, while the filtered back-projection operation leads to the correlation between pixels of reconstructed CT images. Our method employs the pixel-shuffle refinement strategy to suppress the correlation between pixels. In addition, USGF also outperforms the classical supervised learning methods DnCNN and REDCNN on partial quantization metrics when the images are severely affected by noise. Experimental results show that the unsharp structure priors contribute to improving image quality and statistical properties. Compared to DnCNN and REDCNN, we can see that the USGF framework helps to suppress the excessive blurring effect that may be caused by supervised learning [55], [56]. Other supervised learning strategies also can be used to reduce noise at the cost of losing critical features. The associated PSNR and SSIM metrics are slightly increased compared to LDCT, but they are much lower than the results produced by our method. In theory, self-supervised learning often suffers from content blindness and may produce images that are severely distorted or blurry in detail. This is why structure priors should be added for edge enhancement.

It should be noted that we did not perform additional processing on the projection data to ensure its integrity. The strong fitting ability of neural networks should be applied to the projection domain to further improve the performance. In addition, the proposed self-supervised learning relies on rich structure prior knowledge. The structure priors generator based on mean box filtering may not be able to remove severe artifacts, resulting in the transfer of artifacts from the guidance image to the reconstructed image. This could be the next step in our work.

VI. Acknowledgment

The authors would like to thank Dr. C. McCollough of the Mayo Clinic (Rochester, MN, USA) for providing clinical data, as agreed under the American Association of Physicists in Medicine.

References

[1] I. Shuryak, R. K. Sachs, and D. J. Brenner, "Cancer Risks After Radiation Exposure in Middle Age," *JNCI: Journal of the National Cancer Institute*, vol. 102, no. 21, pp. 1628–1636, 11 2010.

[2] E. Marfo, N. G. Anderson, A. P. H. Butler, N. Schleich, P. Carbonez, J. Damet, C. Lowe, J. Healy, A. I. Chernoglazov, M. Moghiseh, M. Collaboration, and A. Y. Raja, "Assessment of material identification errors, image quality, and radiation doses using small animal spectral photon-counting ct," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 4, pp. 578–587, 2021.

[3] H. Barrett, K. Myers, C. Hoeschen, M. Kupinski, and M. Little, "Task-based measures of image quality and their relation to radiation dose and patient risk," *Physics in Medicine and Biology*, vol. 60, no. 2, pp. R1–R75, 2015.

[4] J. Liu, Y. Hu, J. Yang, Y. Chen, H. Shu, L. Luo, Q. Feng, Z. Gui, and G. Coatrieux, "3d feature constrained reconstruction for low-dose ct imaging," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 5, pp. 1232–1247, 2018.

[5] D. Hu, W. Wu, M. Xu, Y. Zhang, J. Liu, R. Ge, Y. Chen, L. Luo, and G. Coatrieux, "Sister: Spectral-image similarity-based tensor with enhanced-sparsity reconstruction for sparse-view multi-energy ct," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 477–490, 2020.

[6] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105–117, 1972. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022519372901804

[7] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Coatrieux, and W. Chen, "Artifact suppressed dictionary learning for low-dose ct image processing," *IEEE Transactions on Medical Imaging*, vol. 33, no. 12, pp. 2271–2292, 2014.

[8] J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, and W. Chen, "Low-dose computed tomography image restoration using previous normal-dose scan," *Medical Physics*, vol. 38, no. 10, pp. 5713–5731, 2011.

[9] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.

[10] M. Hoffmann, A. Brost, M. Koch, F. Bourier, A. Maier, K. Kurzidim, N. Strobel, and J. Hornegger, "Electrophysiology catheter detection and reconstruction from two views in fluoroscopic images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 567–579, 2016.

[11] 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 Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[12] 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 Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, 2018.

[13] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose ct via convolutional neural network," *Biomedical Optics Express*, vol. 8, pp. 679–694, 01 2017.

[14] D. Won, E. Jung, S. An, P. Chikontwe, and S. H. Park, "Low-dose ct denoising using pseudo-ct image pairs," in *Predictive Intelligence in Medicine*, I. Rekik, E. Adeli, S. H. Park, and J. Schnabel, Eds. Cham: Springer International Publishing, 2021, pp. 1–10.

[15] A. Haque, A. Wang, and A.-A.-Z. Imran, "Window-level is a strong denoising surrogate," in *Machine Learning in Medical Imaging*, C. Lian, X. Cao, I. Rekik, X. Xu, and P. Yan, Eds. Cham: Springer International Publishing, 2021, pp. 457–466.

[16] N. Moran, D. Schmidt, Y. Zhong, and P. Coady, "Noisier2noise: Learning to denoise from unpaired noisy data," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12 061–12 069.

[17] C. Niu, M. Li, X. Guo, and G. Wang, "Self-supervised dual-domain network for low-dose CT denoising," in *Developments in X-Ray Tomography XIV*, B. Müller and G. Wang, Eds., vol. 12242, International Society for Optics and Photonics. SPIE, 2022, p. 122420H. [Online]. Available: https://doi.org/10.1117/12.2633197

[18] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[19] B. Zhang and J. P. Allebach, "Adaptive bilateral filter for sharpness enhancement and noise removal," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 664–678, 2008.

[20] J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 428–438, 2016.

[21] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep joint image filtering," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 154–169.

[22] L. Yijun, H. JiaBin, A. Narendra, and Y. Ming-Hsuan, "Joint image filtering with deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1909–1923, 2019.

[23] G. Deng, "A generalized unsharp masking algorithm," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1249–1261, 2011.

[24] W. Ye and K.-K. Ma, "Blurriness-guided unsharp masking," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4465–4477, 2018.

[25] Silver and D. Michael, "A method for including redundant data in computed tomography," *Medical Physics*, vol. 27, no. 4, p. 773, 2000.

[26] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, 1984.

[27] AAPM, "Low dose ct grand challenge," http://www.aapm.org/GrandChallenge/LowDoseCT.

[28] R. Neshatavar, M. Yavartanoo, S. Son, and K. M. Lee, "Cvf-sid: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from image," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[29] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2neighbor: Self-supervised denoising from single noisy images," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 14 776–14 785.

[30] L. A. Shepp and B. F. Logan, "The fourier reconstruction of a head section," *IEEE Transactions on Nuclear Science*, vol. 21, no. 3, pp. 21–43, 1974.

[31] J. Wang, T. Li, J. Liang, and L. Xing, "Dose reduction in kilovotage conebeam computed tomography for radiation therapy," *Medical Physics*, vol. 35, no. 6, pp. 2938–2939, 2008.

[32] J. Anderson, B. Mair, M. Rao, and C.-H. Wu, "Weighted least-squares reconstruction methods for positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 159–165, 1997.

[33] P. Sukovic and N. Clinthorne, "Penalized weighted least-squares image reconstruction for dual energy x-ray transmission tomography," *IEEE Transactions on Medical Imaging*, vol. 19, no. 11, pp. 1075–1081, 2000.

[34] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.-L. Coatrieux, and C. Toumoulin, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Physics in Medicine and Biology*, vol. 58, no. 16, pp. 5803–5820, aug 2013. [Online]. Available: https://doi.org/10.1088/0031-9155/58/16/5803

[35] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler, "Pwls-ultra: An efficient clustering and learning-based approach for low-dose 3d ct image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1498–1510, 2018.

[36] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose ct denoising with convolutional neural network," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 143–146.

[37] Z. Huang, Z. Chen, Q. Zhang, G. Quan, M. Ji, C. Zhang, Y. Yang, X. Liu, D. Liang, H. Zheng, and Z. Hu, "Cagan: A cycle-consistent generative adversarial network with attention for low-dose ct imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1203–1218, 2020.

[38] 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 Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[39] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109–119, 2019.

[40] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo, "Adn: Artifact disentanglement network for unsupervised metal artifact reduction," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 634–643, 2020.

[41] J. Lee, J. Gu, and J. C. Ye, "Unsupervised ct metal artifact learning using attention-guided -cyclegan," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3932–3944, 2021.

[42] K. Lee and W.-K. Jeong, "Iscl: Interdependent self-cooperative learning for unpaired image denoising," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3238–3248, 2021.

[43] K. Kim, S. Soltanayev, and S. Y. Chun, "Unsupervised training of denoisers for low-dose ct reconstruction without full-dose ground truth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1112–1125, 2020.

[44] K. Liang, L. Zhang, Y. Yang, H. Yang, and Y. Xing, "A self-supervised deep learning network for low-dose ct reconstruction," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, 2018, pp. 1–4.

[45] D. Zeng, L. Wang, M. Geng, S. Li, Y. Deng, Q. Xie, D. Li, H. Zhang, Y. Li, Z. Xu, D. Meng, and J. Ma, "Noise-generating-mechanism-driven unsupervised learning for low-dose ct sinogram recovery," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 6, no. 4, pp. 404–414, 2022.

[46] J. He, Y. Wang, and J. Ma, "Radon inversion via deep learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2076–2087, 2020.
[47] J. Pan, J. Dong, J. S. Ren, L. Lin, J. Tang, and M.-H. Yang, "Spatially variant linear representation models for joint filtering," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1702–1711.
[48] Y. Zhou, J. Jiao, H. Huang, Y. Wang, J. Wang, H. Shi, and T. Huang, "When awgn-based denoiser meets real noises," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 34, no. 7, 2020, pp. 13 074–13 081.
[49] W. Lee, S. Son, and K. M. Lee, "Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
[50] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
[52] I. Elbakri and J. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 89–99, 2002.
[53] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
[54] 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 Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.
[55] Y. Zhang, D. Hu, Q. Zhao, G. Quan, J. Liu, Q. Liu, Y. Zhang, G. Coatrieux, Y. Chen, and H. Yu, "Clear: Comprehensive learning enabled adversarial reconstruction for subtle structure enhanced low-dose ct imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3089–3101, 2021.
[56] Z. Huang, Z. Chen, Q. Zhang, G. Quan, M. Ji, C. Zhang, Y. Yang, X. Liu, D. Liang, H. Zheng, and Z. Hu, "Cagan: A cycle-consistent generative adversarial network with attention for low-dose ct imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1203–1218, 2020.