

# CT-SDM: A Sampling Diffusion Model for Sparse-View CT Reconstruction Across Various Sampling Rates

Liutao Yang<sup>ID</sup>, Jiahao Huang<sup>ID</sup>, Guang Yang,<sup>ID</sup> Senior Member, IEEE,  
and Daoqiang Zhang<sup>ID</sup>, Senior Member, IEEE

**Abstract**—Sparse views X-ray computed tomography has emerged as a contemporary technique to mitigate radiation dose. Because of the reduced number of projection views, traditional reconstruction methods can lead to severe artifacts. Recently, research studies utilizing deep learning methods have made promising progress in removing artifacts for Sparse-View Computed Tomography (SVCT). However, given the limitations on the generalization capability of deep learning models, current methods usually train models on fixed sampling rates, affecting the usability and flexibility of model deployment in real clinical settings. To address this issue, our study proposes an adaptive reconstruction method to achieve high-performance SVCT reconstruction at various sampling rates. Specifically, we design a novel imaging degradation operator in the proposed sampling diffusion model for SVCT (CT-SDM) to simulate the projection process in the sinogram domain. Thus, the CT-SDM can gradually add projection views to highly undersampled measurements to generalize the full-view sinograms. By choosing an appropriate starting point in diffusion inference, the proposed model can recover the full-view sinograms from various sampling rates with only one trained model. Experiments on several datasets have verified the effectiveness and robustness of our approach,

Received 20 January 2025; accepted 6 February 2025. Date of publication 13 February 2025; date of current version 3 June 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62136004 and Grant 62276130; in part by the ERC IMI under Grant 101005122; in part by the H2020 under Grant 952172; in part by the MRC under Grant MC/PC/21013; in part by the Royal Society under Grant IEC\NSFC\211235; in part by the NVIDIA Academic Hardware Grant Program; in part by SABER project through Boehringer Ingelheim Ltd., NIHR Imperial Biomedical Research Centre under Grant RDA01; in part by the Wellcome Leap Dynamic Resilience, UKRI Guarantee funding for Horizon Europe MSCA Postdoctoral Fellowships under Grant EP/Z002206/1; and in part by the UKRI Future Leaders Fellowship under Grant MR/V023799/1. (Corresponding author: Daoqiang Zhang.)

Liutao Yang is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China, and also with the Bioengineering Department and Imperial-X, Imperial College London, W12 7SL London, U.K.

Jiahao Huang is with the Bioengineering Department and Imperial-X, Imperial College London, W12 7SL London, U.K. (e-mail: j.huang21@imperial.ac.uk).

Guang Yang is with the Bioengineering Department and Imperial-X, Imperial College London, W12 7SL London, U.K., also with the National Heart and Lung Institute, Imperial College London, SW7 2AZ London, U.K., also with the National Heart Cardiovascular Research Centre, Royal Brompton Hospital, SW3 6NP London, U.K., and also with the School of Biomedical Engineering and Imaging Sciences, King's College London, WC2R 2LS London, U.K. (e-mail: g.yang@imperial.ac.uk).

Daoqiang Zhang is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: dqzhang@nuaa.edu.cn).

Digital Object Identifier 10.1109/TMI.2025.3541491

demonstrating its superiority in reconstructing high-quality images from sparse-view CT scans across various sampling rates.

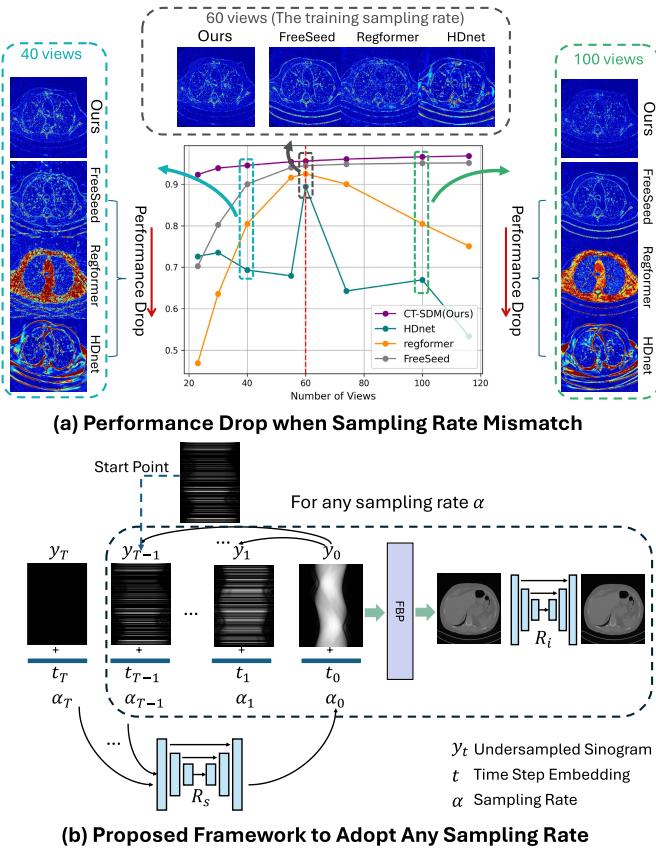
**Index Terms**—Sparse-view CT, sampling rate adaptive reconstruction, deep learning, diffusion model.

## I. INTRODUCTION

IN THE fields of medical imaging, while X-ray computed tomography (CT) is widely used for its ability to generate high-quality and detailed images, it also raises a significant concern regarding potential cancer risks attributed to radiation exposure [1]. Sparse-view CT (SVCT) reconstruct images from a reduced number of projection views, can effectively control the radiation dose under a reasonable threshold. Furthermore, sparse-view CT can intriguing more possibilities in emerging fields like spectral CT, where innovative techniques such as alternating kVp switching [2] and dynamic beam blocker [3] are employed. Moreover, in applications such as C-arm CT or dental CT, where scanning time is predominantly determined by the slower flat-panel detector, sparse-view CT emerges as a promising solution for speeding the scanning process [4].

However, while the number of projection views is reduced during scanning, sever artifacts would appear in reconstructed images which can strongly undermine clinical diagnosis. To improve the image quality of SVCT, model based methods are proposed in the early researches. Tian et al. [5] proposed EPTV (edge-preserving TV norm), which enhances edge information by incorporating adaptive weights. Yang et al. developed a high-order total variation (TV) minimization algorithm. Wang et al. [6] introduced an iterative reweighted anisotropic total variation (RwATV) method. As the developments of the deep learning methods in recent years, deep reconstruction methods have achieved promising performance for sparse-view CT imaging. Some of these methods [7], [8], [9], [10] train an image model to remove artifacts from reconstructed SVCT images, while others unfold the iterative reconstruction [11], [12], [13] into a deep network and learn the reconstruction parameters in end-to-end training.

However, the prevailing trend among deep learning reconstruction methods is their training and testing within specific sampling rate configurations, often overlooking the model's generalization performance across diverse sampling rates. This oversight is exemplified in Fig 1(a), where deep learning-based



**Fig. 1.** (a) DL based SVCT reconstructions are face performance drop when train and test on different sampling rates (i.e., sampling mismatch). (b) The proposed method replace the image degradation in forward process of diffusion model as the projection view sampling. Choosing a certain sampling rate as a start point at inference, the proposed model can reconstruct SVCT images under various sampling rates.

SVCT reconstructions manifest performance degradation when confronted with sampling rate mismatches, presenting a significant challenge for SVCT deployment in clinical settings. Given the paramount importance of CT imaging as a diagnostic tool, it is imperative for it to exhibit flexibility across various examination requirements. Yet, accommodating this versatility by training and storing model parameters across all requisite setups is rendered impractical due to prohibitive training times and memory overheads.

For this reason, this paper introduces a sampling-rate-adaptive method based on diffusion models, which employs a single model capable of accommodating various sampling rates. Specifically, the proposed approach substitutes the conventional degradation process in the forward step of the diffusion model (i.e., usually the Gaussian noise) with the projection view sampling. Subsequently, the inference stage (i.e., the reverse process) of the diffusion model is designed to gradually recover full-sampled data. As shown in 1(b), by selecting a specific sampling rate as the starting point during inference, the proposed model can effectively reconstruct SVCT images across various sampling rates.

Moreover, challenge also lies in effectively allocating the sampling rate and specific projection views to each iteration step. This task is critical for achieving optimal reconstruction performance across various sampling rates. To address this challenge, we introduce a grouped-random sampling strategy.

This strategy is designed to dynamically adjust the sampling rate and select projection views during the training process. It divides all sampling views into orderly and equally spaced groups, facilitating comprehensive and partially random angle coverage. By sequentially selecting whole groups of sampling views and resorting to random selection within groups when necessary, our strategy ensures both diversity and uniformity in views selection. This not only mitigates bias but also provides valuable data augmentation during training, promising more robust and accurate results across diverse sampling scenarios.

Our contributions can be summarised as follows:

- We propose a sampling-rate-adaptive method based on diffusion models, which employs a single model capable of accommodating various sampling rates.
- We introduce a novel grouped-random sampling strategy for dynamically adjusting the sampling rate and selecting projection views during training, addressing biases and providing valuable data augmentation.
- We demonstrate the effectiveness of our approach through comprehensive experimentation, showcasing improved reconstruction performance across diverse sampling scenarios in sparse-view CT reconstruction.

## II. RELATED WORKS

### A. Deep Learning Based CT Reconstruction

The exploration of CT reconstruction has been significantly enriched by the deployment of deep learning strategies, as evidenced in a variety of studies [8], [14], [15], [16], [17], [18]. These approaches can be divided into two categories: image domain enhancement and dual-domain reconstruction. The former leverages denoising techniques from low-level vision to develop networks aimed at artifact reduction and image clarification. Jin et al. [14] proposed a CNN framework inspired by the U-net architecture, designed to refine images reconstructed via FBP (Filtered Back Projection) through the use of CNNs. Similarly, Han et al. [7] target the restoration of high-frequency edges in sparse-view CT, employing wavelet transforms within deep learning models to effectively diminish artifacts. Chen et al. [19] introduced a residual encoder-decoder CNN (REDCNN) tailored for low-dose scenarios.

On the flip side, dual-domain reconstruction methods utilize data from both the sinogram and image domains to overcome the over smoothing tendencies of CNN architectures. Hu et al. [15] proposed the HDNet, a hybrid-domain network that deconstructs the SVCT reconstruction problem into simpler, sequential tasks. Yang et al. [20] conceived the Sinogram Inner-Structure Transformer, capitalizing on the inherent structures within the sinogram domain to mitigate noise in low-dose CT (LDCT) images. Additionally, AUTOMAP [21] and IRadonMap [22] both present convolutional network-based solutions that facilitate the translation of measurement data directly into image representations.

### B. Diffusion Model in Image Reconstruction

To date, diffusion models have been found to be highly powerful across various domains, ranging from generative modeling tasks such as image generation [23], image super-resolution [24], and image inpainting [25] to discriminative

tasks such as image segmentation [26], classification [27], and anomaly detection [28]. Recently, in medical imaging, researchers have also proposed diffusion model-based methods. Kim et al. [29] proposed DiffuseMorph, a diffusion-based method for medical image registration. DiffuseMorph integrates the diffusion network and the deformation network into an end-to-end training framework. The diffusion network scores the deformations between the moving and fixed images, while the deformation network uses this information to estimate the deformation field. Packhäuser et al. [30] utilized a latent variable diffusion model [31] to generate high-quality class-conditional chest X-rays, preserving the privacy of sensitive biological features while generating samples.

Meng et al. [32] proposed a unified multi-modal conditional generation method (UMM-CSGM) to generate missing modalities in multi-modal data. This method uses all available modalities as conditions to synthesize the missing modalities, utilizing a conditional model based on SDE [33] to learn cross-modal conditional distributions. Chung et al. [34] applied score-based diffusion models to solve the inverse problem of image reconstruction in fast MRI scans. This method first uses denoising score matching to train a continuously time-varying score function on magnitude images; then, in the reverse SDE process, it employs the pre-trained score model and uses the VE-SDE method [35], taking undersampled data as conditional input to generate reconstructed images. Liu et al. [36] used the DDPM model in the limited-angle CT reconstruction problem and proposed the DOLCE method. This method takes the FBP reconstructed images from limited-angle sinograms as prior information to modulate the diffusion model and uses the consistency condition of the sinogram with an  $l_2$  norm loss in the iterative steps of inference to ensure consistency.

### III. METHODOLOGY

#### A. Problem Formulation

The imaging of X-ray CT can be viewed as a Radon transform projecting the CT images  $x$  into measurements  $y$  at different imaging angles:

$$y = Ax + \eta, \quad (1)$$

where  $x \in \mathbb{R}^n$ ,  $n = w \times h$ , and  $w, h$  are the width and height of  $x$ .  $A \in \mathbb{R}^{m \times n}$  is the Radon transform matrix, where  $m = v \times d$ ,  $v$  is the number of projection views, and  $d$  is the number of detectors.  $\eta$  is the noise. Thus, CT reconstruction is to solve  $x$  from the measurements  $y$ . With sufficient sampling rate (i.e., sufficient projection views  $v$ ) and high Signal-Noise Ratio (i.e., low noise  $\eta$ ), image  $x$  can be reconstructed with good quality for clinical usage. However, more projection views cause higher radiation dose which raise the potential cancer risks. Thus, Sparse-View CT technique is trying to reconstruct images from undersampled measurements:

$$y_s = A_s x + \eta_s, \quad (2)$$

where  $A_s$  represents the Radon transform matrix with a reduced number of projections  $v_s < v$ ,  $\eta_s$  is the noise under sparse view sampling.

#### B. Cold Diffusion Model

The Cold Diffusion Model (CDM) [37] represents a significant extension of traditional diffusion models, which are fundamentally based on the interplay between degradation and restoration processes of images. At its core, the standard diffusion model employs an initial process degrades images by introducing Gaussian noise, followed by a denoising process developed through rigorous training for image restoration. This iterative application of degradation and restoration mechanisms facilitates the generation of images and achieves promising performance.

In contrast to the conventional diffusion models that limit the degradation process to the Gaussian noise, the CDM extends the degradation operator to include a wide variety of transformations such as blurring, animorphing, and masking, thus accommodating a broader spectrum of degradation effects. This approach allows for a more diverse and realistic representation of image degradation.

Consider an image  $x_0 \in \mathbb{R}^N$ . The degradation of  $x_0$  by an operator  $D$ , under sampling step  $t$ , is expressed as:

$$x_t = D(x_0, t), \quad (3)$$

where  $D(x_0, t)$  is expected to vary continuously with  $t$ , ensuring that  $D(x_0, 0) = x_0$ .

To recover images from a degraded condition, the CDM employs a restoration operator  $R$  that aims to approximately reverse the effects of  $D$ . The operator  $R$  seeks to fulfill the condition:

$$R(x_t, t) \approx x_0. \quad (4)$$

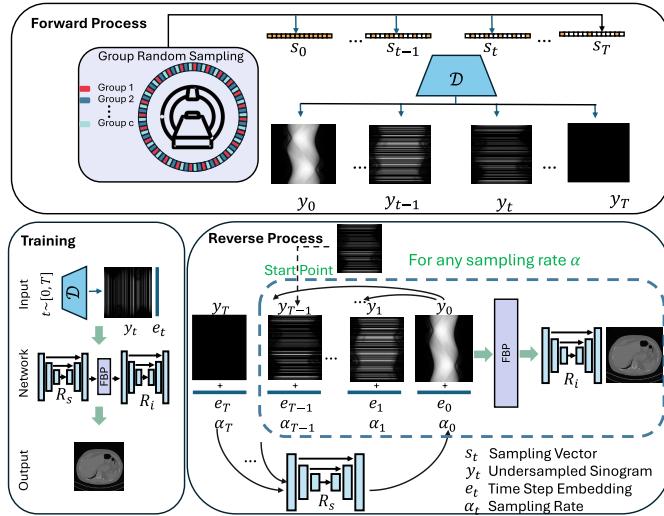
In practice, this restoration function is implemented through a neural network parameterized by  $\theta$ , which is optimized to minimize the difference between the degraded and original images, represented as:

$$\min_{\theta} \|R_{\theta}(x_t, t) - x_0\|. \quad (5)$$

This generalized framework of image degradation and restoration forms the foundation of the Cold Diffusion Model, providing a versatile and powerful tool for low-level vision tasks such as deblurring, inpainting and super-resolution.

#### C. Sampling Diffusion Model for CT Reconstruction

Following the principles of the Cold Diffusion Model, we conceptualize the sampling processes of CT imaging as a category of image degradation operators, namely Sampling Diffusion Model for SVCT (CT-SDM). In CT-SDM, both the forward and reverse process are designed in sinogram domain to allow the simulation of imaging data sampling. This approach allows us to endow the degradation operator with practical physical meanings that directly correspond to data sampling practices in CT imaging. Consequently, each step of the Sampling Diffusion Model can be analogously mapped to a specific sampling rate within the sparse-view CT imaging, thus bridging a methodological link between image degradation and X-ray projections. The overall architecture are shown in Fig 2.



**Fig. 2.** The overall framework of the proposed methods. The forward process of sampling diffusion model is to determine the undersampled CT measurements (i.e., sinograms) at each sampling step  $t$ . And a sampling step  $t$  is corresponded to a specific sampling rate  $\alpha_t$ . The reverse process is aims to recover the full-view sinogram  $y_0$  from the measurements  $y_T$  obtained at a certain start sampling rate  $\alpha_T$ , utilizing the networks trained for undersampled data recovering. Both the forward and reverse process diffusion are designed in sinogram domain to allow the simulation of accoutre data sampling.

**1) Degradation Operator of Sampling Diffusion:** The forward process of Sampling Diffusion Model is to determine the undersampled CT measurements (i.e., sinograms) at each sampling step  $t$ . And a sampling step  $t$  is corresponded to a specific sampling rate  $\alpha_t$ , which represents the number of sampled views during CT imaging. Given the Full-sampled CT image  $x$ , the degradation operator of forward process can be expressed as:

$$y_t = D(y, t) = A_{\alpha_t}x + \eta_t, \quad t = 0, 1, \dots, T, \quad (6)$$

where,  $y_t$  signifies the measurements obtained under the sampling step  $t$ ,  $\alpha_t$  is the sampling rate corresponding to the step  $t$ , and  $A_{\alpha_t} \in \mathbb{R}^{m_{\alpha_t} \times n}$  represents the Radon transform adapted to the sparse-view setup characterized by the reduced number of projections  $v_{\alpha_t} < v$ . The term  $\eta_t$  denotes the noise associated with the sparse-view imaging process at the corresponding sampling rate.

By defining the forward diffusion process in such a manner, we not only adhere to the Cold Diffusion Model's theoretical framework but also directly align it with practical imaging scenarios in CT, particularly emphasizing the importance of managing radiation exposure through the adjustment of sampling rates.

**2) Restoration Operator of Sampling Diffusion:** The reverse process is aims to recover the full-view sinogram  $y_0$  from the measurements  $y_t$  obtained at a certain start sampling rate  $\alpha_t$ . Thus, a restoration operator is designed to reverse the effects of degradation, in this case, the under-sampling inherent in sparse-view CT imaging. A CNN  $R_s$  (Detailed in Section. III-E) is defined as the restoration operator, which takes a undersampled sinogram  $y_t$  at sampling step  $t$  as input,

### Algorithm 1 TACoS

```

1: Input: undersampled sinogram  $y_t$ 
2: for  $s = t, t - 1, \dots, 1$  do
3:    $\hat{y}_0 \leftarrow R_s(y_s, s)$ 
4:   if  $s > 1$  then
5:      $\hat{y}_{s-1} \leftarrow y_s - D(\hat{y}_0, s) + D(\hat{y}_0, s - 1)$ 
6:   end if
7: end for

```

to estimate the full-view sinogram:

$$\hat{y}_0 = R_s(y_t, t), \quad (7)$$

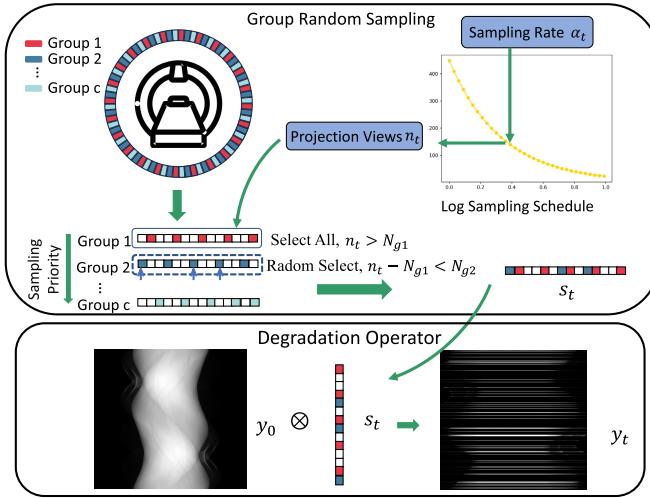
where  $\hat{y}_0$  is a coarse estimation of full-view sinogram. In the reverse process, we take a iterative approach to generally make  $\hat{y}_0$  to approximate  $y_0$ . Starting form a very low sampling rate  $\alpha_T$ , the restoration operator first recover a coarse estimation  $\hat{y}_0$ . As illustrated in Fig. 2, we introduce a time step embedding  $e_t$  to encode the sampling rate information into the network. Following the approach used in Denoising Diffusion Probabilistic Models (DDPM) [38], the time step  $t$  is first transformed into a vector using sinusoidal embeddings, which capture periodic information. This vector is then processed through a Multi-Layer Perceptron (MLP) layer to generate the final embedding.

Then, iterative reverse process is conducted in sinogram domain. Since each step of recovering is not perfect, we use the Transformation Agnostic Cold Sampling (TACoS) [37], [39] during inference to avoid noise accumulation. The TACoS is shown in Algorithm 1. Then, the final estimated  $\hat{y}_0$  is then reconstructed into images and refined by the image domain network  $R_i$ .

### D. Grouped-Random Sampling Strategy

To effectively allocate the sampling rate  $\alpha_t$  and specific projection views corresponding to each iteration step  $t$ , we introduces a grouped-random sampling strategy as a dynamic training method. Initially, we gradually backtrack the reverse process by setting the mapping of the sampling rate and iteration steps exponentially [39] to adjust the variation in sampling rate, as shown in Figure 3. Since the process of restoring images  $\hat{y}_0$  at low sampling rates is relatively more complex, the exponential relationship causes a slower increase in the sampling rate at lower values and a faster increase at higher values.

Subsequently, the grouped-random sampling method divides all sampling angles into  $c$  orderly and equally spaced groups  $g_1, g_2, \dots, g_c$ . Let  $Y = [y_1, y_2, y_3, \dots, y_v]$ , then  $g_1 = [y_1, y_c, y_{2c}, \dots, y_{\lfloor \frac{v}{c} \rfloor}]$ ,  $g_2 = [y_2, y_{c+1}, y_{2c+1}, \dots, y_{\lfloor \frac{v}{c} \rfloor+1}]$ , and so on.  $\lfloor \cdot \rfloor$  denotes the floor function, which rounds down to the nearest integer. The number of projections is determined according to the target sampling rate of the given iteration step  $t$ . Whole groups of sampling views are selected sequentially from  $g_1, g_2, \dots, g_c$ . When the remaining unselected views are fewer than the size of a full group, views are randomly selected from within the current group. This procedure ensures comprehensive and partially random angle coverage, preventing bias from uniform or deterministic selection processes and



**Fig. 3.** The group-random sampling schedule. The grouped-random sampling method divides all sampling angles into  $c$  orderly and equally spaced groups to ensure comprehensive and partially random angle coverage.

providing data augmentation for the training process. This process can be described by the following algorithm:

#### Algorithm 2 Grouped-Random Sampling Strategy

```

1: Input: Number of sampling views  $k$ , fully sampling views  $v$ , number of group  $c$ 
2: for  $s = 1, 2, \dots, c$  do
3:   if  $s \times \lfloor \frac{v}{c} \rfloor \leq k$  then
4:     Select the entire group  $g_s$ 
5:   else
6:     Randomly select  $k - (s - 1) \times \lfloor \frac{v}{c} \rfloor$  projection views
        within group  $g_s$ 
7:   BREAK
8:   end if
9: end for
```

Here, the number of projections  $k$  is determined by the sampling rate  $\alpha_t$ ,  $k = \alpha_t \times v$ .

The main advantage of the grouped-random sampling strategy lies in its ability to simulate the variability of available projection views in real-world scenarios, which is common in clinical settings due to patient movement, equipment limitations, or specific diagnostic needs. By introducing randomness within a structured grouping framework, the model extensively simulates potential imaging conditions, thereby enhancing its adaptability and performance in various sparse-view CT configurations. Additionally, this strategy serves as a powerful form of data augmentation, expanding the diversity of training data without the need for additional physical scans. This augmentation is crucial for training deep learning models to accurately reconstruct images from sparsely sampled sinogram data, effectively improving reconstruction accuracy and reducing artifacts in the final images.

#### E. Network Training

The proposed method contains two CNN to handle the SVCT reconstruction task in both the sinogram domain and

the image domain. The network in sinogram domain is the restoration operator  $R_s$  to estimate the full-view sinogram  $\hat{y}_0$  from the input  $y_t$ . Thus, the network parameters  $s$  are optimized by minimizing the loss function:

$$\begin{aligned} L_s &= \|R_s(y_t) - y\|_1, \\ &= \|R_s(D(y, t), t) - y\|_1, \quad t = 1, 2, \dots, T. \end{aligned} \quad (8)$$

The estimated full-view sinogram  $\hat{y}_0$  is then reconstructed into images and a image domain network  $R_i$  is used for image domain refinements. The loss function can be defined as follows:

$$L_i = \|R_i(A^\dagger \hat{y}_0) - x\|_1, \quad (9)$$

where  $A^\dagger$  is the pseudo-inversion of Radon transform  $A$ , which can be implemented by Filtered Back-Projection (FBP) [13]. These two networks are jointly trained together:

$$L = L_s + \lambda L_i, \quad (10)$$

where  $\lambda$  is the control the ratio of two parts.

## IV. EXPERIMENTS

### A. Datasets

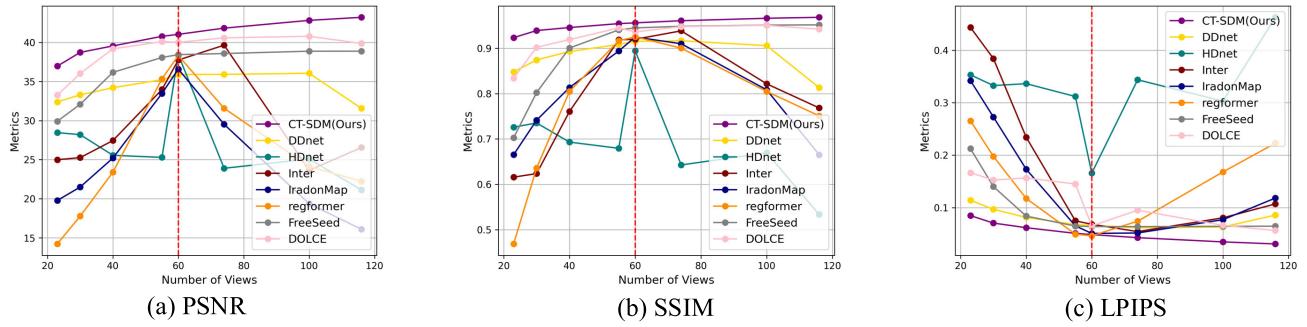
To validate the performance of the proposed methods, *Low-Dose CT Image and Projection Datasets* (LDCT Datasets) [40] and the *TCGA-KIRC* [41], [42] dataset are used.

The LDCT Dataset consists of CT scans from three common exam types: non-contrast head CT scans acquired for acute cognitive or motor deficit, low-dose non-contrast chest scans acquired to screen high-risk patients for pulmonary nodules, and contrast-enhanced CT scans of the abdomen acquired to look for metastatic liver lesions. It contains 25,141 CT images in total from 150 patients. 120 patients are randomly selected into the training set and remains are for testing.

The TCGA-KIRC dataset consists both CT and MRI of scans for Kidney Renal Clear Cell Carcinoma. In our experiments, we use 15153 CT images from 50 participants to generate sinograms with the same configuration as LDCT Datasets. 40 patients are randomly split into the training set and remains are for testing.

### B. Implementation Details

In this paper, we use an attention-based UNet with time embedding module [35] for the restoration operator. In image domain, we use one ResNet Block [43] implemented in original CDM [37] for a light weight image refinements. As radiation dose of each view is not reduced in our experimental settings, the noise  $\eta_s$  is small and we ignore it for simplification. During training,  $\lambda$  is set to 1. The number of groups  $c$  in Grouped-Random Sampling Strategy is set to 5 for all the experiments. We also discussed the influence of  $c$  in ablation studies. The Torch-Radon [44] toolbox is used for sinogram generation. We use a fan-beam CT geometry with 448 projection views and 736 detectors as the full view sampling setup. The source-to-center distance was set to 1000 mm, and the detector-to-center distance was 512 mm. The reconstructed image resolution is set to  $256 \times 256$  pixels.



**Fig. 4.** Performance comparison of various methods at different sampling rates on the LDCT dataset. Each method is trained under 60 views(red line), and the results are evaluated on test data with varying numbers of projections: 116, 100, 74, 60, 55, 40, 30, and 23.

### C. Evaluation Metrics

For the quantitative evaluation, we employ the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) [45] as metrics to gauge the quality of reconstruction. It's worth mentioning that LPIPS, with its foundation in deep learning-based perceptual metrics, mirrors human visual perception closely, offering a detailed assessment of the fidelity in image reconstruction.

### D. Performance Under Different Sampling Rates

To validate the generalization ability of the proposed method on data with different sampling rates, this section compares seven methods: DDNet [46], View-Inter [16], HDNet [15], IradonMap [22], RegFormer [13], FreeSeed [47], and DOLCE [36].

DDNet [46] is a representative image-domain method aimed at removing artifacts and improving image quality from FBP reconstructed images. View-Inter [16] is a sinogram-domain method that completes the full projection angle sinogram only in the sinogram domain. HDNet [15] is a typical dual-domain method that interpolates sparse view sinogram data and uses two CNNs to refine image quality. IradonMap [22] is essentially a learnable Iradon transform combined with an image-domain CNN. RegFormer [13] is a deep unfolding reconstruction method combined with Swin-Transformer [48]. FreeSeed [47] is an image-domain post-processing method that combines frequency-domain adaptive bandpass filtering with an image-domain reconstruction network. DOLCE [36] is a diffusion model based method using iterative correction while reverse process.

During the training of the comparative methods, the number of projection is set to 60 (i.e., sampling rate of  $60/448 = 13.4\%$ ) was selected to generate the corresponding sparse-view sinograms for network training. The proposed method, using SDM to gradually add projection angles, does not require setting a sampling rate during training. Instead, it can achieve reconstruction at the corresponding sampling rate during testing by selecting the diffusion starting point corresponding to the desired sampling rate.

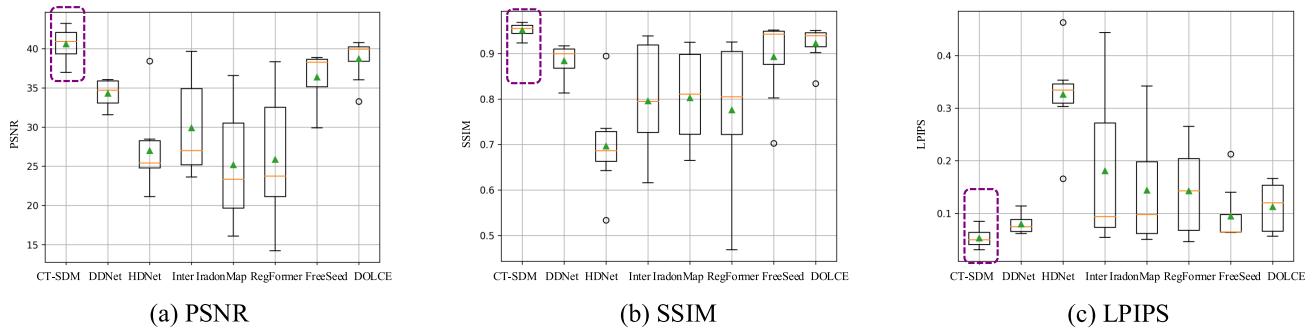
During testing, the performance of the methods was evaluated on test images by generating sinograms with projection

numbers of 116, 100, 74, 60, 55, 40, 30, and 23, respectively, to test their performance at different sampling rates.

**1) Experimental Results on the LDCT Dataset:** The experimental results are shown in Figure 4. Through these experiments, we observed that while deep learning methods often exhibit excellent performance at the sampling rates set during training, their limited network generalization capability makes it difficult to maintain performance at non-training sampling rates. Particularly for dual-domain methods like RegFormer and HDNet, these algorithms perform excellently at the training sampling rate. However, errors in the sinogram domain tend to accumulate during the image reconstruction process, leading to a sharp decline in performance once the sampling rate deviates from the training setting. In contrast, pure image-domain methods exhibit better generalization ability, but their performance still noticeably degrades in experiments. Notably, increasing the sampling rate sometimes resulted in a decline in reconstruction quality, further highlighting the significant limitations of deep learning methods in terms of generalization. Diffusion model method (i.e. DOLCE) shows better performance when sampling rate changes. However, there still a performance degradation when changes are large (e.g. at 116 views and 30 views).

The proposed SDM reduces the reliance on model parameters trained at specific sampling rates, significantly enhancing the model's generalization ability. This improvement enables the model to meet the image reconstruction requirements at different sampling rates, greatly expanding its applicability and practicality. As shown in Figure 5, box plots of three key performance metrics are presented for various methods at different sampling rates. The figure clearly demonstrates that the proposed method not only maintains high-quality reconstructed images under varying sampling rates but also exhibits excellent stability. In contrast, other methods like HDNet and Inter, although achieving high peak performance at specific sampling rates, exhibit large performance fluctuations between different sampling rates, indicating a lack of stability. Methods like FreeSeed and DOLCE, while showing better stability, still fall short of the overall performance achieved by the proposed method.

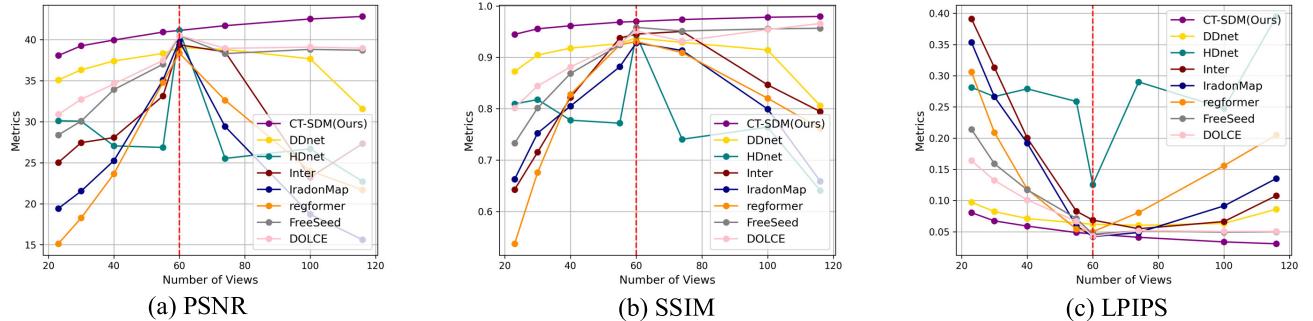
**2) Experimental Results on the TCGA-KIRC Dataset:** The experimental results on the TCGA-KIRC dataset are shown in Figure 6. The performance trends of each method with



**Fig. 5.** Box plots showing the performance of various methods at different sampling rates on the LDCT dataset. Each method is trained under 60 projections, and the results are evaluated on test data with varying numbers of projections: 116, 100, 74, 60, 55, 40, 30, and 23.

**TABLE I**  
MEAN AND STANDARD DEVIATION OF RECONSTRUCTION PERFORMANCE OF DIFFERENT METHODS AT VARIOUS SAMPLING RATES ON THE LDCT DATASET AND TCGA-KIRC DATASET. THE BEST RESULTS ARE HIGHLIGHTED, AND THE SECOND RESULTS ARE UNDERLINED. ↓ (↑) INDICATES THAT LOWER (HIGHER) IS BETTER. \* DENOTES RESULTS THAT ARE SIGNIFICANTLY DIFFERENT FROM THE BEST RESULTS BY THE WILCOXON TEST ( $p < 0.05$ )

Methods	LDCT(Averaged over Sampling Rates)			TCGA-KIRC(Averaged over Sampling Rates)		
	PSNR(↑)	SSIM(↑)	LPIPS(↓)	PSNR(↑)	SSIM(↑)	LPIPS(↓)
DDnet	$34.314 \pm 1.630^*$	$0.884 \pm 0.035^*$	$0.079 \pm 0.018^*$	$36.783 \pm 2.321^*$	$0.901 \pm 0.041^*$	$0.073 \pm 0.013^*$
Inter	$29.909 \pm 5.864^*$	$0.796 \pm 0.120^*$	$0.181 \pm 0.145^*$	$30.265 \pm 5.682^*$	$0.832 \pm 0.106^*$	$0.161 \pm 0.120^*$
HDnet	$27.007 \pm 4.834^*$	$0.697 \pm 0.095^*$	$0.326 \pm 0.076^*$	$28.776 \pm 5.179^*$	$0.783 \pm 0.079^*$	$0.268 \pm 0.069^*$
iRadonMap	$25.192 \pm 6.862^*$	$0.803 \pm 0.098^*$	$0.144 \pm 0.103^*$	$25.611 \pm 7.983^*$	$0.800 \pm 0.098^*$	$0.149 \pm 0.106^*$
RegFormer	$25.875 \pm 7.910^*$	$0.776 \pm 0.147^*$	$0.143 \pm 0.070^*$	$26.062 \pm 7.750^*$	$0.798 \pm 0.128^*$	$0.147 \pm 0.084^*$
FreeSeed	$36.376 \pm 3.253^*$	$0.893 \pm 0.086^*$	$0.095 \pm 0.051^*$	$35.718 \pm 4.162^*$	$0.894 \pm 0.080^*$	$0.095 \pm 0.059^*$
DOLCE	$38.724 \pm 2.499^*$	$0.922 \pm 0.037^*$	$0.113 \pm 0.044^*$	$36.658 \pm 3.262^*$	$0.907 \pm 0.055^*$	$0.082 \pm 0.042^*$
CT-SDM(Ours)	<b><math>40.612 \pm 1.978</math></b>	<b><math>0.952 \pm 0.014</math></b>	<b><math>0.053 \pm 0.017</math></b>	<b><math>40.805 \pm 1.526</math></b>	<b><math>0.966 \pm 0.011</math></b>	<b><math>0.051 \pm 0.016</math></b>



**Fig. 6.** Performance comparison of various methods at different sampling rates on the TCGA-KIRC dataset. Each method is trained with 60 projections, and the results are evaluated on test data with varying numbers of projections: 116, 100, 74, 60, 55, 40, 30, and 23.

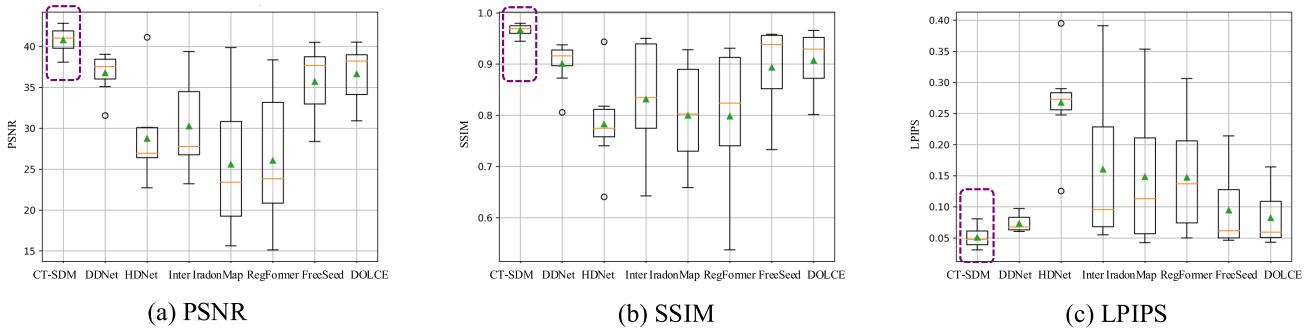
varying sampling rates are generally consistent with those observed on the LDCT dataset. Dual-domain methods exhibit significant performance degradation regardless of whether the sampling rate decreases or increases, while single-domain methods experience more severe degradation at low sampling rates. Figure 7 shows the box plots of performance metrics for various methods at different sampling rates on the TCGA-KIRC dataset. With The proposed method still demonstrates superior stability and higher average performance compared to other methods.

#### E. Performance at the Training Sampling Rate

Table II presents the reconstruction results of various methods tested at the model's set sampling rate (60 projections and 100 projections). At 60 projection views, dual-domain

methods such as HDNet, iRadonMap, RegFormer, and DOLCE exhibit superior reconstruction performance compared to single-domain methods like image-domain DDNet and projection-domain Inter. FreeSeed, which can be considered a post-processing method since it does not use original sinogram data, achieves good performance by performing joint processing in the frequency domain. We also train the comparison method on 100 projection views. As result shows, the comparison methods work well when sampling rate are matched. However, this would require large amount of computational resources and storage for training different models to ensure those methods work.

The proposed method not only ensures that a single model performs well across various sampling rates, but also has no compromise on performance at a single sampling rate, achieving the comparable performance across multiple metrics.



**Fig. 7.** Box plots showing the performance of various methods at different sampling rates on the TCGA-KIRC dataset. Each method is trained with 60 projections, and the results are evaluated on test data with varying numbers of projections: 116, 100, 74, 60, 55, 40, 30, and 23.

TABLE II

QUANTITATIVE RESULTS OF DIFFERENT METHODS AT A SINGLE SAMPLING RATE (60 PROJECTION VIEWS) ON THE LDCT DATASET AND TCGA-KIRC DATASET. THE BEST RESULTS ARE HIGHLIGHTED, AND THE SECOND RESULTS ARE UNDERLINED. ↓ (↑) INDICATES THAT LOWER (HIGHER) IS BETTER. \* DENOTES RESULTS THAT ARE SIGNIFICANTLY DIFFERENT FROM THE OURS RESULTS BY THE WILCOXON TEST ( $p < 0.05$ )

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	PSNR(↑)	SSIM(↑)	LPIPS(↓)	
trained on LDCT dataset (60 Projection Views)				trained on TCGA-KIRC dataset (60 Projection Views)			
DDNet	$35.890 \pm 1.634^*$	$0.915 \pm 0.018^*$	$0.066 \pm 0.018^*$	$39.046 \pm 1.876^*$	$0.938 \pm 0.014^*$	$0.062 \pm 0.021^*$	
Inter	$37.746 \pm 1.698^*$	$0.920 \pm 0.020^*$	$0.068 \pm 0.021^*$	$39.385 \pm 1.722^*$	$0.945 \pm 0.015^*$	$0.069 \pm 0.022^*$	
HDNet	$38.426 \pm 3.313^*$	$0.894 \pm 0.062^*$	$0.166 \pm 0.088^*$	$40.147 \pm 2.484^*$	$0.944 \pm 0.041^*$	$0.126 \pm 0.053^*$	
IradonMap	$36.587 \pm 1.582^*$	$0.925 \pm 0.017^*$	$0.051 \pm 0.014^*$	$39.866 \pm 1.766^*$	$0.928 \pm 0.015^*$	$0.049 \pm 0.019^*$	
RegFormer	$38.316 \pm 2.050^*$	$0.925 \pm 0.024^*$	$\underline{\underline{0.046 \pm 0.020}}$	$38.374 \pm 1.641^*$	$0.931 \pm 0.021^*$	$0.050 \pm 0.019^*$	
FreeSeed	$38.439 \pm 2.228^*$	$0.945 \pm 0.024^*$	$0.064 \pm 0.027^*$	$40.494 \pm 1.969^*$	$0.958 \pm 0.013^*$	$\underline{\underline{0.046 \pm 0.014}}$	
DOLCE	$40.012 \pm 1.303^*$	$0.936 \pm 0.009^*$	$0.065 \pm 0.016^*$	$40.532 \pm 1.815^*$	$0.951 \pm 0.014^*$	$0.049 \pm 0.016^*$	
CT-SDM (Ours)	$\underline{\underline{41.032 \pm 2.907}}$	$\underline{\underline{0.956 \pm 0.023}}$	$0.049 \pm 0.023$	$\underline{\underline{41.130 \pm 1.769}}$	$\underline{\underline{0.970 \pm 0.015}}$	$0.047 \pm 0.020$	
trained on LDCT dataset (100 Projection Views)				trained on TCGA-KIRC dataset (100 Projection Views)			
DDNet	$37.801 \pm 1.748^*$	$0.925 \pm 0.019^*$	$0.054 \pm 0.018^*$	$40.420 \pm 1.721^*$	$0.946 \pm 0.014^*$	$0.050 \pm 0.023^*$	
Inter	$39.697 \pm 1.846^*$	$0.928 \pm 0.020^*$	$0.055 \pm 0.020^*$	$40.877 \pm 1.865^*$	$0.952 \pm 0.014$	$0.057 \pm 0.021^*$	
IradonMap	$38.244 \pm 1.494^*$	$0.936 \pm 0.016^*$	$0.039 \pm 0.013^*$	$41.392 \pm 1.758^*$	$0.937 \pm 0.016^*$	$0.039 \pm 0.017^*$	
HDNet	$40.195 \pm 2.994^*$	$0.903 \pm 0.056^*$	$0.154 \pm 0.095^*$	$42.655 \pm 2.319^*$	$0.951 \pm 0.040^*$	$0.112 \pm 0.050^*$	
regformer	$39.748 \pm 1.931^*$	$0.936 \pm 0.026^*$	$\underline{\underline{0.035 \pm 0.021}}$	$39.607 \pm 1.701^*$	$0.939 \pm 0.021^*$	$0.036 \pm 0.020^*$	
FreeSeed	$40.107 \pm 2.412^*$	$0.956 \pm 0.021^*$	$0.052 \pm 0.028^*$	$41.698 \pm 1.811^*$	$0.966 \pm 0.013^*$	$0.033 \pm 0.015^*$	
DOLCE	$41.486 \pm 1.289^*$	$0.945 \pm 0.019^*$	$0.050 \pm 0.016^*$	$41.924 \pm 1.942^*$	$0.958 \pm 0.013^*$	$\underline{\underline{0.031 \pm 0.016^*}}$	
CT-SDM(Ours)	$\underline{\underline{42.819 \pm 3.277}}$	$\underline{\underline{0.966 \pm 0.020}}$	$\underline{\underline{0.035 \pm 0.019}}$	$\underline{\underline{42.532 \pm 1.758}}$	$\underline{\underline{0.978 \pm 0.012}}$	$0.034 \pm 0.015$	

Figures 8 present visualized results and error maps on the LDCT dataset. We first compare the reconstructed images at the same sampling rate used for training. Although our method does not specify a particular sampling rate during training, it still achieves better or comparable performance to state-of-the-art methods that train and test at a fixed sampling rate. When the sampling rates are varied, all comparison methods exhibit a significant performance drop, especially the dual-domain methods (i.e., IradonMap, HDNet, and RegFormer). In contrast, the proposed method maintains good performance, demonstrating robustness and stability across different sampling rates. Further visualization of reconstruction results on lesion slices at 100 views is provided in Fig. 14, with additional details in Section IV-F2.

#### F. Performance on Clinically Relevant Task

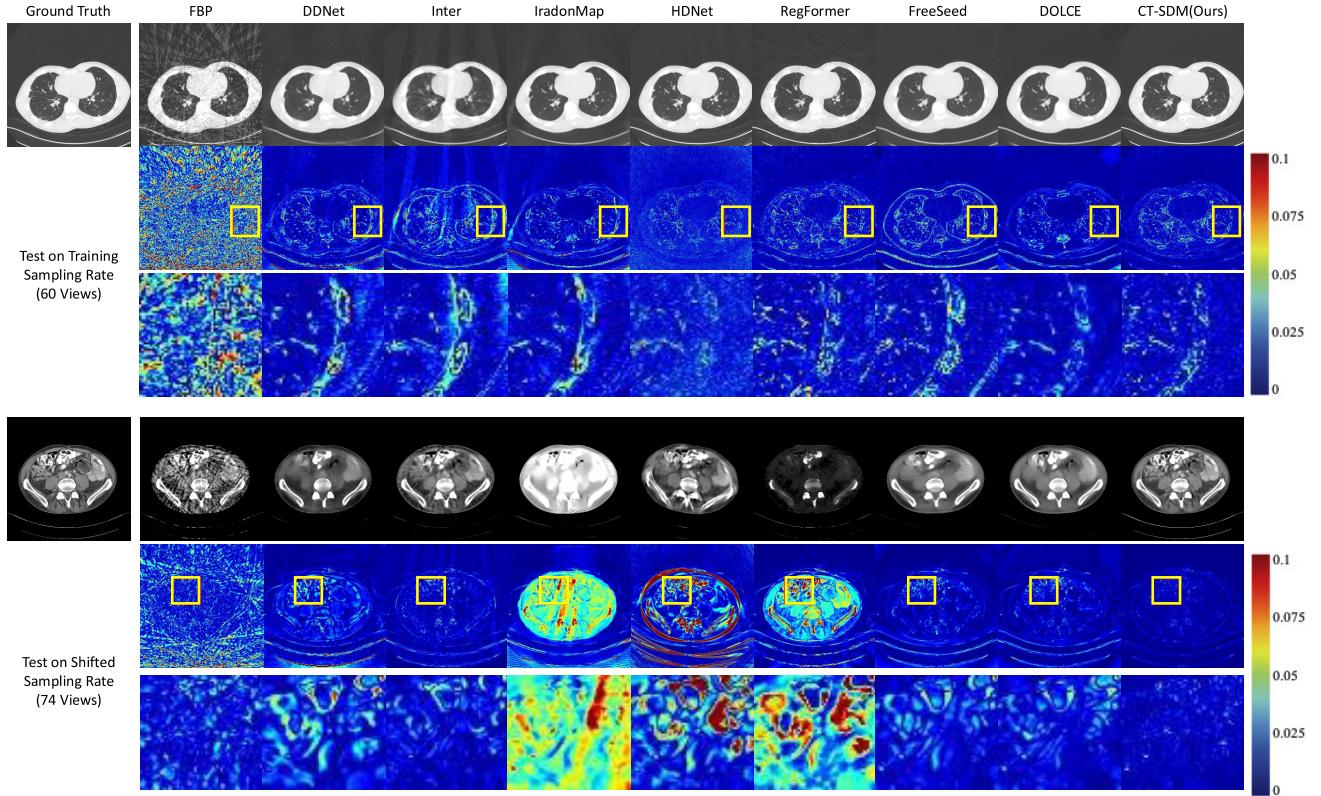
1) **4D CBCT Reconstruction:** Due to breathing-induced tissue motion, four-dimensional (4D) imaging techniques, such as 4D cone-beam CT (4D CBCT), are often necessary to accurately characterize motion dynamics [49]. Considering the variability in patients' breathing cycles and patterns, the sampling rate for 4D CBCT can vary significantly across

different scans. In this experiment, we evaluate the proposed method by performing 4D CBCT reconstruction across varying sampling rates to verify its effectiveness.

We utilize the 4D-Lung Dataset [49], which includes 4D FBCT scans and 4D CBCT scans acquired during chemoradiotherapy. Since raw projection data for 4D CBCT and respiratory signals are not available, our experiments use high-quality 4D FBCT as the ground truth and simulate 4D CBCT data at various sampling rates. The respiratory signals for simulation are obtained from the BIDMC PPG and Respiration Dataset [50], which provides impedance respiratory signals and manually annotated breath phases by expert annotators.

For training, data from 15 patients were used, while 5 patients were reserved for testing. As shown in Fig 9, each respiratory cycle is divided into 10 phases, labeled from Phase 0% to Phase 90%. For each patient, 32 slices from the 4D FBCT are used to generate sparse-view CBCT projections. The exact number of projections in each phase varies due to the diversity in respiratory patterns.

To extend our method for 4D CBCT data, we adapted the backbone network in the sinogram domain to a 3D-UNet for CBCT sinogram recovery and employed a 3D version of the residual block for image domain recovery. The

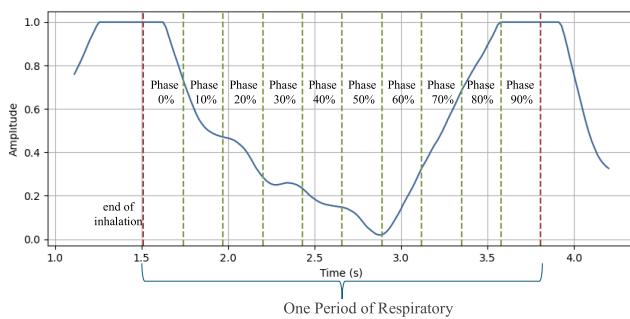


**Fig. 8.** Visualization of reconstruction results on the LDCT dataset. Reconstructions are compared at the training sampling rate (60 views) and the shifted sampling rate (74 views). Chest images (top row) are displayed using a window of  $[-1350, 150]$ , and abdomen images (bottom row) are displayed using a window of  $[-240, 160]$ .

TABLE III

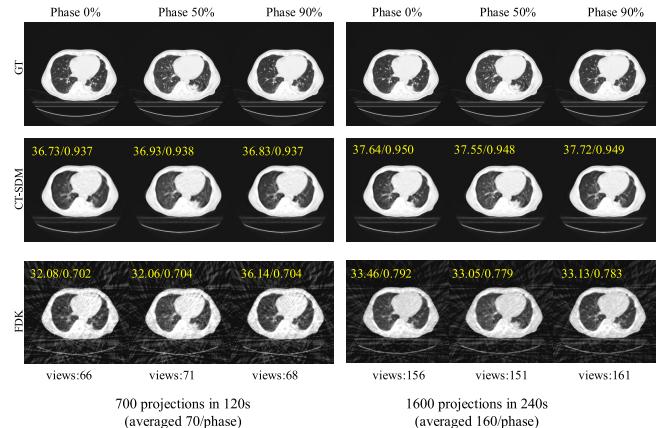
ABLATION STUDY ON DIFFERENT PART OF NETWORK. THE BEST RESULTS ARE HIGHLIGHTED.  $\downarrow$  ( $\uparrow$ ) INDICATES THAT LOWER (HIGHER) VALUES ARE BETTER. \* DENOTES RESULTS THAT ARE SIGNIFICANTLY DIFFERENT FROM THE BEST RESULTS BY THE WILCOXON TEST ( $p < 0.05$ )

Method	PSNR( $\uparrow$ )	SSIM( $\uparrow$ )	LPIPS( $\downarrow$ )
Sinogram Domain	$38.613 \pm 2.614^*$	$0.947 \pm 0.026^*$	$0.071 \pm 0.030^*$
Image Domain	$28.830 \pm 1.460^*$	$0.617 \pm 0.053^*$	$0.309 \pm 0.049^*$
Dual Domain	<b><math>41.032 \pm 2.907</math></b>	<b><math>0.956 \pm 0.023</math></b>	<b><math>0.049 \pm 0.023</math></b>



**Fig. 9.** Impedance respiratory signal and manual annotations. Each respiratory cycle is divided into 10 phases marked as Phase 0% to Phase 90%.

Astra-toolbox [51] was utilized for sinogram projection and FDK reconstruction during both training and testing phases.



**Fig. 10.** The 15th slice of reconstructed volumes for 4DCBCT. Group 1 use 700 projections in a period of 2 minutes scan, Group 2 use 1,600 projections in a period of 4 minutes scans. The sampling rates and PSNR/SSIM are marked on the images. All images are displayed on window  $[-1100, 100]$ .

Two groups of reconstructions are presented in Fig. 10. Group 1 uses 700 projections over a 2-minute scan period, while Group 2 uses 1,600 projections over a 4-minute scan period. The sampling rates, along with the corresponding PSNR and SSIM values, are annotated directly on the images for clarity. As results shows, our method can effectively improve the image quality no matter the slight sampling rate

TABLE IV

COMPARISON OF DIFFERENT SAMPLING STRATEGIES ON 60 PROJECTION VIEWS. THE BEST RESULTS ARE HIGHLIGHTED. ↓ (↑) INDICATES THAT LOWER (HIGHER) VALUES ARE BETTER. \* DENOTES RESULTS THAT ARE SIGNIFICANTLY DIFFERENT FROM THE BEST RESULTS BY THE WILCOXON TEST ( $p < 0.05$ )

Method	PSNR(↑)	SSIM(↑)	LPIPS(↓)
Fixed Sampling	$40.412 \pm 2.332^*$	$0.921 \pm 0.021^*$	$0.053 \pm 0.021^*$
Random Sampling	$38.012 \pm 2.667^*$	$0.901 \pm 0.056^*$	$0.066 \pm 0.043^*$
Grouped-Random Sampling	<b><math>41.032 \pm 2.907</math></b>	<b><math>0.956 \pm 0.023</math></b>	<b><math>0.049 \pm 0.023</math></b>

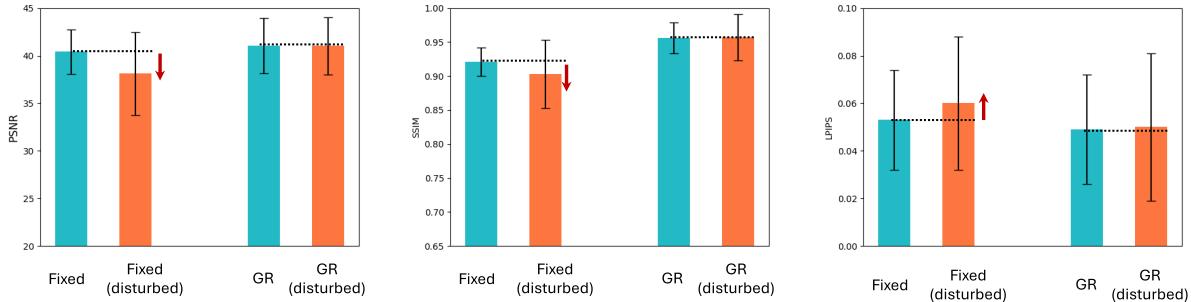


Fig. 11. Robustness comparison over sampling views disturbance. Models trained with a fixed sampling strategy show a noticeable performance drop, while the proposed grouped-random sampling (GR) strategy demonstrates much more stable performance.

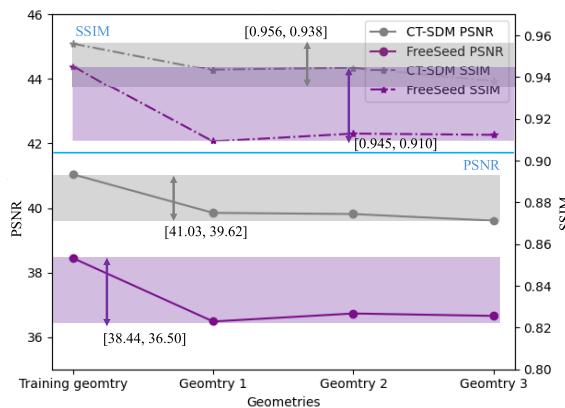


Fig. 12. Performance over different CT geometry. The default equal-distance fan-beam configuration is 1,000 mm source-to-center distance, 512 mm detector-to-center distance, and 1.5 mm detector spacing (training geometry). We tested on three additional geometries, Geometry 1 (600 mm source-to-center distance, 580 mm detector-to-center distance, and 1.5 mm detector spacing), Geometry 2 (1000 mm source-to-center distance, 380 mm detector to center distance, and 1 mm detector spacing), and Geometry 3 (1000 mm source-to-center distance, 180 mm detector-to-center distance, and 0.8 mm detector spacing).

shifting between respiratory phases or large sampling rate changes between different respiratory signal and experimental settings.

2) *Performance on Lesion Area:* To validate the method's relevance to diagnosis-related tasks, we selected slices with annotated lesion areas from the LDCT dataset. All evaluations were conducted at the same sampling rate (100 views) during training and testing. As shown in Fig. 14, the first row shows the reconstruction performance for a liver cyst. Several methods (e.g., DDNet, Inter, and iRadonMap) failed to capture the lesion details, while others (e.g., HDNet, RegFormer, FreeSeed, and DOLCE) produced oversmoothed shapes of the

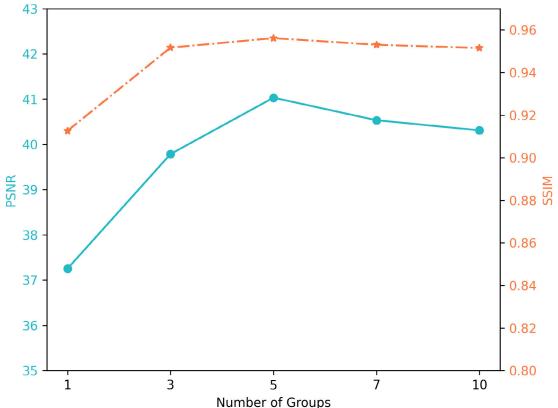


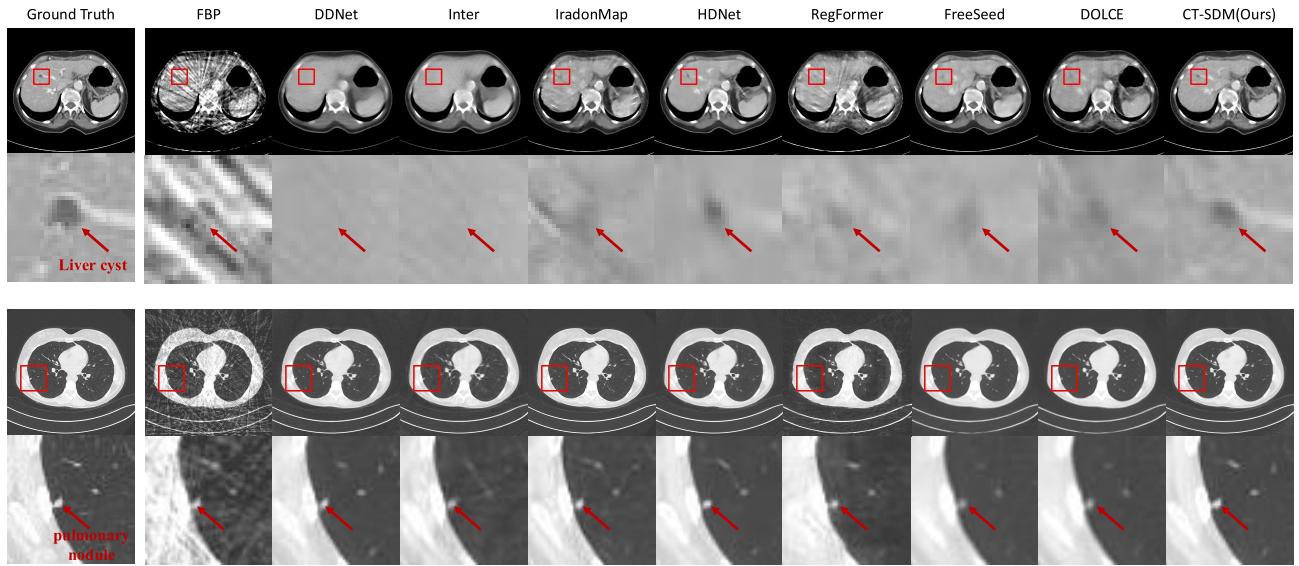
Fig. 13. Performance over different number of groups. The best results are observed around  $c = 5$ , where the balance between randomness and data augmentation is optimal.

lesion. By contrast, the proposed method provides a clearer depiction of the cyst's intensity and shape, enhancing its diagnostic value. The third row shows results for a pulmonary nodule. While all methods were able to detect the nodule, the proposed method demonstrated superior detail and shape fidelity.

### G. Ablation Study

To verify the effectiveness of the proposed method, we conduct several ablation studies in this section.

1) *Performance Over Different CT Geometry:* As the group-random sampling during training, the proposed method shows a better performance when imaging settings are changed. In these experiments, we compare the PSNR and SSIM using different CT Geometry settings and results are shown on Fig. 12. Both methods are trained on the



**Fig. 14.** Visualization of reconstruction results on the lesion slices. We compare the reconstruction images at the training sampling rate (100 views) of all methods. Abdomen images (the first and second rows) are displayed on window  $[-240, 160]$ , chest images (the third and fourth rows) are displayed on window  $[-1350, 150]$ .

same equal-distance fan-beam, with  $1000\text{mm}$  source-to-center distance,  $512\text{mm}$  detector-to-center distance, and  $1.5\text{ mm}$  detector spacing (Training geometry). We test on 3 other geometry, Geometry 1 ( $600\text{mm}$  source-to-center distance,  $580\text{mm}$  detector-to-center distance, and  $1.5\text{ mm}$  detector spacing), Geometry 2 ( $1000\text{mm}$  source-to-center distance,  $380\text{mm}$  detector to center distance, and  $1\text{mm}$  detector spacing), and Geometry 3 ( $1000\text{mm}$  source-to-center distance,  $180\text{mm}$  detector-to-center distance, and  $0.8\text{ mm}$  detector spacing).

As shown in Fig. 12, the proposed method demonstrates superior performance and stability compared to other methods across all geometry configurations. This highlights the model’s robustness and adaptability to variations in imaging settings, further supported by the group-random sampling during training.

**2) Ablation Study of Sinogram and Image Domain Network.** In this experiment, we analyzed the contributions of different parts of our network to the reconstruction process. To evaluate the effectiveness of the sinogram and image-domain networks, we performed the following comparisons: Sinogram Domain: Reconstructing images directly from the output of the sinogram domain (diffusion model) using FBP. Image-Domain: Feeding input sparse-view FBP images directly into the image-domain network without using the sinogram domain network. Dual-Domain (Proposed CT-SDM): Combining both sinogram and image-domain networks as designed. As presented in Tab. III, the sinogram domain network provides the primary contributions to reconstruction quality. In contrast, the lightweight image-domain network alone does not significantly improve performance, as it is specifically designed to complement the sinogram domain. When combined, the dual-domain design achieves the highest overall reconstruction performance, underscoring the importance and effectiveness of the integrated approach.

**3) Number of Groups in Grouped-Random Sampling:** In this part, we conducted an ablation study to evaluate the impact of the number of groups  $c$  on the performance of Grouped-Random Sampling. The results are summarized in Fig. 13. For smaller  $c$ , higher randomness in training leads to instability and lower performance. The extreme case of  $c = 1$ , where views are chosen entirely randomly, results in a significant performance drop due to the lack of structured sampling. As  $c$  increases, the randomness is better distributed, and the performance stabilizes. The best results are observed around  $c=5$ , where the balance between randomness and data augmentation is optimal. When  $c$  becomes too large, the sampling strategy closely approximates a fixed sampling list. This reduces the augmentation effect, slightly degrading performance beyond  $c = 5$ .

**4) Effectiveness of Grouped-Random Sampling:** As illustrated in Fig 3, the grouped-random sampling strategy serves as a powerful form of data augmentation to improve reconstruction performance. To validate its effectiveness, we compare the performance of three sampling strategies: fixed sampling (i.e., projection views at each sampling rate are fixed and uniformly aligned), random sampling (projection views are randomly drawn from a uniform distribution), and the proposed Grouped Random Sampling. The results are shown in Table IV. The naive random sampling disrupts the uniformity of sampling views, undermining reconstruction performance. In contrast, the grouped-random sampling strategy introduces randomness while maintaining uniformity, thereby significantly enhancing performance compared to fixed sampling.

**5) Robustness Over Sampling Views Disturbance:** The main advantage of the grouped-random sampling strategy lies in its ability to simulate the variability of available projection views in real-world scenarios, which is common in clinical settings

due to patient movement, equipment limitations, or specific diagnostic needs. To verify the robustness of the proposed method, we conduct experiments using the grouped-random sampling strategy to generate the testing dataset, simulating sampling view disturbances. The performance of fixed sampling and Grouped Random Sampling strategies is compared on view-disturbed testing data, as shown in Fig 11. The results demonstrate that the grouped-random Sampling strategy exhibits superior robustness to sampling view disturbances compared to fixed sampling.

## V. CONCLUSION

This paper proposes a sparse-view CT reconstruction method based on Diffusion Models, aimed at achieving high-performance reconstruction of sparse-view CT using deep learning networks at arbitrary sampling rates. While deep learning models have shown promising performance in artifact suppression for sparse-view CT, their generalization limitations make it challenging for a model trained at one sampling rate to perform well at other sampling rates. Consequently, existing methods often require models to be trained for specific sampling rates, limiting their flexibility and usability in clinical environments. To address this issue, this paper introduces the CT-SDM and employs a stepwise inference approach, enabling the same model to perform reconstructions at various sampling rate. In CT-SDM, a projection degradation operator is used to simulate the projection process in the sinogram domain, allowing the model to gradually add new projection data to highly undersampled projection data, thereby reconstructing a complete projection sinogram. By selecting a diffusion inference starting point corresponding to a specific sampling rate, our model can recover a complete projection sinogram from various sampling rate using a single trained model. Through comparative experiments on multiple public datasets, we validate that our method can achieve high-quality reconstructions of SVCT at different sampling rates, providing a flexible and reliable reconstruction solution for sparse-view imaging.

This work primarily focuses on the technical novelty of the proposed CT-SDM and its ability to address generalization issues across varying sampling rates. As a result, this study does not include full-blown clinical trials or validations. Future work will aim to bridge this gap by conducting clinical studies to evaluate the performance and reliability of CT-SDM in practical applications, particularly for tasks requiring high diagnostic accuracy.

## ACKNOWLEDGMENT

Guang Yang and Daoqiang Zhang are co-last authors.

## REFERENCES

- [1] N. B. Shah and S. L. Platt, "ALARA: Is there a cause for alarm? Reducing radiation risks from computed tomography scanning in children," *Curent Opinion Pediatrics*, vol. 20, no. 3, pp. 243–247, 2008.
- [2] K. Kim et al., "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE Trans. Med. Imag.*, vol. 34, no. 3, pp. 748–760, Mar. 2015.
- [3] T. Lee, C. Lee, J. Baek, and S. Cho, "Moving beam-blocker-based low-dose cone-beam CT," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 5, pp. 2540–2549, Oct. 2016.
- [4] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Problems*, vol. 25, no. 12, Dec. 2009, Art. no. 123009.
- [5] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, "Low-dose CT reconstruction via edge-preserving total variation regularization," *Phys. Med. Biol.*, vol. 56, no. 18, p. 5949, 2011.
- [6] T. Wang, K. Nakamoto, H. Zhang, and H. Liu, "Reweighted anisotropic total variation minimization for limited-angle CT reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 64, no. 10, pp. 2742–2760, Oct. 2017.
- [7] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.
- [8] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, Oct. 2017.
- [9] G. Zang, M. Aly, R. Idoughi, P. Wonka, and W. Heidrich, "Super-resolution and sparse view CT reconstruction," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 137–153.
- [10] Z. Zhao, Y. Sun, and P. Cong, "Sparse-view CT reconstruction via generative adversarial networks," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2018, pp. 1–5.
- [11] 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.
- [12] 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, Feb. 2018.
- [13] 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.
- [14] 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, no. 9, pp. 4509–4522, Sep. 2017.
- [15] D. Hu et al., "Hybrid-domain neural network processing for sparse-view CT reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 1, pp. 88–98, Jan. 2021.
- [16] H. Lee, J. Lee, and S. Cho, "View-interpolation of sparsely sampled sinogram using convolutional neural network," *Proc. SPIE*, vol. 10133, pp. 617–624, Feb. 2017.
- [17] L. Yang, R. Ge, S. Feng, and D. Zhang, "Learning projection views for sparse-view CT reconstruction," in *Proc. 30th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 2645–2653, doi: [10.1145/3503161.3548204](https://doi.org/10.1145/3503161.3548204).
- [18] H. Zhang, B. Liu, H. Yu, and B. Dong, "MetaInv-net: Meta inversion network for sparse view CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 621–634, Feb. 2021.
- [19] 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.
- [20] L. Yang, Z. Li, R. Ge, J. Zhao, H. Si, and D. Zhang, "Low-dose CT denoising via sinogram inner-structure transformer," *IEEE Trans. Med. Imag.*, vol. 42, no. 4, pp. 910–921, Apr. 2023.
- [21] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, pp. 487–492, Mar. 2018.
- [22] J. He, Y. Wang, and J. Ma, "Radon inversion via deep learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2076–2087, Jun. 2020.
- [23] 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.
- [24] X. Li, J. Thickstun, I. Gulrajani, P. S. Liang, and T. B. Hashimoto, "Diffusion-LM improves controllable text generation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 4328–4343.
- [25] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. van Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11461–11471.
- [26] T. Amit, T. Shahrabany, E. Nachmani, and L. Wolf, "SegDiff: Image segmentation with diffusion probabilistic models," 2021, *arXiv:2112.00390*.
- [27] R. Zimmermann, L. Schott, Y. Song, B. Dunn, and D. Klindt, "Score-based generative classifiers," 2021, *arXiv:2110.00473*.

- [28] J. Wolleb, F. Bieder, R. Sandkuhler, and P. C. Cattin, “Diffusion models for medical anomaly detection,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 35–45.
- [29] B. Kim, I. Han, and J. C. Ye, “DiffuseMorph: Unsupervised deformable image registration using diffusion model,” in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 347–364.
- [30] K. Packhäuser, L. Folle, F. Thamm, and A. Maier, “Generation of anonymous chest radiographs using latent diffusion models for training thoracic abnormality classification systems,” in *Proc. IEEE 20th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2023, pp. 1–5.
- [31] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 10684–10695.
- [32] X. Meng et al., “A novel unified conditional score-based generative framework for multi-modal medical image completion,” 2022, *arXiv:2207.03430*.
- [33] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” 2020, *arXiv:2011.13456*.
- [34] H. Chung and J. C. Ye, “Score-based diffusion models for accelerated MRI,” *Med. Image Anal.*, vol. 80, Aug. 2022, Art. no. 102479.
- [35] J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” 2020, *arXiv:2010.02502*.
- [36] J. Liu et al., “DOLCE: A model-based probabilistic diffusion framework for limited-angle CT reconstruction,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 10498–10508.
- [37] A. Bansal et al., “Cold diffusion: Inverting arbitrary image transforms without noise,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 41259–41282.
- [38] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
- [39] J. Huang, A. I. Avilés-Rivero, C. Schönlieb, and G. Yang, “CDiffMR: Can we replace the Gaussian noise with K-space undersampling for fast MRI?” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, Jan. 2023, pp. 3–12.
- [40] T. R. Moen et al., “Low-dose CT image and projection dataset,” *Med. Phys.*, vol. 48, no. 2, pp. 902–911, 2021.
- [41] K. Clark et al., “The cancer imaging archive (TCIA): Maintaining and operating a public information repository,” *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.
- [42] O. Akin et al., “The cancer genome atlas kidney clear cell carcinoma collection (TCGA-KIRC) (version 3) [data set],” *Cancer Imag. Arch.*, 2016. [Online]. Available: <https://doi.org/10.7937/K9/TCAA.2016.V6PBVTD>
- [43] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [44] M. Ronchetti, “TorchRadon: Fast differentiable routines for computed tomography,” 2020, *arXiv:2009.14788*.
- [45] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
- [46] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, “A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution,” *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1407–1417, Jun. 2018.
- [47] C. Ma, Z. Li, J. Zhang, Y. Zhang, and H. Shan, “FreeSeed: Frequency-band-aware and self-guided network for sparse-view CT reconstruction,” in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, 2023, pp. 250–259.
- [48] Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.
- [49] G. D. Hugo et al., “A longitudinal four-dimensional computed tomography and cone beam computed tomography dataset for image-guided radiation therapy research in lung cancer,” *Med. Phys.*, vol. 44, no. 2, pp. 762–771, Feb. 2017.
- [50] A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. 215–220, Jun. 2000.
- [51] W. van Aarle et al., “Fast and flexible X-ray tomography using the ASTRA toolbox,” *Opt. Exp.*, vol. 24, no. 22, pp. 25129–25147, 2016.