Iterative Reconstruction Based on Latent Diffusion Model for Sparse Data Reconstruction

Linchao He[*1,2], **Hongyu Yan**[*1,2], **Mengting Luo**[1,2], **Kunming Luo**[3], **Wang Wang**[4], **Wenchao Du**[2], **Hu Chen**[†2] **Hongyu Yang**[2] **Yi Zhang**[5]
[1]Department of National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu, China
[2]College of Computer Science, Sichuan University, Chengdu, China
[3]The Hong Kong University of Science and Technology, Hong Kong, China
[4]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China
[5]School of Cyber Science and Engineering, Sichuan University, Chengdu, China
{hlc, hongyuyan, lmt}@stu.scu.edu.cn; kluoad@connect.ust.hk; wwwwhu520@gmail.com; wenchaodu.scu@gmail.com; huchen@scu.edu.cn; yanghongyu_scu@163.com; yzhang@scu.edu.cn

Abstract

Reconstructing Computed tomography (CT) images from sparse measurement is a well-known ill-posed inverse problem. The Iterative Reconstruction (IR) algorithm is a solution to inverse problems. However, recent IR methods require paired data and the approximation of the inverse projection matrix. To address those problems, we present Latent Diffusion Iterative Reconstruction (LDIR), a pioneering zero-shot method that extends IR with a pre-trained Latent Diffusion Model (LDM) as a accurate and efficient data prior. By approximating the prior distribution with an unconditional latent diffusion model, LDIR is the first method to successfully integrate iterative reconstruction and LDM in an unsupervised manner. LDIR makes the reconstruction of high-resolution images more efficient. Moreover, LDIR utilizes the gradient from the data-fidelity term to guide the sampling process of the LDM, therefore, LDIR does not need the approximation of the inverse projection matrix and can solve various CT reconstruction tasks with a single model. Additionally, for enhancing the sample consistency of the reconstruction, we introduce a novel approach that uses historical gradient information to guide the gradient. Our experiments on extremely sparse CT data reconstruction tasks show that LDIR outperforms other state-of-the-art unsupervised and even exceeds supervised methods, establishing it as a leading technique in terms of both quantity and quality. Furthermore, LDIR also achieves competitive performance on nature image tasks. It is worth noting that LDIR also exhibits significantly faster execution times and lower memory consumption compared to methods with similar network settings. Our code will be publicly available.

1 Introduction

Computed tomography (CT) is a crucial medical imaging technique in contemporary medicine to aid physicians in diagnosing relevant conditions. Measurements in CT are obtained by X-rays

[*]Equal contribution.
[†]Corresponding Authors.

Preprint. Under review.

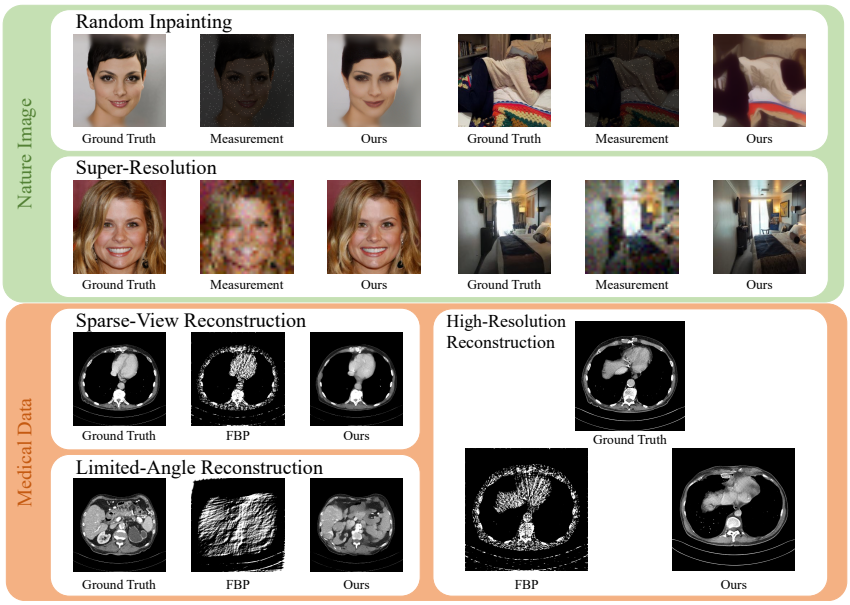

Figure 1: LDIR is capable of reconstructing both nature images and medical data in a zero-shot manner. In this paper, we demonstrate the reconstruction ability of our method on various datasets and measurements. Different from previous works, LDIR can reconstruct high-resolution images (512 x 512) from sparse measurements using unconditional latent diffusion models. FBP is filtered backprojection. The display window of CT images is set to [-150, 256] HU.

projections of an object from different views. However, the use of X-rays in CT exposes the human body to potentially harmful doses of radiation, raising concerns among the public regarding radiation-induced diseases. Therefore, reducing the exposure dose, such as sparse-view and limited-angle imaging, while maintaining the quality of imaging has beneficial implications for both public health and medical diagnosis, specifically in intraoperative CT. Due to the sparse information, the CT reconstruction processes are well-known ill-posed inverse problems. In the past decade, numerous works have focused on Iterative Reconstruction (IR), which is considered a solution to CT reconstruction Donoho (2006); Candès et al. (2006). Iterative reconstruction aims to recover signals x from noisy measurements $y = \mathbf{A}x + n$, where n represents the noise in the measuring process and $\mathbf{A}$ is the linear projection matrix that typically maps x to a lower dimension. As a result, a typical IR process can be formulated as follows:

$$\hat{x} = \arg\min_{x} ||\mathbf{A}x - y||_2^2 + \lambda R(x), \tag{1}$$

here, $||\mathbf{A}x - y||_2^2$ is the data-fidelity term that ensures the reconstructed results are consistent with the measurements, while $\lambda R(x)$ is a prior term that ensures the reconstructed results are realistic and follow the distribution $p(x)$ of the ground truth images.

The key challenge of IR is to find appropriate data prior or sparse transformation to generate the prior term. Traditional IR methods Beck, Teboulle (2009); Kim et al. (2016); Zhao et al. (2000) leverage the total variation, nonlocal means, or wavelets to gain the reconstructed results by a handcrafted prior and the approximation of the inverse projection matrix. Inspired by the success of deep learning and neural network, recent data-driven methods Bora et al. (2017); Chen et al. (2018) achieve impressive performance by learning the prior and data-fidelity terms. However, most of these methods require large-scale paired data to train their networks. Besides, they directly project the measurement to the results and they also need to retrain while the detector geometry changes. Although method Song et al. (2022) solves these problems mentioned above by introducing a score-based generative model, it still needs to know the approximation of the inverse projection matrix which is difficult to be obtained in the real-world application. Additionally, these methods use direct projection to replace the data-fidelity term which makes they make them need to design different sampling procedures for different detector geometries.

In this paper, we introduce a novel Latent Diffusion Iterative Reconstruction (LDIR) for sparse CT data reconstruction in a zero-shot manner. The LDIR is the first to extend traditional IR techniques by incorporating a latent diffusion model. Specifically, we train a latent unconditional diffusion model to learn the data distribution. In the reverse process, the trained diffusion model is utilized to replace the prior term in the regular IR. In particular, to generate specified prior from the unconditional diffusion model, the gradient from the data-fidelity term is applied to guide the sampling process of the diffusion model. Therefore, we do not need any paired data or inverse of the measurement matrix to train our network. Moreover, LDIR can solve various CT reconstruction tasks with a single model. In addition, by guiding the reverse diffusion sampling process in the latent space, our zero-shot method can generate high-resolution images with impressive performance. Since LDIR does not make any assumption on the data-fidelity term, we can use any differentiable measurement function to keep the consistency of data. Further on, we propose a novel guidance strategy to adaptively adjust the sample-level gradient by fusing the history gradient, thereby improving the performance of our zero-shot diffusion model.

Extensive experiments demonstrate our method outperforms state-of-the-art supervised and unsupervised methods for extremely sparse CT data reconstruction. Additionally, due to the commonality of iterative reconstruction on CT images and natural images, we extend LDIR to the natural image restoration task, and our approach achieves competitive performance compared to other state-of-the-art zero-shot methods. Our approach provides a valuable zero-shot tool for solving inverse problems with latent diffusion models, allowing us to leverage the vast amount of available latent diffusion models. Fig. 1 shows some visual results of the proposed method.

2 Background

2.1 Diffusion models

Consider a T-step Gaussian diffusion process, where $x_t \in \mathbb{R}^n, t \in [0, \ldots, T-1]$ and initial x_0 is sampled from the original data distribution P_{data}. We define the forward diffusion process using stochastic differential equation (SDE) Song et al. (2021):

$$dx = f(x,t)\ dt + g(t)\ dw, \tag{2}$$

where $f(\cdot,t) : \mathbb{R}^d \to \mathbb{R}^d$ is a drift coefficient function, $g(t) \in \mathbb{R}$ is defined as a diffusion coefficient function, and $w \in \mathbb{R}^n$ is a standard n-dimensional Brownian motion. Thus, the reverse SDE of Eq. (2) can also defined as:

$$dx = \left[f(x,t) - g(t)^2 \nabla_x \log p_t(x)\right] dt + g(t)\ dw, \tag{3}$$

where dt is a negative infinitesimal time step. The reverse SDE defines a generative process that transforms standard Gaussian noise into meaningful content. To accomplish this transformation, the score function $\nabla_x \log p_t(x)$ needs to be matching, which is typically replaced with $\nabla_x \log p_{0|t}(x_t|x_0)$ in practice. Therefore, we can train a score model $s_\theta(x,t)$, so that $s_\theta(x,t) \approx \nabla_x \log p_t(x) \approx \nabla_x \log p_{0|t}(x_t|x_0)$ using the following score-matching objective:

$$\min_\theta \mathbb{E}_{t\in[0,\ldots,T-1],x_0\sim P_{data},x_t\sim p_{0|t}(x_t|x_0)} \left[\left|\left|s_\theta(x,t) - \nabla_x \log p_{0|t}(x_t|x_0)\right|\right|_2^2\right]. \tag{4}$$

Therefore, the reverse SDE can yield meaningful contents $x_0 \sim P_{data}$ from random noises $x_{T-1} \sim \mathcal{N}(0, \mathbf{I})$ by iteratively using $s_\theta(x,t)$ to estimate the scores $\nabla_x \log p_t(x)$. In our experiments, we adopt the standard Denoising Diffusion Probabilistic Models (DDPM) Ho et al. (2020) which is equivalent to the above variance preserving SDE (VP-SDE Song et al. (2021)).

2.2 Diffusion model for inverse problem solving

To solve the inverse problems using the diffusion model, various workarounds are proposed Rombach et al. (2022); Saharia et al. (2022); Gao et al. (2023); Luo et al. (2023). These methods use conditional diffusion models to iteratively denoise Gaussian noise and obtain reconstructions. However, these approaches have limitations, as they rely on conditional diffusion models that require paired data for training and can only handle specific tasks without retraining. To address these issues, several zero-shot diffusion-based inverse solvers Lugmayr et al. (2022); Song et al. (2022); Kawar et al. (2022);

Chung et al. (2022); Wang et al. (2023) have been proposed. Typically, assuming $n \equiv 0$, for each denoising step, they Lugmayr et al. (2022); Song et al. (2022); Wang et al. (2023) unconditionally estimate new denoised samples based on the previous step, followed by replacing the corresponding items in the denoised samples using the measurement $\mathbf{A}^{-1}y$, which is also known as range-null space decomposition Wang et al. (2023). This approach ensures data consistency, but it fails in the case of noisy measurements, since $\mathbf{A}^{-1}y$ is not a correct corresponding item for the denoised samples. To address this limitation, alternative methods Chung et al. (2023b,a) have been proposed to solve the inverse problems with noised measurements. Rather than directly replace items, these approaches use the gradient of $||y - \mathbf{A}x||_2^2$ to conditionally guide the generative process. These methods are robust to noise and can process nonlinear projection operators. However, these methods try to solve inverse problems on the pixel space and make strong assumptions on the data-fidelity term, which significantly underestimates the complexity of real-world problems and can not reconstruct high-resolution results.

3 Method

3.1 Diffusion iterative reconstruction

Generally, it is possible to transform the IR methods presented in Eq. (1) into a more generic form:

$$\hat{x} = \arg\min_x E(x) \tag{5}$$

$$= \arg\min_x \mathcal{U}(\mathbf{A}x, y) + \lambda R(x), \tag{6}$$

where $\mathcal{U}$ is a measurement function that ensures data consistency. Assuming that both the measurement function and prior term are differentiable, it is possible to apply gradient descent to Eq. (6). This yields an ordinary differential equation (ODE) for iterative reconstruction:

$$x_{t-1} = x_t - \frac{\partial E(x)}{\partial x_t} \tag{7}$$

$$= x_t - (\epsilon \nabla_{x_t} \mathcal{U}(\mathbf{A}x_t, y) + \lambda_t \nabla_{x_t} R(x_t)), \tag{8}$$

where guidance rate ϵ and λ_t are used to balance the consistency and realness. In practice, we can replace $\lambda \nabla_{x_t} R(x_t)$ with a step-dependent prior function $\lambda_t \nabla_{x_t} R(x_t, t)$ to balance realness and data consistency, as demonstrated by Chen et al. (2018):

$$x_{t-1} = (x_t - \lambda_t \nabla_{x_t} R(x_t, t)) - \epsilon \nabla_{x_t} \mathcal{U}(\mathbf{A}x_t, y). \tag{9}$$

As demonstrated by Song et al. (2021), the score function is a powerful tool for representing probability distributions. The score function does not require the computation of a tractable normalizing constant or its approximation. This makes it possible to estimate $\log p(x)$ using the score function. By replacing the prior term $\nabla_{x_t} R(x_t, t)$ with the score function $\nabla_{x_t} \log p_t(x)$, it is possible to ensure that the new prior term accurately represents the probability distribution without requiring heavy computation:

$$x_{t-1} = (x_t - \lambda_t \nabla_x \log p_t(x)) - \epsilon \nabla_{x_t} \mathcal{U}(\mathbf{A}x_t, y), \tag{10}$$

In Eq. (4), we can use a score model $s_\theta(x_t, t)$ to approximate $\nabla_{x_t} \log p_t(x)$ using the score matching technique:

$$x_{t-1} \simeq (x_t - \lambda_t s_\theta(x_t, t)) - \epsilon \nabla_{x_t} \mathcal{U}(\mathbf{A}x_t, y). \tag{11}$$

In fact, the first term $x_t - \lambda_t s_\theta(x, t)$ is Eq. (3) with a noise-free constraint [3]. Thus, we can relax the noise-free constraint and use Eq. (3) to replace this term:

$$x_{t-1} \simeq x_t - \underbrace{(\lambda_t s_\theta(x_t, t) - g(t) z)}_{\text{Prior term}} - \underbrace{\epsilon \nabla_{x_t} \mathcal{U}(\mathbf{A}x_t, y)}_{\text{Data-fidelity term}}, z \sim \mathcal{N}(0, \mathbf{I}). \tag{12}$$

We propose a new iterative reconstruction method called Diffusion Iterative Reconstruction (DIR). The model uses a score model to represent the original data distribution and employs it as a learned prior. Algo. 1 demonstrates the pixel guidance process of DIR.

[3] $f(x_t, t) = g(t) z = 0$ and $g(t)^2 = \lambda_t$

It is worth noting that our DIR has a mathematical form that is similar to the approach proposed by Chung et al. (2023b). However, our approach is more general and can be applied to a wider range of scenarios. Diffusion Posterior Sampling (DPS) Chung et al. (2023b) models the data consistency term $\mathcal{U}(\cdot,\cdot)$ as a noise measurement problem. The evaluation function for the data-fidelity term is derived **based on the prior assumption** of the noise distribution (such as the $L2$ function for Gaussian noise). This approach works well when the noise type is known in the measurement process. However, it is challenging to estimate the noise distribution and derive the correct evaluation function in real-world applications. In contrast, our DIR models the data-fidelity term as a quality evaluation function **without any assumptions** about the noise distribution. Thus, the evaluation function $\mathcal{U}(\cdot,\cdot)$ can be any differentiable evaluation function, allowing for more flexibility.

Algorithm 1 Pixel-space data reconstructing

Require: : N, y
$x_T \sim \mathcal{N}$
for do $\{i = T-1, \ldots, 0\}$
  $\hat{s} \leftarrow s_\theta(x_i, i)$
  $x_{0|i} \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}\left(x_i + (1-\bar{\alpha}_i)\,\hat{s}\right)$
  $z \sim \mathcal{N}(0, \mathbf{I})$
  $x'_{i-1} \leftarrow \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i} x'_{0|i} + \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i} x_i + g(i)\, z$
  $x_{i-1} \leftarrow x'_{i-1} - \epsilon \nabla_{x_i} \mathcal{U}\left(\mathbf{A} x'_{i-1}, y\right)$
end for
return $x_{0|0}$

Algorithm 2 Latent-space data reconstructing

Require: : $N, y, \mathcal{D}$
$\ell_T \sim \mathcal{N}$
for do $\{i = T-1, \ldots, 0\}$
  $\hat{s} \leftarrow s_{\theta_l}(\ell_i, i)$
  $\ell_{0|i} \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}\left(\ell_i + (1-\bar{\alpha}_i)\,\hat{s}\right)$
  $z \sim \mathcal{N}(0, \mathbf{I})$
  $\ell'_{i-1} \leftarrow \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i} \ell'_{0|i} + \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i} \ell_i + g(i)\, z$
  $\ell_{i-1} \leftarrow \ell'_{i-1} - \epsilon \nabla_{\ell_i} \mathcal{U}(\mathbf{A}\mathcal{D}\left(\ell'_{i-1}\right), y)$
end for
$x_{0|0} \leftarrow \mathcal{D}\left(\ell_{0|0}\right)$
return $x_{0|0}$

3.2 History gradient update

In what follows, we show that introducing history gradient update policies can provide better reconstruction results. The basic formulation of gradient guidance in Eq. (12) corresponds to a simple gradient descent scheme. The guidance rate ϵ can be thought of as the learning rate value in the Stochastic Gradient Descent (SGD). Thus, in order to further improve the data consistency, we adopt gradient information from the previous steps. Because the history gradient information can provide sample-level information to decide the optimization direction of the guidance process. This is also known as the *first-order* gradient-based optimization. Here, we demonstrate two variants of gradient update policies based on two typical optimizers.

Momentum-like gradient update policy. Similar to the Momentum optimizer Sutskever et al. (2013), we consider to use the moving averages m^t of the gradients $\nabla_{x_t}\mathcal{U}(\mathbf{A}x_t, y)$ to perform gradient descent:

$$m^t = \eta m^{t-1} + (1-\eta)\,\nabla_{x_t}\mathcal{U}(\mathbf{A}x_t, y), \tag{13}$$

$$x_{t-1} = x_t - \epsilon m^t, \tag{14}$$

where η is a hyper-parameter to adjust the factor of momentum.

Adam-like gradient update policy. Adam optimizer Kingma, Ba (2014) uses momentum and adaptive learning rate to perform gradient descent:

$$m^t = \eta_1 m^{t-1} + (1-\eta_1)\,\nabla_{x_t}\mathcal{U}(\mathbf{A}x_t, y), \tag{15}$$

$$v^t = \eta_2 v^{t-1} + (1-\eta_2)\,\nabla_{x_t}\mathcal{U}(\mathbf{A}x_t, y)^2, \tag{16}$$

$$x_{t-1} = x_t - \epsilon \frac{\hat{m}^t}{\sqrt{\hat{v}^t} + \varepsilon}, \tag{17}$$

where (η_1, η_2) are the coefficients used to calculate the exponentially weighted moving averages of gradient and its square, while ε helps improve the numerical stability. With the help of the moving averages m^t and v^t, we can efficiently locate the flat minima.

It is worth noting that the choice of ϵ and gradient policy is dependent on the evaluation function $\mathcal{U}$ (Details and ablation studies can be found in Appendix. C). Our gradient update policies are

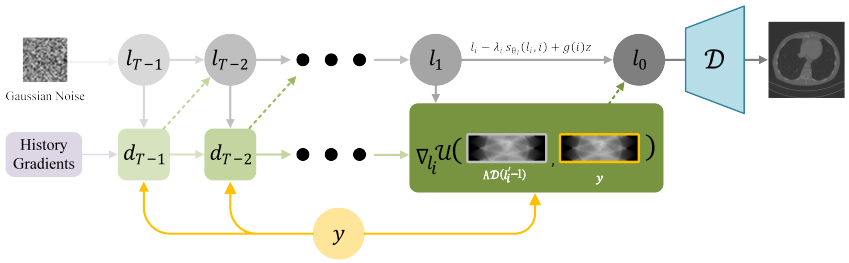

Figure 2: Our LDIR can handle both CT reconstruction and nature image restoration in the same pipeline. We guide latent diffusion models by using the gradients $d_i = \nabla_{\ell_i} \mathcal{U}\left(\mathbf{A}\mathcal{D}\left(\ell_i\right), y\right)$ along with history gradients. The guidance is obtained through gradients, rather than direct projection.

compatible with DIR, LDIR, and DPS whether on pixel space or latent space. The details are demonstrated in the ablation studies. The detail of LDIR algorithms with the above gradient update policies are presented in Appendix A.

3.3 Latent diffusion iterative reconstruction

Our review of previous works on diffusion-based data reconstruction, including Chung et al. (2023b, 2022, 2023a,c); Song et al. (2022); Wang et al. (2023), reveals that they all perform reconstruction in the pixel space, which requires significant computational resources. To address this limitation, we draw inspiration from the Latent Diffusion Models (LDMs) proposed by Rombach et al. (2022) and our DIR. We introduce a novel data reconstruction diffusion model called Latent Diffusion Iterative Reconstruction (LDIR). LDIR offers several advantages over previous methods: (i) Instead of processing images in the pixel space, we encode images into a low-dimensional latent space, enabling us to process them more efficiently with fewer computational demands. (ii) The latent space contains significantly higher information density compared to the pixel space, allowing us to incorporate natural priors such as sparsity and improve the quality of the restored data.

From pixels to latents. To encode pixels to latents, we construct an autoencoder comprising an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$. Specifically, given an input image x in the pixel space, $\mathcal{E}$ maps x to a low-dimensional latent vector $\ell = \mathcal{E}\left(x\right)$. $\mathcal{D}$ reconstructs the image $\bar{x} = \mathcal{D}\left(\mathcal{E}\left(x\right)\right)$ from ℓ. To incorporate the sparsity prior of x, we use the *Vector Quantized Variational Autoencoder* (*VQ-VAE*) proposed by Kingma, Welling (2013); Rezende et al. (2014); Esser et al. (2021) with the quantization layer Van Den Oord et al. (2017).

Score matching for latents. With our semantic compression model $\mathcal{E}$ and $\mathcal{D}$, we can now establish the score of latents $\nabla_\ell \log p_t\left(\ell\right)$ and use score model s_{θ_ℓ} to approx it with the following objective:

$$\min_{\theta} \mathbb{E}_{t\in[0,\ldots,T-1],\ell_0\sim p_\ell,\ell_t\sim p_{0|t}(\ell_t|\ell_0)}\left[\left|\left|s_{\theta_\ell}\left(\ell,t\right)-\nabla_\ell \log p_{0|t}\left(\ell_t|\ell_0\right)\right|\right|_2^2\right], \tag{18}$$

where $p\left(\ell\right) = p\left(\mathcal{E}\left(x\right)\right)$ and $x \sim p\left(x\right)$.

Conditional guidance process on the latent space. Similar to the conditional guidance process in the pixel space, we begin by using the score model s_{θ_ℓ} to generate latents from standard Gaussian noise:

$$\ell'_{t-1} = \ell_t - f\left(\ell,t\right) - g\left(t\right)^2 s_{\theta_\ell}\left(\ell,t\right) + g\left(t\right) z, z \sim \mathcal{N}\left(0,\mathbf{I}\right), \tag{19}$$

We need to use $\mathcal{D}$ to decode ℓ to pixel space, and compute the data consistency term $\mathcal{U}\left(\mathbf{A}\mathcal{D}\left(\ell\right), y\right)$, which can be derived to:

$$\ell_{t-1} = \ell'_{t-1} - \epsilon\nabla_{\ell_t}\mathcal{U}\left(\mathbf{A}\mathcal{D}\left(\ell_t\right), y\right). \tag{20}$$

Thus, we can get the conditional guidance algorithm of LDIR as:

$$\ell_{t-1} \simeq \ell_t - \underbrace{\left(\lambda_t\, s_{\theta_\ell}\left(\ell_t,t\right) - g\left(t\right) z\right)}_{\text{Prior term}} - \underbrace{\epsilon\nabla_{\ell_t}\mathcal{U}\left(\mathbf{A}\mathcal{D}\left(\ell_t\right), y\right)}_{\text{Data-fidelity term}}, z \sim \mathcal{N}\left(0,\mathbf{I}\right). \tag{21}$$

The final results can be obtained by decoding the final latent $x_0 = \mathcal{D}\left(\ell_0\right)$ using the decoder $\mathcal{D}$.

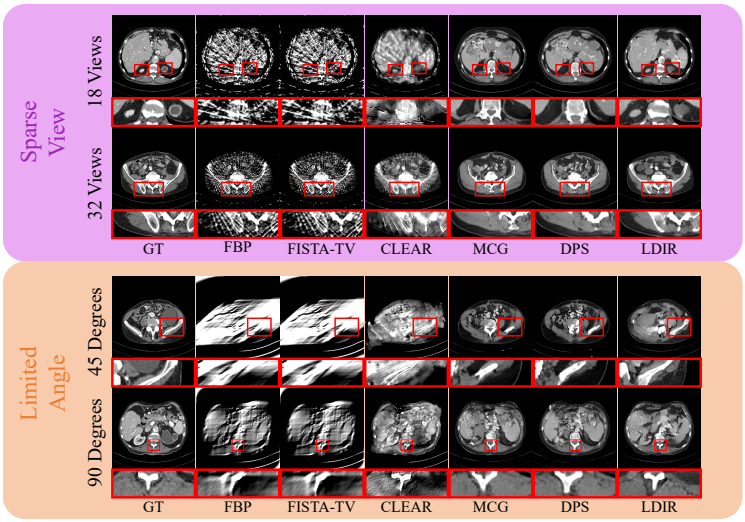

Figure 3: Qualitative results of medical sparse data reconstruction. The resolution of CT images is 256×256 The display window of CT images is set to $[-150, 256]$ HU. The standard measurement of CT is 512 views around 180 degrees.

4 Experiments

4.1 Experimental setup

Models and datasets. For medical image reconstruction, we train our DDPM and LDM model on the 2016 American Association of Physicists in Medicine (AAPM) grand challenge dataset. The dataset has normal-dose data from 10 patients. 9 patients' data are used for training, and 1 for validation which contains 526 images. To simulate low-dose imaging, a parallel-beam imaging geometry with 180 degrees was employed. Regarding inpainting and super-resolution tasks, we test our method on CelebAHQ 1k 256×256 dataset Liu et al. (2015) and LSUN-bedroom 256×256 dataset Yu et al. (2015). We utilize pretrained DDPM and LDM models from the open-source model repository from Ho et al. (2020); Rombach et al. (2022). All the images are normalized to range $[0, 1]$. More details including the hyper-parameters are listed in Appendix. B.

Measurement operators. For sparse-view CT reconstruction, we uniformly sample 18 and 32 views. For limited-angle CT reconstruction, we restrict the imaging degree range to 45 and 90 degrees with 128 views using parallel beam geometry. For random inpainting, following Chung et al. (2022, 2023b), we mask out 99% of the total pixels (including all the channels). For super-resolution, we use $8\times$ bilinear downsampling. Gaussian noise is added in the nature image evaluation after a forward operation performed with $\sigma = 0.05$. The medical data are evaluated without noise.

4.2 Evaluation on medical data

To assess the performance of LDIR in reconstructing medical sparse data, we compare it with several recent state-of-the-art methods: manifold constrained gradients (MCG) Chung et al. (2022), diffusion posterior sampling (DPS) Chung et al. (2023b), comprehensive learning enabled adversarial reconstruction (CLEAR) Zhang et al. (2021), fast iterative shrinkage-thresholding algorithm with total-variation (FISTA-TV), and the analytical reconstruction method, filtered backprojection (FBP). Peak-signal-to-noise-ratio (PSNR) and structural similarity index measure (SSIM) are used for quantitative evaluation.

The quantitative results of medical sparse data reconstruction are demonstrated in Tab. 1 and Tab. 2. Our method outperforms all other state-of-the-art methods by a significant margin across all experiment settings. We also compare our method in the high-resolution CT image reconstruction task with zero-shot methods. However, due to the large memory consumption of DDPM, it is challenging

Table 1: Quantitative evaluation (PSNR, SSIM) of medical image reconstruction on AAPM test 256×256 dataset. We mark **bold** for the best and underline for the second best. CLEAR Zhang et al. (2021) is a supervised method.

Method	Sparse view				Limited angle			
	18		32		45		90	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FBP	24.76	0.5296	28.03	0.6779	16.65	0.5422	20.35	0.5113
FISTA-TV	24.86	0.5408	28.14	0.6888	16.66	0.5463	20.40	0.5241
CLEAR	32.28	0.8798	36.24	0.9257	25.71	0.8559	31.60	**0.9223**
MCG	28.54	0.8135	28.98	0.8242	26.08	0.7418	28.44	0.8079
DPS	28.55	0.8140	28.97	0.8242	28.25	0.8204	28.25	0.8088
LDIR	**39.01**	**0.9552**	**39.77**	**0.9612**	**30.05**	**0.8747**	**32.68**	0.9032

Table 2: Quantitative evaluation (PSNR, SSIM) of medical image reconstruction on AAPM test 512×512 dataset for zero-shot methods. We mark **bold** for the best and underline for the second best.

Method	Sparse view				Limited angle			
	18		32		45		90	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
FBP	23.48	0.5096	26.70	0.6423	16.53	0.5480	19.88	0.4932
FISTA-TV	23.93	0.5566	27.11	0.6768	16.59	0.5695	20.08	0.5348
LDIR	**27.30**	**0.8443**	**27.33**	**0.8441**	**26.18**	**0.8355**	**26.70**	**0.8381**

to train DDPM models for high-resolution reconstruction. Thus, we exclude MCG and DPS which rely on DDPM from Tab. 2. The results show that LDIR provides noise-free reconstruction results, although there is still a significant gap between the reconstructed images and the ground truth. In contrast, other zero-shot methods fail to reconstruct meaningful results.

The qualitative results of medical sparse image reconstruction are demonstrated in Fig. 3 which are consistent with the quantitative results reported in Tab. 1. In Fig. 3, we compare our method with the state-of-the-art zero-shot unsupervised and supervised methods. We observe that our method can provide high-quality reconstructions, especially for the sparse view reconstruction task. Specifically, LDIR can provide better overall structure and nearly artifact-free reconstruction. Additionally, our method also provides better reconstructions than other methods in limited angle reconstruction tasks. (More qualitative results of medical sparse data reconstruction can be found in Appendix. D).

4.3 Evaluation on nature images

In order to further test the performance of our method, we compare our method against state-of-the-art methods, namely, MCG, DPS, and plug-and-play alternating direction method of multipliers (PnP-ADMM) Chan et al. (2016). For quantitative analysis, we utilize three widely-used perceptual evaluation metrics: LPIPS distance, PSNR, and SSIM.

The quantitative results of nature image reconstruction are illustrated in Tab. 3. Our method achieves competitive results compared to the previous state-of-the-art. Specifically, we observe that our method is able to accurately reconstruct the original data and preserve the most data consistency, even when dealing with highly sparse measurements such as 99% random inpainting. Additionally, we note that LDIR gains some advantages over the previous best method on the super-resolution task.

The qualitative results of nature sparse image reconstruction are demonstrated in Fig. 4. Notably, the traditional iterative method PnP-ADMM failed to produce satisfactory results for both the inpainting and super-resolution tasks due to its limited prior terms. In contrast, our method outperforms the comparison methods, particularly in terms of color and structure in the inpainting task. In the super-resolution task, the results obtained by MCG exhibit many artifacts, which are likely due to the projection stepChung et al. (2023b). Our method, on the other hand, achieves competitive results

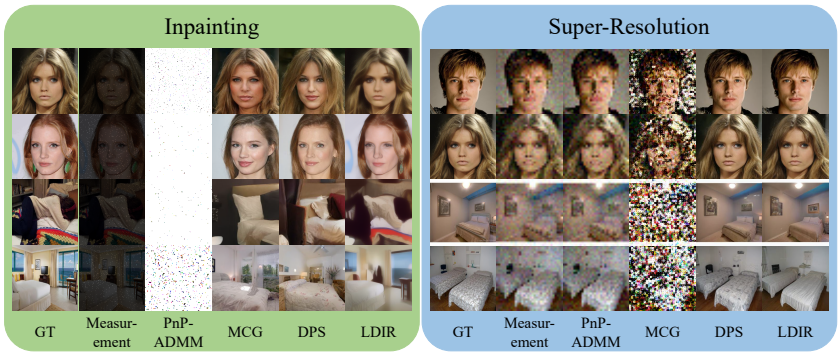

Figure 4: Qualitative results of nature sparse image reconstruction with Gaussian noise ($\sigma = 0.05$).

Table 3: Quantitative evaluation (PSNR, SSIM, LPIPS) of nature image reconstruction on CelebAHQ and LSUN-bedroom dataset. We mark **bold** for the best and underline for the second best.

Method	Type	CelebAHQ						LSUN-bedroom					
		Inpaint			SR (8×)			Inpaint			SR (8×)		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
PnP-ADMM	Traditional IR	3.97	0.3017	0.8916	22.94	0.6303	0.6820	5.059	0.3236	0.8940	**20.14**	0.5458	0.7944
MCG	Pixel Diffusion	18.91	0.5600	0.2544	12.47	0.1655	0.6713	16.89	0.4555	0.5486	9.39	0.0606	0.8698
DPS	Pixel Diffusion	18.95	0.5614	0.2543	24.36	0.7116	0.1089	17.03	0.4587	0.5414	19.15	0.5614	**0.3074**
LDIR	Latent Diffusion	**22.14**	**0.6647**	**0.2280**	**25.27**	**0.7530**	**0.0878**	**20.33**	**0.5845**	**0.4858**	19.83	**0.5762**	0.3253

with DPS, the most advanced method, with small gaps. (More qualitative results of nature sparse image reconstruction can be found in Appendix. E).

4.4 Ablation studies

We conducted ablation studies to validate the effectiveness of our approach. we compared the performance of our latent-based iterative reconstruction approach against a pixel-based iterative reconstruction approach. To ensure a fair comparison, we conducted these ablation studies on the medical image reconstruction task, as both the DDPM and LDM models were trained using the same protocol.

In Table 4, we can observe that our LDIR outperforms both pixel-based iterative reconstruction methods, DPS and DIR, by a large margin. This result confirms that the latent-based approach is superior to the pixel-based approach in terms of both speed and accuracy. Additionally, we can see that changing the evaluation function to $L1$ and using a Momentum-like gradient update policy in the pixel space can improve the performance of DIR, allowing it to surpass DPS. Compared to pixel-space models, LDIR achieves significant speed-up with less memory consumption. Although LDIR decodes latent into pictures at each step, it still has a greater advantage than processing directly in pixel space.

5 Conclusion

In this paper, we propose Latent Diffusion Iterative Reconstruction (LDIR) as a novel approach for reconstructing CT sparse data in a zero-shot manner. We show theoretically that utilizing the latent diffusion models as the prior term is able to ignore the depends for the paired data. In addition, we generate the prior term in the latent space instead of the pixel space, which encourages us to form a high-resolution image with lower computational complexity and sampling time. By using sample-level historical gradient information from the data-fidelity term, we can guide the reconstruction process in the latent space. Our experimental results demonstrate that LDIR outperforms state-of-the-art methods including the supervised method on sparse CT data reconstruction and achieves competitive results on nature image restoration. We believe that our work offers the community a promising tool for leveraging the rapidly growing field of latent diffusion models to restore high-quality and high-resolution data from degraded measurements.

Table 4: Ablation evaluation (PSNR, SSIM) on the effect of latent-based iterative reconstruction. We mark **bold** for the best and underline for the second best. DPS Chung et al. (2023b) is a special case of DIR where $\mathcal{U}(\cdot) \triangleq ||\cdot||_2$ and use naive gradient update policy.

Method	Type	Sparse view				Limited angle				Speed(iter/s) ↑	Memory(MB) ↓
		18		32		45		90			
		PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑		
DPS	Pixel Diffusion	28.55	0.8140	28.97	0.8242	28.25	0.8204	28.25	0.8088	20.88	6338
DIR (Ours)	Pixel Diffusion	31.45	0.8654	32.82	0.8898	27.31	0.8133	28.98	0.8280	20.75	6338
LDIR (Ours)	Latent Diffusion	**39.01**	**0.9552**	**39.77**	**0.9612**	**29.60**	**0.8779**	**32.89**	**0.9116**	**36.67**	**4268**

References

Beck Amir, Teboulle Marc. A fast iterative shrinkage-thresholding algorithm for linear inverse problems // SIAM journal on imaging sciences. 2009. 2, 1. 183–202.

Bora Ashish, Jalal Ajil, Price Eric, Dimakis Alexandros G. Compressed sensing using generative models // International Conference on Machine Learning. 2017. 537–546.

Candès Emmanuel J, Romberg Justin, Tao Terence. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information // IEEE Transactions on information theory. 2006. 52, 2. 489–509.

Chan Stanley H, Wang Xiran, Elgendy Omar A. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications // IEEE Transactions on Computational Imaging. 2016. 3, 1. 84–98.

Chen Hu, Zhang Yi, Chen Yunjin, Zhang Junfeng, Zhang Weihua, Sun Huaiqiang, Lv Yang, Liao Peixi, Zhou Jiliu, Wang Ge. LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT // IEEE transactions on medical imaging. 2018. 37, 6. 1333–1347.

Chung Hyungjin, Kim Jeongsol, Kim Sehui, Ye Jong Chul. Parallel Diffusion Models of Operator and Image for Blind Inverse Problems // IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023a.

Chung Hyungjin, Kim Jeongsol, Mccann Michael Thompson, Klasky Marc Louis, Ye Jong Chul. Diffusion Posterior Sampling for General Noisy Inverse Problems // The Eleventh International Conference on Learning Representations. 2023b.

Chung Hyungjin, Ryu Dohoon, Mccann Michael T, Klasky Marc L, Ye Jong Chul. Solving 3D Inverse Problems using Pre-trained 2D Diffusion Models // IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023c.

Chung Hyungjin, Sim Byeongsu, Ryu Dohoon, Ye Jong Chul. Improving Diffusion Models for Inverse Problems using Manifold Constraints // Advances in Neural Information Processing Systems. 2022.

Donoho David L. Compressed sensing // IEEE Transactions on information theory. 2006. 52, 4. 1289–1306.

Esser Patrick, Rombach Robin, Ommer Bjorn. Taming transformers for high-resolution image synthesis // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. 12873–12883.

Gao Sicheng, Liu Xuhui, Zeng Bohan, Xu Sheng, Li Yanjing, Luo Xiaoyan, Liu Jianzhuang, Zhen Xiantong, Zhang Baochang. Implicit Diffusion Models for Continuous Super-Resolution // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

Ho Jonathan, Jain Ajay, Abbeel Pieter. Denoising diffusion probabilistic models // Advances in Neural Information Processing Systems. 2020. 33. 6840–6851.

Kawar Bahjat, Elad Michael, Ermon Stefano, Song Jiaming. Denoising Diffusion Restoration Models // Advances in Neural Information Processing Systems. 2022.

Kim Hojin, Chen Josephine, Wang Adam, Chuang Cynthia, Held Mareike, Pouliot Jean. Non-local total-variation (NLTV) minimization combined with reweighted L1-norm for compressed sensing CT reconstruction // Physics in Medicine & Biology. 2016. 61, 18. 6878.

Kingma Diederik P, Ba Jimmy. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.

Kingma Diederik P, Welling Max. Auto-encoding variational bayes // arXiv preprint arXiv:1312.6114. 2013.

Liu Ziwei, Luo Ping, Wang Xiaogang, Tang Xiaoou. Deep Learning Face Attributes in the Wild // Proceedings of International Conference on Computer Vision (ICCV). December 2015.

Lugmayr Andreas, Danelljan Martin, Romero Andres, Yu Fisher, Timofte Radu, Van Gool Luc. Repaint: Inpainting using denoising diffusion probabilistic models // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 11461–11471.

Luo Ziwei, Gustafsson Fredrik K, Zhao Zheng, Sjölund Jens, Schön Thomas B. Image Restoration with Mean-Reverting Stochastic Differential Equations // International Conference on Machine Learning. 2023.

Rezende Danilo Jimenez, Mohamed Shakir, Wierstra Daan. Stochastic backpropagation and approximate inference in deep generative models // International conference on machine learning. 2014. 1278–1286.

Rombach Robin, Blattmann Andreas, Lorenz Dominik, Esser Patrick, Ommer Björn. High-resolution image synthesis with latent diffusion models // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. 10684–10695.

Saharia Chitwan, Ho Jonathan, Chan William, Salimans Tim, Fleet David J, Norouzi Mohammad. Image super-resolution via iterative refinement // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022.

Song Yang, Shen Liyue, Xing Lei, Ermon Stefano. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models // International Conference on Learning Representations. 2022.

Song Yang, Sohl-Dickstein Jascha, Kingma Diederik P, Kumar Abhishek, Ermon Stefano, Poole Ben. Score-Based Generative Modeling through Stochastic Differential Equations // International Conference on Learning Representations. 2021.

Sutskever Ilya, Martens James, Dahl George, Hinton Geoffrey. On the importance of initialization and momentum in deep learning // International conference on machine learning. 2013. 1139–1147.

Van Den Oord Aaron, Vinyals Oriol, others . Neural discrete representation learning // Advances in neural information processing systems. 2017. 30.

Wang Yinhuai, Yu Jiwen, Zhang Jian. Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model // International Conference on Learning Representations. 2023.

Yu Fisher, Seff Ari, Zhang Yinda, Song Shuran, Funkhouser Thomas, Xiao Jianxiong. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop // arXiv preprint arXiv:1506.03365. 2015.

Zhang Yikun, Hu Dianlin, Zhao Qianlong, Quan Guotao, Liu Jin, Liu Qiegeng, Zhang Yi, Coatrieux Gouenou, Chen Yang, Yu Hengyong. CLEAR: Comprehensive Learning Enabled Adversarial Reconstruction for Subtle Structure Enhanced Low-Dose CT Imaging // IEEE Transactions on Medical Imaging. 2021. 40, 11. 3089–3101.

Zhao Shiying, Robeltson DD, Wang Ge, Whiting Bruce, Bae Kyongtae T. X-ray CT metal artifact reduction using wavelets: an application for imaging total hip prostheses // IEEE transactions on medical imaging. 2000. 19, 12. 1238–1247.