PrideDiff: Physics-Regularized Generalized Diffusion Model for CT Reconstruction

Zexin Lu, Qi Gao, *Graduate Student Member, IEEE*, Tao Wang, Ziyuan Yang, Zhiwen Wang, Hui Yu, Hu Chen, Jiliu Zhou, *Senior Member, IEEE*, Hongming Shan, *Senior Member, IEEE*, and Yi Zhang, *Senior Member, IEEE*

***Abstract*—Achieving a lower radiation dose and a faster imaging speed is a pivotal objective of computed tomography (CT) reconstruction. However, these often come at the cost of compromised reconstruction quality. With the advent of deep learning, numerous CT reconstruction methods rooted in this field have significantly improved the reconstruction performance. Recently, diffusion models have further enhanced training stability and imaging quality for CT. However, many of these methods only focus on CT image domain features, ignoring the intrinsic physical information of the imaging process. Although compressive sensing-based iterative reconstruction algorithms utilize physical prior information, their intricate iterative process poses challenges in training, subsequently influencing their efficiency. Motivated by these observations, we introduce a novel physics-regularized generalized diffusion model for CT reconstruction (PrideDiff). On the one hand, our method further improves the quality of reconstructed images by fusing physics-regularized iterative reconstruction methods with diffusion models. On the other hand, we propose a prior extraction module embedded with temporal features, which effectively improves the performance of the iteration process. Extensive experimental results demonstrate that PrideDiff outperforms several state-of-the-art methods in low-dose and sparse-view CT reconstruction tasks on different datasets, with faster reconstruction speed. We further discuss the effectiveness of relevant components in PrideDiff and validate the stability of the iterative reconstruction process, followed by detailed analysis of computational cost and inference time.**

***Index Terms*—Deep learning (DL), diffusion models, iterative reconstruction, low-dose computed tomography (CT), sparse-view CT.**

I. Introduction

COMPUTED tomography (CT) is an imaging technique extensively used in clinical settings. However, the radiation from X-rays poses potential risks to human health [1]. Consequently, various low-dose CT techniques have emerged. Typical methods include reducing the X-ray tube current and sparse sampling. Nonetheless, these alternatives often inevitably introduce noticeable noise and artifacts, compromising clinical diagnostic accuracy. As a result, how to balance radiation dose and imaging quality has become a key issue in the field of CT reconstruction.

Recently, researchers have proposed a variety of algorithms to solve this conundrum [2], [3], [4], [5], [6], [7]. Typically, these techniques can be categorized into three primary groups: 1) sinogram domain processing [8], [9], [10]; 2) image domain processing [11], [12], [13], [14]; and 3) iterative reconstruction [15], [16], [17].

For sinogram domain processing methods, the main concept involves utilizing raw sinogram data with a specific filter, followed by an analytic image reconstruction method, such as filtered back projection (FBP). Typical methods include penalized weighted least square [18], bilateral filtering [19], and structural filtering [20]. Recently, rapid advances in deep learning (DL) have catalyzed numerous algorithms based on convolution neural networks (CNNs) for low-dose CT reconstruction [4], [21]. Typical methods utilize different network architectures to process the sinogram before FBP is applied [3], [8], [22], [23]. However, such algorithms usually suffer from a loss of spatial resolution and blurred edges.

For image domain processing methods, traditional methods usually introduce natural image restoration methods into low-dose CT denoising, such as nonlocal means [24] and BM3D [25]. Many DL-based methods were developed due to their fast inference speed. Chen et al. [26] introduced a residual network aided by an encoder–decoder structure, termed RED-CNN. Jin et al. [27] employed a U-Net [28] denoiser, termed FBPConvNet, which effectively suppresses image noise and artifacts. Nevertheless, many of these

Received 15 June 2024; revised 23 August 2024; accepted 26 September 2024. Date of publication 1 October 2024; date of current version 5 February 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62271335 and Grant 62101136; in part by the Sichuan Science and Technology Program under Grant 2021JDJQ0024; and in part by the Sichuan University "From 0 to 1" Innovative Research Program under Grant 2022SCUH0016. *(Corresponding author: Yi Zhang.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Zexin Lu, Tao Wang, Ziyuan Yang, Zhiwen Wang, Hui Yu, Hu Chen, and Jiliu Zhou are with the College of Computer Science, Sichuan University, Chengdu 610065, China, and also with the Key Laboratory of Data Protection and Intelligent Management, Ministry of Education, Sichuan University, Chengdu 610207, China (e-mail: zexinlu.scu@gmail.com; scuer_wt@stu.scu.edu.cn; cziyuanyang@gmail.com; zwwang1228@gmail.com; smileeudora@163.com; huchen@scu.edu.cn; zhoujl@scu.edu.cn).

Qi Gao and Hongming 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).

Yi Zhang is with the School of Cyber Science and Engineering, Sichuan University, Chengdu 610065, China, and also with the Key Laboratory of Data Protection and Intelligent Management, Ministry of Education, Sichuan University, Chengdu 610207, China (e-mail: yzhang@scu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TRPMS.2024.3471677.

Digital Object Identifier 10.1109/TRPMS.2024.3471677

2469-7311 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

models usually employ the ℓ_2 norm as their loss function, leading to oversmoothing in the predicted images. To alleviate this, Yang et al. [29] introduced a model built upon generative adversarial networks (GANs) with perceptual loss, further enhancing the quality of reconstructed images. Geng et al. [30] also introduced a content-noise complementary learning (CNCL) strategy, which employs GANs to learn the content and noise of the image dataset in a complementary manner. However, challenges, such as mode collapse during GANs training, render the training of GANs arduous [31]. At the same time, transformer architectures have recently garnered significant attention in the image processing field, demonstrating impressive efficacy [32], [33], [34]. However, training a transformer-based model necessitates not only a large volume of training data but also a considerable computational cost [35]. A key limitation of these image domain processing techniques is their sole reliance on image domain attributes, neglecting the measurement in the sinogram domain. This oversight often leads to suboptimal reconstruction results without the constraint of data consistency.

Iterative reconstruction methods synergize the feature information of the sinogram domain with the prior knowledge inherent in the image domain to yield improved reconstruction performance. Classic methods include total variation [36], dictionary learning [37], and low-rank minimization [38]. However, these methods are usually time-consuming and need to manually adjust parameters for different target images. DL-based iteration reconstruction methods are potential solutions to fix these limitations. Chen et al. [39] unrolled the steepest gradient descent algorithm and proposed the learned experts' assessment-based reconstruction network (LEARN) for sparse-view CT. He et al. [40] proposed a parameterized plug-and-play alternating direction method of multipliers framework to achieve robust low-dose CT reconstruction. Xia et al. [41] introduced a manifold and graph integrative convolutional reconstruction network for low-dose CT. Although DL-based iterative reconstruction methods have achieved promising results, these techniques face new challenges: they require extensive iterative models, which, in turn, demand substantial computational resources, amplifying the training burden. Additionally, the model's intricate iterative structure can also prolong the reconstruction time.

Recently, the denoising diffusion probabilistic model (DDPM) has garnered considerable success in the realm of image generation [42], [43], [44], [45], [46]. Compared to other generative models, diffusion models demonstrate more consistent training and better performance in image generation [43], [47], [48], [49]. DDPM estimates and eliminates noise from a Gaussian distribution, which is achieved by training a noise estimation network progressively via continuous iterations. Usually, DDPM requires numerous iterative steps to achieve a satisfactory image generation result. For instance, vanilla DDPM requires nearly 1000 steps for complete sampling. To improve the applicability of the DDPM method to the field of medical imaging, Xia et al. [50] adapted the fast ordinary differential equation solver [51] to improve the sampling efficiency in CT reconstruction. While these enhanced DDPM algorithms deliver commendable results, they remain rooted in vanilla diffusion model theory. Inspired by [52], Qi et al. [53] proposed a contextual error-modulated generalized diffusion (CoreDiff) model for low-dose CT denoising. Chung et al. [54], [55], [56], [57] further explored the score-based method for image reconstruction under ultrasparse sampling conditions by incorporating manifold constraints and prior features into the diffusion process. Dou and Song [58] proposed sequential Monte Carlo methods to ensure asymptotic accuracy in Bayesian posterior sampling for diffusion models, which demonstrated excellent performance in zero-shot tasks. However, these methods primarily focus on image domain features while neglecting crucial sinogram consistency, which can lead to unexpected artifacts. Additionally, most DDPM-based methods necessitate a large number of iterations, which reduces the efficiency of image restoration.

In this study, to further improve the performance of CT reconstruction tasks, we propose a novel fusion model synergizing iterative reconstruction and DDPM. Different from CoreDiff [53] which focuses on image domain features, our method further integrates prior feature learning with the diffusion model efficiently. Inspired by the LEARN model [39], we incorporate the regularization prior learning network into the diffusion process, further promoting the reconstruction quality. To this end, we develop a prior extraction module (PEM) that fuses time-embedding features, thereby unifying the diffusion process with regularization learning. Furthermore, to balance the number of iteration steps and reconstruction performance, and better fit the network with CT reconstruction tasks, we establish the diffusion process based on cold diffusion, thereby achieving superior reconstruction performance with fewer iteration steps.

In summary, the contributions of this article are as follows.

1) We develop a novel iterative CT reconstruction network model, termed PrideDiff, which serves as a bridge between the general diffusion model and the physics-regularized iterative reconstruction model. To the best of our knowledge, PrideDiff is the first work in CT reconstruction that synergizes Cold Diffusion and prior feature learning effectively.
2) To integrate the CT sinogram physical prior features into the diffusion process, we devise a novel restoration–reconstruction framework, which allows effective integration for a trainable regularization network and the diffusion process.
3) To validate our proposed PrideDiff, both sparse-view and low-dose experiments are conducted on two datasets with different sampling conditions. The results indicate that our method not only produces superior reconstruction results but also accelerates the sampling process.

The remainder of this article is structured as follows. We introduce the relevant principles used in PrideDiff and describe the detailed workflow of PrideDiff in Section II. Experimental setups, results, and discussions are presented in Section III. At last, the conclusion is drawn in Section IV.

II. Method

To achieve efficient feature extraction in the image domain, we utilize a general diffusion model, cold diffusion, and

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

show its basic principles in this section. However, in CT reconstruction tasks, the guidance of physical priors plays a pivotal role, which is exactly what the diffusion model lacks. In order to effectively exploit the priors, we introduce a regularization learning method. Finally, we present our PrideDiff to integrate the physics-regularized learning process into the diffusion model, yielding a general and efficient CT reconstruction network.

A. Cold Diffusion

Diffusion has been understood as a random walk under the guidance of Langevin dynamics [46], [59]. The diffusion process can be described as gradually adding Gaussian noise from a low noise state (cold state) to a high noise state (hot state). Cold diffusion [52] generalizes this process, relaxing it to arbitrary degradation operations, which implements the transition from the cold state to the hot state. Similar to the vanilla diffusion model, cold diffusion is also composed of two phases: 1) forward (diffusion) process and 2) reverse (sampling) process. In the forward process, cold diffusion aims to transform the image from the cold state to the hot state. Within this framework, the clean image x_0 can be seen as being in a cold state, which is randomly sampled from a training data distribution P_{data}. x_T can be seen as an image in a hot state sampled from a random distribution P_{rand}. By using a degradation operator $\mathcal{D}(\cdot)$, image at any time t can be determined as

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

where T represents the total number of steps for the diffusion process. α_t are hyperparameters, which dictate the variance of the noise introduced at time t.

In the reverse process, cold diffusion transforms the sampled image from a random hot state to a cold state. We introduce a restoration operator $\mathcal{R}(\cdot)$, tasked with recovering the clear image x_0 from any degraded image x_t at time t. This process can be described as follows:

$$\hat{x}_0 = \mathcal{R}(x_t, t) \approx x_0. \tag{2}$$

We utilize a neural network with parameters θ to serve as the restoration operator, $\mathcal{R}_\theta(\cdot)$, which can be optimized by

$$\min_\theta \mathbb{E}_{\substack{x_0\sim P_{\text{data}} \\ x_T\sim P_{\text{rand}} \\ t\sim U(0,T)}} \|\mathcal{R}_\theta(\mathcal{D}(x_0, x_T, t), t) - x_0\| \tag{3}$$

where U denotes a uniform distribution. However, researchers suggest that such a one-step learning procedure significantly diminishes image quality [52], [53]. To address this issue, cold diffusion employs a naive sampling algorithm, which progressively restores the degraded image over T iterations

$$x_{t-1} = \mathcal{D}\big(\hat{x}_0, x_T, t-1\big). \tag{4}$$

This technique yields more favorable reconstruction results than the one-step version. However, Bansal et al. [52] found that when $\mathcal{R}(\cdot)$ is not perfectly equal to the inverse of $\mathcal{D}(\cdot)$, it will bring prediction errors between x_0 and $\hat{x}_0$. During the sampling process, errors are accumulated, which will further undermine the reconstruction result. To effectively avoid this accumulative error, an improved sampling algorithm is used

$$x_{t-1} = x_t - \mathcal{D}\big(\hat{x}_0, x_T, t\big) + \mathcal{D}\big(\hat{x}_0, x_T, t-1\big). \tag{5}$$

B. Physics-Regularized Learning

To guide image reconstruction with the physical priors of the image, inspired by LEARN [39], we adopted a learnable regularization network structure. To ensure self-contained and comprehensiveness, we provide a brief overview of the LEARN model. A typical reconstruction model can be formulated as

$$\min_x \frac{1}{2}\|\boldsymbol{A}x - y\|_2^2 + \lambda G(x) \tag{6}$$

where $x \in \mathbb{R}^{mn}$ denotes the vectorization of an image matrix of size $m \times n$, $y \in \mathbb{R}^M$ represents the measured sinogram data, and the system matrix, $\boldsymbol{A} \in \mathbb{R}^{M\times mn}$, aims to map image pixels to sinogram data. The function $G(\cdot)$ indicates the physical prior derived regularization term and λ serves as a tradeoff hyperparameter to balance the data fidelity and regularization terms. Following the LEARN model, we employ a generalized regularization term known as "field of experts" (FoE) [60]

$$G(x) = \sum_{k=1}^{K} \varphi_k(\phi_k(x)) \tag{7}$$

where K represents the number of regularizers. The terms $\varphi_k(\cdot)$ and $\phi_k(\cdot)$ represent the transform function and potential function, respectively. In the FoE model, these functions can be substituted with convolution operators, which can be learned in a data-driven manner. Based on (6), x can be obtained by minimizing the following objective function:

$$\bar{x} = \operatorname*{argmin}_x \frac{1}{2}\|\boldsymbol{A}x - y\|_2^2 + \sum_{k=1}^{K} \lambda_k \varphi_k(\phi_k(x)). \tag{8}$$

Equation (8) can be solved iteratively using the gradient descent method

$$x^{t+1} = x^t - \eta\left[\boldsymbol{A}^T\big(\boldsymbol{A}x^t - y\big) + \sum_{k=1}^{K} \lambda_k \phi_k' \varphi_k\big(\phi_k(x^t)\big)\right] \tag{9}$$

where $\phi'(\cdot)$ signifies the conjugate operator of $\phi(\cdot)$, and η stands for the iteration step size. When making the terms in (9) to be iteration-dependent, we have

$$x^{t+1} = x^t - \eta^t \boldsymbol{A}^T\big(\boldsymbol{A}x^t - y\big) - \sum_{k=1}^{K} \lambda_k^t \phi_k'' \varphi_k^t\big(\phi_k^t(x^t)\big). \tag{10}$$

The last term in (10) can be implemented as a neural network, $\Omega_\xi(\cdot)$. Consequently, the entire process can be represented as

$$x^{t+1} = x^t - \eta^t \boldsymbol{A}^T\big(\boldsymbol{A}x^t - y\big) - \Omega_\xi\big(x^t\big). \tag{11}$$

C. Proposed PrideDiff Model

Building upon the foundational principles discussed earlier and taking into account the physical characteristics of CT images, we develop a novel iterative diffusion model for CT reconstruction, termed PrideDiff. This new model employs a

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

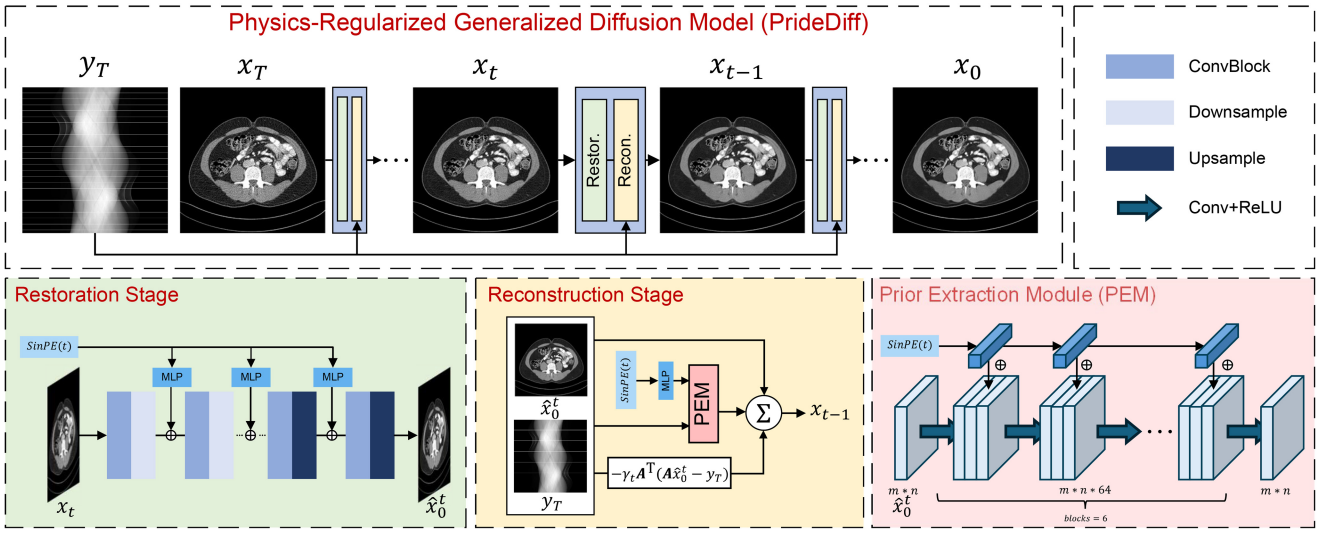

Fig. 1. Overview of the proposed PrideDiff. PrideDiff enhances image reconstruction by integrating CT image sinogram domain prior features into the diffusion process during each iterative using a two-stage network. To achieve a more effective feature representation, we propose a PEM embedded with time embedding features.

general diffusion model to integrate CT physical prior with image domain features in an efficient manner. The whole structure of the model is shown in Fig. 1.

1) Improved Diffusion Process: In the context of a CT image reconstruction task with a vanilla diffusion model, one usually transitions a fully sampled CT image to a whole random Gaussian distribution and uses an undersampled CT image as a condition to predict the fully sampled CT image. However, real clinical CT images are usually contaminated by hybrid noise, which cannot be precisely characterized by any statistical distribution. As a result, we leverage cold diffusion to make the degradation process more general, aligning more closely with the statistical features of CT images. In PrideDiff, we use undersampled CT images as the starting point and fully sampled CT images as the endpoint for the sampling process. To make this process close to the CT image degradation process, we draw inspiration from CoreDiff [53] and adopt a consistent strategy for our experiments. Consequently, (1) is adjusted as follows:

$$x_t = \mathcal{D}(x_0, x_t, t) = \alpha_t x_0 + (1 - \alpha_t) x_T \tag{12}$$

where x_T represents the undersampled CT image, x_0 is the fully sampled CT image, and x_t denotes the intermediate result obtained at time t. Indeed, x_T can be perceived as an intermediary state between a fully sampled CT (cold state) and random noise (hot state), termed a warm state. Clearly, the transition from the cold state to the warm state is simpler than from the hot state. Thus, by using an undersampled CT image instead of random noise as the sampling starting point, we can effectively reduce the number of iteration steps T.

2) Restoration–Reconstruction Mechanism: To incorporate the sinogram domain prior data of CT images with the diffusion process, we are inspired by the LEARN structure and redesign the iterative process to a "restoration–reconstruction" mechanism, illustrated in Fig. 1. In the restoration stage, we get the degraded image x_t generated by (12) and then use the restoration operator $\mathcal{R}_\theta(\cdot)$ to obtain the restored image $\hat{x}_0$. In the reconstruction stage, we take the restored CT image $\hat{x}_0$ and the corresponding undersampled sinogram data y_T as input, then we can obtain the reconstruction result $\tilde{x}_0$. At this stage, we construct a PEM that incorporates time embedding features as a regularization term to effectively utilize physical priors. In order to extract effective prior features at different time steps t, we need to use time embeddings in PEM for guidance. The time step t is embedded using the time embedding module, as shown in Fig. 1. This process can be denoted as: MLP(SinPE(t)), where SinPE($\cdot$) represents sinusoidal position encoding for time step t, followed by a multilayer perceptron (MLP) to get the time embedding features.

The whole restoration–reconstruction process can be described as

$$\begin{cases} \hat{x}_0^t = \mathcal{R}_\theta(x_t, t) \\ \tilde{x}_0^t = \mathcal{F}\left(\hat{x}_0^t, y_T, t, \Omega_\xi\right) \\ x_{t-1} = x_t - \mathcal{D}\left(\tilde{x}_0^t, x_T, t\right) + \mathcal{D}\left(\tilde{x}_0^t, x_T, t-1\right) \end{cases} \tag{13}$$

where $\hat{x}_0^t$ and $\tilde{x}_0^t$ are the outputs of the restoration stage and reconstruction stage at time t, respectively. $\mathcal{F}(\cdot)$ represents the reconstruction stage, and $\Omega_\xi(\cdot)$ is the PEM with learnable parameters ξ. The $\mathcal{F}(\cdot)$ process combined with (10) can be expanded as follows:

$$\mathcal{F}\left(\hat{x}_0^t, y_T, t, \Omega_\xi\right) = \hat{x}_0^t - \gamma_t \boldsymbol{A}^T\left(\boldsymbol{A}\hat{x}_0^t - y_T\right) - \Omega_\xi\left(\hat{x}_0^t, t\right) \tag{14}$$

where y_T is the projection data corresponding to the undersampled CT image x_T, and γ is a learnable tradeoff factor. In order to further improve the performance of the reconstruction stage, according to [16] and [61], we replace $\boldsymbol{A}^T$ with an inverse transformation of $\boldsymbol{A}$ (e.g., FBP) to make this process more efficient. Then, the formula can be rewritten as

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Training Stage of PrideDiff

Input: Total sample steps T. Paired fully sampled and undersampled CT images $(x_0, x_T, y_T) \sim P_{\text{data}}$.

Output: R_θ, Ω_ξ

while *not converged* **do**
  Sample (x_0, x_T); Sample $t \sim U(1, \ldots, T)$;
  $x_t \leftarrow \alpha_t x_0 + (1-\alpha_t)x_T$; // Eq. (1)
  $\hat{x}_0^t \leftarrow \mathcal{R}_\theta(x_t, t)$; // Restor.
  $\tilde{x}_0^t \leftarrow \mathcal{F}(\hat{x}_0^t, y_T, t, \Omega_\xi)$; // Recon.
  Update θ, ξ using Eq. (16)
end

Algorithm 2: Sampling Stage of PrideDiff

Input: Total sample steps T. An undersampled CT image with sinogram (x_T, y_T)

Output: Reconstruction CT image x_{out}.

for $t = T, T-1, \ldots, 1$ **do**
  $\hat{x}_0^t \leftarrow \mathcal{R}_\theta(x_t, t)$; $\tilde{x}_0^t \leftarrow \mathcal{F}(\hat{x}_0^t, y_T, t, \Omega_\xi)$;
  $x_{t-1} = x_t - \mathcal{D}(\tilde{x}_0^t, x_T, t) + \mathcal{D}(\tilde{x}_0^t, x_T, t-1)$;
end
$x_{out} \leftarrow x_0$

$$\tilde{x}_0^t = \hat{x}_0^t - \gamma_t \,\text{FBP}\big(\boldsymbol{A}\hat{x}_0^t - y_T\big) - \Omega_\xi\big(\hat{x}_0^t, t\big). \tag{15}$$

Additionally, according to the CT imaging algorithm shown in [16], $\text{FBP}(p) = B(h * p)$, where B denotes the backprojection matrix and h represents the *R–L* filter. Since the FBP algorithm involves only convolution and matrix multiplication operations, which are linearly differentiable, (15) can still be optimized using a gradient-based optimization algorithm.

It is worth noting that both the diffusion model and iterative network model are time-dependent. The use of the PEM structure well integrates these two types of models. On the one hand, we can benefit from the advantages of both iterative network model and diffusion model. On the other hand, we can avoid the additional computational consumption required by the iterative reconstruction method as much as possible.

Based on the above description, the training objective function of our PrideDiff is defined as follows:

$$\min_{\theta,\xi} \mathbb{E}_{x_0,t}\Big[\|\hat{x}_0^t - x_0\|^2 + \|\tilde{x}_0^t - x_0\|^2\Big] \tag{16}$$

where $\hat{x}_0^t = \mathcal{R}_\theta(x_t, t)$ represents the restoration result obtained at time t and $\tilde{x}_0^t = \mathcal{F}(\hat{x}_0^t, y_T, t, \Omega_\xi)$ represents the corresponding reconstruction result.

In summary, during the training process, we continuously randomly sample data at time t to train the network parameters. During the sampling process, we fix the parameters of the corresponding network and iteratively perform the reconstruction T times. After T iterations, we obtain the final reconstruction results. The training and sampling processes of the proposed PrideDiff are listed in Algorithms 1 and 2, respectively.

III. Experiments

A. Datasets

To validate the effectiveness of the proposed PrideDiff, we conducted experiments on two different tasks: 1) low-dose and 2) sparse-view CT reconstructions. The data were obtained from the "2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge (Mayo-2016)" and the "Low-dose CT image and projection dataset (Mayo-2020)" [62]. The "Mayo-2016" data includes 5936 CT slices of 1-mm slice thickness from ten patients. We selected data from seven patients as the training set, data from one patient as the validation set, and the rest as the test set. The "Mayo-2020" dataset includes data from the head, lungs, and abdomen. Here, we selected eight sets of lung data as a supplement to the "Mayo-2016" data, including 2378 slices.

The physical parameters of CT simulation were set as follows: the physical height and width of the pixels are 0.7433 mm, and there are a total of 768 detector units, each with a size of 1.2858 mm. The projection data were simulated using the distance-driven method [63], [64] with fan beam geometry. The distance between the X-ray source and the scanner rotation center is 595.0 mm, and the distance between the scanner rotation center and the detector is 490.6 mm. To meet the requirements of different tasks, we processed the data separately. For the low-dose CT reconstruction task, we conducted experiments using 25% dose data from the Mayo-2016 dataset and 10% dose data from the Mayo-2020 dataset. All low-dose data used for our experiments were provided by the Mayo Clinic [62]. For the sparse-view CT reconstruction task, the "Mayo-2016" data were used. We downsampled the 1024-views fully sampled CT projection data to 120 views and 60 views, respectively, to obtain sparse-view CT data for training.

B. Implementation Details

In the data preprocessing stage, we conducted experiments using CT images of size 512×512, and normalized the data to a range of [0, 1]. The basic operators in CT reconstruction, including projection, back-projection, and FBP, were implemented using the CUDA-based DeepRecon package [65]. In the training stage, Adam optimizer was employed with a learning rate of 1.0×10^{-5} across 400K iterations. The model training took approximately three days. The whole experiment was conducted on PyTorch v1.12 and ran on Ubuntu 20.04, using a single NVIDIA RTX 4090 (24 GB). We will release the associated training weights and codes later at https://github.com/phoenixyu/PrideDiff.

We evaluated PrideDiff against several state-of-the-art DL-based CT reconstruction methods, including: 1) *DL-based image domain processing methods:* FBPConvNet [27], RED-CNN [26], and CNCL [30]; 2) *DL-based iterative reconstruction methods:* LEARN [39]; 3) *Diffusion methods:* DDPM [45], DDPM with DPM-solver (DDPM-DPM) [51], Cold Diffusion [52], and CoreDiff [53]; and 4) *Unconditional score-based methods:* MCG [55] and DiffMBIR [57]. Specifically, except for unconditional score-based methods, all other methods are based on supervised learning and require

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

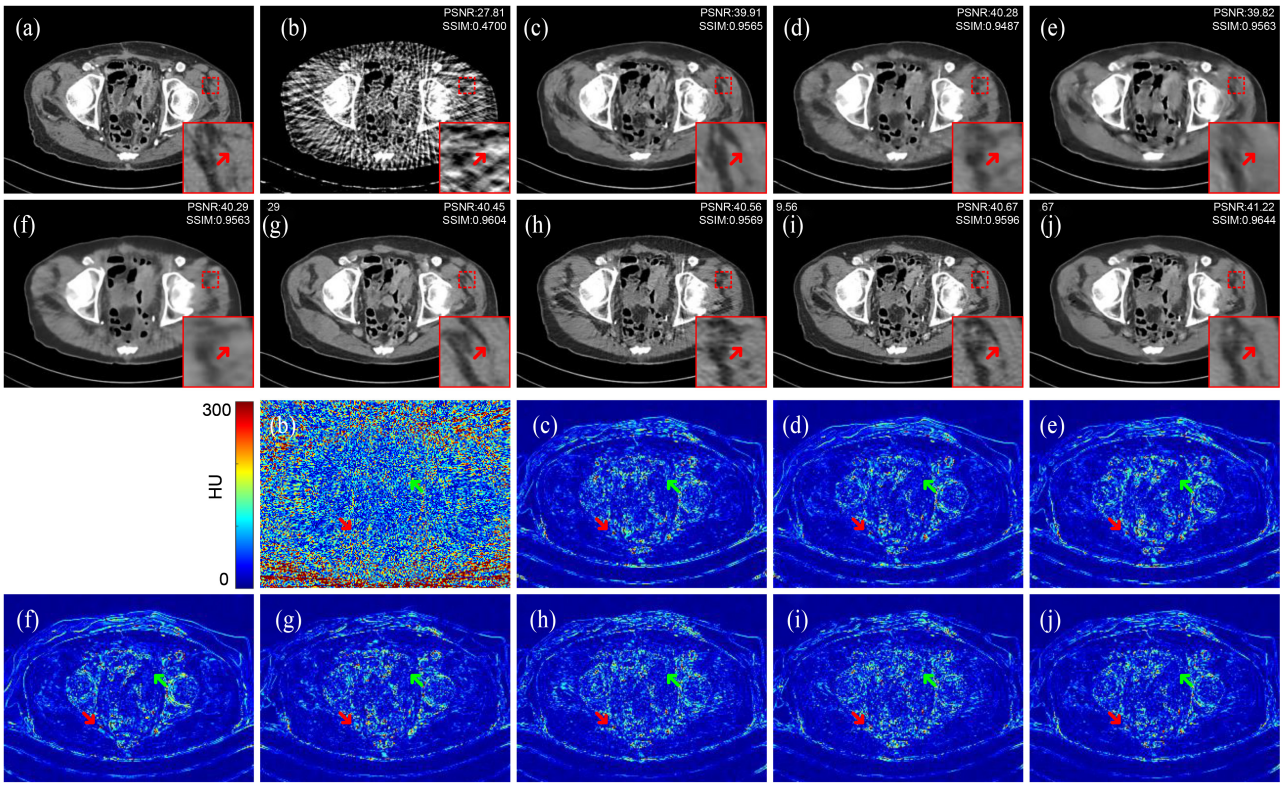

Fig. 2. Qualitative results of 60-view abdomen sparse-view CT image from Mayo 2016 by different methods. (a) GT, (b) FBP, (c) FBPConvNet, (d) RED-CNN, (e) CNCL, (f) LEARN, (g) DDPM-1000, (h) Cold Diffusion, (i) CoreDiff, and (j) PrideDiff (**ours**). The display window is [−160, 240] HU.

paired data for training. For all comparative methods, we used the officially released codes with the hyperparameters or inference weights suggested in the original papers.

For image domain processing methods, we set the batch size to 6, the maximum epoch number to 200, and the learning rate to 1.0×10^{-4}. For CNCL, we adopted the PatchGAN [66] as the discriminator, as used in the original paper, with a learning rate 1.0×10^{-5}. For LEARN, we set the number of iteration blocks to 50 with a learning rate of 1.0×10^{-4}. For DDPM-based methods, we set the iteration step T to 1000, the learning rate was 1.0×10^{-5}, and generated $\alpha_1, \ldots, \alpha_T$ from 0.999 to 0 linearly. We also utilized the DPM-Solver to accelerate the DDPM inference phase, reducing the iteration steps to 50 (termed DDPM-DPM-50). For the cold diffusion, we set the iteration T to 10, as recommended in [52] and [53]. In the PEM module, we configured six convolution blocks and integrated temporal features, while maintaining an image feature size of 64 throughout. All the methods employed the Adam optimizer, with other parameters set to their default values.

We chose three quantitative metrics commonly employed in image reconstruction for comparison: 1) peak signal-to-noise ratio (PSNR); 2) structural similarity index measure (SSIM); and 3) root mean-square error (RMSE).

C. Comparisons With State-of-the-Art Methods

1) Sparse-View CT Task: Fig. 2 shows one representative slice reconstructed with the 60-view data using different methods. Red arrows point to the region of interest (ROI). The lower half of Fig. 2 presents the difference maps with the ground truth (GT). Fig. 2 shows that for image domain methods, although most of the artifacts are removed, some details of the image are oversmoothed. In the result of LEARN, some image details are lost, which is caused by the limited capabilities in network feature extraction. For DDPM-1000, some structures in the image are distorted due to error accumulation during iterations. Although the use of the DPM-Solver can effectively reduce the number of iteration steps in the inference stage, the experiments show that the reconstruction results were unsatisfactory for this task. For the cold diffusion model, it is noticed that while some structures are well preserved, some streak artifacts are not effectively removed. For the CoreDiff model, the image details have been further improved, but there are still some shortcomings in suppressing some streak artifacts. Our PrideDiff provides the best visual result and the smallest difference map than other methods. PrideDiff also obtains the best scores in terms of both PSNR and SSIM.

Fig. 3 displays another slice reconstructed with the 120-view data using different methods. From the region indicated by the red arrows, it can be observed that DL-based image domain processing methods can not recover the structural details effectively. LEARN falls in eliminating the artifacts. DDPM-1000 tends to distort the structures. For the acceleration version DDPM-DPM-50, although it can remove most of the artifacts, many noises still remain. Cold diffusion and CoreDiff remain challenged in addressing streak artifacts

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

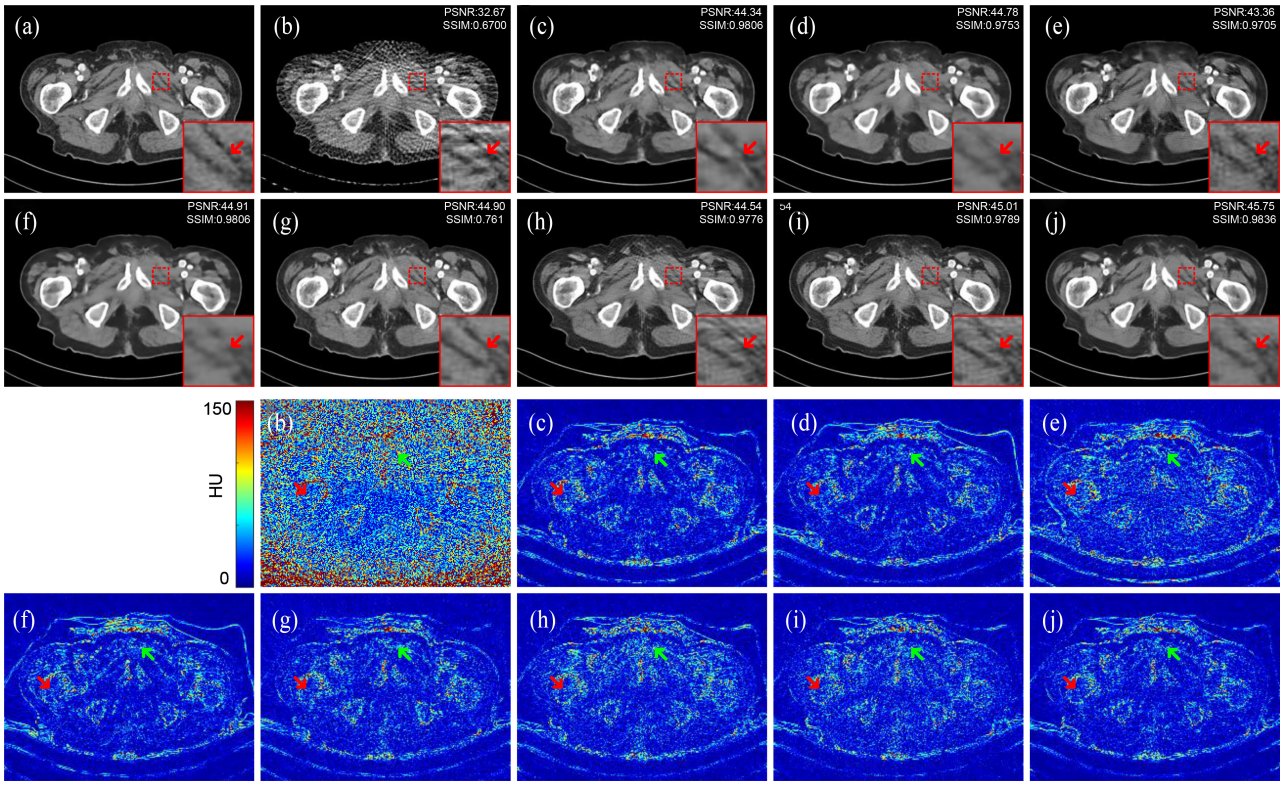

Fig. 3. Qualitative results of 120-view pelvic sparse-view CT image from Mayo 2016 by different methods. (a) GT, (b) FBP, (c) FBPConvNet, (d) RED-CNN, (e) CNCL, (f) LEARN, (g) DDPM-1000, (h) Cold Diffusion, (i) CoreDiff, and (j) PrideDiff (**ours**). The display window is [−160, 240] HU.

TABLE I
QUANTITATIVE RESULTS (MEAN±SD) ON SPARSE-VIEW TESTING DATA. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Views	60 Views			120 Views		
	PSNR ↑	SSIM ↑	RMSE ↓	PSNR ↑	SSIM ↑	RMSE ↓
FBP	28.13 ± 0.59	0.5123 ± 0.0402	0.0393 ± 0.0027	32.94 ± 0.61	0.7039 ± 0.0339	0.0226 ± 0.0016
FBPConvNet [27]	40.91 ± 0.80	0.9640 ± 0.0058	0.0090 ± 0.0008	44.92 ± 0.87	0.9811 ± 0.0033	0.0057 ± 0.0006
RED-CNN [26]	41.12 ± 0.83	0.9561 ± 0.0061	0.0088 ± 0.0008	45.31 ± 0.83	0.9772 ± 0.0037	0.0054 ± 0.0005
CNCL [30]	40.75 ± 0.97	0.9582 ± 0.0114	0.0092 ± 0.0010	44.45 ± 0.97	0.9755 ± 0.0062	0.0060 ± 0.0007
LEARN [39]	41.08 ± 0.95	0.9613 ± 0.0058	0.0089 ± 0.0010	45.41 ± 0.89	0.9809 ± 0.0036	0.0054 ± 0.0006
DDPM-1000 [45]	41.94 ± 0.97	0.9674 ± 0.0050	0.0081 ± 0.0009	45.52 ± 0.79	0.9782 ± 0.0024	0.0053 ± 0.0005
DDPM-DPM-50 [51]	23.36 ± 2.17	0.2031 ± 0.0823	0.0700 ± 0.0163	36.87 ± 5.77	0.7719 ± 0.1936	0.0177 ± 0.0115
Cold Diffusion [52]	41.74 ± 1.08	0.9649 ± 0.0064	0.0082 ± 0.0010	45.22 ± 0.93	0.9795 ± 0.0035	0.0055 ± 0.0006
CoreDiff [53]	42.31 ± 1.12	0.9684 ± 0.0058	0.0077 ± 0.0009	45.95 ± 0.94	0.9816 ± 0.0032	0.0050 ± 0.0005
PrideDiff (**ours**)	**42.54 ± 1.07**	**0.9710 ± 0.0050**	**0.0075 ± 0.0009**	**46.43 ± 0.91**	**0.9846 ± 0.0028**	**0.0048 ± 0.0005**

effectively. In contrast, our PrideDiff not only preserves the structural details but also demonstrates superior suppression of artifacts and noise. The difference maps further reveal that in the bone region, PrideDiff achieves the closest result to the GT.

Table I lists the quantitative results of the whole test set. It is evident that our method yields the best results on the test set in all the metrics.

2) Low-Dose CT Task: Fig. 4 illustrates one thoracic image reconstructed with the low-dose data using different methods at 10% dose level. The corresponding PSNR and SSIM results are listed in the caption, and the related difference images are shown below. In the ROIs indicated by the red arrows, taking the vessel region as an example, FBPConvNet, RED-CNN, and CNCL produce blurred details. Oversmoothing effects in certain regions can be noticed in the result of LEARN. The difference maps suggest that DDPM exhibits suboptimal reconstruction within the lung regions. The method cold diffusion, CoreDiff, and PrideDiff gain more desirable results. Compared to CoreDiff, our method achieves better preservation of image details. The quantitative results show an inherent trend to the visual inspection and PrideDiff gets the best scores in all the competing methods.

Fig. 5 shows another abdomen image reconstructed with the low-dose data using different methods at 25% dose level. Since the low-dose image contains less noise, most methods can remove most noise. However, as indicated by the arrows in Fig. 5, the results of FBPConvNet, RED-CNN, CNCL, and LEARN exhibit oversmoothing effects to varying degrees. Diffusion-based methods demonstrate better performance in detail preservation. In particular, in this task, the input image can provide a more effective feature because it has less noise, most models can achieve noise removal well, and our PrideDiff shows superior reconstruction around the spine. Additionally,

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

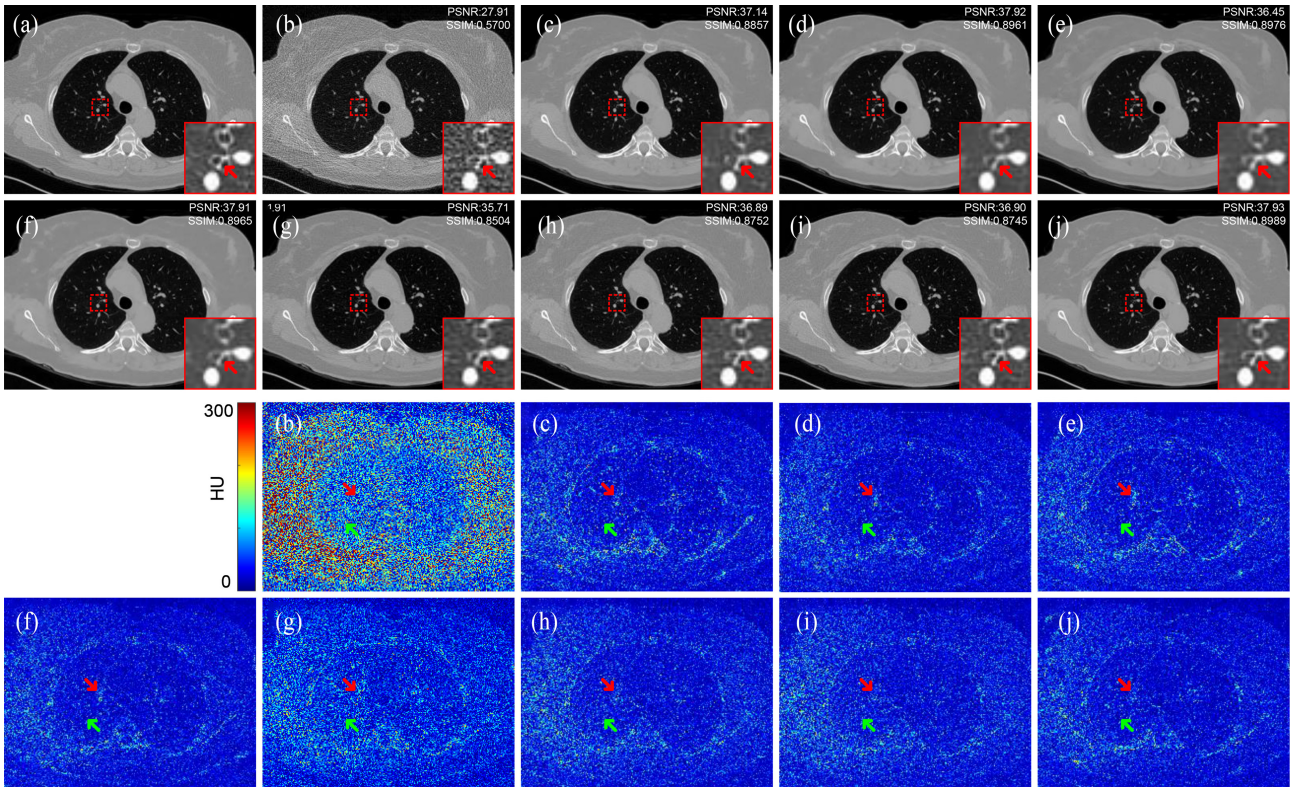

Fig. 4. Qualitative results of 10% dose thoracic CT image from Mayo 2020 by different methods. (a) GT, (b) FBP, (c) FBPConvNet, (d) RED-CNN, (e) CNCL, (f) LEARN, (g) DDPM-1000, (h) Cold Diffusion, (i) CoreDiff, and (j) PrideDiff (**ours**). The display window is [−200, 1000] HU.

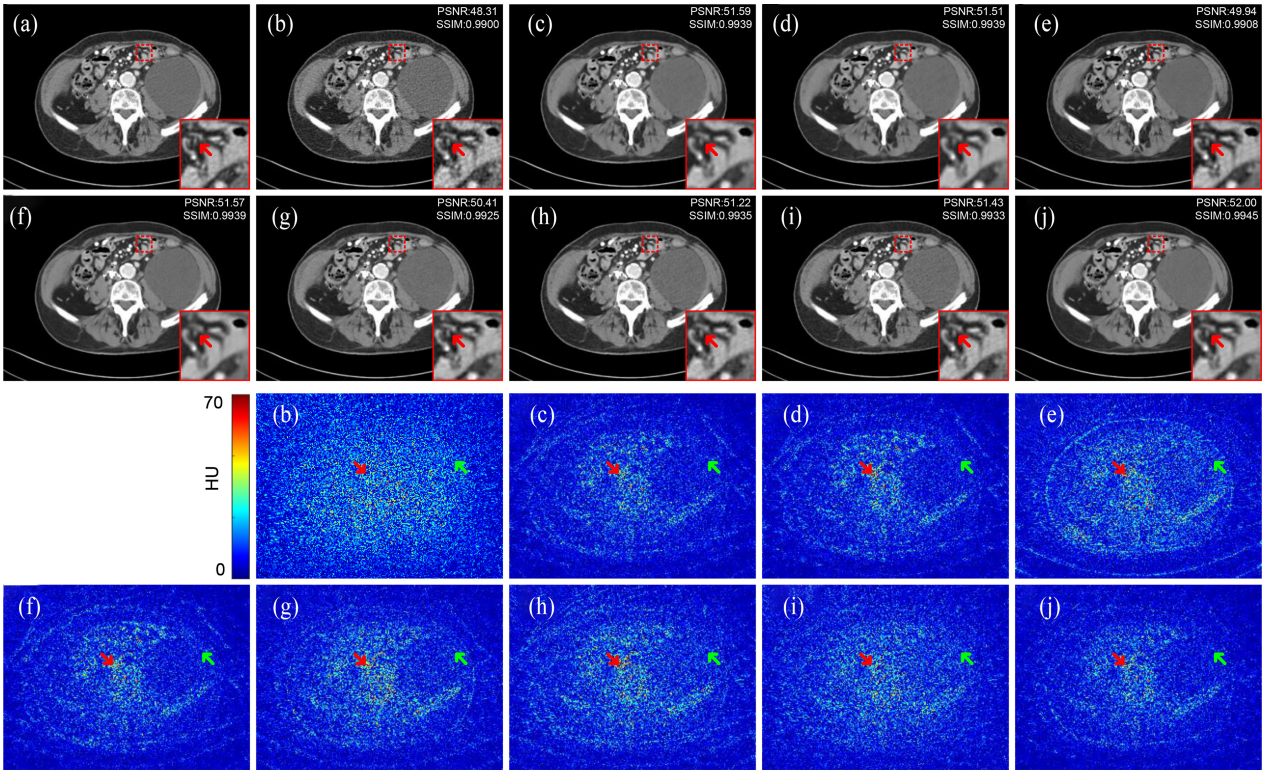

Fig. 5. Qualitative results of 25% dose abdomen CT image from Mayo 2016 by different methods. (a) GT, (b) FBP, (c) FBPConvNet, (d) RED-CNN, (e) CNCL, (f) LEARN, (g) DDPM-1000, (h) Cold Diffusion, (i) CoreDiff, and (j) PrideDiff (**ours**). The display window is [−160, 240] HU.

in the right part of the stomach, PrideDiff exhibits smaller discrepancies than other methods, indicating better performance.

Table II shows that the quantitative results on the whole test set support the visual results and the proposed PrideDiff has superior reconstruction performance over other SOTA models.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

TABLE II
QUANTITATIVE RESULTS (MEAN±SD) ON LOW-DOSE TESTING DATA. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Dose	25%			10%		
	PSNR ↑	SSIM ↑	RMSE ↓	PSNR ↑	SSIM ↑	RMSE ↓
FBP	46.31 ± 1.55	0.9783 ± 0.0080	0.0049 ± 0.0009	25.83 ± 2.26	0.4773 ± 0.0987	0.0530 ± 0.0148
FBPConvNet [27]	50.35 ± 1.08	0.9923 ± 0.0023	0.0031 ± 0.0004	35.77 ± 1.74	0.8408 ± 0.0597	0.0166 ± 0.0034
RED-CNN [26]	42.37 ± 6.87	0.9152 ± 0.0760	0.0100 ± 0.0062	36.52 ± 1.75	0.8560 ± 0.0545	0.0152 ± 0.0032
CNCL [30]	48.93 ± 1.33	0.9869 ± 0.0075	0.0036 ± 0.0006	36.27 ± 1.60	0.8586 ± 0.0540	0.0156 ± 0.0031
LEARN [39]	50.60 ± 1.09	0.9926 ± 0.0022	0.0030 ± 0.0004	36.59 ± 1.68	0.8615 ± 0.0522	0.0149 ± 0.0030
DDPM-1000 [45]	49.56 ± 0.94	0.9905 ± 0.0026	0.0033 ± 0.0004	34.75 ± 1.58	0.8032 ± 0.0641	0.0186 ± 0.0035
DDPM-DPM-50 [51]	50.15 ± 1.12	0.9910 ± 0.0030	0.0031 ± 0.0004	23.99 ± 0.60	0.8459 ± 0.0527	0.0633 ± 0.0045
Cold Diffusion [52]	49.87 ± 1.24	0.9913 ± 0.0030	0.0032 ± 0.0005	35.30 ± 1.83	0.8249 ± 0.0627	0.0176 ± 0.0039
CoreDiff [53]	49.81 ± 1.38	0.9906 ± 0.0033	0.0032 ± 0.0005	35.35 ± 1.89	0.8246 ± 0.0637	0.0175 ± 0.0039
PrideDiff (**ours**)	**50.72 ± 1.17**	**0.9928 ± 0.0023**	**0.0029 ± 0.0004**	**36.65 ± 1.78**	**0.8603 ± 0.0536**	**0.0150 ± 0.0032**

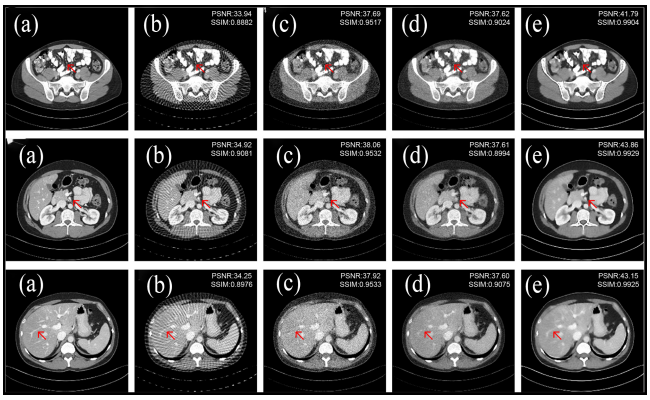

Fig. 6. Qualitative results of 60-view sparse-view CT image from Mayo 2016 by different methods. (a) GT, (b) FBP, (c) MCG, (d) DiffMBIR, and (e) PrideDiff (**ours**). The display window is [−160, 240] HU.

D. Comparison With Unconditional Diffusion Methods

We further assessed the performance differences between our PrideDiff and unconditional diffusion methods by comparing it with the MCG [55] and DiffMBIR [57] on 60-view sparse-view CT reconstruction. Although MCG and DiffMBIR demonstrate strong generative capabilities, particularly at the ultrasparse-view tasks, their high diversity can make it challenging to maintain content consistency, leading to deviations from the GTs [67]. As shown in Fig. 6, the red arrow in the second row highlights significant structural distortions caused by these methods. Our method, however, not only preserves detailed structures but also achieves excellent performance at 60-view settings. Furthermore, our method has a notable efficiency advantage. Table III shows that our PrideDiff can restore an image in just 0.69 s, meeting the practical requirements for medical applications.

E. Ablation Study

1) Impact of Iteration Steps: To further investigate the impact of the number of iteration steps T on network performance, we conducted three experiments with several different values of T on sparse-view CT with 120 views. The statistical quantitative results in Table IV show that when $T = 5$, PrideDiff already achieves a relatively satisfactory reconstruction result. At $T = 10$, we notice an unremarkable improvement in quantitative scores. However, as we continue to increase the value of T, the improvement in results becomes quite marginal. As a result, we empirically chose $T = 10$ in this study.

TABLE III
QUANTITATIVE RESULT (MEAN±SD) AND INFERENCE TIME ANALYSIS WITH DIFFERENT METHODS

	PSNR	SSIM	Time [s]
MCG [55]	38.02 ± 0.34	0.9536 ± 0.0016	1138.2
DiffMBIR [57]	36.98 ± 0.87	0.8895 ± 0.0017	718.9
PrideDiff (**ours**)	42.94 ± 0.79	0.9923 ± 0.0012	0.69

2) Effect of Restoration–Reconstruction Mechanism: To validate the effectiveness of the proposed restoration–reconstruction mechanism, we conducted an ablation study. In the 120-view CT reconstruction task, we kept other parameters the same and obtained results by removing the reconstruction stage in Fig. 1. Table V shows the quantitative results and inference time. In Table V, we observe that by introducing the reconstruction stage, the inference time only increases by 0.26 s, while there is a significant improvement in PSNR and SSIM. Based on this observation, the proposed restoration–reconstruction mechanism can effectively enhance the reconstruction quality while ensuring model efficiency.

F. Model Analysis

To delve deeper into whether PrideDiff meets expectations during the iterative reconstruction phase, we showcased the outputs of sparse-view and 25% dose level reconstruction in Fig. 7. It is evident that as the iterative process continues, noise and artifacts in the image are gradually removed, culminating in the final reconstruction after T iterations. Additionally, we observed a continuous enhancement in image details, further suggesting that PrideDiff provides positive feedback on image details during the iterative process. As we all know, diffusion-based methods are difficult to apply in clinical practice due to their long inference time. Here, we analyze the different models, as shown in Table VI. We categorize the methods based on whether they are diffusion-based. CNN-based methods, such as FBPConvNet and RED-CNN, have the fastest test speeds. GAN-based CNCL requires longer test time. As an iterative reconstruction network model, LEARN contains multiple projection and back-projection operations, which significantly increase the computational time.

For diffusion-based methods, we only used a simple network structure during the reconstruction process, as shown

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

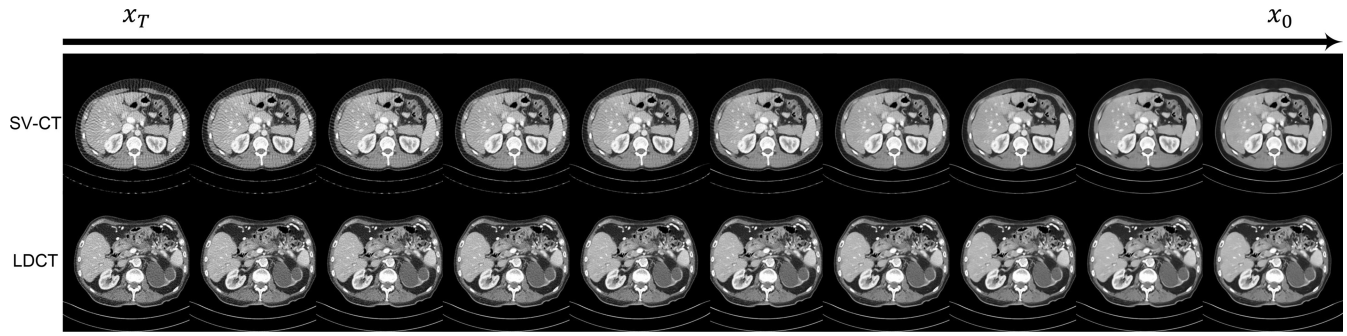

Fig. 7. Iterative process visualization of PrideDiff in sparse-view and low-dose tasks with iteration step $T = 10$ on Mayo 2016. The display window is [−160, 240] HU.

TABLE IV
QUANTITATIVE RESULT (MEAN±SD) OF DIFFERENT NUMBERS OF ITERATION STEPS

	$T = 5$	$T = 10$	$T = 20$
PSNR	46.28 ± 0.93	46.43 ± 0.91	**46.44 ± 0.91**
SSIM	0.9843 ± 0.0028	**0.9846 ± 0.0028**	0.9845 ± 0.0028

TABLE V
ABLATION STUDY OF RESTORATION–RECONSTRUCTION MECHANISM

	PSNR	SSIM	Time [s]
PrideDiff w/o recon.	45.22 ± 0.93	0.9795 ± 0.0035	**0.43**
PrideDiff w/ recon.	**46.43 ± 0.91**	**0.9846 ± 0.0028**	0.69

TABLE VI
MODEL SCALE AND INFERENCE TIME ANALYSIS

	Param. [M]	FLOPs [G]	Time [s]
FBPConvNet [27]	34.56	194.15	0.01
RED-CNN [26]	1.85	229.32	0.01
CNCL [30]	43.83	290.90	0.10
LEARN [39]	1.08	285.66	1.16
DDPM-1000 [45]	3.46	131.33	223.47
DDPM-DPM-50 [51]	3.46	131.33	116.63
Cold Diffusion [52]	4.35	135.92	0.43
CoreDiff [53]	4.35	152.65	0.64
PrideDiff (**ours**)	4.61	194.39	0.69

in Fig. 1. Therefore, in terms of the number of parameters and computational complexity, diffusion-based methods have a distinct advantage. However, since these methods require many iterations to complete the reconstruction, they demand extra test time. As shown in Table VI, vanilla DDPM requires approximately 223 s to process one single slice, heavily hindering its promotion in clinical practice. Even with the expedited DPM-Solver algorithm, the inference time hovers around 116 s. However, the cold diffusion model yields a remarkable reduction in test time. The proposed PrideDiff also benefits from this, slightly increasing the testing time while effectively improving performance. Our method improves the testing speed by 63% compared to the LEARN.

IV. CONCLUSION

In this article, we introduced PrideDiff, a novel CT reconstruction network that merges the iterative reconstruction network with a generalized diffusion model. Our PrideDiff serves as a bridge between the diffusion model and the physics-regularized iterative reconstruction network model. On the one hand, PrideDiff effectively integrates physical prior into the diffusion process, significantly enhancing the imaging quality. On the other hand, PrideDiff efficiently leverages the advantages of cold diffusion and fuses time-embedding features into the regularization network, effectively reducing the number of iteration steps and enhancing network performance. Moreover, we verified the impact of different components on network performance. We also analyzed the intermediate results produced by PrideDiff during iterations to demonstrate the stability and effectiveness. The model analysis also demonstrates the superiority of PrideDiff in terms of model scale and inference time. In the future, we aim to enhance the generative capabilities of our method, explore its performance under conditions of ultra-sparse sampling, and further refine the quality of image restoration.

ETHICAL STATEMENTS

This research study was conducted retrospectively using human subject data made available in open access by Mayo Clinic. Ethical approval was not required as confirmed by the license attached with the open access data.

ACKNOWLEDGMENT

All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this article.

REFERENCES

[1] D. J. Brenner and E. J. Hall, "Computed tomography—An increasing source of radiation exposure," *N. Engl. J. Med.*, vol. 357, no. 22, pp. 2277–2284, 2007.
[2] J. Jing et al., "Training low dose CT denoising network without high quality reference data," *Phys. Med. Biol.*, vol. 67, no. 8, 2022, Art. no. 84002.
[3] T. Okamoto, T. Ohnishi, and H. Haneishi, "Artifact reduction for sparse-view CT using deep learning with band patch," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 8, pp. 859–873, Nov. 2022.
[4] W. Xia, H. Shan, G. Wang, and Y. Zhang, "Physics-/model-based and data-driven methods for low-dose computed tomography: A survey," *IEEE Signal Process. Mag.*, vol. 40, no. 2, pp. 89–100, Mar. 2023.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

[5] Z. Huang et al., "CaGAN: A cycle-consistent generative adversarial network with attention for low-dose CT imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1203–1218, Aug. 2020.

[6] Y. Zhang et al., "DREAM-Net: Deep residual error iterative minimization network for sparse-view CT reconstruction," *IEEE J. Biomed. Health. Inf.*, vol. 27, no. 1, pp. 480–491, Jan. 2023.

[7] Y. Liu et al., "Cross-domain unpaired learning for low-dose CT imaging," *IEEE J. Biomed. Health. Inf.*, vol. 27, no. 11, pp. 5471–5482, Nov. 2023.

[8] 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.

[9] T. Wang et al., "SemiMAR: Semi-supervised learning for CT metal artifact reduction," *IEEE J. Biomed. Health. Inf.*, vol. 27, no. 11, pp. 5369–5380, Nov. 2023.

[10] M. Meng et al., "DDT-Net: Dose-agnostic dual-task transfer network for simultaneous low-dose CT denoising and simulation," *IEEE J. Biomed. Health. Inf.*, vol. 28, no. 6, pp. 3613–3625, Jun. 2024.

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

[12] J. Gu and J. C. Ye, "AdaIN-based tunable CycleGAN for efficient unsupervised low-dose CT denoising," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 73–85, Jan. 2021.

[13] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2289–2301, Jul. 2020.

[14] 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, Feb. 2022, Art. no. 4500512.

[15] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, Jun. 2018.

[16] W. Xia, Z. Yang, Z. Lu, Z. Wang, and Y. Zhang, "RegFormer: A local-nonlocal regularization-based model for sparse-view CT reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 8, no. 2, pp. 184–194, Feb. 2024.

[17] D. Hu et al., "PRIOR: Prior-regularized iterative optimization reconstruction for 4D CBCT," *IEEE J. Biomed. Health. Inf.*, vol. 26, no. 11, pp. 5551–5562, Nov. 2022.

[18] 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.

[19] A. Manduca et al., "Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT," *Med. Phys.*, vol. 36, no. 11, pp. 4911–4919, 2009.

[20] M. Balda, J. Hornegger, and B. Heismann, "Ray contribution masks for structure adaptive sinogram filtering," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1228–1239, Jun. 2012.

[21] 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.

[22] X. Dong, S. Vekhande, and G. Cao, "Sinogram interpolation for sparse-view micro-CT with deep learning neural network," in *Proc. SPIE Med. Imag., Phys. Med. Imag.*, 2019, pp. 692–698.

[23] S. Li, W. Ye, and F. Li, "LU-Net: Combining LSTM and U-Net for sinogram synthesis in sparse-view SPECT reconstruction," *Math. Biosci. Eng.*, vol. 19, no. 4, pp. 4320–4340, 2022.

[24] Y. Zhang, K. Yang, Y. Zhu, W. Xia, P. Bao, and J. Zhou, "NOWNUNM: Nonlocal weighted nuclear norm minimization for sparse-sampling CT reconstruction," *IEEE Access*, vol. 6, pp. 73370–73379, 2018.

[25] L. Chen, S. Gou, Y. Yao, J. Bai, L. Jiao, and K. Sheng, "Denoising of low dose CT image with context-based BM3D," in *Proc. IEEE Region 10 Conf. (TENCON)*, 2016, pp. 682–685.

[26] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

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

[28] Y. Weng, T. Zhou, Y. Li, and X. Qiu, "NAS-Unet: Neural architecture search for medical image segmentation," *IEEE Access*, vol. 7, pp. 44247–44257, 2019.

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

[30] M. Geng et al., "Content-noise complementary learning for medical image denoising," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 407–419, Feb. 2022.

[31] D. Bau et al., "Seeing what a GAN cannot generate," in *Proc. ICCV*, 2019, pp. 4502–4511.

[32] Z. Zhang, L. Yu, X. Liang, W. Zhao, and L. Xing, "TransCT: Dual-path transformer for low dose computed tomography," in *Proc. MICCAI*, 2021, pp. 55–64.

[33] D. Wang, Z. Wu, and H. Yu, "TED-Net: Convolution-free T2T vision transformer-based encoder-decoder dilation network for low-dose CT denoising," in *Proc. IWMLMI*, 2021, pp. 416–425.

[34] D. Wang, F. Fan, Z. Wu, R. Liu, F. Wang, and H. Yu, "CTformer: Convolution-free Token2Token dilated vision transformer for low-dose CT denoising," *Phys. Med. Biol.*, vol. 68, no. 6, 2023, Art. no. 65012.

[35] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. ICCV*, 2021, pp. 10012–10022.

[36] A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," *Numer. Math.*, vol. 76, pp. 167–188, Apr. 1997.

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

[38] W. Wu, F. Liu, Y. Zhang, Q. Wang, and H. Yu, "Non-local low-rank cube-based tensor factorization for spectral CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 1079–1093, Apr. 2019.

[39] H. Chen et al., "LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1333–1347, Jun. 2018.

[40] J. He et al., "Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 371–382, Feb. 2019.

[41] W. Xia et al., "MAGIC: Manifold and graph integrative convolutional network for low-dose CT reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3459–3472, Dec. 2021.

[42] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. 34th Conf. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[43] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. ICML*, 2021, pp. 8162–8171.

[44] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. ICLR*, 2021, pp. 1–22. [Online]. Available: https://openreview.net/forum?id=St1giarCHLP

[45] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "ILVR: Conditioning method for denoising diffusion probabilistic models," in *Proc. ICCV*, 2021, pp. 14347–14356.

[46] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 11918–11930.

[47] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. 35th Int. Conf. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 8780–8794.

[48] G. Müller-Franzes et al., "Diffusion probabilistic models beat GANs on medical images," 2022, *arXiv:2212.07501*.

[49] C. Saharia et al., "Palette: Image-to-image diffusion models," in *Proc. ACM SIGGRAPH Conf.*, 2022, pp. 1–10.

[50] W. Xia, Q. Lyu, and G. Wang, "Low-dose CT using denoising diffusion probabilistic model for 20× speedup," 2022, *arXiv:2209.15136*.

[51] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "DPM-solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps," in *Proc. 36th Int. Conf. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 5775–5787.

[52] A. Bansal et al., "Cold diffusion: Inverting arbitrary image transforms without noise," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2023, pp. 41259–41282.

[53] G. Qi, Z. Li, J. Zhang, Y. Zhang, and H. Shan, "CoreDiff: Contextual error-modulated generalized diffusion model for low-dose CT denoising and generalization," *IEEE Trans. Med. Imag.*, vol. 43, no. 2, pp. 745–759, Feb. 2024.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.

[54] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *Proc. ICLR*, 2023, pp. 1–30. [Online]. Available: https://openreview.net/forum?id=OnD9zGAGT0k

[55] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," in *Proc. 36th Conf. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 25683–25696.

[56] H. Chung, S. Lee, and J. C. Ye, "Decomposed diffusion sampler for accelerating large-scale inverse problems," in *Proc. ICLR*, 2024, pp. 1–28. [Online]. Available: https://openreview.net/forum?id=DsEhqQtfAG

[57] 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. CVPR*, 2023, pp. 22542–22551.

[58] Z. Dou and Y. Song, "Diffusion posterior sampling for linear inverse problem solving: A filtering perspective," in *Proc. ICLR*, 2024, pp. 1–30. [Online]. Available: https://openreview.net/forum?id=tplXNcHZs1

[59] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. ICML*, 2015, pp. 2256–2265.

[60] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, pp. 205–229, Apr. 2009.

[61] G. Chen et al., "AirNet: Fused analytical and iterative reconstruction with deep neural network regularization for sparse-data CT," *Med. Phys.*, vol. 47, no. 7, pp. 2916–2930, 2020.

[62] T. R. Moen et al., "Low-dose CT image and projection dataset," *Med. Phys.*, vol. 48, no. 2, pp. 902–911, 2021.

[63] B. De Man and S. Basu, "Distance-driven projection and backprojection," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 3, 2002, pp. 1477–1480.

[64] B. De Man and S. Basu, "Distance-driven projection and backprojection in three dimensions," *Phys. Med. Biol.*, vol. 49, no. 11, p. 2463, 2004.

[65] W. Xia, C. Niu, W. Cong, and G. Wang, "Cube-based 3D denoising diffusion probabilistic model for cone beam computed tomography reconstruction with incomplete data," 2023, *arXiv:2303.12861*.

[66] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018, *arXiv:1803.07422*.

[67] C. He et al., "Diffusion models in low-level vision: A survey," 2024, *arXiv:2406.11138*.

Authorized licensed use limited to: IEEE Xplore. Downloaded on October 21,2025 at 18:02:58 UTC from IEEE Xplore. Restrictions apply.