

One-Sample Diffusion Modeling in Projection Domain for Low-Dose CT Imaging

Bin Huang, Shiyu Lu, Liu Zhang, Boyu Lin, Weiwen Wu^{ID}, Member, IEEE,
and Qiegen Liu^{ID}, Senior Member, IEEE

Abstract—Low-dose computed tomography (CT) is crucial in clinical applications for reducing radiation risks. However, lowering the radiation dose will significantly degrade the image quality. In the meanwhile, common deep learning methods require large data, which are short for privacy leaking, expensive, and time-consuming. Therefore, we propose a fully unsupervised one-sample diffusion modeling (OSDM) in projection domain for low-dose CT reconstruction. To extract sufficient prior information from a single sample, the Hankel matrix formulation is employed. Besides, the penalized weighted least-squares and total variation are introduced to achieve superior image quality. First, we train a score-based diffusion model on one sinogram to capture the prior distribution with input tensors extracted from the structural-Hankel matrix. Then, at inference, we perform iterative stochastic differential equation solver and data-consistency steps to obtain sinogram data, followed by the filtered back-projection algorithm for image reconstruction. The results approach normal-dose counterparts, validating OSDM as an effective and practical model to reduce artifacts while preserving image quality.

Index Terms—Diffusion model, Hankel matrix, low-dose computed tomography (CT), sinogram domain.

I. INTRODUCTION

COMPUTED tomography (CT) techniques have been extensively utilized in clinical settings for diagnosis and intervention, including imaging, image-guided needle biopsy, image-guided intervention, and radiotherapy, with significant benefits over the past few decades [1], [2]. However, the

Manuscript received 19 February 2024; revised 26 March 2024; accepted 16 April 2024. Date of publication 22 April 2024; date of current version 5 November 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61871206 and Grant 62201628; in part by the National Key Research and Development Program of China under Grant 2022YFA1204200; and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515011780. (*Bin Huang and Shiyu Lu are co-first authors.*) (*Corresponding authors:* Weiwen Wu; Qiegen Liu.)

This work did not involve human subjects or animals in its research.

Bin Huang, Shiyu Lu, and Liu Zhang are with the School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China (e-mail: huangbin@email.ncu.edu.cn; lushiyu@email.ncu.edu.cn; zhangliu@email.ncu.edu.cn).

Boyu Lin is with the Second Clinical School, Nanchang University, Nanchang 330031, China (e-mail: linboyu@email.ncu.edu.cn).

Weiwen Wu is with the School of Biomedical Engineering, Sun Yat-Sen University, Shenzhen, Guangdong 528406, China (e-mail: wuweiw7@mail.sysu.edu.cn).

Qiegen Liu is with the School of Information Engineering, Nanchang University, Nanchang 330031, China (e-mail: liuqiegen@ncu.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TRPMS.2024.3392248>.

Digital Object Identifier 10.1109/TRPMS.2024.3392248

associated X-ray radiation dose poses a lifelong risk of cancer, which has raised widespread public concern [3]. Generally, reducing the X-ray exposure by adjusting the tube current is an effective approach for lowering the radiation dose [4], [5], [6], but it can lead to sharp artifacts and noise [7]. Sparse-view CT reconstruction is another attractive solution to address this issue [8].

Low-dose CT reconstruction is a classic inverse problem, and numerous algorithms have been proposed, which can be grouped into three categories: 1) image post-processing methods [9], [10]; 2) sinogram domain methods [11], [12], [13]; and 3) iterative reconstruction methods [14], [15], [16]. Image post-processing methods can directly process low-quality images without raw projection data. Zheng et al. [9] employed sparse representation and image decomposition theory to reconstruct high-quality low-dose CT images. Chen et al. [10] proposed a patch-based dictionary learning method based on the popular concept of sparse representation, effectively suppressing mottled noise and streak artifacts. However, it is challenging to remove severe streak artifacts and accurately recover image details and features without projection data. Sinogram domain-based reconstruction methods are beneficial in solving this problem. Yin et al. [11] introduced a domain progressive 3-D residual convolution network, while Humphries et al. [12] studied the performance of a simple CNN-based approach to low-dose CT using low-intensity and sparse view scans. Ghani and Karl [13] applied deep learning to denoise the original sinograms. Iterative reconstruction focuses on solving the low-dose CT problem iteratively by extracting prior information from target images [3], [14], [15]. Various priors were developed, with total variation (TV)-based reconstructions [16], [17], [18] being the most well-known. Liu et al. [16] defined a local TV and improved wavelet residual convolutional neural network denoising model. Additionally, Deng et al. [17] proposed a denoising model for projection data utilizing noise level-weighted TV regularization terms. Sagheer and George [18] considered tensor TV and developed a method based on low-rank approximation to improve global smoothness. Nonetheless, iterative reconstruction methods severely impact computational costs.

Diffusion models have provided a new perspective to address image processing tasks [19], [20], [21]. For instance, Lyu and Wang [22] proposed a conditional DDPM method for low-dose CT reconstruction. Song et al. [23] incorporated forward and backward diffusion processes into the stochastic

differential equation (SDE) framework to design a score-based generative model. The model has been widely applied in clinical scenarios and achieved excellent results [24], [25], [26]. Moreover, compared with generative adversarial networks (GANs), diffusion models achieve superior performance.

It is a complex and expensive process to collect medical data. Especially, privacy and security concerns constantly challenge clinical data sharing, which leads to the scarcity and infeasibility of clinical data. Consequently, an increasing number of researchers have turned their attention toward one-shot or few-shot reconstruction studies. In order to overcome this challenge, changing the scale of data has become a popular solution. Large images are successively subsampled to capture internal statistical properties at different scales and attain sufficient small image patches as the training set. The Hankel matrix is commonly used in image denoising [27], artifact removal [28] and deconvolution [29], owing to its data redundancy and structural reproducibility. Wang et al. [30] designed an encoder-decoder network based on Hankel matrix decomposition, which can reconstruct normal-dose images from low-dose images using a low-rank Hankel matrix formation via one-shot or few-shot learning technology. However, while studies on one-sample or few-sample natural images have emerged, there has been a lack of research on medical images [31], [32], [33].

In this study, we introduce one-sample diffusion modeling (OSDM), a one-sample or one-shot learning technology using diffusion modeling in the projection domain for low-dose CT reconstruction. Our approach combines the low-rank structural-Hankel matrix with the diffusion model to generate an ideal sinogram from the low-dose projection data. We also employ penalized weighted least-squares (PWLSs) and TV regularization to achieve superior image quality and accelerating iteration speed. Unlike previous supervised learning methods, our proposed unsupervised method does not require retraining on low-dose/normal-dose CT image pairs when the dosage of projection changes. Furthermore, only one data is required for training. The results show that OSDM accomplishes high performance compared with the state-of-the-art, which supports the feasibility of employing only one data to conduct the low-dose CT reconstruction study. Training the model in the projection domain is more conducive to generalization. For example, while the brain and lungs are vastly different in the image domain, they are similar in the projection domain. Therefore, we develop an OSDM reconstruction method in the projection domain rather than the image domain to enhance the performance of our model.

The theoretical and practical contributions of this work can be summarized as follows.

- 1) We propose an innovative score-based diffusion model for projection domain training, which maintains great generalization without requiring the retraining of models when the training or test set changes.
- 2) To address the scarcity of medical samples, we construct a low-rank Hankel matrix to infer internal statistics within the CT sinogram. Only one sample is required to extract sufficient prior information.

- 3) By incorporating PWLS and TV techniques into the iterative loop of our model, OSDM achieves significantly improved iteration speed and convergence speed compared to other diffusion models.

The remaining sections of the manuscript are organized as follows. Section II provides background information on score-based diffusion models and the construction process of the Hankel matrix. In Section III, we present the detailed procedure and algorithm of the proposed method. Section IV includes the experimental results, implementation specifications, and details about the experiments. Finally, we draw conclusions in Section V.

II. PRELIMINARY

A. Low-Dose CT Images

Low-dose CT reconstruction is a classic inverse problem, which seeks to reconstruct the fuzzy part from low-dose CT data. In particular, assuming that $x \in \mathbb{R}^N$ is the degraded sinogram, the forward formulation of sinogram reconstruction problem could be given by

$$y = x + n \quad (1)$$

where $n \in \mathbb{R}^M$ represents for additive noise. y denotes a low-dose sinogram. Note that the inverse problem amounts to generate x from y .

For avoiding ill-posed, the problem of reconstructing CT sinograms is formulated as the optimization equation with constraints

$$\min_x \|x - y\|_2^2 + \mu R(x) \quad (2)$$

where $\|x - y\|_2^2$ is the data fidelity term. $R(x)$ denotes the regulation prior knowledge term, which is chosen to be a TV semi-norm. $\|\cdot\|_2^2$ represents the l_2 norm. Besides, μ is the factor to keep a good balance between the data-consistency (DC) term and regularization term.

Taking the TV term into consideration, this optimization process differentiates infinite solutions to (2) and picks out the best one with desired image properties as the reconstructed sinogram. Generally, the TV term is defined as

$$R(x) = \|x\|_{\text{TV}}^2 = \int_{\Omega} |\nabla x| dx \quad (3)$$

where Ω is the bounded domain. ∇x represents the gradient of the sinogram x . The TV term has been proven to be robust in removing noise and artifacts from denoising sinogram [3], [34].

B. Construction of Hankel Matrix

Data-redundancy property of the Hankel matrix helps to extract internal statistics within low-dose CT sinograms. Based on the idea that a low-rank matrix could express projection data, the new data formulation is designed here to manifest the inner relationship in sinograms.

As illustrated in Fig. 1, the sinogram can be transformed to a new data formulation by sliding a window $a \times a$ across entire sinogram $W_x \times W_y$. After the transformation H , the new

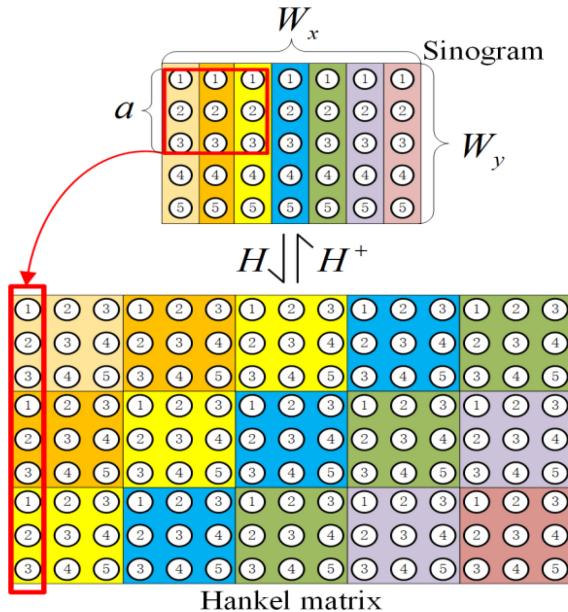


Fig. 1. Constructing a new data formulation from one sinogram (H) and vice versa (H^+).

data formulation with a size of $a^2 \times (W_x - a + 1)(W_y - a + 1)$ is generated. Fig. 1 demonstrates the case of $W_x = 7$, $W_y = 5$, and $a = 3$. Besides, the step size of sliding window is set to be 1.

Individual blocks data in the projection domain are vectorized as columns in the new data formulation. The linear operator H is defined as generating new data formulation from the sinogram concatenated in a vector form

$$H := R^{W_x \times W_y} \rightarrow R^{a^2 \times (W_x - a + 1)(W_y - a + 1)}. \quad (4)$$

When reversely forming the sinogram from the new data formulation, multiple antidiagonal entries are averaged and stored in the projection domain. Hence, the reverse operator H^+ generates the corresponding projection dataset from the new data formulation, and is given as follows:

$$H^+ := R^{a^2 \times (W_x - a + 1)(W_y - a + 1)} \rightarrow R^{W_x \times W_y} \quad (5)$$

where $+$ stands for a pseudo-inverse operator. It is equivalent to averaging the anti-diagonal elements and placing them in the appropriate locations.

C. Score-Based SDE

The great success of diffusion models, especially score-based SDE in creating realistic and variable image samples has aroused widespread concern [23]. Score-based SDE comprises the forward process and the reverse-time process.

Given a continuous diffusion process $\{x(t)\}_{t=0}^T$ with $x(t) \in \mathbb{R}^N$, which is indexed by $t \in [0, T]$ as the progression time variable. N stands for the sinogram dimension. The forward diffusion process can be formulated as the solution to the following SDE:

$$dx = f(x, t)dt + g(t)dw \quad (6)$$

where $f(x, t) \in \mathbb{R}^N$ and $g(t) \in \mathbb{R}$ correspond to the drift coefficient and diffusion coefficient, respectively. dt corresponds

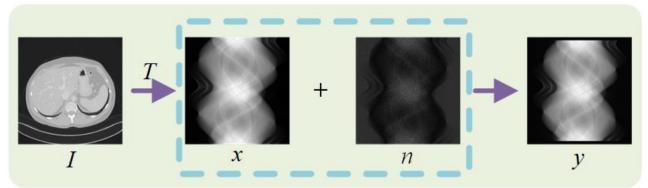


Fig. 2. Linear measurement process for low-dose CT.

to an infinitesimal positive time step and $w \in \mathbb{R}^N$ induces the Brownian motion.

Via reversing the above process, samples can be attained. Notably, the reverse-time SDE is also a diffusion process, which could be expressed as follows:

$$dx = \left[f(x, t) - g(t)^2 \nabla_x \log p_t(x) \right] dt + g(t)d\bar{w} \quad (7)$$

where \bar{w} is a standard Wiener process when time flows backward from T to 0, and dt is an infinitesimal negative time step.

III. PROPOSED METHOD

A. Low-Dose CT Imaging Model

As shown in Fig. 2, the linear measurement process for low-dose CT imaging is visualized. Intuitively, n can be defined as the low-dose noise on the sinogram. I represents CT images. $T(\cdot)$ corresponds to the Radon transform. If the ideal sinogram x of size 768×768 is measured in the presence of n , the CT reconstruction problem can be formulated as solving the following equation:

$$y = T(I) + n = x + n \quad (8)$$

where y with size 768×768 is the low-dose CT sinogram. In addition, $T^{-1}(\cdot)$ denotes the inverse Radon transform.

Fueled by a large number of experiments using real projection data, low-dose sinograms can be approximated as ideal projection data contaminated by additive noise [35], [36], [37], [38], [39]. Suppose merely a monochromatic source is employed, the additive noise follows the Poisson distribution. Especially, the Poisson model for the intensity measurement is presented by:

$$L_i \sim \text{Poisson}\{a_i e^{-|x|_i}\} + r_i, \quad i = 1, \dots, N_m \quad (9)$$

where L_i denotes the number of transmitted photons, a_i is the X-ray source intensity of the i th ray, and r_i indicates the background contributions of scatter and electrical noise. x corresponds to a vector for the representation of attenuation coefficients with units of inverse length, N_m stands for the number of measurements and N_v stands for the number of image voxels. In (9), the noise level is handled by a_i , whose specific value is given in Section IV.

Via taking the logarithm operation, the measurement data is transformed to the weighted Gaussian formulation [40]

$$y_i \sim N\left([x]_i, \bar{L}_i / (\bar{L}_i - r_i)^2\right) \quad (10)$$

where $\bar{L}_i = E[L_i]$. In fact, the low-dose CT problem is expressed as the typical inverse problem with ill-posedness. In

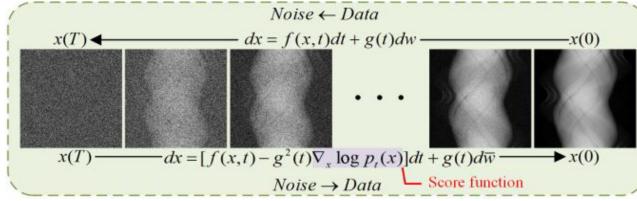


Fig. 3. Perturbed data by noise is smoothed according to the trajectory of an SDE. By estimating the score function $\nabla_x \log p_t(x)$ with SDE, it is possible to approximate the reverse SDE and then solve it to generate sinogram samples from noise.

order to address such the problem, the posterior distribution $p(x|y)$ is introduced by the theory of Bayesian inversion [41]. Thereby, the inverse problem is transformed to a problem conditioned on the measurement y .

B. Data Preprocessing

During training, we process the Hankel matrix before feeding it into the network. This process is illustrated in Fig. 3. To store information within the data, we slide a window of size over the initial data to construct a Hankel matrix of size 579121×64 . After applying the Hankel transformation, the same information appears at different locations within the matrix, using its redundancy to capture internal statistics. Remarkably, we only need one normal-dose sinogram for the training process. The construction process for the Hankel matrix from the initial projection data is as follows: suppose the initial normal-dose projection data is denoted as x . We construct the Hankel matrix H as follows:

$$H_s = H(x) \quad (11)$$

where $H(\cdot)$ corresponds to the Hankel transformation and H_s represents constructed Hankel matrix.

To better utilize prior learning knowledge, we crop the Hankel matrix into high-dimensional data with size $64 \times 64 \times 9048$. We also randomly extract a large number of small patches from the matrix, as shown in Fig. 4. Specifically, we split the constructed Hankel matrix into patches in a random manner using the following process:

$$X = S(H_s). \quad (12)$$

The random split operation is denoted as $S(\cdot)$ and applied to the high-dimensional tensor X to extract numerous patches. These patches are then treated as inputs to the network, resulting in the construction of many tensors. This technique of data amplification enhances the training set and enables us to acquire sufficient prior knowledge.

The score-based model, described in Section II-C, learns a prior distribution by leveraging an SDE. This is achieved through a forward SDE that smoothly transforms a complex data distribution to a known prior distribution by gradually injecting noise. Fig. 3 demonstrates these two processes.

During the training phase, the peak performance of network can be reached by optimizing the parameters θ^* of the score-based network. The objective function can be described as

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathbb{E}_t \left\{ \lambda(t) \mathbb{E}_{x(0)} \mathbb{E}_{x(t|x(0))} \left[\| s_{\theta}(x(t), t) - \nabla_{x(t)} \log p_t(x(t)|x(0)) \|_2^2 \right] \right\} \quad (13)$$

where $\lambda : [0, T] \rightarrow \mathbb{R}_{>0}$ is a positive weighting function and t is uniformly sampled over $[0, T]$. $p_t(x(t)|x(0))$ is the Gaussian perturbation kernel centered at $x(0)$. Once the network satisfies $s_{\theta}(x(t), t) \simeq \nabla_x \log p_t(x) \nabla_x \log p_t(x)$ will be known for all t by solving $s_{\theta}(x(t), t)$.

C. OSDM—Iterative Reconstruction

The iterative reconstruction process of OSDM is described in this part. First, the sinogram data x is transformed from the CT image I via T^{-1} operation, which corresponds to forward projection (FP)

$$x = T^{-1}(I). \quad (14)$$

Same as Section III-B, the Hankel transform (HT) and the random split operation are conducted on the sinogram data. After that, the high-dimensional tensor X with numerous patches are acquired.

The score-based diffusion model is employed to estimate the prior distribution of sinogram data $p_t(X)$, allowing for noise suppression and richer information. Unlike perturbing data with a finite number of noise distributions, this method considers a continuous distribution over time during the forward diffusion process. By reversing the SDE, random noise can be converted into data for sampling. This article introduces predictor-corrector (PC) sampling at the samples updating step, as suggested in [42]. The predictor is viewed as a numerical solver for the reverse-time SDE in PC sampling. Once the reverse-time SDE process concludes, samples are generated based on the discretized prior distribution, which can be discretized as follows:

$$X^i \leftarrow X^{i+1} + (\sigma_{i+1}^2 - \sigma_i^2) s_{\theta}(X^{i+1}, \sigma_{i+1}) + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} z \quad i = N-1, \dots, 0 \quad (15)$$

where $z \sim N(0, 1)$, $X(0) \sim p_0$, and $\sigma_0 = 0$ is chosen to simplify the notation. The above formulation is repeated for $i = N-1, \dots, 0$. Thus, previous discrete process turns into continuous diffusion process. With adding the conditional constraints to (15), it can be rewritten as follows:

$$X^i = X^{i+1} + (\sigma_{i+1}^2 - \sigma_i^2) \nabla_X \left[\log p_t(y|X^{i+1}) + \log p_t(X_{\text{LR}}^{i+1}) + \log p_t(X_{\text{TV}}^{i+1}) \right] + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} z. \quad (16)$$

In the above equation, $\log p_t(y|X)$ stems from sinogram data knowledge. $\log p_t(X_{\text{LR}})$ is derived from low-rank (LR) prior and $\log p_t(X_{\text{TV}})$ comes from TV prior.

When it comes to the corrector, it refers to the Langevin dynamics via transforming any initial sample $x(t)$ to the final sample $x(0)$ with the following procedure:

$$X^{i,j} \leftarrow X^{i,j-1} + \varepsilon_i s_{\theta}(X^{i,j-1}, \sigma_i) + \sqrt{2\varepsilon_i} z \quad j = 1, 2, \dots, M, \quad i = N-1, \dots, 0 \quad (17)$$

where $z \sim N(0, 1)$ is the step size, and $z \sim N(0, 1)$ refers to a standard normal distribution. The above formulation is

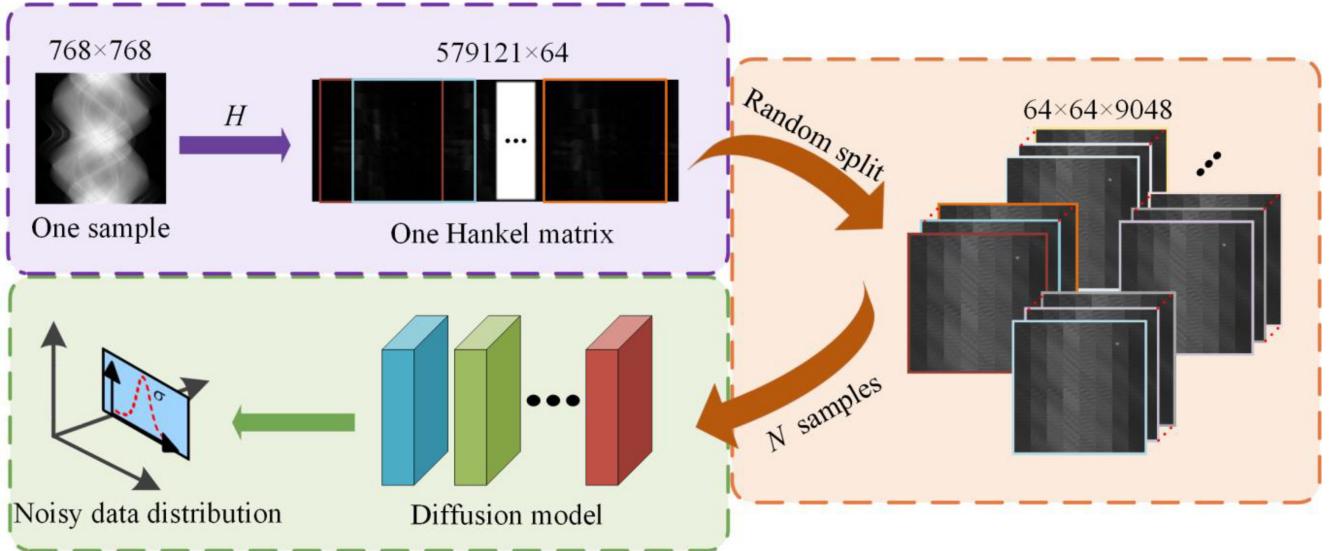


Fig. 4. OSDM training process follows the illustrated pipeline. First, we gradually inject multiscale noise into a single normal-dose sinogram. Next, we construct a low-rank Hankel matrix using Hankel transformation to capture redundant information. Finally, we randomly split the Hankel matrix into multiple tensors (N samples) and use denoising score matching to train the network for learning the gradient distribution. This prior learning of the score-based network is an essential step in the training process.

repeated for $j = 1, 2, \dots, M$, $i = N - 1, \dots, 0$. The theory of Langevin dynamics guarantees that when $M \rightarrow \infty$ and $\varepsilon \rightarrow 0$, x^i is a sample from $p_t(X)$ under designated conditions.

Sampling is performed not directly from the distribution $p(X)$, but from the posterior distribution $p(X|y)$ by utilizing SDE as explained in Section II. The DC operation can be considered as a conditional term that is incorporated into the sampling procedure of (27), and it yields

$$\begin{aligned} X^{i,j} = & X^{i,j-1} + \varepsilon_i \nabla_x \left[\log p_t(y|X^{i,j-1}) + \log p_t(X_{\text{LR}}^{i,j-1}) \right. \\ & \left. + \log p_t(X_{\text{TV}}^{i,j-1}) \right] + \sqrt{2\varepsilon_i} z. \end{aligned} \quad (18)$$

Then, the updated sample x is transformed to a new data formulation through HT operation $H(\cdot)$

$$H_p^i \leftarrow H(X^i). \quad (19)$$

LR Step: For facilitate manipulation and analysis, the Hankel matrix can be decomposed using the singular value decomposition (SVD)

$$[U \Delta V^T] = \text{svd}(H^i) \quad (20)$$

where U is an orthogonal matrix, Δ is a diagonal matrix with non-negative diagonal elements, and V is an orthogonal matrix. Precisely, H is a 579121×64 matrix of rank L while $U_{[k]}$, $V_{[k]}$, and $\Delta_{[k]}$ represents the first K columns of U , V , and Δ , respectively

$$\begin{aligned} U_{[k]} &= [u_1, \dots, u_k, \dots, u_K] \\ V_{[k]} &= [v_1, \dots, v_k, \dots, v_K] \\ \Delta_{[k]} &= [\delta_1, \dots, \delta_k, \dots, \delta_K]. \end{aligned} \quad (21)$$

The hard-threshold (hard-THR) singular value process can be expressed as

$$H_{[k]}^i = U_{[k]} \Delta_{[k]} V_{[k]}^T \quad (22)$$

where $H_{[k]}$ represents the matrix H reconstructed from the first K eigenvectors. The SVD is particularly useful for ill-defined linear problems with almost degenerate matrices, as it provides the best approximation with a smaller rank. After the low-rank process, the Hankel matrix is transformed back into a sinogram using the inverse HT (IHT) operation denoted by $H^+(\cdot)$ with

$$x^i \leftarrow H^+(H_{[k]}^i). \quad (23)$$

TV Step: The TV minimization is also conducted for removing noise and artifacts. Suppose $\Delta x = \|x - x^i\|$, and TV minimization can be stated as follows:

$$\text{TV}(x^i) = x^{i+1} - \alpha \times \Delta x \times \frac{\nabla \|x^i\|_{\text{TV}}}{\|\nabla \|x^i\|_{\text{TV}}\|} \quad (24)$$

where α is the length of each gradient-descent step.

DC Step: To improve noise immunity, statistical properties of the projected data can be utilized in the object function [35], [36], [37], [38], [39], [43]. A statistical approach for sinogram denoising also exists, which leverages the PWLS method to determine an optimal valuation from noisy sinograms. The PWLS prior is integrated into a regularized objective function expressed as

$$x^i = \arg \min_x \left[\|y - x^{i+1}\|_W^2 + \lambda_1 \|x^{i+1} - H^+(H_{[K]}^{i+1})\|_2^2 + \lambda_2 \|x^{i+1}\|_{\text{TV}}^2 \right] \quad (25)$$

where hyperparameter λ_1 , λ_2 balances the tradeoff among the terms of PWLS, LR, and TV. $i = N - 1, \dots, 0$ denotes the iteration of outer loops. Specifically, the standard PWLS can be described as follows:

$$x^i = \arg \min_x \left[(x^{i+1} - y)^T W (x^{i+1} - y) + \mu R(x^{i+1}) \right] \quad (26)$$

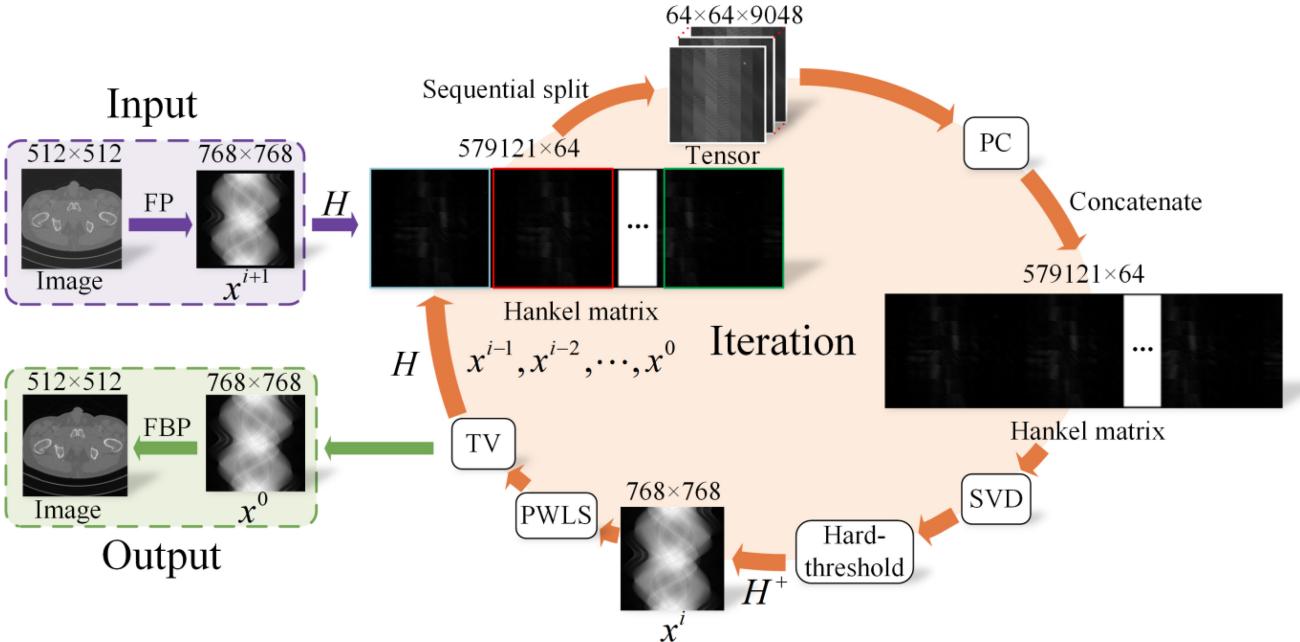


Fig. 5. Pipeline of iterative reconstruction procedure in OSDM.

where superscript T represents the transposing operation. Equation (24) can be further solved as

$$x^i = \frac{W(y - x^{i+1}) + \mu R'(x^{i+1})}{W + \mu}. \quad (27)$$

In order to decrease the influence of noise, the scale coefficient η for system calibration is set to 22 000

$$W = \text{diag}\{w_i\} = \text{diag}\left\{\frac{1}{\sigma_{x_i}^2}\right\} = \text{diag}\left\{\frac{1}{l_1 \exp(x_i/\eta)}\right\}. \quad (28)$$

Once we obtain the reconstructed projection x , the final image \tilde{I} is obtained

$$\tilde{I} = T(x) \quad (29)$$

where $T(\cdot)$ stands for filtered back-projection (FBP).

The flowchart of the iterative reconstruction process is demonstrated in Fig. 5. At the inference stage, the numerical SDE solver and data fidelity step is iteratively updated to achieve the normal-dose projection data. Since the Langevin dynamics updating can guarantee the convergence of the algorithm, the overall OSDM will be convergent after finite iterations.

Algorithm 1 provides a detailed demonstration of the training and reconstruction processes of OSDM. During the training phase, only a single sinogram sample is demanded to capture sufficient prior information. The OSDM reconstruction procedure comprises two loops. In the outer loop, predictor is performed with trained network. In the inner loop, correction is applied for error reduction. The predictor and the corrector work together as a whole to generate final samples. Moreover, either after the predictor operation or the corrector operation, the data fidelity term is enforced to ensure the quality of the final image.

IV. EXPERIMENTS

A. Data Specification

AAPM Challenge Data: The simulated data from human abdomen images provided by Mayo Clinic for the AAPM Low-Dose CT Grand Challenge [44] are used for evaluation. The dataset contains 2588 3mm 512x512 normal-dose CT images from 10 patients, which is handled to 7764 1mm 512x512 normal-dose CT images. The corresponding projection data with 1e5, 5e4 and 1e4 noise levels are made by adding Poisson noise into the sinograms simulated from the normal-dose CT images. Artifact-free images generated from normal-dose projection data by FBP algorithm, can be considered as the ground truth. In this study, the dataset is divided into two parts: 1) training set and 2) test set. During the training phase, it is notable that only one projection data is used for OSDM while 4839 projection data are used for NCSN++ [23] and 4839 paired datasets for U-Net [51] and CNN [50]. In the meanwhile, there are 537 sinograms picked as validation set and 12 sinograms picked as test set for all models. In fan-beam CT reconstruction, the Siddon's ray-driven algorithm [45], [46] is utilized to generate the projection data. The distance from rotation center to the source and detector are set to 40 cm, respectively. The detector width is 41.3 cm, including 720 detector elements and a total of projection views, are evenly distributed over 360°.

CQ500 Dataset: We test OSDM on Qure.ai's CQ500 dataset [47], which contains 491 CT scans of human brains. Among them, 12 CT scans are picked for testing. The resolution of the selected subset of data is 512x512 per slice, with 101 to 645 slices per volume. The volumes are given in Hounsfield units in the range of [0, 4095] and normalized to [0,1]. The neural network's reconstruction parameters are identical to those of the AAPM data.

TABLE I
RECONSTRUCTION PSNR/SSIM/MSEs OF AAPM CHALLENGE DATA USING DIFFERENT METHODS AT 1E5, 5E4, AND 1E4 NOISE LEVEL

Noise level	FBP	SART-TV	CNN	U-Net	NCSN++	OSDM
$a_i = 1e5$	34.62/0.9252/3.66e-4	41.03/0.9892/8.65e-5	41.36/0.9904/8.20e-5	42.52/ 0.9921 /6.28e-5	41.53/0.9869/7.17e-5	42.62 /0.9899/ 5.51e-5
$a_i = 5e4$	32.43/0.8866/5.81e-4	38.72/0.9786/1.39e-4	39.26/0.9872/1.28e-4	41.34 / 0.9897 /7.83e-5	40.17/0.9810/9.75e-5	41.20/0.9857/ 7.74e-5
$a_i = 1e4$	25.78/0.6897/2.69e-3	29.58/0.8710/1.18e-3	37.58/ 0.9795 /1.81e-4	38.50 /0.9782/ 1.44e-4	37.32/0.9653/1.90e-4	37.43/0.9683/1.83e-4

Algorithm 1 Training and Reconstruction Stages

Training stage

Dataset: One sample x in the projection domain

- 1: **Repeat**
- 2: $x \sim p(x)$, $t \sim \mathcal{U}([0, T])$, $\varepsilon \sim \mathcal{N}(0, I)$
- 3: $x(t) = x(0) + \varepsilon\sigma(t)$
- 4: Take a gradient descent step on $\nabla_\theta \|s_\theta(x(t), t) + \varepsilon\|_2^2$
- 5: **Until** converged
- 6: Trained OSDM

Reconstruction stage

Setting: s_θ , N , M , σ , ε

- 1: Initial data $x = T^{-1}(I)$ (**FP**)
- 2: $x^N \sim \mathcal{N}(0, \sigma_{\max}^2 I)$
- 3: $H_s = H(x)$
- 4: **For** $i = N - 1$ to 0 **do** (**Outer loop**)
- 5: $X = S(H_s)$
- 6: $X^i \leftarrow Predictor(X^{i+1}, \sigma_i, \sigma_{i+1})$
- 7: $H_p^i \leftarrow H(X^i)$ (**HT**)
- 8: $[U \Delta V^T] = svd(H_p^i)$ (**SVD**)
- 9: $H_{[k]}^i = U_{[k]} \Delta_{[k]} V_{[k]}^T$ (**hard-THR**)
- 10: $x^i \leftarrow H^+(H_{[k]}^i)$ (**IHT**)
- 11: $x^i = \frac{W(y-x^{i+1})+\mu R'(x^{i+1})}{W+\mu}$ (**PWLS**)
- 12: $x^i = TV(x^i)$ (**TV**)
- 13: **For** $j = 1$ to M **do** (**Inner loop**)
- 14: $X^{i,j} \leftarrow Corrector(X^{i,j-1}, \sigma_i, \varepsilon_i)$
- 15: **Repeat from step 5 to step 10**
- 16: **End for**
- 17: **End for**
- 18: Final image $\tilde{I} = T(x)$ (**FBP**)
- 19: **Return** \tilde{I}

B. Model Training and Parameter Selection

In the experiments, OSDM is implemented in Python using operator discretization library (ODL) [48] and PyTorch on a personal workstation with a GPU card (NVIDIA-GeForce-GTX 1080 Ti-11GB). For simulating low-dose projection data, Poisson noise is added into testing slices by applying three different X-ray source intensity $a_i = 1e5$, $a_i = 5e4$, and $a_i = 1e4$. At the training stage, OSDM uses the Adam algorithm with the learning rate 10^{-3} and Kaiming initialization is picked as the weight initialization method. During the reconstruction stage, the outer iteration number and inner iteration number are set to $N = 1000$ and $M = 2$, respectively. Once the prediction process of the outer loop is executed, the correction process of the inner loop is iterated twice by annealing Langevin.

Referring to [35], [36], [37], and [38], η in PWLS scheme is set to a constant 22 000. The length of each gradient-descent step α in TV minimization is set to 10. The singular value thresholding in SVD operation is 38 and the sliding window size is 8×8 . Source code of OSDM can be accessed at: <https://github.com/yqx7150/OSDM>.

C. Quantitative Indices

To evaluate the quality of the reconstructed data, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and mean squared error (MSE) are used for quantitative assessment. All quantitative metrics are calculated by the average of 12 test data.

PSNR describes the maximum possible power of the signal in relation to the noise corrupting power. Higher PSNR means better-reconstruction quality. Denoting I and \tilde{I} to be the estimated reconstruction and ground-truth, PSNR is expressed as

$$\text{PSNR}(I, \tilde{I}) = 20 \log_{10} [\text{Max}(\tilde{I}) / \|I - \tilde{I}\|_2]. \quad (30)$$

The SSIM value is used to measure the similarity between the ground-truth and reconstruction. SSIM is defined as

$$\text{SSIM}(I, \tilde{I}) = \frac{(2\mu_I\mu_{\tilde{I}} + c_1)(2\sigma_{I\tilde{I}} + c_2)}{(\mu_I^2 + \mu_{\tilde{I}}^2 + c_1)(\sigma_I^2 + \sigma_{\tilde{I}}^2 + c_2)} \quad (31)$$

where μ_I and σ_I^2 are the average and variances of I . $\sigma_{I\tilde{I}}$ is the covariance of I and \tilde{I} . c_1 and c_2 are used to maintain a stable constant. MSE is employed to evaluate the errors and it is defined as

$$\text{MSE}(I, \tilde{I}) = \sum_{i=1}^W \|I_i - \tilde{I}_i\|_2/W \quad (32)$$

where W is the number of pixels within the reconstruction result. If MSE approaches to zero, the reconstructed image is closer to the reference image.

D. Experimental Comparison

Quick Comparison: We compare the proposed unsupervised model OSDM with four baseline techniques in low-dose CT reconstruction, including FBP [2], SART-TV [49], CNN [50], U-Net [51], and NCSN++ [23]. CNN, U-Net, and NCSN++ are trained by 4839 projection data while OSDM is trained by one. The involved parameters are set by the guidelines in their original studies.

There are three different noise levels, we set 1e5, 5e4 and 1e4 photons along each path of X-ray for low-dose CT reconstruction. Table I shows the PSNR, SSIM, and MSE

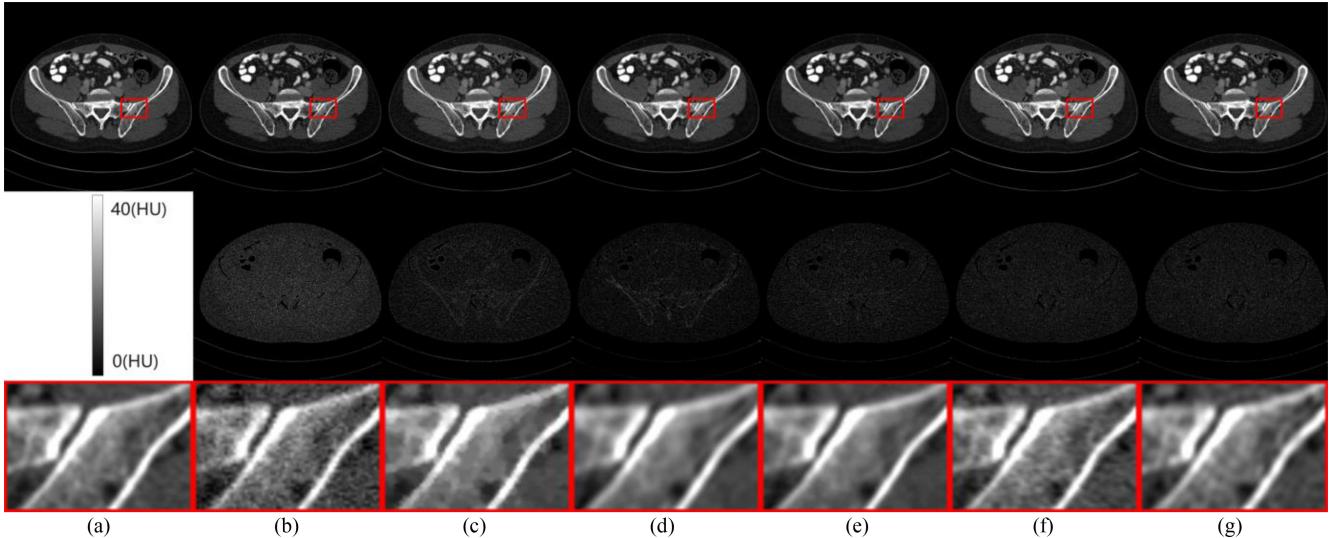


Fig. 6. Reconstruction results from $1e5$ noise level using different methods. (a) Reference image versus the images reconstructed by (b) FBP, (c) SART-TV, (d) CNN, (e) U-Net, (f) NCSN++, and (g) OSDM. The display windows are $[-250, 600]$ HU. The second row depicts the residuals between the reference and reconstructed images.

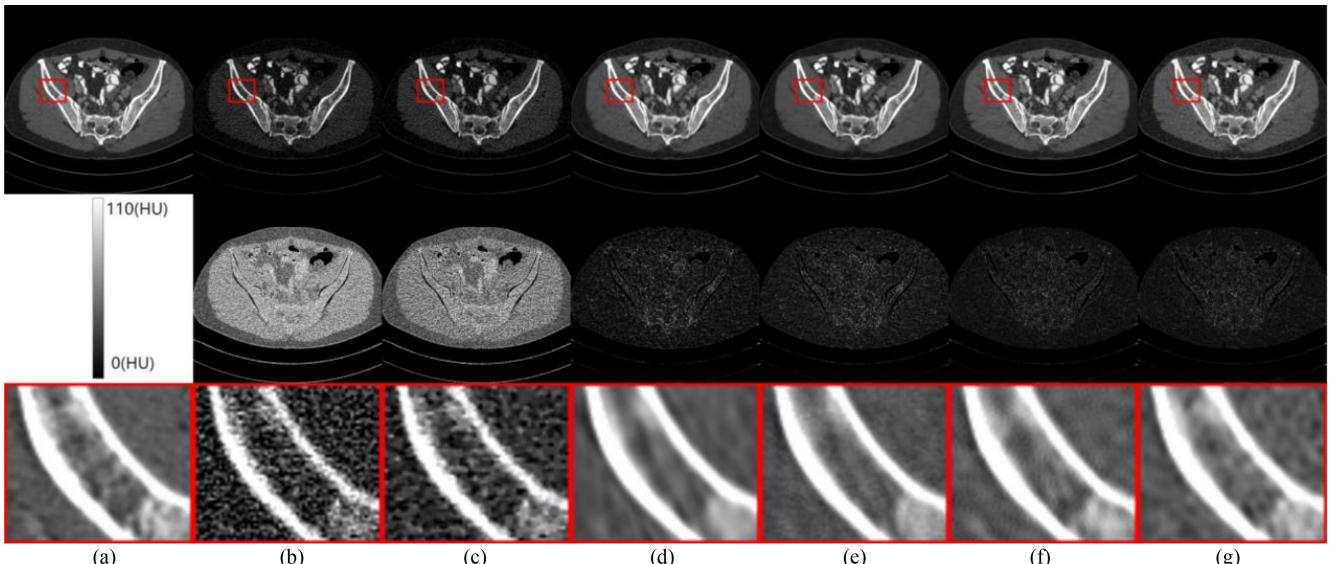


Fig. 7. Reconstruction results from $1e4$ noise level using different methods. (a) Reference image versus the images reconstructed by (b) FBP, (c) SART-TV, (d) CNN, (e) U-Net, (f) NCSN++, and (g) OSDM. The display windows are $[-250, 600]$ HU. The second row depicts the residuals between the reference and reconstructed images.

values of the reconstructed results from the AAPM Challenge Dataset, with the best PSNR and SSIM values highlighted in bold. OSDM method outperforms the other methods in terms of detail preservation, as seen in the reconstructed images with fewer artifacts and less noise. Particularly impressive are the average MSE values achieved by OSDM of $5.51e-5$ and $7.74e-5$ at $1e5$ and $5e4$ noise levels, respectively. In fact, the OSDM method can achieve an impressive 42.62 dB at $1e5$ noise level.

To further demonstrate its advantages, Figs. 6 and 7 depict the reconstructed and residual images. The analytical FBP method performs the worst, as its results are heavily dependent on noise. The SART-TV method outperforms FBP, but lacks geometric details. The deep prior reconstruction approach produces sharper results, but also misses some details. CNN,

U-Net and NCSN++ achieve acceptable results, but still lose some edge details. However, the images reconstructed by OSDM are closely resemble to the ground truth, preserving distinct details and structure.

Compared to the results reconstructed from a $1e4$ noise level, the image quality of the results reconstructed from a $1e5$ noise level shows significant improvements as depicted in Figs. 6 and 7. The FBP algorithm produces images with noticeable artifacts and missing structural features. The U-Net images still exhibit severe streaking artifacts, while the CNN over-smooths the edges. The reconstruction result of NCSN++ also demonstrates over-smooth textures, although it was not as pronounced as that of the CNN. On the other hand, the OSDM methodology preserves more structural

TABLE II
RECONSTRUCTION PSNR/SSIM/MSEs OF AAPM CHALLENGE DATA
USING DIFFERENT TRAINING MODEL AT 1E5 NOISE LEVEL

Training set	CNN	U-Net	NCSN++	OSDM
$a_i=1e5$	41.36/0.990 4/8.20e-5	42.52/0.992 1/6.28e-5	\	\
$a_i=5e4$	41.00/0.989 8/1.10e-4	41.97/0.990 7/7.54e-5	\	\
$a_i=1e4$	36.40/0.983 0/2.87e-4	40.15/0.986 8/1.11e-4	\	\
$a_i=0$	\	\	41.53/0.986 9/7.17e-5	42.62/0.989 9/5.51e-5

details while suppressing streaking artifacts. Although OSDM dissatisfies the performance in high-noise situations compared with CNN and U-Net, OSDM only demands one sample for training to capture abundant prior knowledge. On the contrary, CNN, U-Net, and NCSN++ require thousands of data for training, which largely consumes precious time and expensive data. Apart from that, unsupervised OSDM merely needs single training once before it can be adapted to any other circumstances. As members of the end-to-end family, CNN and U-Net can achieve high-performance reconstruction at the cost of tedious and homologous model training. Even compared with the unsupervised method NCSN++, OSDM also exhibits superior performance. In a word, OSDM shows promising effects in noise removal.

Generalization and Robustness Test: To compare the generalization of CNN, U-Net, NCSN++, and OSDM, we pretrain the models separately for each noise level in this experiment. Instead, we use the models trained at 1e5 noise level to implement this task. Two supervised methods CNN and U-Net are selected as baselines. Remarkably, OSDM is trained by a single sample while CNN and U-Net are trained by thousands of samples. It has crucial implications for alleviating the shortage of medical samples. Table II shows the proposed OSDM performs the best on most occasions, achieving the highest PSNR and SSIM with the smallest MSE, and it is an unsupervised method that requires only single training.

It is impossible to train for every noise level in real circumstances, so CNN, U-Net, or any other supervised methods cannot be used for different noise doses. From Table II, we can see that the results of CNN and U-Net deteriorate rapidly as the training noise changes.

To qualitatively and quantitatively evaluate the generalization of our proposed unsupervised learning scheme, we apply OSDM to the CQ500 dataset at the noise level of 1e4. We also utilize prior knowledge from the AAPM challenge data to train OSDM and test it on the CQ500 dataset. The results are summarized in Table III, where OSDM outperforms other methods in terms of PSNR and MSE with scores of 38.15 dB and 1.55e-4, respectively. Additionally, the image quality evaluation metrics are consistently good when testing on the CQ500 dataset, indicating the adaptability of OSDM against dataset changes.

For visually illustrating the superior performance of OSDM, we conducted a qualitative comparison of the different test datasets. We also included zoomed regions-of-interest, marked

TABLE III
RECONSTRUCTION PSNR/SSIM/MSEs OF CQ500 DATASET USING
DIFFERENT METHODS AT 1E4 NOISE LEVEL

$a_i=1e4$	FBP	CNN	U-Net	NCSN++	OSDM
PSNR	19.09	36.65	32.56	36.78	38.15
SSIM	0.5269	0.9836	0.9545	0.9734	0.9818
MSE	1.35e-2	2.25e-4	5.68e-4	2.32e-4	1.55e-4

TABLE IV
RECONSTRUCTION PSNR/SSIM/MSEs OF CQ500 DATASET USING
DIFFERENT METHODS AT 1E3 NOISE LEVEL

$a_i=1e3$	FBP	CNN	U-Net	NCSN++	OSDM
PSNR	11.04	18.16	18.38	32.36	33.65
SSIM	0.1879	0.6219	0.5817	0.9470	0.9608
MSE	7.95e-2	1.54e-2	1.46e-2	6.32e-4	4.38e-4

by red rectangles in Fig. 8, to better evaluate image quality. As shown in Fig. 8(c), OSDM achieved the best performance in terms of suppressing noise artifacts and preserving tissue features. In contrast, the reconstruction results by FBP are unsatisfactory. The results obtained by CNN, U-Net, and NCSN++ are inferior to OSDM, as they introduce more noise and artifacts. This experiment conclusively demonstrates the excellent generality of OSDM.

To better explore the generalization of OSDM, we utilize the pretrained model at 1e4 noise level to reconstruct the data at 1e3 noise level for CNN and U-Net, while OSMD still uses the old one. Quantitative results reconstructed from different reconstruction methods are tabulated in Table IV. It can be observed that OSDM outperforms the other methods in a trend similar to what we have seen from the reconstruction images and produces the highest PSNR. Anyhow varying the test data and noise level, OSMD is better than CNN, U-Net, and NCSN++. Remarkably, even compared with the unsupervised method NCSN++, OSDM also exhibits better generalization.

E. Convergence Analysis

This section presents an empirical investigation on the convergence behavior of NCSN++ and OSDM in relation to the number of iterations. We perform the reconstruction on a randomly selected example from the test set. Fig. 9 displays the PSNR curve of OSDM. With an increase in the number of iterations, both OSDM and NCSN++ show a rapid rise in terms of the PSNR curve. However, OSDM exhibits a more rapid convergence rate and consistently outperforms NCSN++. All 12 test images follow the same iterative pattern. Notably, the PSNR of OSDM arrives at the peak at the 215th step while NCSN++ demands 417 steps. Moreover, OSDM can produce an impressive result in approximately the first 50 iteration steps. As we all know, NCSN++ requires a high-inference cost, which is hardly acceptable in practice. For faster convergence and low-time cost, PWLS and TV operations are enforced in each iterative reconstruction step of OSDM to ensure that the output is consistent with the original projection data. PWLS can be considered as the data

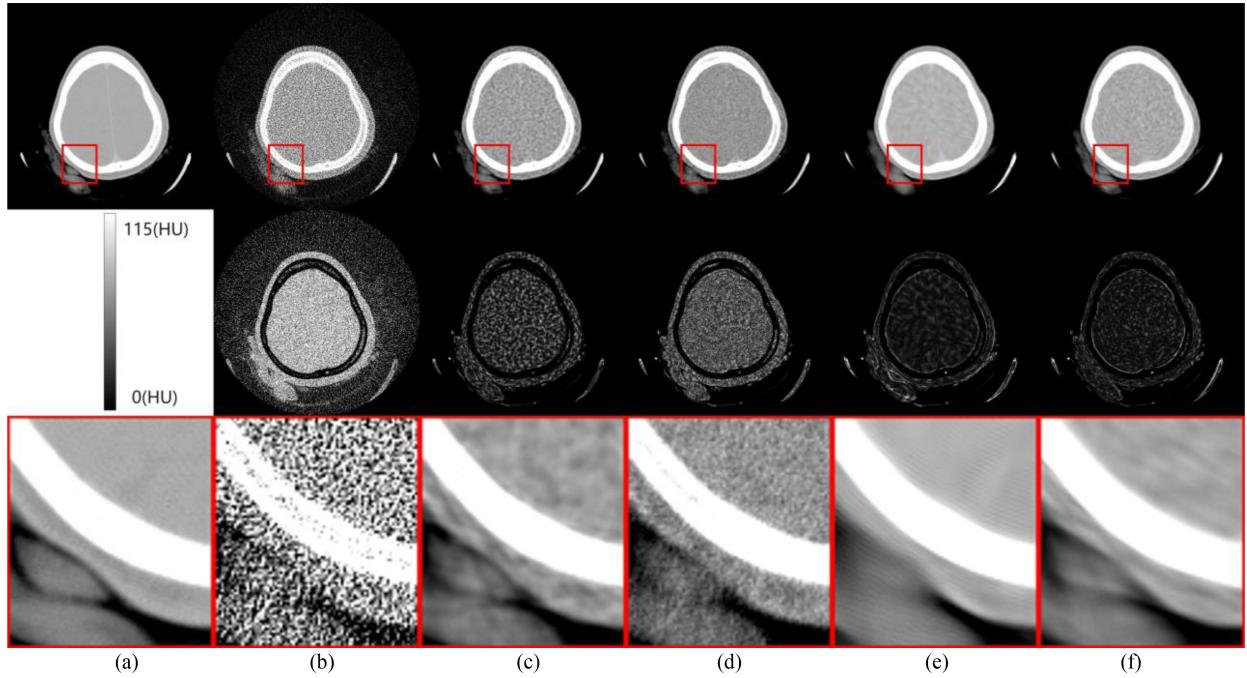


Fig. 8. Reconstruction results from $1e4$ noise level using different methods. (a) Reference image versus the images reconstructed by (b) FBP, (c) CNN, (d) U-Net, (e) NCSN++, and (f) OSDM. The display windows are $[-250, 600]$ HU. The second row depicts the residuals between the reference and reconstructed images.

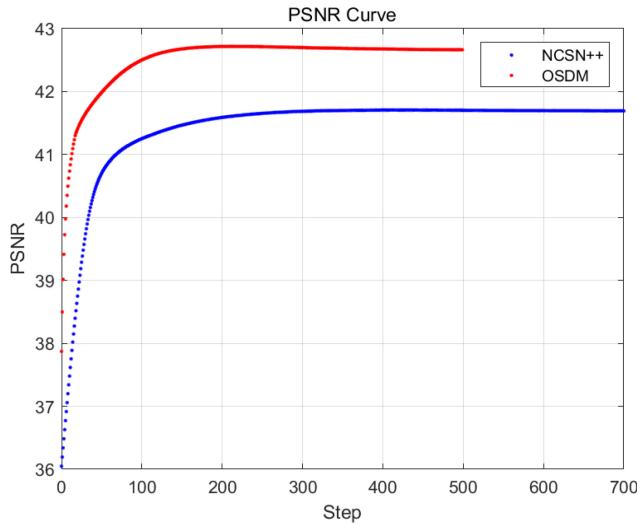


Fig. 9. Convergence curves of OSDM and NCSN++ in terms of PSNR versus iterations when reconstructing the random AAPM challenge data.

consistency module to generate fixed data. Statistical weight W in PWLS helps to decrease the influence of noise on the effect of reconstruction. TV is able to effectively remove noise and eliminate blur from the generated data. Thus, using PWLS and TV simultaneously could accelerate iteration speed and convergence rate, while improving the sampling quality.

F. Profile Lines Analysis

For the comparison of edge-preserving performances, profile lines of reconstructed results from various methods are analyzed in this section. In Fig. 10, the upper part compares

FBP and SART-TV with the ground truth, revealing significant pixel value fluctuations in traditional methods. Conversely, the deep learning methods on the lower part exhibit better performance, but CNN and U-Net do not match the profile of the ground truth very well. In the meanwhile, NCSN++ and OSDM acquire the profiles closer to the ground truth, which verifies the strong profile-keeping capability of the diffusion models. In comprehensive comparison, OSDM generates the most accurate profile. Fig. 10 demonstrates that excellent agreement can be observed from image profiles obtained by our method, which indicates strong profile holding ability of OSDM.

G. Ablation Study

Different Components in OSDM: In order to verify the necessity of corresponding modules in OSDM, we compare it qualitatively and quantitatively with Hankel-based SVD (H-SVD) [52] and H-SVD with diffusion model at noise levels of $1e5$, $5e4$, and $1e4$. All methods utilize PWLS as the data fidelity item to improve image reconstruction quality. It is worth noting that OSDM employs the TV minimization technique embedding into diffusion model to suppress the noise generation in the reconstruction process. In contrast, H-SVD is a conventional algorithm without a network. We calculate the average value of the test set and present the results in Table V to verify the effectiveness of OSDM.

Table V shows the quantitative results for the reconstructions from $1e5$, $5e4$ and $1e4$ noise level and exhibit the benefit from H-SVD and TV steps. The H-SVD step denoising capability can be directly observed from Table V while the difference between the H-SVD with Diffusion Model and

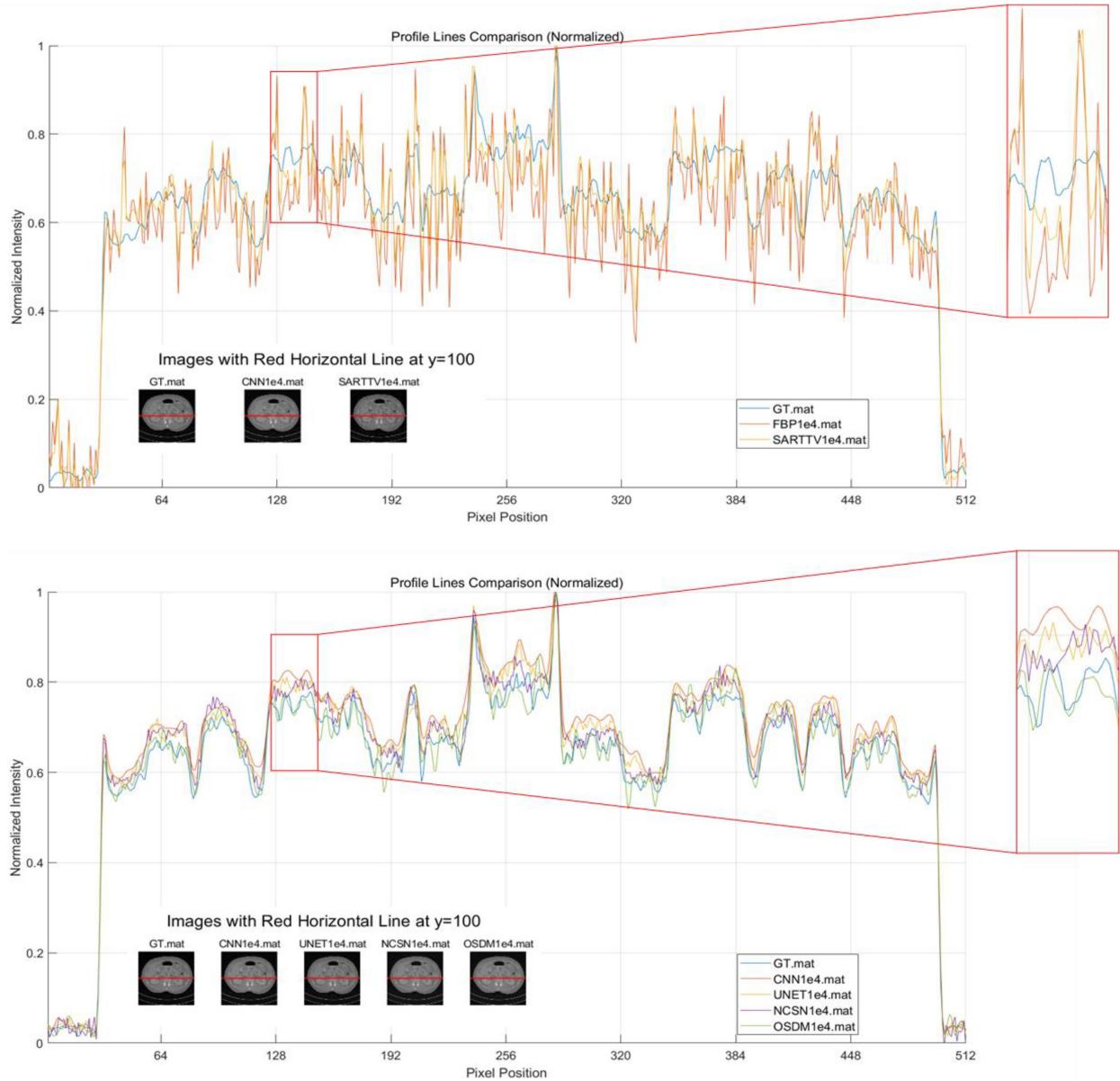


Fig. 10. Comparison of CT image profiles between ground truth and other methods on abdomen images. Red lines represent the location where the line profile is measured.

TABLE V
RECONSTRUCTION PSNR/SSIM/MSES OF AAPM CHALLENGE DATA USING DIFFERENT METHODS AT 1E5, 5E4, AND 1E4 NOISE LEVEL

Noise level	H-SVD	H-SVD with Diffusion Model	OSDM
$a_i = 1e5$	$41.65 \pm 0.93 / 0.9883 \pm 0.0014 / 6.99e-5$	$42.02 \pm 0.86 / 0.9884 \pm 0.0014 / 6.40e-5$	$42.62 \pm 0.51 / 0.9899 \pm 0.0011 / 5.51e-5$
$a_i = 5e4$	$40.24 \pm 1.12 / 0.9839 \pm 0.0024 / 9.80e-5$	$40.79 \pm 0.60 / 0.9841 \pm 0.0019 / 8.40e-4$	$41.20 \pm 0.93 / 0.9857 \pm 0.0026 / 7.74e-5$
$a_i = 1e4$	$35.70 \pm 0.72 / 0.9551 \pm 0.0050 / 2.73e-4$	$36.01 \pm 0.76 / 0.9615 \pm 0.0053 / 2.54e-4$	$37.43 \pm 0.78 / 0.9683 \pm 0.0062 / 1.83e-4$

OSDM refers to the benefit from TV step. It can be observed that OSDM yields the best results in terms of PSNR, SSIM and MSE values, which are consistent with the visual effects. Specifically, compared to H-SVD, OSDM has a significant PSNR gain of 0.97 dB, 0.96 dB and 1.73 dB in cases of 1e5, 5e4 and 1e4 noise levels. The results of OSDM are also impressive in terms of SSIM and MSE.

Figs. 11 and 12 show the reconstruction results of three methods clearly. The quality of H-SVD image is relatively poor that some details and important structures cannot be distinguished. Although H-SVD with diffusion model can remove some artifacts while some important details are lost and the edges of the results are blurred in Fig. 11. In contrast, OSDM achieves the best performance in terms of noise

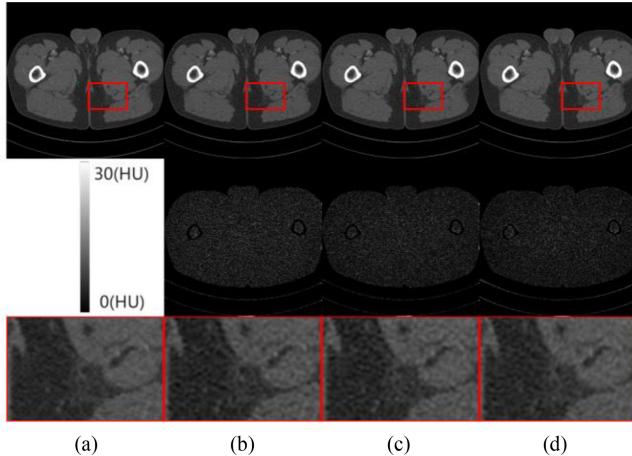


Fig. 11. Reconstruction results from $1e5$ noise level using different methods. (a) Reference image versus the images reconstructed by (b) H-SVD, (c) H-SVD with diffusion model, and (d) OSDM. The display windows are $[-250, 600]$. The second row depicts the residuals between the reference and reconstructed images.

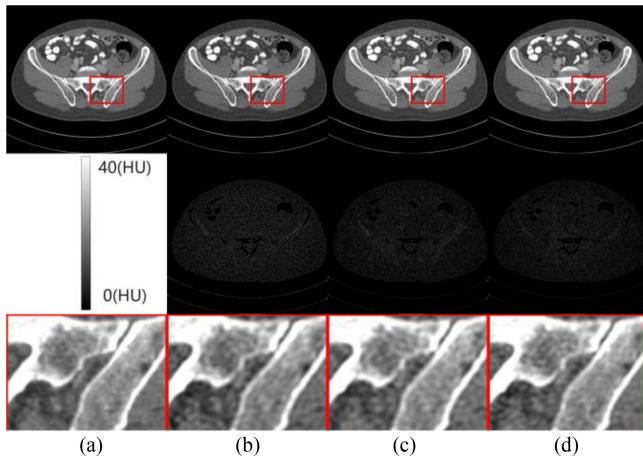


Fig. 12. Reconstruction results from $5e4$ noise level using different methods. (a) Reference image versus the images reconstructed by (b) H-SVD, (c) H-SVD with diffusion model, and (d) OSDM. The display windows are $[-250, 600]$. The second row depicts the residuals between the reference and reconstructed images.

suppression and structural detail preservation (Fig. 11). It can be seen from the residual images that the proposed method provides the best performance as shown in Fig. 11.

Fig. 12 depicts the reconstructed images from a $5e4$ noise level with different methods. Compared with the images in Fig. 11, the artifacts in the image reconstructed by all algorithms are increased as shown in Fig. 12. The H-SVD compromises structural details and suffers from notorious blocky artifacts. Some details in H-SVD with diffusion model are over-smoothed. However, OSDM can retain more structural information and details than other methods. From the ablation study above, we can easily conclude that each module, including H-SVD, diffusion model, and TV minimization technique, plays an important role in OSDM.

V. DISCUSSION

It is very common in medical imaging that medical samples are scarce, which hinders the development of reconstruction

TABLE VI
RECONSTRUCTION PSNR/SSIM/MSES OF AAPM CHALLENGE DATA
USING DIFFERENT NUMBER OF SAMPLES AT $1e5$ NOISE LEVEL

$a_s = 1e5$	1	50	100
PSNR	42.616	42.950	42.953
SSIM	0.9899	0.9899	0.9903
MSE	$5.51e-5$	$5.13e-5$	$5.14e-5$

tasks. Therefore, it is important for deep learning-based models to capture sufficient prior knowledge with limited medical samples. In this article, a OSDM called OSDM in the projection domain is proposed for low-dose CT reconstruction. Although the training process only uses a single sample, OSDM outperforms traditional approaches and also achieves comparable reconstruction compared to DL-based methods. Besides, OSDM manifest great generalization compared with supervised methods. Thus, there is no need for OSDM for retraining when the test set changes.

One-Sample Versus Few-Sample: To verify the performance of OSDM with different number of samples, here we choose 1, 50, and 100 samples as the training set, respectively. The results are shown in Table VI. Although using only one-sample for training does not produce optimal results for all metrics, there is not a significant difference between them. Notably, all metrics begin to stabilize when using 50 samples. This indicates that using only a few samples or even just one sample is feasible. The PSNR values are presented with three decimal places for a more detailed comparison. From Table VI, we could observe that each metric starts to stabilize when the number of training samples reaches 50. Therefore, employing few-samples or even one-sample is exactly feasible. Compared with one-sample training, few-sample leads to higher performance. Further investigation on few-sample reconstruction is planned in our future study.

Practical Limitations: Despite promising experimental results, there still has some limitations of OSDM. As described in this article, data is transformed into the Hankel matrix for processing. Due to the property of redundancy on the Hankel matrix, the amount of data and corresponding storage space are largely amplified. This leads to vaster computational capacity, longer calculation time, and higher-hardware requirements. Encouragingly, some researches have devoted to improving the sampling speed of diffusion models. For that, Xia et al. [53] adapted the fast ordinary differential equation solver to achieve 20x speedup without compromising image quality for low-dose CT reconstruction. Anyway, faster generation retains an ongoing area of low-dose CT reconstruction.

VI. CONCLUSION

Although deep learning-based CT reconstruction methods have made significant progress in recent years, the generalizability and robustness of trained networks remained open challenges. In this study, we introduced a score-based diffusion model for low-dose CT reconstruction using PC sampling to generate the sinogram. Our approach used a fully unsupervised technique to train the score-based diffusion model to capture

the prior distribution of sinogram data. During iterative inference, we performed the numerical SDE solver, PWLS strategy, TV minimization, and LR operation alternatively to achieve image reconstruction. The predictor referred to the numerical solver for reverse-time SDE and Langevin dynamics acted as the corrector. We validated the effectiveness and generalizability of OSDM using the AAPM challenge dataset and CQ500 dataset. Our experimental results showed that OSDM can effectively suppress streaking artifacts and preserve important sinogram details in low-dose CT reconstruction. Furthermore, given challenges with data collection in medical fields due to privacy and security concerns, one-sample or few-sample-based CT reconstruction was crucial. OSDM was a promising approach that achieved adequate performance with a single sample and outperformed with few samples.

ACKNOWLEDGMENT

The authors sincerely thank the anonymous referees for their valuable comments on this work. All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships.

REFERENCES

- [1] M. Bakator and D. Radosav, "Deep learning and medical diagnosis: A review of literature," *Multimodal Tech. Interact.*, vol. 2, no. 3, p. 47, Aug. 2018.
- [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] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, p. 4777, Aug. 2008.
- [4] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2791–2805, 2009.
- [5] Y. Chen et al., "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive weighting nonlocal prior," *Comput. Med. Imag. Graph.*, vol. 33, no. 7, pp. 495–500, 2009.
- [6] J. Liu et al., "3D feature constrained reconstruction for low dose CT Imaging," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 5, pp. 1232–1247, May 2018.
- [7] X. Zheng, X. Lu, S. Ravishankar, Y. Long, and J. Fessler, "Low dose CT image reconstruction with learned sparsifying transform," in *Proc. IEEE Workshop Image, Video, Multidimensional Signal Process*, 2016, pp. 1–5.
- [8] T. Wang, W. Xia, J. Lu, and Y. Zhang, "A review of deep learning CT reconstruction from incomplete projection data," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 8, no. 2, pp. 138–152, Feb. 2024, doi: [10.1109/TRPMS.2023.3316349](https://doi.org/10.1109/TRPMS.2023.3316349).
- [9] W. Zheng, B. Yang, Y. Xiao, J. Tian, S. Liu, and L. Yin, "Low-dose CT image post-processing based on learn-type sparse transform," *Sensors*, vol. 22, no. 8, p. 2883, 2022.
- [10] Y. Chen et al., "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, no. 16, pp. 5803–5820, 2013.
- [11] X. Yin et al., "Domain progressive 3D residual convolution network to improve low-dose CT imaging," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2903–2913, Dec. 2019.
- [12] T. Humphries, D. Si, S. Coulter, M. Simms, and R. Xing, "Comparison of deep learning approaches to low dose CT using low intensity and sparse view data," in *Proc. Med. Imag., Phys. Med. Imag.*, 2019, pp. 1048–1054.
- [13] M. U. Ghani and W. C. Karl, "CNN based sinogram Denoising for low-dose CT," in *Proc. Math. Imag.*, 2018, Art. no. MM2D-5.
- [14] M. Beister, D. Kolditz, and W. A. Kalender, "Iterative reconstruction methods in X-ray CT," *Phys. Med.*, vol. 28, no. 2, pp. 94–108, Apr. 2012.
- [15] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 31, no. 3, pp. 677–688, Mar. 2012.
- [16] Y. Liu, J. Kang, Z. Li, Q. Zhang, and Z. Gui, "Low-dose CT noise reduction based on local total variation and improved wavelet residual CNN," *J. X-Ray Sci. Technol.*, vol. 30, no. 6, pp. 1–14, 2022.
- [17] X. Deng, Y. Zhao, and H. Li, "Projection data smoothing through noise-level weighted total variation regularization for low-dose computed tomography," *J. X-ray Sci. Technol.*, vol. 27, no. 3, pp. 537–557, 2019.
- [18] S. V. M. Sagheer and S. N. George, "Denoising of low-dose CT images via low-rank tensor modeling and total variation regularization," *Artif. Intell. Med.*, vol. 94, pp. 1–17, Mar. 2019.
- [19] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2256–2265.
- [20] L. Yang, Z. Zhang, and S. Hong, "Diffusion models: A comprehensive survey of methods and applications," 2022, *arXiv:2209.00796*.
- [21] Q. Gao and H. Shan, "CoCoDiff: A contextual conditional diffusion model for low-dose CT image denoising," in *Proc. Develop. X-Ray Tomogr. XIV*, 2022, Art. no. pp. 92–98.
- [22] Q. Lyu and G. Wang, "Conversion between CT and MRI images using diffusion and score-matching models," 2022, *arXiv:2209.12104*.
- [23] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021, pp. 1–36.
- [24] H. Chung and J. C. Yea, "Score-based diffusion models for accelerated MRI," 2021, *arXiv:2110.05243*.
- [25] A. Jalal, M. Arvinte, G. Daras, E. Price, A. Dimakis, and J. Tamir, "Robust compressed sensing MRI with deep generative priors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 1–17.
- [26] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," 2021, *arXiv:2111.08005*.
- [27] K. H. Jin and J. C. Ye, "Sparse and low-rank decomposition of a Hankel structured matrix for impulse noise removal," *IEEE Trans. Image Process.*, vol. 27, pp. 1448–1461, 2018.
- [28] K. H. Jin, J.-Y. Um, D. Lee, J. Lee, S.-H. Park, and J. C. Ye, "MRI artifact correction using sparse+low-rank decomposition of annihilating filter-based Hankel matrix," *Magn. Reson. Med.*, vol. 78, no. 1, pp. 327–340, Jul. 2017.
- [29] J. Min, L. Carlini, M. Unser, S. Manley, and J. C. Ye, "Fast live cell imaging at nanometer scale using annihilating filter-based low-rank Hankel matrix approach," in *Proc. SPIE*, 2015, pp. 204–211.
- [30] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1289–1296, Jun. 2018.
- [31] T. R. Shaham, T. Dekel, and T. Michaeli, "Singan: Learning a generative model from a single natural image," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 4570–4580.
- [32] V. Sushko, J. Gall, and A. Khoreva, "One-shot GAN: Learning to generate samples from single images and videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2596–2600.
- [33] Z. Zheng, J. Xie, and P. Li, "Patchwise generative ConvNet: Training energy-based models from a single natural image for internal learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2961–2970.
- [34] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jiang, "GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation," *Med. Phys.*, vol. 37, no. 4, pp. 1757–1760, 2010.
- [35] T. Li et al., "Nonlinear sinogram smoothing for low-dose X-ray CT," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 5, pp. 2505–2513, Oct. 2004.
- [36] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.
- [37] J. Wang, H. Lu, J. Wen, and Z. Liang, "Multiscale penalized weighted least-squares sinogram restoration for low-dose X-ray computed tomography," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1022–1031, Mar. 2008.
- [38] J. Wang, T. Li, and L. Xing, "Iterative image reconstruction for CBCT using edge-preserving priori," *Med. Phys.*, vol. 36, no. 1, pp. 252–260, 2009.
- [39] Y. Zhang, J. Zhang, and H. Lu, "Statistical sinogram smoothing for low-dose CT with segmentation-based adaptive filtering," *IEEE Trans. Nucl. Sci.*, vol. 57, no. 5, pp. 2587–2598, Oct. 2010.

- [40] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, 2017.
- [41] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.
- [42] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 11918–11930.
- [43] 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.
- [44] (Am. Assoc. Phys. Med., Alexandria, VA, USA). *Low Dose CT Grand Challenge*. Accessed: Apr. 6, 2017. [Online]. Available: <http://www.aapm.org/GrandChallenge/LowDoseCT/>
- [45] 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.
- [46] F. Jacobs, E. Sundermann, B. De Sutter, M. Christiaens, and I. Lemahieu, "A fast algorithm to calculate the exact radiological path through a pixel or voxel space," *J. Comput. Inf. Technol.*, vol. 6, no. 1, pp. 89–94, 1998.
- [47] S. Chilamkurthy et al., "Development and validation of deep learning algorithms for detection of critical findings in head CT scans," 2018, *arXiv:1803.05854*.
- [48] J. Adler, H. Kohr, and O. Oktem, "Operator discretization library (ODL)," 2017. [Online]. Available: <https://github.com/odlgroup/odl>
- [49] Y. Censor and T. Elfving, "Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 40–58, 2002.
- [50] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, pp. 3142–3155, 2017.
- [51] H. Lee, J. Lee, H. Kim, 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.
- [52] P. J. Shin et al., "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magn. Reson. Med.*, vol. 72, no. 4, pp. 959–970, 2014.
- [53] W. Xia, Q. Lyu, and G. Wang, "Low-dose CT using Denoising diffusion probabilistic model for 20x times speedup," 2022, *arXiv:2209.15136*.