

CoreDiff: Contextual Error-Modulated Generalized Diffusion Model for Low-Dose CT Denoising and Generalization

Qi Gao, Zilong Li, Junping Zhang, *Senior Member, IEEE*, Yi Zhang, *Senior Member, IEEE* and Hongming Shan, *Senior Member, IEEE*

Abstract—Low-dose computed tomography (CT) images suffer from noise and artifacts due to photon starvation and electronic noise. Recently, some works have attempted to use diffusion models to address the over-smoothness and training instability encountered by previous deep-learning-based denoising models. However, diffusion models suffer from long inference times due to the large number of sampling steps involved. Very recently, cold diffusion model generalizes classical diffusion models and has greater flexibility. Inspired by the cold diffusion, this paper presents a novel COntextual eRror-modulated gEneralized Diffusion model for low-dose CT (LDCT) denoising, termed CoreDiff. First, CoreDiff utilizes LDCT images to displace the random Gaussian noise and employs a novel mean-preserving degradation operator to mimic the physical process of CT degradation, significantly reducing sampling steps thanks to the informative LDCT images as the starting point of the sampling process. Second, to alleviate the error accumulation problem caused by the imperfect restoration operator in the sampling process, we propose a novel Contextual Error-modulAted Restoration Network (CLEAR-Net), which can leverage contextual information to constrain the sampling process from structural distortion and modulate time step embedding features for better alignment with the input at the next time step. Third, to rapidly generalize to a new, unseen dose level with as few resources as possible, we devise a one-shot learning framework to make CoreDiff generalize faster and better using only a single LDCT image (un)paired with NDCT. Extensive experimental results on two datasets demonstrate that our CoreDiff outperforms competing methods in denoising and generalization performance, with a clinically acceptable inference time.

Index Terms—Low-dose CT, denoising, diffusion model, generalized diffusion, one-shot learning

I. INTRODUCTION

COMPUTED tomography (CT) is a widely-used imaging modality in clinical diagnosis. However, X-ray ionizing radiation in CT scans could cause health risks such as hair loss and cancer [1], [2]. One can reduce the radiation dose by lowering the tube current in clinical practice. Unfortunately, the resulting low-dose CT (LDCT) images contain severe noise and artifacts, compromising the radiologists' diagnosis. When the raw data are accessible, vendor-specific sinogram preprocessing

Q. Gao and H. Shan are with the Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai 200433, China (e-mail: qgao21@m.fudan.edu.cn; hmshan@fudan.edu.cn)

Z. Li and J. Zhang are with the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: zilongli21@m.fudan.edu.cn; jpzhang@fudan.edu.cn)

Y. Zhang is with School of Cyber Science and Engineering, Sichuan University, Chengdu, Sichuan 610065, China (e-mail: yzhang@scu.edu.cn)

or iterative reconstruction algorithms can effectively remove noise in LDCT images. However, sinogram preprocessing may cause blurred edges and resolution loss while iterative reconstruction methods suffer from expensive computational costs [3]–[5]. In addition, raw data are typically not available to researchers due to commercial privacy. Unlike them, image post-processing algorithms [6]–[8] directly process the reconstructed images and are gaining popularity due to their plug-and-play nature without access to raw data.

In recent years, many efforts have been made to develop deep learning (DL) techniques for LDCT image post-processing, achieving promising performance [9], [10]. Initially, some researchers optimize the encoder-decoder network by minimizing the pixel-wise loss between the denoised and normal-dose CT (NDCT) images; one representative model is the residual encoder-decoder convolutional neural network (RED-CNN) [11]. Further, Xia *et al.* integrated RED-CNN into a parameter-dependent framework (PDF-RED-CNN) for multiple geometries and dose levels [12]. Despite the outstanding denoising performance, these methods often lead to over-smoothing images [13], [14]. To alleviate this problem, some works use generative adversarial networks (GANs) to preserve more textures and details as close to NDCT images as possible [15]–[17]. For example, Yang *et al.* combined Wasserstein GAN and perceptual loss (WGAN-VGG) to produce more realistic denoised images [18]. Huang *et al.* proposed a dual-domain GAN (DU-GAN) to learn the global and local differences between the denoised and NDCT images [19]. However, GANs are usually difficult to train due to their adversarial nature and require careful design of optimization and network architectures to ensure convergence [20], [21].

Recently, diffusion models have received much attention due to their impressive image generation performance [22]–[27], demonstrating the advantages of multiple generative models: good distribution coverage similar to variational autoencoder and better generation quality than GANs [28]–[30]. However, since the diffusion models generate images progressively from Gaussian noise, it suffers from expensive computational costs for inference due to the multiple iterative sampling; *e.g.* denoising diffusion probabilistic model (DDPM) [23] requires 1,000 sampling steps. This limits its application in various real-time scenarios, especially in the field of medical imaging [31]–[34]. Some works accelerate diffusion models to make them practical. For example, Nichol and Dhariwal enhanced the log-likelihood performance of the DDPM and reduced the sampling

steps to 100 [25]. Xia *et al.* used a fast ordinary differential equation solver to accelerate the DDPM for LDCT image denoising, requiring only 50 sampling steps [32]. Despite these improved diffusion models reducing inference time to some extent, they focus on the trade-off between performance and sampling speed within the theoretical framework of classical diffusion models. Recently, a generalized diffusion model, referred to as cold diffusion, extends classical diffusion models by gradually degrading images through a pre-defined degradation operator, such as adding noise, blurring, downsampling, etc [35], [36]. Cold diffusion uses a learnable restoration operator to reverse the degradation process and generates images through a “restoration-redegradation” sampling process. Although cold diffusion allows customizing the diffusion process, its performance is subject to the learned restoration operator. In practice, the learnable restoration operator may be imperfect, leading to accumulated errors between the restored and ground truth images after multiple sampling iterations and causing non-negligible pixel-wise deviations.

In this paper, we propose a contextual error-modulated generalized diffusion model (CoreDiff) for LDCT denoising inspired by cold diffusion. To accelerate sampling, we develop a mean-preserving degradation operator applicable to the LDCT denoising task with the LDCT images as the endpoint of the diffusion process (forward) and the starting point of the sampling process (reverse). In doing so, the sampling steps can be significantly reduced since LDCT images (warm state) are more informative than random Gaussian noise (hot state). To alleviate the accumulated error caused by imperfect restoration operators during the sampling process, we propose a novel contextual error-modulated restoration network (CLEAR-Net), which can leverage rich contextual information from adjacent slices to mitigate structural distortion in z-axis, and rectify the misalignment between the input image and time-step embedding features through an error-modulated module. Finally, benefiting from the proposed mean-preserving degenerate operator, we devise a one-shot learning framework, which can quickly generalize CoreDiff to a new, unseen dose level using a single LDCT image (un)paired with NDCT.

In summary, the contributions of this work are listed as follows.

- 1) We propose a novel generalized diffusion model CoreDiff for LDCT denoising, in which the resulting diffusion process mimics the physical process of CT image degradation. To the best of our knowledge, this is the first work to extend the cold diffusion model for LDCT denoising.
- 2) We introduce a novel restoration network CLEAR-Net, which can mitigate accumulated errors by constraining the sampling process using contextual information among adjacent slices and calibrating the time step embedding feature using the latest prediction.
- 3) We further devise a one-shot learning framework, which can quickly and easily adapt the trained CoreDiff to a new, unseen dose level. This can be done with a single LDCT image (un)paired with NDCT.
- 4) Extensive experiment results on two testing datasets demonstrate the superior performance of the proposed CoreDiff, with a clinically acceptable inference time of

0.12 seconds per slice.

II. METHOD

In this section, we first introduce the basic principles of the cold diffusion model and the error accumulation issue. Then we present our CoreDiff for LDCT denoising in a generalized diffusion model framework with a new mean-preserving diffusion process and a new contextual error-modulated restoration network (CLEAR-Net), followed by a one-shot learning framework for rapid generalization.

A. Preliminaries: Cold Diffusion

Cold diffusion model is a generalized diffusion model [35], which extends the diffusion and sampling of Gaussian noise to any type of degradation of images such as adding noise, blurring, downsampling, etc. Specifically, given an image \mathbf{x}_0 from the training data distribution Q , a custom *degradation* operator $\mathcal{D}(\cdot)$ is used to gradually degrade the image \mathbf{x}_0 (cold state) into the image \mathbf{x}_T sampled from a random initial distribution P (hot state), e.g. Gaussian distribution, where T is the total time steps for diffusion. Then, image \mathbf{x}_t of any time step t during the diffusion process is defined as:

$$\mathbf{x}_t = \mathcal{D}(\mathbf{x}_0, \mathbf{x}_T, t), \quad (1)$$

where t corresponds to the degree of degradation and the operator $\mathcal{D}(\cdot)$ should be continuous for any t . In the context of LDCT denoising, adding noise is the most related degradation operator. For the diffusion process of adding noise, the degradation operator in the cold diffusion model is the same as the one used in classical diffusion models, defined as:

$$\mathbf{x}_t = \mathcal{D}(\mathbf{x}_0, \mathbf{x}_T, t) = \sqrt{\alpha_t} \mathbf{x}_0 + \sqrt{(1 - \alpha_t)} \mathbf{x}_T, \quad (2)$$

where \mathbf{x}_T is random noise with known distribution and $\alpha_t < \alpha_{t-1}, \forall 1 \leq t \leq T$.

In the reverse process, we first sample \mathbf{x}_T from P , and then use a *restoration* operator $\mathcal{R}(\cdot)$ to reverse the diffusion process, which can be expressed as follows:

$$\hat{\mathbf{x}}_0 = \mathcal{R}(\mathbf{x}_t, t) \approx \mathbf{x}_0. \quad (3)$$

In practice, $\mathcal{R}(\cdot)$ is a neural network parameterized by θ , which can be optimized by the following objective function:

$$\min_{\theta} \mathbb{E}_{\mathbf{x}_0 \sim Q, \mathbf{x}_T \sim P} \|\mathcal{R}_{\theta}(\mathcal{D}(\mathbf{x}_0, \mathbf{x}_T, t), t) - \mathbf{x}_0\|. \quad (4)$$

Note that for any t , $\mathcal{R}_{\theta}(\cdot)$ can directly generate the restored image $\hat{\mathbf{x}}_0$ from \mathbf{x}_T . However, we highlight that such a one-step prediction will produce blurred image $\hat{\mathbf{x}}_0$ with severe detail loss [35]. To address this, following the annealing sampling algorithm in classical diffusion models [23], [25], [28], the cold diffusion model uses a “restoration-redegradation” sampling algorithm to gradually generate images with a total of T sampling steps. Image $\hat{\mathbf{x}}_{t-1}$ at time step $t-1$ can be calculated based on the prediction $\hat{\mathbf{x}}_0$ as follows:

$$\hat{\mathbf{x}}_{t-1} = \mathcal{D}(\hat{\mathbf{x}}_0, \mathbf{x}_T, t-1). \quad (5)$$

Although such an iterative sampling algorithm can produce sharper images than a one-step prediction, the prediction error

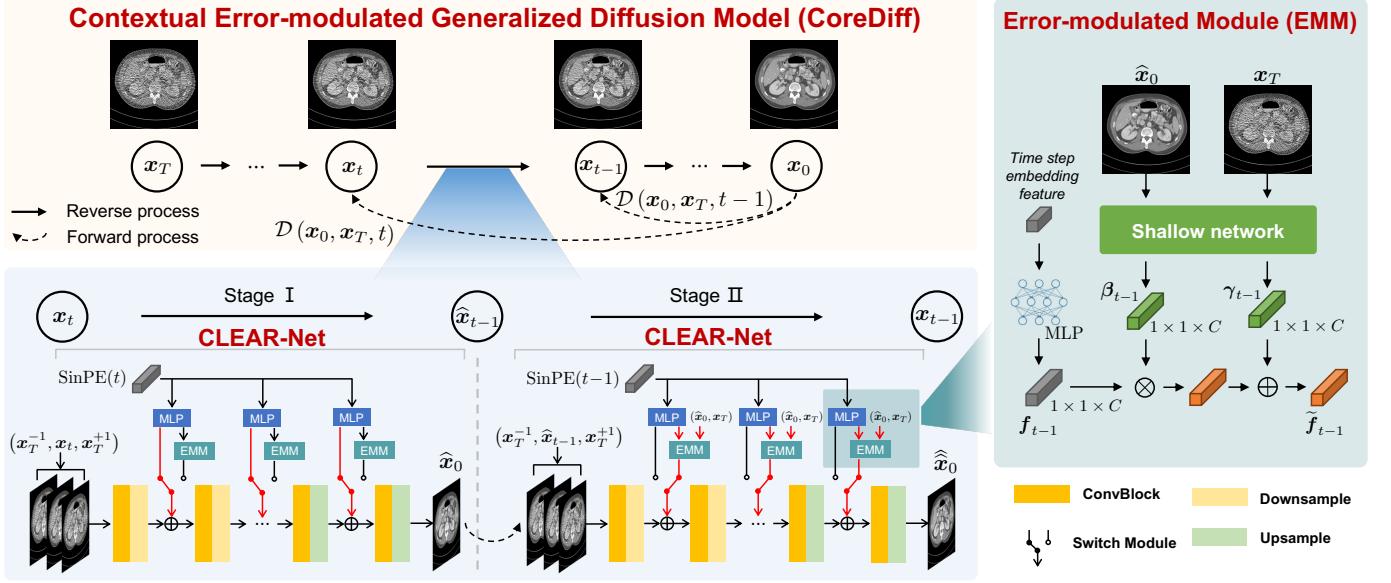


Fig. 1. Overview of the proposed CoreDiff for low-dose CT denoising. The introduced generalized diffusion model leverages a novel degradation operator to mimic the physical process of CT image degradation during the diffusion process. The proposed CLEAR-Net can alleviate the accumulated error and is trained in a two-stage manner for each time step; one key feature of CLEAR-Net is the error-modulated module (EMM) that can calibrate the time step embedding feature with the latest prediction and the given input LDCT image.

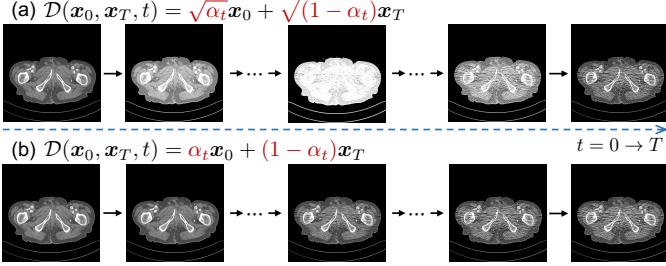


Fig. 2. Comparison of (a) the degradation operator in Eq. (2) and (b) the proposed one in Eq. (7). The proposed operator achieves a mean-preserving process, simulating the physical process of CT degradation.

between x_0 and \hat{x}_0 could introduce misalignment between \hat{x}_{t-1} and time step $t - 1$. As a result, the prediction bias of $\mathcal{R}_\theta(\cdot)$ may be further worsened by the misalignment, as errors are accumulated during the sampling process. Bansal *et al.* proposed an improved sampling algorithm to reduce this accumulated error [35]:

$$x_{t-1} = x_t - \mathcal{D}(\hat{x}_0, \hat{x}_T, t) + \mathcal{D}(\hat{x}_0, \hat{x}_T, t-1), \quad (6)$$

where $\hat{x}_T = (x_t - \sqrt{\alpha_t} \hat{x}_0) / \sqrt{1 - \alpha_t}$. Although the improved sampling algorithm in Eq. (6) mitigates the issue of error accumulation and has been shown to produce higher quality images [35], [36], it does not rectify the misalignment between the input and its corresponding time step, which could cause a non-negligible shift in the pixel value.

B. The Proposed CoreDiff Model

Fig. 1 presents the overall architecture of our CoreDiff, which involves a generalized diffusion model with LDCT images as the endpoint of the diffusion process, a new mean-preserving degradation operator to mimic the physical process

of CT degradation, and a novel CLEAR-Net to address the accumulated errors and misalignment in cold diffusion.

1) *Generalized diffusion model for low-dose CT:* Previous diffusion-based LDCT denoising methods [31], [32] typically characterize the diffusion process as the addition of Gaussian noise and use LDCT images as a condition to predict the corresponding NDCT image. However, it is important to note that the statistical characteristics of noise in CT images are complex and cannot be generalized by a Gaussian distribution. Moreover, we find that the widely-adopted degradation operator in Eq. (2) alters the mean value of the intermediate images in the diffusion process as shown in Fig. 2(a), which departs from the actual physical process of CT degradation due to the dose reduction.

Unlike previous diffusion-based methods that transform the LDCT denoising task into a conditional image generation task with random Gaussian noise at the starting point of the sampling process and require a significant number of steps T to generate an accurately estimated image, we propose a generalized diffusion model for LDCT denoising, which uses LDCT images as the endpoint of the diffusion process, *i.e.* x_T . To make the diffusion process mimic the physical process of CT image degradation, we introduce a new mean-preserving degradation operator $\mathcal{D}(\cdot)$ defined as follows:

$$x_t = \mathcal{D}(x_0, x_T, t) = \alpha_t x_0 + (1 - \alpha_t) x_T, \quad (7)$$

where image x_t at each time step retains the noise statistics specific to LDCT image x_T . In addition, another merit of this operator is that the diffusion process is a mean-preserving process as shown in Fig. 2(b). Such property not only makes the diffusion process of CoreDiff consistent with the LDCT image degradation process but also is important for our one-shot learning framework that will be discussed in Sec. II-C.

Algorithm 1 Training for CoreDiff

Input: Paired ND/LDCT image sets $I = \{(\mathbf{x}_0, \mathbf{x}_T)\}_{i=1}^N$, total time steps T
Output: Trained CLEAR-Net \mathcal{R}_θ

- 1: **Initialization:** Randomly initializes CLEAR-Net \mathcal{R}_θ
- 2: **repeat**
- 3: **Sample** $(\mathbf{x}_0, \mathbf{x}_T) \sim I$
- 4: **Sample** $t \sim \text{Uniform}\{1, \dots, T\}$
- 5: **Calculate** \mathbf{x}_t^c by Eq. (7) and $\text{Concat}(\cdot)$ \triangleright Stage I
- 6: $\hat{\mathbf{x}}_0 \leftarrow \mathcal{R}_\theta(\mathbf{x}_t^c, t)$
- 7: **Calculate** $\hat{\mathbf{x}}_{t-1}^c$ by Eq. (5) and $\text{Concat}(\cdot)$ \triangleright Stage II
- 8: $\hat{\mathbf{x}}_0 \leftarrow \mathcal{R}_\theta(\hat{\mathbf{x}}_{t-1}^c, t-1, \mathcal{F}_\phi(\hat{\mathbf{x}}_0, \mathbf{x}_T))$
- 9: **Update** θ, ϕ by Eq. (10)
- 10: **until** converged

The LDCT image \mathbf{x}_T can be considered as an intermediate state between the cold state (clean image) and the hot state (random noise), which we refer to as the warm state. As a result, the proposed CoreDiff can perform sampling from the warm state using a smaller T , instead of starting from random Gaussian noise.

2) *Contextual Error-modulated Restoration Network (CLEAR-Net):* To mitigate accumulated errors and the misalignment in cold diffusion caused by an imperfect restoration network, we introduce a novel restoration network, called CLEAR-Net. Based on the “restoration-redegradation” sampling algorithm, we split each time step in the training process into two stages, as shown in Fig. 1: 1) in Stage I, we first obtain the degraded image \mathbf{x}_t using Eq. (7), and then use CLEAR-Net $\mathcal{R}_\theta(\cdot)$ to estimate $\hat{\mathbf{x}}_0$; and 2) in Stage II, we perform the redegradation operation in Eq. (5) to compute $\hat{\mathbf{x}}_{t-1}$ based on the latest prediction $\hat{\mathbf{x}}_0$, and then use the same network $\mathcal{R}_\theta(\cdot)$ to predict the NDCT image.

The novelties of our CLEAR-Net are two-fold. On the one hand, inspired by the contextual information used in [5], [31], we introduce the contextual information from adjacent slices of \mathbf{x}_T to mitigate the structural distortion during the sampling process. More specifically, we assume the adjacent slices of \mathbf{x}_T are \mathbf{x}_T^{-1} and \mathbf{x}_T^{+1} , corresponding to its upper and lower slice, respectively, we concatenate $\mathbf{x}_t \in \mathbb{R}^{1 \times H \times W}$ at each step and the adjacent slices at the starting point, \mathbf{x}_T^{-1} and \mathbf{x}_T^{+1} , along the channel dimension, which yields a contextual version of \mathbf{x}_t , i.e. $\mathbf{x}_t^c = \text{Concat}(\mathbf{x}_T^{-1}, \mathbf{x}_t, \mathbf{x}_T^{+1}) \in \mathbb{R}^{3 \times H \times W}$. Since the adjacent slices remain unchanged during sampling, thereby constraining the network $\mathcal{R}_\theta(\cdot)$ to produce continuous z-axis structures.

On the other hand, CLEAR-Net leverages an error-modulated module (EMM) to calibrate the misalignment between the input to the network $\hat{\mathbf{x}}_{t-1}$ and the time-step embedding features of $t-1$. Specifically, our EMM is a feature-wise linear modulation [37]–[39] to modulate the time step embedding features, in which the modulation factors at time step $t-1$ are estimated as follows:

$$\beta_{t-1}, \gamma_{t-1} = \mathcal{F}_\phi(\hat{\mathbf{x}}_0, \mathbf{x}_T), \quad \hat{\mathbf{x}}_0 = \mathcal{R}_\theta(\mathbf{x}_t, t), \quad (8)$$

where $\mathcal{F}_\phi(\cdot)$ is a shallow network parameterized by ϕ to estimate the modulation factors based on the latest prediction

Algorithm 2 Sampling for CoreDiff

Input: A test LDCT image \mathbf{x}_T

Output: Denoised image \mathbf{x}_0

- 1: **Load** the trained CLEAR-Net \mathcal{R}_θ
- 2: $\mathbf{x}_T^c \leftarrow \text{Concat}(\mathbf{x}_T^{-1}, \mathbf{x}_T, \mathbf{x}_T^{+1})$
- 3: **for** $t = T, T-1, \dots, 1$ **do**
- 4: $\mathbf{x}_0 \leftarrow \mathcal{R}_\theta(\mathbf{x}_t^c, t, \mathcal{F}_\phi(\hat{\mathbf{x}}_0, \mathbf{x}_T))$
- 5: **Calculate** $\hat{\mathbf{x}}_{t-1}^c$ by Eq. (6) and $\text{Concat}(\cdot)$
- 6: **end for**

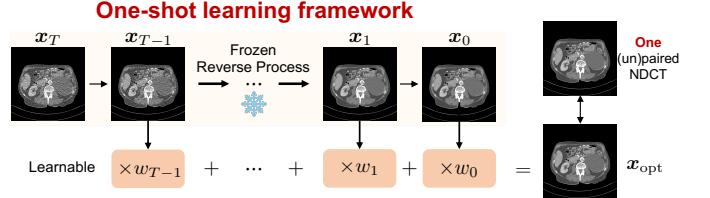


Fig. 3. Framework of one-shot learning for rapid generalization.

$\hat{\mathbf{x}}_0$ and the initial input LDCT image \mathbf{x}_T . Then the time step embedding features of $t-1$ are modulated as follows:

$$\tilde{\mathbf{f}}_{t-1} = \beta_{t-1} \mathbf{f}_{t-1} + \gamma_{t-1}, \quad \mathbf{f}_{t-1} = \text{MLP}(\text{SinPE}(t-1)), \quad (9)$$

where $\text{SinPE}(\cdot)$ represents Sinusoidal position encoding for time step $t-1$, \mathbf{f}_{t-1} is the time step embedding feature from a multi-layer perceptron (MLP), and $\tilde{\mathbf{f}}_{t-1}$ is the modulated one. Note that the proposed EMM is only involved in Stage II and the modulated features are used after each up-/down-sampling operation in $\mathcal{R}_\theta(\cdot)$.

With the proposed CLEAR-Net, the final training objective of our CoreDiff is defined as follows:

$$\min_{\theta, \phi} \mathbb{E} \left[\underbrace{\|\mathcal{R}_\theta(\mathbf{x}_t^c, t) - \mathbf{x}_0\|^2}_{\text{Stage I}} + \underbrace{\|\hat{\mathbf{x}}_0 - \mathbf{x}_0\|^2}_{\text{Stage II}} \right], \quad (10)$$

where $\hat{\mathbf{x}}_0$ denotes the output of $\mathcal{R}_\theta(\cdot)$ at time step $t-1$ in Stage II; i.e. $\hat{\mathbf{x}}_0 = \mathcal{R}_\theta(\hat{\mathbf{x}}_{t-1}^c, t-1, \mathcal{F}_\phi(\hat{\mathbf{x}}_0, \mathbf{x}_T))$ and $\hat{\mathbf{x}}_{t-1}^c = \text{Concat}(\mathbf{x}_T^{-1}, \hat{\mathbf{x}}_{t-1}, \mathbf{x}_T^{+1})$.

Finally, in the sampling process of CoreDiff, the degradation operator and the restoration operator are performed only once at each time step. With the trained CLEAR-Net, we use the improved sampling algorithm in Eq. (6) and replace the coefficients according to our degenerate operator in Eq. (7) to infer the final denoised image. The training and sampling (inference) procedures are shown in Algs. 1 and 2, respectively.

C. One-shot Learning for Rapid Generalization

LDCT images acquired in clinical practice are diverse due to different equipments and protocols. *How to rapidly adapt one trained model to new unseen dose levels using as few as resources* is an important clinical question [12], [40]–[42].

Here, we devise a one-shot learning (OSL) framework specifically designed for the trained CoreDiff with only as few as T learnable parameters and enable training with a single LDCT image, as depicted in Fig. 3. With the introduced mean-preserving degradation operator in Eq. (7), the “restoration-redegradation” process of the CoreDiff is able to progressively

remove the noise and artifacts in the image, yielding a series of denoised images with the same mean and varying degrees of noise. Our idea is to integrate these images to produce a visually optimal denoised image for a new, unseen dose level, which is implemented as:

$$\mathbf{x}_{\text{opt}} = \sum_{t=0}^{T-1} w_t \mathbf{x}_t \text{ s.t. } \sum_{t=0}^{T-1} w_t = 1, \forall t, w_t \geq 0, \quad (11)$$

where $w_t, t = 0, \dots, T - 1$ are the learnable weights used to synergize the images at each step, and \mathbf{x}_{opt} represents the optimal denoised image for the new dose level. When training this one-shot learning framework, the parameters of CLEAR-Net remain frozen while only $w_t, t = 0, \dots, T - 1$ need to be learned. Therefore, we can train the framework by dividing a single image into multiple patches. Even though the new LDCT and NDCT images are not paired, our one-shot learning framework would not introduce structural distortions since all \mathbf{x}_t correspond to the same NDCT image. To ensure the optimal denoised images have a better visual perception and prevent over-smoothing, we use the perceptual loss [18] to guide the learning of those T parameters.

III. EXPERIMENTS AND RESULTS

A. Datasets

1) *Mayo 2016 Dataset*: We use the “2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge” dataset for training and testing [43], which contains 5,936 1mm thickness normal-dose CT slices from 10 patients. We randomly selected 9 patients as the training set and the remaining one as the testing set. To obtain different dose level images, a proven LDCT simulation algorithm with the widely-recognized “Poisson+Gaussian” noise model is used to generate low-dose projection [44], [45]:

$$\mathbf{p}_{\text{ld}} = \ln \frac{I_0}{\text{Poisson}(I_0 \exp(-\mathbf{p}_{\text{hd}})) + \text{Gaussian}(0, \sigma_e^2)}, \quad (12)$$

where \mathbf{p}_{ld} and \mathbf{p}_{hd} represent the low-dose and normal-dose projections, respectively. I_0 is the number of incident photons, which is set to 1.5×10^5 . σ_e^2 is the variance of electronic noise, which is fixed at 10 according to [12]. Then, the filtered back projection (FBP) algorithm is used to reconstruct images. In this experiment, we simulated 50%, 25%, 10% and 5% dose data, of which 5% corresponds to ultra-low-dose situation [46], [47].

2) *Mayo 2020 Dataset*: To verify the generalization performance of different methods, Mayo 2020 dataset [48] is used as external validation, which is the “Low Dose CT Image and Projection Data” latest released by Mayo Clinic. This dataset contains 299 scans from two vendors, which also provided 25% dose data for the head and abdomen and 10% dose data for chest scans. We randomly selected 5 chest and 5 abdomen scans, containing 800 images for mixed dose levels testing.

B. Implementation Details

Following [49], we use a U-Net as the backbone of the proposed CLEAR-Net, which consists of two downsampling blocks, one middle block, and two upsampling blocks. The

input to CLEAR-Net is of size $3 \times 512 \times 512$ containing the contextual information of adjacent slices. We use Adam optimizer to optimize CoreDiff with a learning rate of 2×10^{-4} and a total of 150k iterations for training. $\alpha_1, \dots, \alpha_T$ are set to vary linearly from 0.999 to 0. We conduct data simulations based on the MIRT toolbox [50]. We implement CoreDiff in PyTorch and trained it on one NVIDIA RTX 3090 GPU (24GB) with a mini-batch of size 2.

We compare our CoreDiff with four types of baseline methods, including 1) iterative reconstruction algorithm: Penalized Weighted Least Squares model (PWLS) [4]; 2) RED-CNN-based methods: RED-CNN [11] and PDF-RED-CNN [12]; 3) GAN-based methods: WGAN-VGG [18], DU-GAN [19], and Content-Noise Complementary Learning with U-Net (CNCL-U-Net) [51]; and 4) Diffusion-based methods: Improved DDPM (IDDPMP) [25]. We set the hyperparameters of the compared DL-based methods following the original paper or official source codes, while the hyperparameters of the PWLS were fine-tuned with guidance from [3]. To perform a fair comparison, all deep learning methods were trained and tested on either 25% or 5% doses. 50% dose testing data was used to validate the generalization performance of the methods, while 10% dose corresponded to the chest scan in the Mayo 2020 dataset for testing our one-shot learning framework. We also trained an additional PDF-RED-CNN on all doses of training data, referred to as PDF-RED-CNN*; we adjust the 7 geometry and dose conditional parameters used in the original paper to 1 parameter, *i.e.*, dose level. For the IDDPMP model, we set $T = 1000$ for training as suggested by their original paper, and then use 1000, 50, and 10 sampling steps during reference for comparison; the resultant models are named as IDDPMP-1000, IDDPMP-50, and IDDPMP-10. Our CoreDiff uses the same number of steps for training and inference; unless noted otherwise, $T = 10$ for CoreDiff and the resultant model is named CoreDiff-10.

Three objective image quality assessment metrics are used to quantitatively evaluate the denoising performance: peak signal-to-noise ratio (PSNR), structural similarity (SSIM) index, and root mean square error (RMSE). Higher PSNR and SSIM and lower RMSE denote better performance. Unless noted otherwise, all metrics are calculated based on a CT window of [-1000, 1000]HU.

C. Performance Comparison on Mayo 2016 Dataset

1) *25% dose*: Fig. 4 presents a representative slice of 25% dose testing data denoised by different methods for visual comparison. The orange arrow indicates the location of the lesion in the red region of interest (ROI). Although the RED-CNN-based methods effectively remove noise from the LDCT image, it tends to blur fine details. Among the GAN-based methods, WGAN-VGG introduces velvet artifacts, and DU-GAN provides textures closer to NDCT images. CNCL-U-Net preserves the most details, but its residual map shows a noticeable difference in predicting the bone edge. Both in terms of texture preservation and detail retention, diffusion-based methods surpass other compared methods. For the IDDPMP, lowering the number of sampling steps T to 50 has little impact

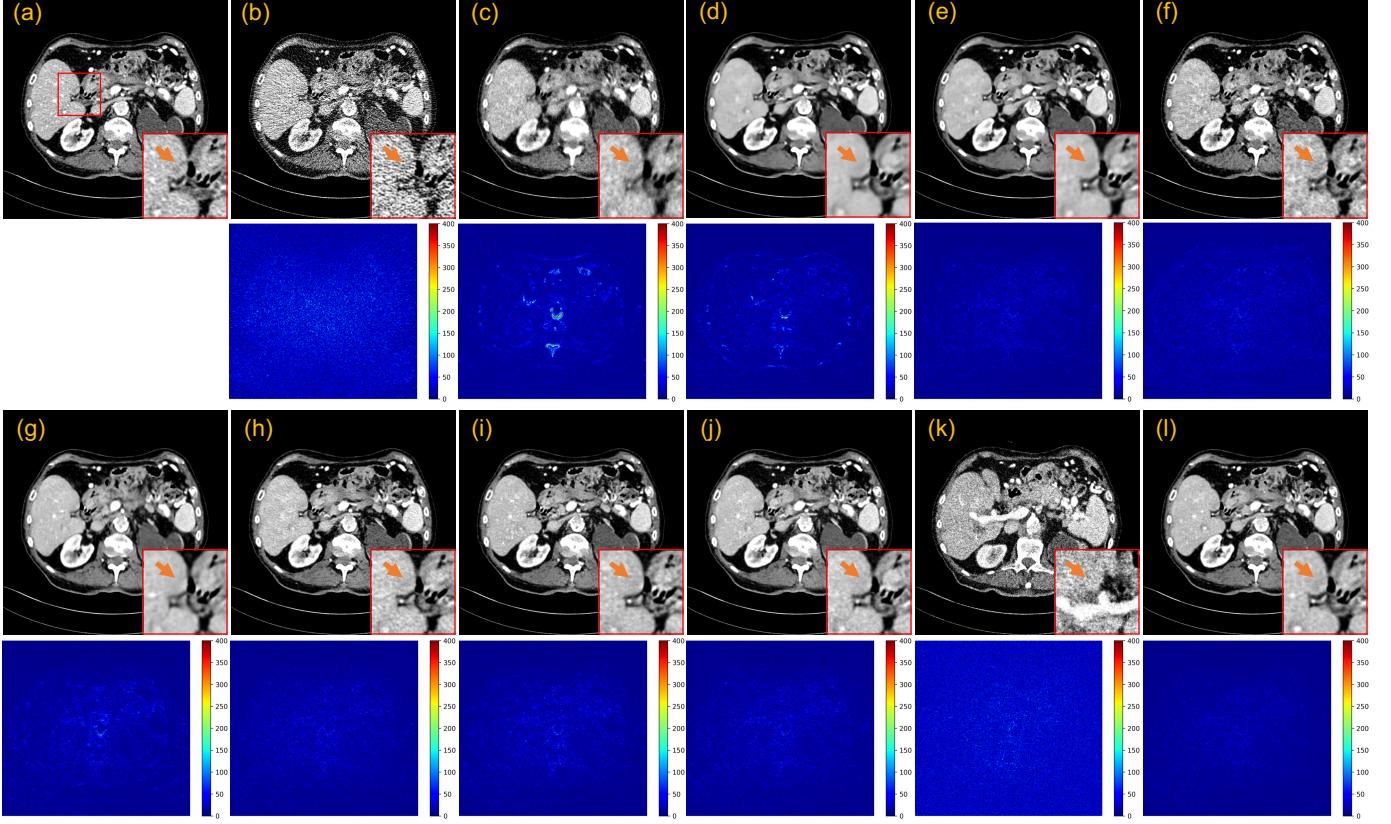


Fig. 4. Qualitative results of a 25% dose abdomen CT image from Mayo 2016 dataset. (a) NDCT image (Ground-truth), (b) FBP, (c) PWLS, (d) RED-CNN, (e) PDF-RED-CNN, (f) WGAN-VGG, (g) CNCL-U-Net, (h) DU-GAN, (i) IDDPMP-1000, (j) IDDPMP-50, (k) IDDPMP-10 and (l) CoreDiff-10 (**ours**). The display window is [-100, 200] HU. The red ROI is zoomed in for visual comparison and the orange arrow points to one lesion.

TABLE I
QUANTITATIVE RESULTS (MEAN \pm SD) ON 25% DOSE TESTING DATA.

	PSNR(dB)	SSIM	RMSE(HU)
FBP	34.16 \pm 1.76	0.7763 \pm 0.0605	40.0 \pm 8.0
PWLS	38.26 \pm 1.29	0.9449 \pm 0.0091	24.7 \pm 3.5
RED-CNN	39.29 \pm 1.53	0.9599 \pm 0.0106	22.1 \pm 4.4
PDF-RED-CNN	42.94 \pm 1.32	0.9685 \pm 0.0103	14.4 \pm 2.3
WGAN-VGG	40.12 \pm 0.98	0.9419 \pm 0.0118	19.9 \pm 2.3
CNCL-U-Net	40.91 \pm 1.05	0.9598 \pm 0.0118	18.2 \pm 2.2
DU-GAN	41.50 \pm 1.22	0.9591 \pm 0.0121	17.0 \pm 2.5
IDDPMP-1000	41.30 \pm 1.20	0.9593 \pm 0.0116	17.4 \pm 2.5
IDDPMP-50	41.49 \pm 1.18	0.9582 \pm 0.0122	17.0 \pm 2.4
IDDPMP-10	33.02 \pm 1.29	0.6934 \pm 0.0912	45.2 \pm 7.0
CoreDiff-10 (ours)	43.92\pm1.33	0.9744\pm0.0087	12.9\pm2.1

on the denoising quality. However, when T is reduced to 10, the model produces the poorest results due to much insufficient sampling. In addition, IDDPMP-1000/-50 erase the critical lesion information, while our CoreDiff retains them well. The residual map confirms that our approach has the least prediction bias.

Table I presents the quantitative results of all methods. Our CoreDiff outperforms all DL-based methods and iterative reconstruction algorithms. Notably, our method outperforms the second best method (PDF-RED-CNN) by a large margin in terms of all metrics.

2) 5% dose: Fig. 5 presents the qualitative results of 5% dose testing data. In the scenario of ultra-low doses, the FBP image suffers from significantly severe noise and streaking

artifacts due to the photon starvation effect, making it unacceptable for clinical diagnosis. The denoising performance of some denoising methods has a sharp decline. Fig. 5 shows that RED-CNN-based methods and CNCL-U-Net produced over-smoothed results. In addition, both PWLS and WGAN-VGG introduced noticeable artifacts to the denoised images. The DU-GAN obtains the best performance besides the diffusion-based methods. However, the denoising result of DU-GAN shrinks the lesion size. Other diffusion-based models, except for IDDPMP-10, consistently exhibit remarkable performance in ultra-low-dose denoising tasks, showing great promise for LDCT denoising. Among them, our CoreDiff exhibits the best denoising performance both in terms of residual maps and zoomed-in ROIs. Furthermore, Fig. 6 shows the profile results of the different methods, as indicated by the blue line in the NDCT images in Fig. 5. The red arrow indicates that our CoreDiff maintains the CT value better than other methods.

Table II presents the quantitative results of 5% dose test data. Our CoreDiff also surpasses all competing methods. On average our CoreDiff achieves around +1.46 dB PSNR, +1.39% SSIM, and -15.45% RMSE over the second best PDF-RED-CNN. In addition, Table II also reports the computational time of denoising a single image by different methods. The inference speed of the CoreDiff is much faster than that of the IDDPMP, which has reached a clinically acceptable level.

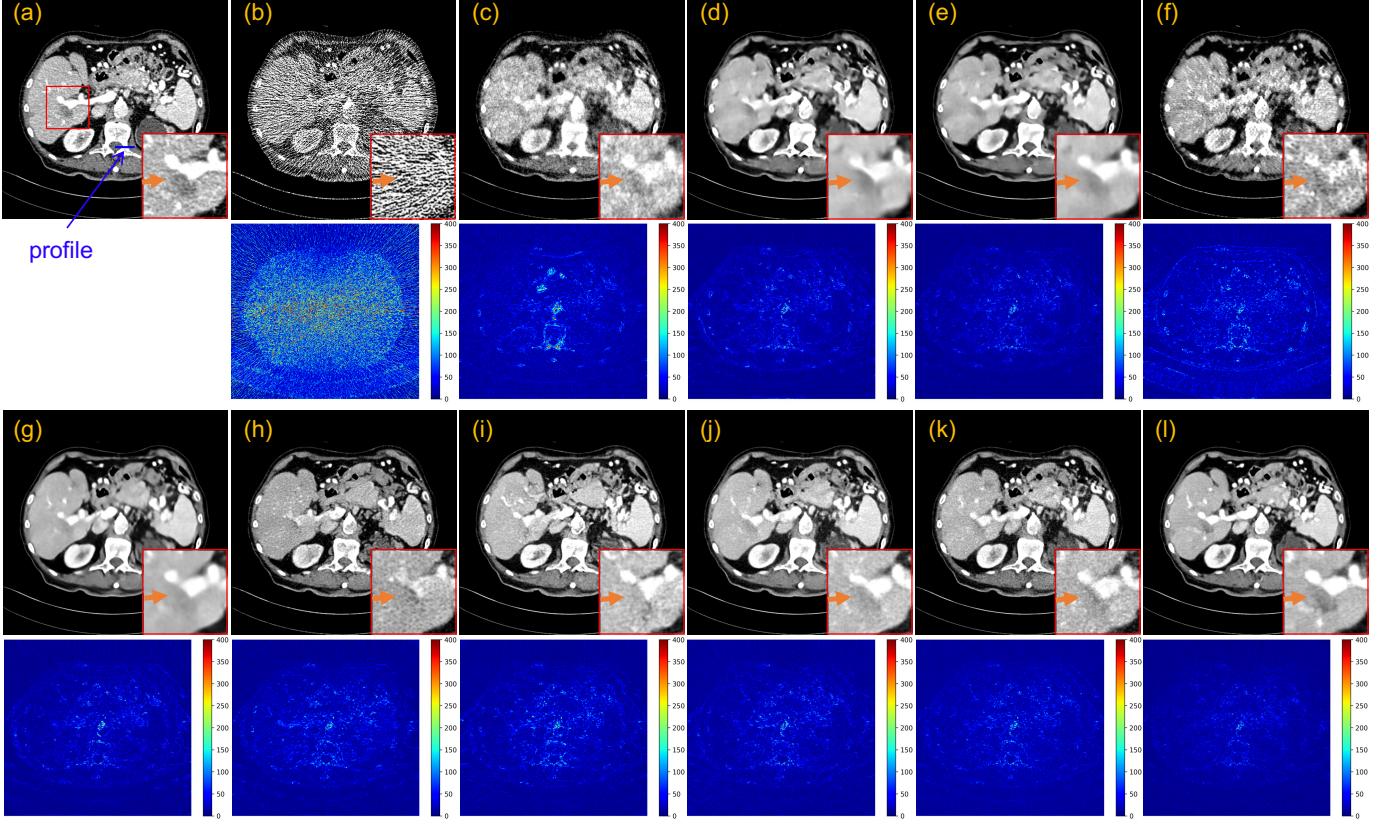


Fig. 5. Qualitative results of a 5% dose CT image from Mayo 2016 dataset. (a) NDCT image (Ground-truth), (b) FBP, (c) PWLS, (d) RED-CNN, (e) PDF-RED-CNN, (f) WGAN-VGG, (g) CNCL-U-Net, (h) DU-GAN, (i) IDDPM-1000, (j) IDDPM-50, (k) IDDPM-10 and (l) CoreDiff-10 (**ours**). The display window is [-100, 200] HU. The red ROI is zoomed in for visual comparison and the orange arrow points to one lesion.

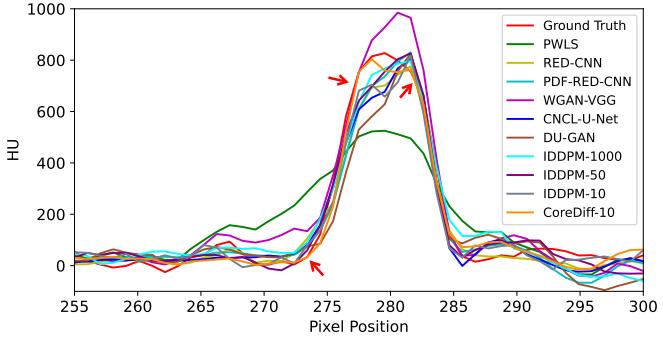


Fig. 6. Profile plots of the blue line in Fig. 5 by different methods.

D. Ablation Study

We conducted ablation studies to investigate the effects of different T settings and all components in CoreDiff.

1) *Ablation on different T settings:* We explored the performance of CoreDiff with $T = 1, 10, 50$, and 250 for training and inference. Fig. 7 shows the denoised images under different T settings. When $T = 1$, CoreDiff reduces to a one-step restoration, resulting in blurred edges in the denoised image. As T increases, the tissue boundaries become gradually sharper, but the inference time also increases accordingly. In addition, the prediction errors of the restoration network are also accumulated as T increases. For example, when $T = 1$, the

TABLE II
QUANTITATIVE RESULTS (MEAN \pm SD) ON 5% DOSE TESTING DATA AND THE COMPUTATIONAL TIME FOR DENOISING A SINGLE IMAGE. THE REPORTED TIME FOR DEEP LEARNING METHODS EXCLUDES THE TIME REQUIRED FOR FBP RECONSTRUCTION.

	PSNR (dB)	SSIM	RMSE (HU)	Time (s)
FBP	25.49 \pm 2.15	0.4310 \pm 0.0908	110.5 \pm 25.4	-
PWLS	34.87 \pm 0.95	0.8736 \pm 0.0084	36.3 \pm 3.9	3.50
RED-CNN	37.43 \pm 0.95	0.9363 \pm 0.0136	27.0 \pm 3.0	0.01
PDF-RED-CNN	39.25 \pm 1.22	0.9445 \pm 0.0147	22.0 \pm 3.1	0.01
WGAN-VGG	34.68 \pm 0.77	0.8821 \pm 0.0199	37.0 \pm 3.3	0.01
CNCL-U-Net	38.34 \pm 1.07	0.9341 \pm 0.0155	24.4 \pm 3.0	0.02
DU-GAN	37.39 \pm 1.13	0.9287 \pm 0.0161	27.2 \pm 3.5	0.01
IDDPM-1000	37.35 \pm 1.15	0.9202 \pm 0.0219	27.4 \pm 3.6	94.2
IDDPM-50	37.93 \pm 1.27	0.9166 \pm 0.0356	25.7 \pm 3.8	4.67
IDDPM-10	34.80 \pm 2.52	0.8063 \pm 0.1125	38.1 \pm 12.8	0.96
CoreDiff-10	40.71\pm1.26	0.9576\pm0.0123	18.6\pm2.7	0.12

PSNR and RMSE metrics are highest, corresponding to fewer pixel-level errors and over-smoothed images. When $T = 10$, the SSIM of CoreDiff is the highest, and the denoised image is visually closest to the ground truth. However, setting $T \geq 50$ results in a gradual decline in the quantization performance of the CoreDiff. Despite the presence of our CLEAR-Net, the accumulated errors cannot be disregarded as T becomes large. Therefore, considering the sharpness of the denoised images, quantitative performance, and inference time of the method, $T = 10$ is a suitable setting for our CoreDiff.

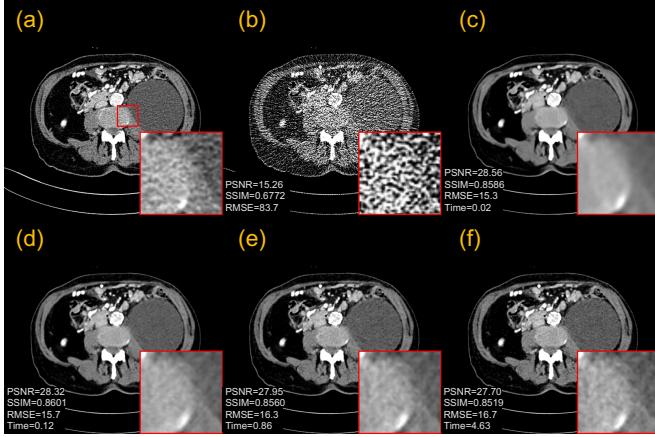


Fig. 7. Ablation study on different T settings for CoreDiff. (a) NDCT image (Ground truth), (b) FBP, (c) $T = 1$, (d) $T = 10$, (e) $T = 50$, and (f) $T = 250$. The display window is [-100, 200] HU. The tissue boundaries in the red ROI are zoomed in for comparison. Quantitative results are provided in the lower left corner and calculated based on the display window. Time (sec.) is the average inference time of CoreDiff on the whole testing set for different T settings.

TABLE III
ABLATION STUDY OF DIFFERENT MODULES ON 5% DOSE TESTING DATA OF MAYO 2016 DATASET.

Warm	CLEAR-Net		PSNR (dB)	SSIM	RMSE (HU)
	CTX	EMM			
-	-	-	39.96±1.46	0.9501±0.0143	26.8±3.5
✓	-	-	41.59±1.57	0.9628±0.0123	22.2±3.1
✓	✓	-	42.56±1.52	0.9680±0.0109	19.9±2.7
✓	✓	✓	43.04±1.55	0.9713±0.0101	18.8±2.7

2) *Ablation on different components*: Then we investigate the effects of our generalized diffusion process with a mean-preserving degradation operator, the introduced context information and EMM in our CLEAR-Net. For a fair comparison, we set the total sampling steps T of the original cold diffusion to 10 as the *Baseline* (*ColdDiff-10*). For simplicity, the proposed generalized diffusion process with the LDCT image (warm state) is abbreviated as *Warm* and the introduction of contextual information is abbreviated as *CTX*.

Fig. 8 shows the denoised images obtained using different components. Since the baseline model (*ColdDiff-10*) uses the mean-varying diffusion process and starts sampling from random Gaussian noise, the denoised image presents many structural distortions with only 10 sampling steps. In contrast, the proposed diffusion process with LDCT images as the starting point of the sampling process (*Warm*) significantly improves the denoising performance of the model with fewer sampling steps, alleviating the structural distortions in the ROI. However, the edges of some small tissues are still unclear. Thanks to the fact that contextual information constrains the structure of the denoised image to maintain continuity in the z-axis. The outline of the detail indicated by the orange arrow is more visible. Finally, our CoreDiff, which integrates all components, demonstrates the best denoising results. The edge of detail in ROI is separated from the surrounding tissues, which is consistent with the ground truth.

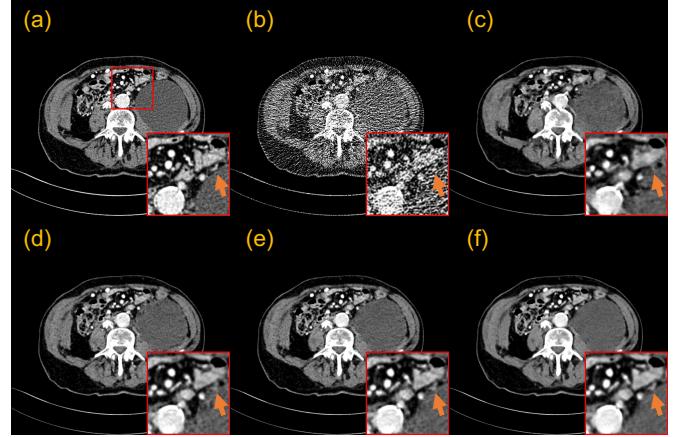


Fig. 8. Ablation study on different components for CoreDiff. (a) NDCT image (Ground truth), (b) FBP, (c) Baseline (*ColdDiff-10*), (d) Baseline+Warm, (e) Baseline+Warm+CTX, and (f) CoreDiff. The display window is [-100, 200] HU. The red ROI is zoomed in for visual comparison and the orange arrow points to key detail.

TABLE IV
THE WEIGHT DISTRIBUTION IN THE ONE-SHOT LEARNING FRAMEWORK USING PAIRED OR UNPAIRED SLICES (PAIR. VS UNPA.).

Dose	Type	$w_0 \rightarrow w_9$
50%	pair.	0.33 0.21 0.14 0.08 0.03 0.03 0.04 0.04 0.05 0.05
	unpa.	0.33 0.22 0.14 0.09 0.03 0.03 0.04 0.04 0.04 0.04
25%	pair.	0.44 0.29 0.12 0.04 0.02 0.01 0.02 0.02 0.02 0.02
	unpa.	0.50 0.28 0.11 0.03 0.01 0.01 0.01 0.01 0.02 0.02

Table III presents the quantitative comparison of different components, which shows that all the proposed components contribute significantly to the overall denoising performance of CoreDiff. On average our CoreDiff with all components achieves around +3.08 dB PSNR, +2.23% SSIM, and -29.85% RMSE over the baseline.

E. One-shot Generalization to New Doses and Datasets

We further perform experiments on two different test datasets to verify the effectiveness of our one-shot learning (OSL) framework. For the new dose from any dataset, we only select one new LDCT image and one (un)paired NDCT image for training the weights of our OSL framework in Eq. (11). If the new LDCT and NDCT images are paired, the resultant OSL model is referred to as CoreDiff+OSL^P. To ease the requirement of the new training data, we also consider the unpaired scenario that LDCT and NDCT images were collected at different times. To achieve this, we also select a NDCT image below two slices of the corresponding LDCT image as the training label to simulate the presence of slight shifts for unpaired training, the resultant OSL model is referred to as CoreDiff+OSL^U.

1) *Generalization to new dose levels on the same dataset*: We examine the generalization of our CoreDiff trained on 5% dose data to 50% and 25% dose testing data on Mayo 2016 dataset. Note that comparing PDF-RED-CNN*, CoreDiff+OSL^P, and CoreDiff+OSL^U to other methods is not fair because they use extra training data of 50% and 25% doses.

Table IV shows the learned weights by CoreDiff+OSL^P and CoreDiff+OSL^U. The distributions of weights trained on paired

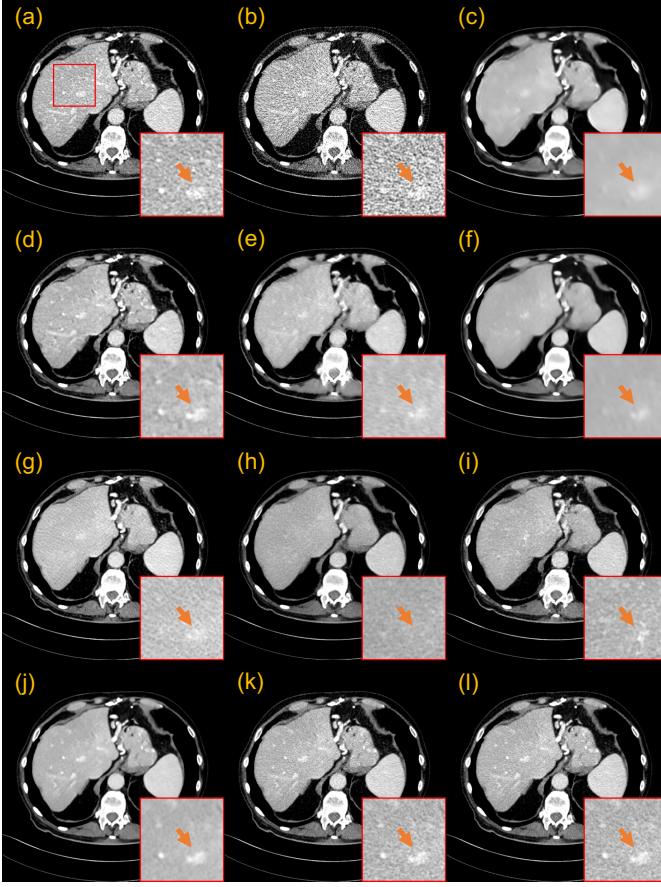


Fig. 9. Qualitative results of a 50% dose CT image denoised by different methods trained on 5% dose training data. (a) NDCT image (Ground truth), (b) FBP, (c) RED-CNN, (d) PDF-RED-CNN*, (e) WGAN-VGG, (f) CNCL-U-Net, (g) DU-GAN, (h) IDDPMP-1000, (i) IDDPMP-50, (j) CoreDiff-10 (**ours**), (k) CoreDiff+OSL^P-10 (**ours**), and (l) CoreDiff+OSL^u-10 (**ours**). The display window is [-100, 200] HU. The red ROI is zoomed in for visual comparison and the orange arrow points to key detail.

and unpaired slices exhibit a high-level of consistency, indicating the effectiveness and flexibility of the OSL framework for clinical use.

Fig. 9 presents the qualitative results of a 50% dose LDCT image. Both RED-CNN and CNCL-U-Net appear to smoothen the image due to the fact that the level of noise in the testing image is higher than that in the training image. While DU-GAN and IDDPMP successfully preserve image texture information, certain details, such as blood vessels, are lost. Although our CoreDiff tends to remove more noise than necessary, it is prone to preserve critical anatomical details. We highlight our CoreDiff+OSL^P and CoreDiff+OSL^u yield a visual perception effect closer to the ground truth. In contrast, PDF-RED-CNN* may introduce some distortions. Fig. 10 presents the qualitative results of a 25% dose LDCT image, which can draw similar conclusions.

2) *Generalization to new dose levels on the external dataset:* We further validate the effectiveness of the proposed CoreDiff and CoreDiff+OSL^u on Mayo 2020 dataset. To simulate the clinical application scenario as closely as possible, we choose a chest slice of 5% dose and a abdomen slice of 25% dose, as well as their unpaired normal-dose slices, to train two our one-shot

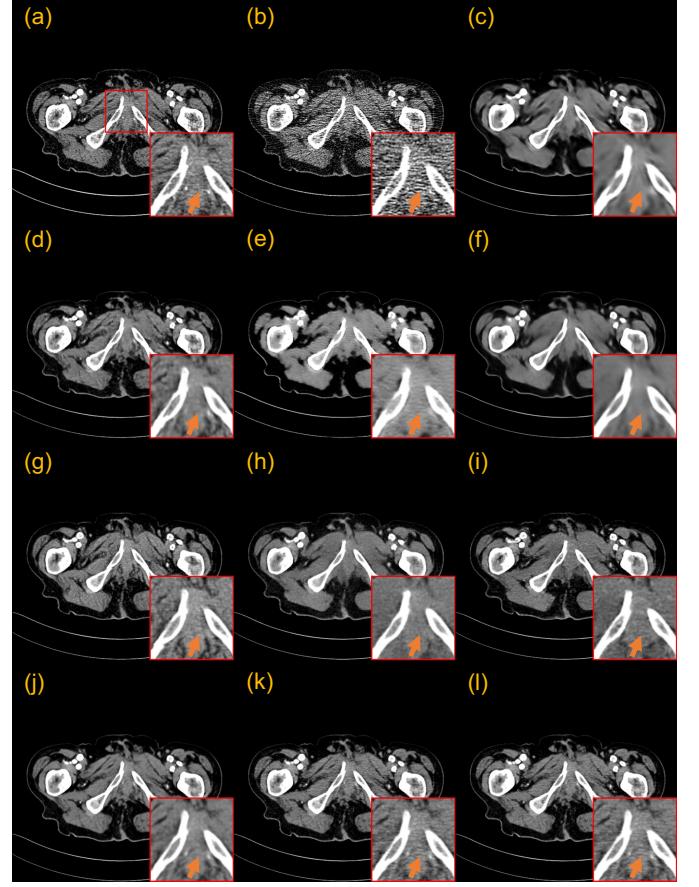


Fig. 10. Qualitative results of a 25% dose CT image denoised by different methods trained on 5% dose training data. (a) NDCT image (Ground truth), (b) FBP, (c) RED-CNN, (d) PDF-RED-CNN*, (e) WGAN-VGG, (f) CNCL-U-Net, (g) DU-GAN, (h) IDDPMP-1000, (i) IDDPMP-50, (j) CoreDiff-10 (**ours**), (k) CoreDiff+OSL^P-10 (**ours**), and (l) CoreDiff+OSL^u-10 (**ours**). The display window is [-100, 200] HU. The red ROI is zoomed in for visual comparison and the orange arrow points to key detail.

models, respectively. By combining with these two separate one-shot models, we can quickly generalize our CoreDiff to the mixed dose levels test data.

Fig. 11 presents the denoising results of a representative chest slice from Mayo 2020 dataset. All methods remove noise to some degree. RED-CNN-based methods smoothen the images, and the numerous details in the lungs were lost in the ROI. PDF-RED-CNN* loses some details partially due to the gap between our incident photon number setting for 10% dose and the one used by Mayo Clinic. While PWLS preserves more information compared to RED-CNN-based methods, it may not provide sufficient noise reduction. Among the GAN-based methods, CNCL-U-Net performs best but also causes the offset of the background value. The diffusion-based approach exhibits a good trade-off between noise suppression and image fidelity. Finally, the proposed OSL framework is proven to be instrumental in enabling CoreDiff to achieve textures that closely resemble the ground truth. Fig. 12 shows the denoising results of a representative abdomen slice. PDF-RED-CNN* preserves more details than RED-CNN. DU-GAN shows the best visual result among GAN-based methods. However, CNCL-U-Net produces over-smoothing results similar to RED-CNN,

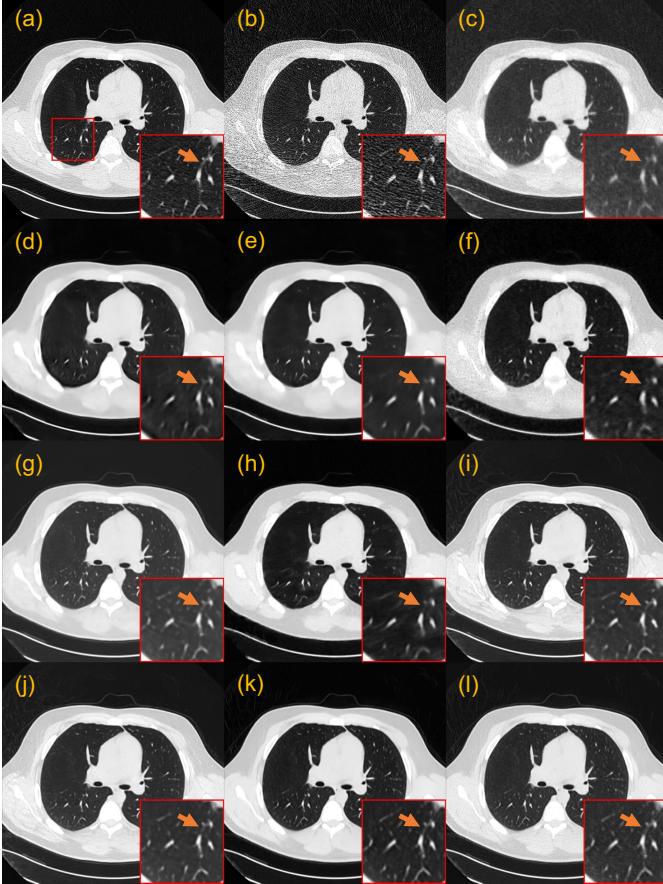


Fig. 11. Qualitative results of a 10% dose chest CT image from Mayo 2020 dataset. (a) NDCT image (Ground truth), (b) FBP, (c) PWLS, (d) RED-CNN, (e) PDF-RED-CNN*, (f) WGAN-VGG, (g) CNCL-U-Net, (h) DU-GAN, (i) IDDPMP-1000, (j) IDDPMP-50, (k) CoreDiff-10 (**ours**), and (l) CoreDiff+OSL^u-10 (**ours**). The display window is [-1350, 150] HU. The red ROI is zoomed in for visual comparison and the orange arrows points to key detail.

and WGAN-VGG introduces extra artifacts to the denoised image. In addition, as indicated by orange arrows, IDDPMP-1000/-50 models fail to retain dense small tissue information while CoreDiff best preserves them. Furthermore, in line with the findings from Fig. 11, when integrated with the one-shot learning framework, CoreDiff can provide faithful texture close to the ground truth.

Table V presents the quantitative results. Among the compared methods, PDF-RED-CNN* has the highest PSNR and lowest RMSE, which benefits from the training with multiple doses of data, while IDDPMP has the highest SSIM, indicating better preservation of the structural information. Surprisingly, our CoreDiff achieves better performance than them, even when trained on only 5% dose data without the additional OSL framework. We think that this is due to a more reasonable diffusion method allowing the model to progressively remove noise from the image and avoid structural distortion. In addition, the contextual error modulation module makes CoreDiff more robust to input shift. The OSL framework further improves PSNR and SSIM, indicating that the CoreDiff+OSL^u can produce more visually realistic texture and structural information by learning an optimal denoised image.

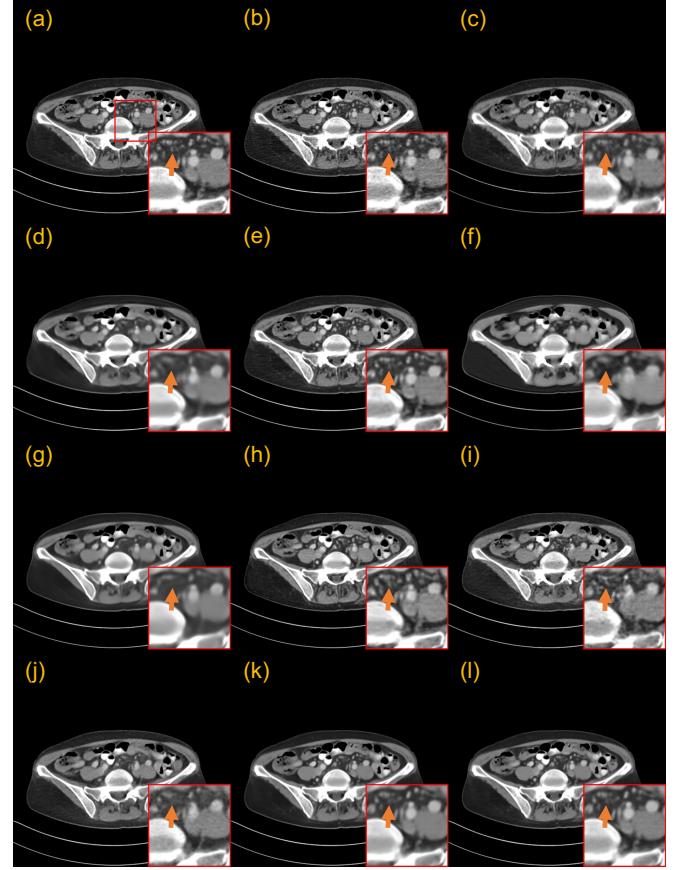


Fig. 12. Qualitative results of a 25% dose abdomen CT image from Mayo 2020 dataset. (a) NDCT image (Ground truth), (b) FBP, (c) PWLS, (d) RED-CNN, (e) PDF-RED-CNN*, (f) WGAN-VGG, (g) CNCL-U-Net, (h) DU-GAN, (i) IDDPMP-1000, (j) IDDPMP-50, (k) CoreDiff-10 (**ours**), and (l) CoreDiff+OSL^u-10 (**ours**). The display window is [-165, 275] HU. The red ROI is zoomed in for visual comparison and the orange arrows points to key detail.

TABLE V
QUANTITATIVE RESULTS (MEAN \pm SD) ON MAYO 2020 DATASET.

	PSNR(dB)	SSIM	RMSE(HU)
FBP	29.06 \pm 8.72	0.6246 \pm 0.3002	108.7 \pm 84.2
PWLS	31.96 \pm 5.29	0.7653 \pm 0.1947	60.3 \pm 32.3
RED-CNN	32.57 \pm 4.83	0.7650 \pm 0.1918	54.7 \pm 27.3
PDF-RED-CNN*	33.69 \pm 6.20	0.7735 \pm 0.1986	52.2 \pm 31.0
WGAN-VGG	31.09 \pm 4.44	0.7475 \pm 0.1922	63.5 \pm 29.9
CNCL-U-Net	33.15 \pm 4.87	0.7612 \pm 0.1954	51.3 \pm 25.9
DU-GAN	32.73 \pm 5.42	0.7546 \pm 0.2030	55.7 \pm 30.6
IDDPMP-1000	32.41 \pm 4.29	0.7849 \pm 0.1540	53.8 \pm 23.8
IDDPMP-50	32.67 \pm 4.48	0.7887 \pm 0.1543	52.7 \pm 24.2
CoreDiff-10	34.10 \pm 5.28	0.7964 \pm 0.1746	46.8\pm24.5
CoreDiff+OSL ^u -10 (ours)	34.13\pm5.37	0.8107\pm0.1626	46.8\pm24.9

IV. CONCLUSION

In this work, we proposed a novel contextual error-modulated generalized diffusion model for LDCT image denoising, termed CoreDiff. The presented CoreDiff utilizes (i) the LDCT images as the informative endpoint of the diffusion process, (ii) the introduced mean-preserving degradation operator to mimic the physical process of CT degradation, (iii) the proposed restoration network CLEAR-Net to alleviate the accumulated error and misalignment, and (iv) the devised one-shot learning framework to boost the generalization. Experimental results

demonstrate the effectiveness of CoreDiff, especially in the ultra-low-dose case. Remarkably, our CoreDiff model only requires 10 sampling steps, making it much faster than classical diffusion models for clinical use.

REFERENCES

- [1] R. Smith-Bindman *et al.*, “Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer,” *Arch. Intern. Med.*, vol. 169, pp. 2078–2086, 2009.
- [2] A. Sodickson *et al.*, “Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults,” *Radiology*, vol. 251, pp. 175–184, 2009.
- [3] Q. Xie *et al.*, “Robust low-dose CT sinogram preprocessing via exploiting noise-generating mechanism,” *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2487–2498, 2017.
- [4] 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, 2006.
- [5] H. Shan *et al.*, “3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network,” *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, 2018.
- [6] G. Wang, J. C. Ye, and B. De Man, “Deep learning for tomographic image reconstruction,” *Nat. Mach. Intell.*, vol. 2, no. 12, pp. 737–748, 2020.
- [7] J. Ma *et al.*, “Low-dose computed tomography image restoration using previous normal-dose scan,” *Med. Phys.*, vol. 38, pp. 5713–5731, 2011.
- [8] Y. Chen *et al.*, “Artifact suppressed dictionary learning for low-dose CT image processing,” *IEEE Trans. Med. Imag.*, vol. 33, pp. 2271–2292, 2014.
- [9] H. Shan *et al.*, “Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction,” *Nat. Mach. Intell.*, vol. 1, pp. 269–276, 2019.
- [10] Z. Chen, C. Niu, G. Wang, and H. Shan, “LIT-Former: Linking in-plane and through-plane transformers for simultaneous CT image denoising and deblurring,” *arXiv:2302.10630*, 2023.
- [11] H. Chen *et al.*, “Low-dose CT with a residual encoder-decoder convolutional neural network,” *IEEE Trans. Med. Imag.*, vol. 36, pp. 2524–2535, 2017.
- [12] W. Xia *et al.*, “CT reconstruction with PDF: Parameter-dependent framework for data from multiple geometries and dose levels,” *IEEE Trans. Med. Imag.*, vol. 40, no. 11, pp. 3065–3076, 2021.
- [13] B. Kim, M. Han, H. Shim, and J. Baek, “A performance comparison of convolutional neural network-based image denoising methods: The effect of loss functions on low-dose CT images,” *Med. Phys.*, vol. 46, pp. 3906–3923, 2019.
- [14] M. Nagare, R. Melnyk, O. Rahman, K. D. Sauer, and C. A. Bouman, “A bias-reducing loss function for CT image denoising,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 1175–1179.
- [15] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, “Generative adversarial networks for noise reduction in low-dose CT,” *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, 2017.
- [16] E. Kang, H. J. Koo, D. H. Yang, J. B. Seo, and J. C. Ye, “Cycle consistent adversarial denoising network for multiphase coronary CT angiography,” *Med. Phys.*, vol. 46, pp. 550–562, 2019.
- [17] S. Bera and P. K. Biswas, “Noise conscious training of non local neural network powered by self attentive spectral normalized Markovian patch GAN for low dose CT denoising,” *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3663–3673, 2021.
- [18] Q. Yang *et al.*, “Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss,” *IEEE Trans. Med. Imag.*, vol. 37, pp. 1348–1357, 2018.
- [19] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, “DU-GAN: Generative adversarial networks with dual-domain U-Net based discriminators for low-dose CT denoising,” *IEEE Trans. Instrum. Meas.*, vol. 71, p. 4500512, 2021.
- [20] S. Zhao, H. Ren, A. Yuan, J. Song, N. Goodman, and S. Ermon, “Bias and generalization in deep generative models: An empirical study,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.
- [21] Z. Xiao, K. Kreis, and A. Vahdat, “Tackling the generative learning trilemma with denoising diffusion GANs,” in *Proc. Int. Conf. Learn. Represent.*, 2022.
- [22] 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.
- [23] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
- [24] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” in *Proc. Int. Conf. Learn. Represent.*, 2021.
- [25] A. Nichol and P. Dhariwal, “Improved denoising diffusion probabilistic models,” in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8162–8171.
- [26] H. Chung, D. Ryu, M. T. McCann, M. L. Klasky, and J. C. Ye, “Solving 3D inverse problems using pre-trained 2D diffusion models,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023.
- [27] H. Chung, B. Sim, D. Ryu, and J. C. Ye, “Improving diffusion models for inverse problems using manifold constraints,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2021.
- [28] 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.*, 2021.
- [29] P. Dhariwal and A. Nichol, “Diffusion models beat GANs on image synthesis,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 8780–8794.
- [30] L. Yang *et al.*, “Diffusion models: A comprehensive survey of methods and applications,” *arXiv:2209.00796*, 2022.
- [31] Q. Gao and H. Shan, “CoCoDiff: a contextual conditional diffusion model for low-dose CT image denoising,” in *Proc. SPIE*, vol. 12242, 2022.
- [32] W. Xia, Q. Lyu, and G. Wang, “Low-dose CT using denoising diffusion probabilistic model for 20× speedup,” *arXiv:2209.15136*, 2022.
- [33] A. Kazerouni *et al.*, “Diffusion models for medical image analysis: A comprehensive survey,” *arXiv:2211.07804*, 2022.
- [34] J. Wu, H. Fang, Y. Zhang, Y. Yang, and Y. Xu, “MedSegDiff: Medical image segmentation with diffusion probabilistic model,” *arXiv:2211.00611*, 2022.
- [35] A. Bansal *et al.*, “Cold diffusion: Inverting arbitrary image transforms without noise,” *arXiv:2208.09392*, 2022.
- [36] H. Yen, F. G. Germain, G. Wichern, and J. L. Roux, “Cold diffusion for speech enhancement,” *arXiv:2211.02527*, 2022.
- [37] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, “FiLM: Visual reasoning with a general conditioning layer,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018.
- [38] X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1501–1510.
- [39] V. Dumoulin, J. Shlens, and M. Kudlur, “A learned representation for artistic style,” in *Proc. Int. Conf. Learn. Represent.*, 2017.
- [40] H. Shan, U. Kruger, and G. Wang, “A novel transfer learning framework for low-dose CT,” in *Proc. SPIE*, vol. 11072, 2019, pp. 513–517.
- [41] D. Li, Z. Bian, S. Li, J. He, D. Zeng, and J. Ma, “Noise characteristics modeled unsupervised network for robust CT image reconstruction,” *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3849–3861, 2022.
- [42] M. Li *et al.*, “Low-dose CT image synthesis for domain adaptation imaging using a generative adversarial network with noise encoding transfer learning,” *IEEE Trans. Med. Imag.*, 2023.
- [43] B. Chen, S. Leng, L. Yu, D. Holmes III, J. Fletcher, and C. McCollough, “An open library of CT patient projection data,” in *Proc. SPIE*, vol. 9783, 2016, pp. 330–335.
- [44] D. Zeng *et al.*, “A simple low-dose x-ray CT simulation from high-dose scan,” *IEEE Trans. Nucl. Sci.*, vol. 62, no. 5, pp. 2226–2233, 2015.
- [45] L. Yu, M. Shiung, D. Jondal, and C. H. McCollough, “Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols,” *J. Comput. Assist. Tomogr.*, vol. 36, no. 4, pp. 477–487, 2012.
- [46] T. Zhao, M. McNitt-Gray, and D. Ruan, “A convolutional neural network for ultra-low-dose CT denoising and emphysema screening,” *Med. Phys.*, vol. 46, no. 9, pp. 3941–3950, 2019.
- [47] T. Zhao, J. Hoffman, M. McNitt-Gray, and D. Ruan, “Ultra-low-dose CT image denoising using modified BM3D scheme tailored to data statistics,” *Med. Phys.*, vol. 46, no. 1, pp. 190–198, 2019.
- [48] T. R. Moen *et al.*, “Low-dose CT image and projection dataset,” *Med. Phys.*, vol. 48, no. 2, pp. 902–911, 2021.
- [49] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, “Toward convolutional blind denoising of real photographs,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1712–1722.
- [50] L. Shi *et al.*, “Review of CT image reconstruction open source toolkits,” *J. X-Ray Sci. Technol.*, vol. 28, no. 4, pp. 619–639, 2020.
- [51] M. Geng *et al.*, “Content-noise complementary learning for medical image denoising,” *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 407–419, 2021.