Fast-DDPM: Fast Denoising Diffusion Probabilistic Models for Medical Image-to-Image Generation

Hongxu Jiang, Muhammad Imran, Teng Zhang, Yuyin Zhou, Muxuan Liang, Kuang Gong, *Member, IEEE*, and Wei Shao, *Senior Member, IEEE*

***Abstract*—Denoising diffusion probabilistic models (DDPMs) have achieved unprecedented success in computer vision. However, they remain underutilized in medical imaging, a field crucial for disease diagnosis and treatment planning. This is primarily due to the high computational cost associated with the use of large number of time steps (e.g., 1,000) in diffusion processes. Training a diffusion model on medical images typically takes days to weeks, while sampling each image volume takes minutes to hours. To address this challenge, we introduce Fast-DDPM, a simple yet effective approach capable of simultaneously improving training speed, sampling speed, and generation quality. Unlike DDPM, which trains the image denoiser across 1,000 time steps, Fast-DDPM trains and samples using only 10 time steps. The key to our method lies in aligning the training and sampling procedures to optimize time-step utilization. Specifically, we introduced two efficient noise schedulers with 10 time steps: one with uniform time step sampling and another with non-uniform sampling. We evaluated Fast-DDPM across three medical image-to-image generation tasks: multi-image super-resolution, image denoising, and image-to-image translation. Fast-DDPM outperformed DDPM and current state-of-the-art methods based on convolutional networks and generative adversarial networks in all tasks. Additionally, Fast-DDPM reduced the training time to 0.2× and the sampling time to 0.01× compared to DDPM.**

***Index Terms*—Conditional diffusion models, deep learning, fast-DDPM, image-to-image generation.**

Received 19 December 2024; revised 23 March 2025 and 23 April 2025; accepted 25 April 2025. Date of publication 28 April 2025; date of current version 7 October 2025. This work was supported by the Department of Medicine and the Intelligent Clinical Care Center at the University of Florida College of Medicine. *(Corresponding author: Wei Shao.)*

This study used only publicly available human imaging datasets and did not involve any data collected directly from human subjects by the authors.

Hongxu Jiang and Teng Zhang are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32610 USA (e-mail: hongxu.jiang@ufl.edu; teng.zhang@ufl.edu).

Muhammad Imran is with the Department of Medicine, University of Florida, Gainesville, FL 32610 USA (e-mail: muhammad.imran@ufl.edu).

Yuyin Zhou is with the Department of Computer Science and Engineering, University of California, Santa Cruz, CA 95064 USA (e-mail: yzhou284@ucsc.edu).

Muxuan Liang is with the Department of Biostatistics, University of Florida, Gainesville, FL 32610 USA (e-mail: muxuan.liang@ufl.edu).

Kuang Gong is with the Department of Biomedical Engineering, University of Florida, Gainesville, FL 32610 USA (e-mail: kgong@bme.ufl.edu).

Wei Shao is with the Department of Electrical and Computer Engineering, Department of Medicine, and Intelligent Clinical Care Center, University of Florida, Gainesville, FL 32610 USA (e-mail: weishao@ufl.edu).

Our code is publicly available at: https://github.com/mirthAI/Fast-DDPM.

Digital Object Identifier 10.1109/JBHI.2025.3565183

I. Introduction

DIFFUSION models [10], [19], [47], [49] have become powerful tools for high-quality image generation [6]. The forward diffusion process incrementally adds noise to a high-quality image until it becomes pure Gaussian noise. An image denoiser is then trained to learn the reverse process, progressively removing noise from random Gaussian noise until it reconstructs a noise-free image. This enables the generation of new high-quality images that match the distribution of the training dataset. Beyond computer vision, diffusion models are increasingly applied to medical imaging [12], [16], [37], [52] for improved disease diagnosis, treatment planning, and patient monitoring.

Training and sampling diffusion models are computationally expensive and time-consuming due to the large number of discrete time steps used to approximate the continuous diffusion process, as well as the need for large datasets to model complex image distributions. For 2D diffusion models, training can take days on a single GPU, limiting experimentation with different model architectures and hyperparameters. Sampling a diffusion model can take several minutes per 2D image, presenting challenges for real-time usability when applying 2D diffusion models to 3D image volumes consisting of hundreds of 2D slices, where the process may take several hours per volume.

Current research mainly focuses on accelerating the sampling process, using either training-free or training-based methods [60]. Training-free methods utilize efficient numerical solvers for stochastic or ordinary differential equations to reduce the number of time steps required for sampling. This approach can decrease the number of sampling time steps from 1,000 to 50–100, without compromising the quality of the generated images. Alternatively, training-based methods, such as progressive knowledge distillation [32], [43], can further reduce the sampling steps to as few as 10. However, this approach requires the training of additional student models, which considerably

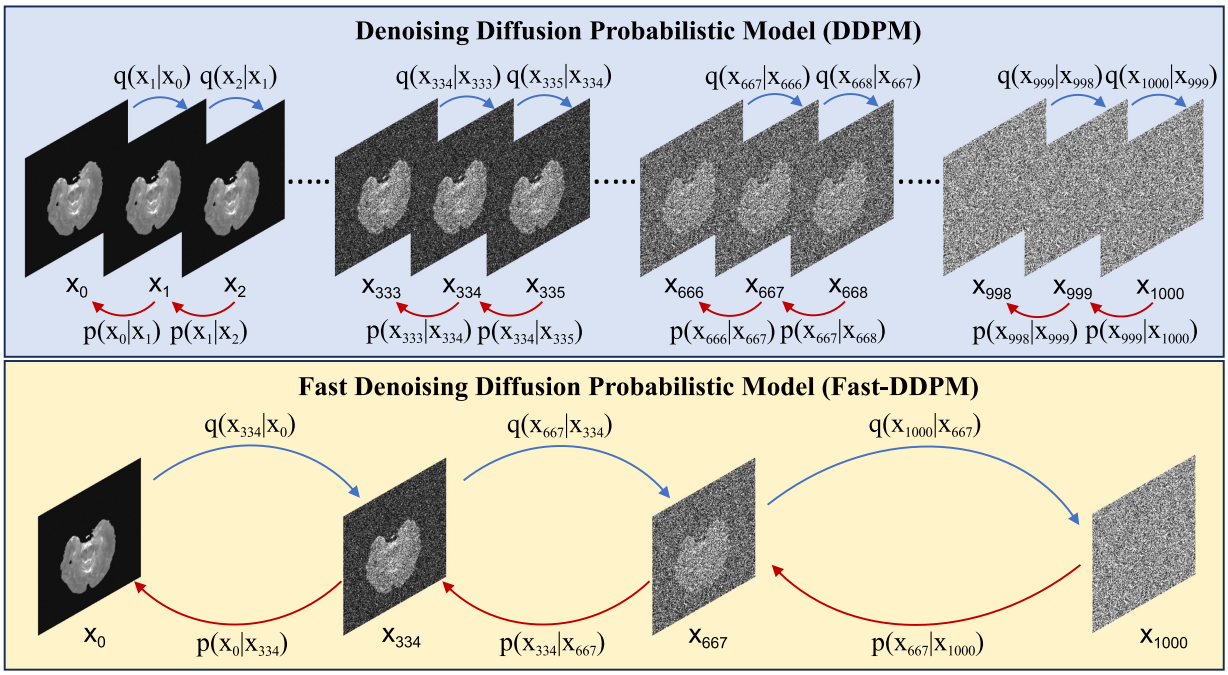


Fig. 1. Fast-DDPM uses significantly fewer time steps than DDPM. For illustration purposes, we used only 3 time steps for Fast-DDPM in this example.

increases the overall training cost and may not be practical for medical image analysis.

In this paper, we propose Fast-DDPM, a model designed to reduce both training and sampling times by optimizing time-step utilization. We observed that the majority of time steps involved in training are not utilized by faster diffusion samplers during sampling. For instance, a Denoising Diffusion Probabilistic Model (DDPM) trained with 1,000 time steps is often sampled by the Denoising Diffusion Implicit Model (DDIM) using only 100 time steps. This practice is resource-inefficient, as the image denoiser is trained on 900 time points that are ultimately skipped during sampling. To address this inefficiency in training, we propose training and sampling the DDPM model at the same 10 time steps, as illustrated in Fig. 1.

We evaluated Fast-DDPM on three medical image-to-image tasks, including multi-image super-resolution for prostate MRI, denoising low-dose CT scans, and image-to-image translation for brain MRI. Fast-DDPM achieved state-of-the-art performance across all three tasks. Notably, Fast-DDPM is approximately 100 times faster than the DDPM model during sampling and about 5 times faster during training. We summarize the major contributions of this paper as the following:

- Introduction of Fast-DDPM: A simple yet effective approach that simultaneously improves training speed, sampling speed, and generation quality.
- Novel Training Strategy: Optimizes time-step utilization by aligning the training and sampling procedures, significantly enhancing both speed and accuracy.
- Efficient Noise Schedulers: Designed two noise schedulers with 5–10 time steps—one utilizing uniform sampling and the other non-uniform sampling—tailored for specific tasks to enhance performance.

II. Related Work

A. Fast Sampling of Diffusion Models

Current methods to accelerate the sampling of diffusion models can be broadly classified into two categories: training-free and training-based. Training-free methods primarily focus on developing efficient diffusion solvers to solve reverse stochastic differential equations (SDEs) or their equivalent ordinary differential equations (ODEs). For instance, the Denoising Diffusion Implicit Models (DDIM) approach [47] models the reverse diffusion process using an ODE and significantly reduces the number of required sampling steps to 50–100 without impacting the quality of the generated samples. Building on this, Fast-DPM [21] establishes a bijective mapping between continuous diffusion steps and continuous noise levels and uses this mapping to construct an approximation of the reverse process between DDPM and DDIM with fewer steps. DPM-Solver [26] introduces a high-order solver for diffusion ODEs based on an exact solution formulation that simplifies to an exponentially weighted integral, achieving high-quality sample generation with just 10–50 sampling steps. The Differentiable Diffusion Sampler Search [54] sampler uses the reparametrization trick and gradient rematerialization to optimize over a parametric family of fast samplers for diffusion models by differentiating through sample quality scores. In addition to these solver-based methods, quantization emerges as another promising training-free method for improved diffusion model sampling. PTQ4DM [45] introduces Post-Training Quantization (PTQ) with a novel calibration method named NDTC, aimed at reducing the computational cost of noise estimation. Concurrently, Q-Diffusion [23] proposes timestep-aware calibration alongside split shortcut quantization, further refining the quantization process for efficiency gains. DeepCache [30] avoids computing deep features repeatedly in

the network by caching them, thereby accelerating the sampling process.

A common training-based approach utilizes knowledge distillation, wherein a teacher diffusion model is first trained with a large number of time steps, followed by the sequential training of multiple student models with progressively fewer time steps, mirroring the teacher model's behavior [32], [43]. This methodology reduces inference time steps from 1,000 to as few as 16 while maintaining sampling quality. Recently, adversarial diffusion distillation [44] has enabled real-time, single-step image synthesis by combining score distillation and adversarial loss, achieving state-of-the-art quality with just 1–4 sampling steps. Furthermore, the rectified flow technique achieves high-quality sampling in a single time step by straightening the trajectory of the reverse-time ODE [25], and consistency models [48] produce superior sampling quality within a single step by directly mapping noise to an image, either through distillation from pre-trained diffusion models or as standalone generative models.

B. Diffusion Models in Medical Imaging

Diffusion models have been applied to various medical imaging tasks, including segmentation [12], [17], [57], anomaly detection [14], [55], [58], denoising [5], [8], [29], [59], reconstruction [2], [9], [24], registration [16], super-resolution [5], [29], [52], [56], and image-to-image translation [18], [28], [37]. Among these, the DDPM model is the most widely used in medical imaging due to its simplicity. Diffusion models can be categorized as 2D, 3D, or 4D based on input image dimensions. Currently, 2D models are the predominant choice in medical imaging due to their ease of implementation and lower memory requirements. However, when implementing diffusion models for 3D and 4D applications, training can take significantly longer, lasting anywhere from weeks to months. Improving the training and sampling processes is crucial for a broader adoption of diffusion models in medical imaging.

C. Unconditional and Conditional Diffusion Models

Diffusion models can be broadly categorized into unconditional and conditional models. Unconditional diffusion models generate data samples based solely on the learned data distribution, without incorporating external information or conditions [10], [36], [47]. In contrast, conditional diffusion models generate data samples guided by additional input conditions, such as text, class labels, or images, enabling controlled and contextually relevant outputs instead of random samples. Recent advancements in text-guided diffusion models, such as Glide [35], DALL-E 2 [38], and Imagen [41], have demonstrated exceptional capabilities in generating high-quality, contextually relevant images based on textual descriptions. These models leverage semantic understanding to bridge the gap between language and visual representation, enabling applications in creative design, personalized content creation, and advertising. Image-guided diffusion models, including SR3 [42], Palette [40], and Srdiff [22], specialize in transforming or enhancing existing images through super-resolution, inpainting, and style transfer. Their ability to generate realistic and detailed outputs makes them valuable for tasks requiring dense image-to-image predictions. The latent diffusion model [39] combines the strengths of both text and image guidance by operating in a lower-dimensional latent space. This approach improves computational efficiency while maintaining high generation quality, making it well-suited for resource-intensive tasks and expanding practical applications in high-quality image synthesis.

III. Methods

A. Background

a) Diffusion Models: Starting from a noise-free image x_0, the forward diffusion process is a continuously-time stochastic process from time $t = 0$ to $t = 1$, incrementally introducing noise to the image x_0 until it becomes pure Gaussian noise at $t = 1$. The distributions of the intermediate noisy images $x(t)$ are given by [19]:

$$q(x(t)|x_0) = \mathcal{N}(\alpha(t)x_0, \sigma^2(t)\mathbb{I}) \tag{1}$$

where $x(0) = x_0$, $\alpha(t)$ and $\sigma(t)$ are differentiable functions, and the signal-to-noise ratio function $\text{SNR}(t) := \frac{\alpha^2(t)}{\sigma^2(t)}$ decreases monotonically from $+\infty$ at $t = 0$ to 0 at $t = 1$. An equivalent formulation of $x(t)$ is given by the following stochastic differential equation (SDE):

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

where $f(t) = \frac{\alpha'(t)}{\alpha(t)}$, $g(t) = \sqrt{(\sigma^2(t))' - 2\frac{\alpha'(t)}{\alpha(t)}\sigma^2(t)}$, and $W(t)$ is a standard Wiener process from $t = 0$ to $t = 1$. It has been shown that the reverse process of $x(t)$ is given by the following reverse-time SDE [49]:

$$dx(t) = [f(t) - g^2(t)\nabla_{x(t)} \log p(x(t))dt + g(t)d\bar{W}(t) \tag{3}$$

where $x(1) = \mathcal{N}(0, \mathbb{I})$, $d\bar{W}(t)$ the standard Wiener process backwards from $t = 1$ to $t = 0$, and $p(x(t))$ denote the probability density of $x(t)$. If the score function $\nabla_{x(t)} \log p(x(t))$ is known for every $t \in [0, 1]$, one can solve the reverse-time SDE to sample new images.

We can train a model $s_\theta(x(t), t)$ to estimate the score function using the following loss function:

$$\mathop{\mathbb{E}}_{t\in[0,1],x_0\sim p_{\text{data}},x(t)\sim q(x(t)|x_0)} \lambda(t) \left\| s_\theta(x(t), t) - \nabla_{x(t)} \log p(x(t)|x_0) \right\|_2^2 \tag{4}$$

where $p(x(t)|x_0)$ denote the density function of $x(t)$ generated in (1) and $\lambda(t)$ is a scalar-valued weight function. The training of $s_\theta(x(t), t)$ is straightforward since $q(x(t)|x_0)$ is a Gaussian distribution and its score function has a closed form:

$$\nabla_{x(t)} \log p(x(t)|x_0) = -\frac{x(t) - \alpha(t)x_0}{\sigma^2(t)} = -\frac{\epsilon}{\sigma(t)} \tag{5}$$

where ϵ is a random noise sample from $\mathcal{N}(0, \mathbb{I})$ to generate $x(t)$ from x_0. Equivalently, we can use a U-Net denoiser $\epsilon_\theta(x(t), t)$ to estimate the random noise ϵ (i.e., the score function scaled by $-\sigma(t)$). Following [10], in this paper, we trained the U-Net denoiser using a simple loss function:

$$\mathop{\mathbb{E}}_{i\in[1,\ldots,T],x_0\sim p_{\text{data}},\epsilon\sim\mathcal{N}(0,\mathbb{I})} [\epsilon - \epsilon_\theta(\alpha(t)x_0 + \sigma(t)\epsilon, t)]^2 \tag{6}$$

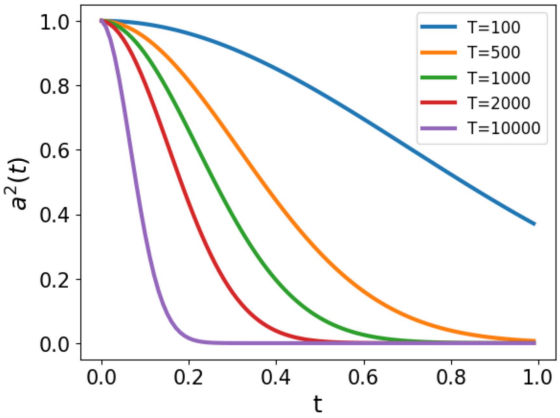


Fig. 2. Impact of the number of time steps, T, on the noise scheduler.

where T is the total number of time steps and $t = \frac{i}{T}$.

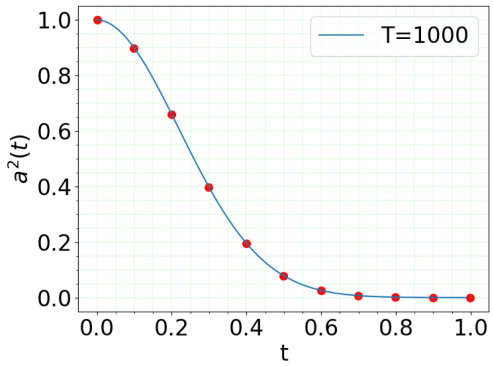
Linear-β Noise Scheduler Any noise scheduler, defined by $\alpha(t)$ and $\sigma(t)$), should ensure that $\text{SNR}(1) = \frac{\alpha^2(1)}{\sigma^2(1)} \approx 0$ so that $x(1)$ is pure Gaussian noise. The variance-preserving noise schedulers [19] meet this requirement, where $\alpha(t)$ is a monotonically decreasing function, $\alpha(0) = 1$, $\alpha(1) = 0$, and $\alpha^2(t) + \sigma^2(t) = 1$ for all t. This formulation guarantees that the variance of any latent image $x(t)$ is bounded. The most popular variance-persevering noise scheduler is the linear-β noise scheduler used in the DDPM model [10]. In this approach, the forward diffusion process was defined as a Markovian process at discrete time steps $i \in \{1, 2, \ldots, T\}$ using the following transition kernel:

$$q\left(x\left(\frac{i}{T}\right)|x\left(\frac{i-1}{T}\right)\right) = \mathcal{N}\left(\sqrt{1-\beta\left(\frac{i}{T}\right)}x\left(\frac{i-1}{T}\right), \beta\left(\frac{i}{T}\right)\mathbb{I}\right) \tag{7}$$

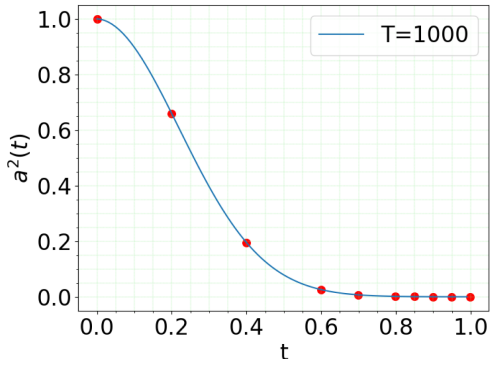
where $t = \frac{i}{T}$, $\beta(t) = 0.0001 + (0.02 - 0.0001)t$, and the corresponding α noise scheduler $\alpha(t)$ at discrete time steps is given by

$$\alpha^2\left(\frac{i}{T}\right) = \prod_{j=1}^{i}\left(1-\beta\left(\frac{j}{T}\right)\right) \tag{8}$$

To establish feasible noise schedulers using (8), the total number of time steps T should be carefully chosen. Fig. 2 plots $\alpha^2(t)$ for different choices of T: 100, 500, 1000, 2000, and 10000. As we can see, the total number of time steps cannot be too small (e.g., $T = 100$), as this results in the endpoint of the forward diffusion process $x(1)$ not being purely Gaussian. Conversely, if the total number of time steps is too large (e.g., $T = 10000$), $x(t)$ becomes pure Gaussian noise at an early time point (e.g., $t = 0.2$). This makes the training of $\epsilon_\theta(x(t), t)$ at time steps $t > 0.2$ unnecessary since there is no change in the noise level of $x(t)$ beyond this point. The plots in Fig. 2 suggest that a good choice of T lies between 500 and 2000. This explains why a linear-β noise scheduler with $T = 1000$ is currently the most popular choice [10]. In other words, we cannot reduce the number of time steps to 10 using (8). In this study, we first define the continuous function $\alpha^2(t)$ and then sample it at discrete time points.



(a) The Uniform Sampling



(b) The Non-Uniform Sampling

Fig. 3. Proposed α noise schedulers with uniform and non-uniform sampling between [0,1].

B. Fast Denoising Diffusion Probabilistic Model (Fast-DDPM)

We have developed the Fast-DDPM model to accelerate the training and sampling of DDPM by utilizing task-specific noise schedulers with only 5–10 time steps.

Task-Specific α Noise Scheduler Previous studies indicate that images of different sizes require different noise schedulers for optimal results [4], [11]. Inspired by this, we propose two α noise schedulers for various medical image-to-image tasks. We first define a smooth, monotonically decreasing function $\alpha^2(t)$ from $t = 0$ to $t = 1$, with boundary conditions $\alpha^2(0) = 1$ and $\alpha^2(1) = 0$. In this paper, we use $\alpha^2(t)$ as defined by the linear-β noise scheduler with $T = 1,000$ in (8) (see the green curve in Fig. 2). Unlike the linear-β noise scheduler used in the DDPM model, which uniformly samples 1,000 time points between [0,1], our first proposed noise scheduler uniformly samples 10 time steps between [0,1] (Fig. 3(a)). The second proposed noise scheduler non-uniformly samples 10 time points between [0,1](Fig. 3(b)). To optimize efficiency, we designate time step 699 as a boundary, sampling 60% of the total time steps uniformly from $t > 699$ and the remaining 40% from $t \leq 699$. While we do not explicitly define a closed-form formula for the sampling rate as a function of noise level, this heuristic approach

Algorithm 1: Training Fast-DDPM.

repeat
  $i \in \{1, ..., T\}$
  $(x_0, c) \sim p_{\text{joint}(x_0,c)}$
  $\epsilon \sim \mathcal{N}(0, \mathbb{I})$
  $t = \frac{i}{T}$
  Gradient descent: $\nabla_\theta ||\epsilon - \epsilon_\theta(\alpha(t)x_0 + \sigma(t)\epsilon, c, t)||^2$
until *converged*;

Algorithm 2: Sampling Fast-DDPM.

$x(1) \sim \mathcal{N}(0, \mathbb{I})$
$c \sim p_c(c)$
for $i = T, \cdots, 1$ **do**
  $t = \frac{i}{T}$
  $x(t - \frac{1}{T}) = \frac{\alpha(t-\frac{1}{T})}{\alpha(t)} x(t) + [\sigma(t-\frac{1}{T}) - \frac{\alpha(t-\frac{1}{T})}{\alpha(t)}\sigma(t)]\epsilon_\theta\big(x(t), c, t\big)$
end
return $x(0)$

follows prior observations [4] that later time steps contribute more to perceptual quality. Our formulation of noise schedulers satisfies all the desirable properties outlined earlier. Due to the reduced training and sampling time, users have the opportunity to try both noise schedulers for each image-to-image task to determine which one fits their task better.

We provide a rationale for our design. In a related context of diffeomorphic image registration, 10 time points are sufficient to solve an ODE for accurately estimating a complex time-dependent image flow between two images [34], [46]. Since we can sample at only 10 discrete time points during inference, we hypothesize that we only need to train the denoiser ϵ_θ at those 10 time steps, not necessarily at all the 1,000 time points.

Training and Sampling of Fast-DDPM In this paper, Fast-DDPM is applied to conditional image-to-image generation. For each image x_0 sampled from $p_{\text{data}}(x_0)$, we assume there are one or more associated condition images, denoted by c. Here, c may represent either a single image or multiple images, depending on the specific application. We denote the joint distribution of x_0 and c as $p_{\text{joint}}(x_0, c)$ and the marginal distribution of the conditional image as $p_c(c)$. We will train a conditional image denoiser $\epsilon_\theta(x(t), c, t)$ that takes the condition c as an additional input to guide the estimation of the score function at each time t. We used the following simple loss function that assigns equal weights to different times t:

$$\mathop{\mathbb{E}}_{i,(x_0,c),\epsilon} ||\epsilon - \epsilon_\theta(\alpha(t)x_0 + \sigma(t)\epsilon, c, t)||^2 \tag{9}$$

where $T = 10, i \in \{1, 2, \ldots, T\}, t = \frac{i}{T}, (x_0, c) \sim p_{\text{joint}}(x_0, c)$, and $\epsilon \sim \mathcal{N}(0, \mathbb{I})$.

The sampling process starts by randomly sampling a Gaussian noise $x(1) \sim \mathcal{N}(0, \mathbb{I})$ and a condition image $c \sim p_c(c)$. The DDIM sampler [47] was used to solve the reverse-time ODE to obtain the noise-free output x_0 that corresponds to c. We summarize the training and sampling processes of Fast-DDPM in Algorithm 1 and Algorithm 2, respectively.

In conditional diffusion-based models, noisy ground truth is used only during training to help the model learn the data distribution across different noise levels. However, during inference, the model generates the output by denoising a random Gaussian noise sample, while being guided at each step by a condition image that supplies structural information. Thus, the use of noisy ground truth in training does not impact real-world applicability.

IV. Experiments

We evaluated our Fast-DDPM model on three medical image-to-image tasks: multi-image super-resolution, image denoising, and image-to-image translation.

A. Datasets

Prostate MRI Dataset for Multi-Image Super-Resolution: We used T2-weighted (T2w) prostate MRI scans from the publicly available Prostate-MRI-US-Biopsy dataset [50]. The in-plane resolution of each MRI is 0.547 mm × 0.547 mm, with a 1.5 mm distance between adjacent slices. An image triplet, comprising three consecutive MRI images, serves as a single training or testing example. In this setup, the first and third slices serve as inputs to the model, while the middle slice is considered as the ground truth. The goal of this task is to predict the missing information in 3D between any two adjacent slices, thereby enhancing through-plane image resolution. During training, triplets of consecutive MRI slices are extracted, with the first and third slices serving as inputs to the model and the middle slice as the ground truth. A forward diffusion process adds noise to the middle slice, and the denoiser is trained to reconstruct it from the noisy version using the first and third slices as conditions. We used a total of 6979 image triplets from 120 MRI volumes for training and 580 image triplets from 10 MRI volumes for testing. All MRI slices were resized to 256 × 256 and normalized to the range [−1, 1].

Low-Dose and Full-Dose Lung CT Dataset for Image Denoising We used paired low-dose and normal-dose chest CT image volumes from the publicly available LDCT-and-Projection-data dataset [31]. The normal-dose CT scans were acquired at routine dose levels, while the low-dose CT scans were reconstructed using simulated lower dose levels, specifically at 10% of the routine dose. In this task, the goal is to generate a normal-dose CT scan from its corresponding low-dose scan. The low-dose CT serves as the condition, and the normal-dose scan is used as the ground truth. During training, a noisy version of the ground truth is created via the forward diffusion process and provided to the denoiser as input, with the corresponding low-dose scan serving as the condition. We randomly selected 38 patients for training and the 10 patients for evaluation, comprising 13,211 and 3,501 2D low-dose and normal-dose CT image pairs, respectively. All images were resized to 256 × 256 and normalized to [−1, 1].

BraTS Brain MRI Dataset for Image-to-Image Translation We used registered T1-weighted (T1w) and T2-weighted (T2w) MR images from the publicly available BraTS 2018 dataset [33]. All images were resampled to a resolution of 1 mm^3 and skull-stripped. In this task, the goal is to convert T1w MRI into T2w MRI, where the T1w MRI serves as the condition and the T2w scan as the ground truth. During training, the denoiser is trained using the noisy ground truth as input, conditioned on the corresponding T1w MRI. The dataset, obtained from [20], comprised 5760 pairs of T1w and T2w MR images for training and 768 pairs for testing. All MRI slices were padded from 224 $\times$ 224 to 256 $\times$ 256 and normalized to $[-1, 1]$.

B. Evaluation Metrics

In line with prior work in medical image-to-image generation [1], [20], [51], we used the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) to evaluate the similarity between the generated images and the ground truth images across all three tasks. PSNR measures the ratio between the maximum possible value (power) of the ground truth image and the power of distorting noise:

$$PSNR = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{10}$$

where MAX_Idenotes the maximum possible pixel value of the image and MSE denotes the mean squared error. SSIM [53] is a widely used metric that measures the structural similarity between two images and aligns better with human perception of image quality. SSIM is defined as:

$$SSIM(x, \hat{x}) = \frac{2\mu_x\mu_{\hat{x}} + c_1}{\mu^2 x + \mu^2 \hat{x} + c_1} \cdot \frac{2\sigma_{x\hat{x}} + c_2}{\sigma^2 x + \sigma^2 \hat{x} + c_2} \tag{11}$$

where $\mu_x, \mu_{\hat{x}}$ and $\sigma_x, \sigma_{\hat{x}}$ are the means and standard deviations of image x and image $\hat{x}$, respectively; $\sigma_{x\hat{x}}$ denotes the covariance of x and $\hat{x}$. The value of c_1 is $(k_1 L)^2$ and $c_2 = (k_2 L)^2$, where k_1 is 0.01, k_2 is 0.03, and L denotes the largest pixel value of the image x.

C. Training Details

The hyperparameter settings and the U-Net architectures of Fast-DDPM follow those of DDIM [47]. Training was conducted using the Adam optimizer with a learning rate of 2×10^{-4}. A linear β noise scheduler with $T = 1000$, $\beta_1 = 10^{-4}$, and $\beta_T = 0.02$, was used to generate the uniform and non-uniform noise schedulers presented in Fig. 3. For the image super-resolution task, we used the noise scheduler with non-uniform sampling, and for the denoising and translation tasks, we used the noise scheduler with uniform sampling. We used a training batch size of 16 and trained the Fast-DDPM model for 400,000 iterations and the DDPM model for 2 million iterations. All experiments were conducted on a computing node with 4 NVIDIA A100 GPUs, each with 80 GB memory. We used Python 3.10.6 and PyTorch 1.12.1 for all experiments .

V. Results

For each image-to-image task, we compare Fast-DDPM with a GAN-based method, a CNN-based method, and DDPM [10] model with ancestral sampling, and two fast sampling methods, DDIM [47] and DPM-Solver++ [27]. Additionally, to facilitate a more granular analysis, we conduct ablation studies to investigate the impact of the number of time steps and the type of noise scheduler used across all three tasks.

TABLE I
Comparison of the Performance and Inference Time Per Image Volume (An Average of 58 Slices) for Various Multi-Image Super-Resolution Methods

Method	PSNR	SSIM	Training time	Inference Time
miSRCNN [7]	26.5	0.87	1 h	0.01 s
miSRGAN [51]	26.8	0.88	40 h	0.04 s
DDPM [10]	25.3	0.83	136 h	3.7 m
DDIM [47]	26.5	0.88	26 h	2.3 s
DPM-Solver++ [27]	26.5	0.87	26 h	2.3 s
Fast-DDPM	**27.1**	**0.89**	26 h	2.3 s

TABLE II
Comparison of the Performance and Inference Time Per Image Volume (An Average of 360 Slices) for Various CT Image Denoising Methods

Method	PSNR	SSIM	Training time	Inference Time
REDCNN [3]	36.4	0.91	3 h	0.5 s
DU-GAN [13]	36.3	0.90	20 h	3.8 s
DDPM [10]	35.4	0.87	141 h	21.4 m
DDIM [47]	37.4	0.91	26 h	12.5 s
DPM-Solver++ [27]	37.0	0.90	26 h	12.5 s
Fast-DDPM	**37.5**	**0.92**	26 h	12.5 s

A. Multi-Image Super-Resolution

Results in Table I show that Fast-DDPM outperformed miSRCNN [7] and miSRGAN [51], as well as all three diffusion-based methods: DDPM, DDIM, and DPM-Solver++. Notably, Fast-DDPM significantly improved upon DDPM, increasing PSNR from 25.3 to 27.1 and SSIM from 0.83 to 0.89, while reducing training time from 136 hours to 26 hours and per-volume inference time from 3.7 minutes to 2.3 seconds. Although DPM-Solver++ and DDIM offer similar computational efficiency to Fast-DDPM, their performance in terms of PSNR and SSIM is inferior.

Fig. 4 shows the results of various image super-resolution methods on a representative subject. The models received the previous and next slices as inputs, with the center slice serving as the ground truth. It is evident that the miSRCNN and miSRGAN methods failed to accurately reconstruct the shadowed area highlighted by the arrows. In contrast, Fast-DDPM most effectively reconstructed the information missing between the previous and next 2D slices. Interestingly, although DDPM is qualitatively inferior to miSRCNN and miSRGAN, it visually outperformed both methods. While DDIM and DPM-Solver++ also demonstrated improvements over DDPM, Fast-DDPM consistently exhibited superior visual and quantitative performance.

B. Image Denoising

The results presented in Table II indicate that Fast-DDPM significantly outperformed DDPM, as well as two other prominent CT denoising methods: REDCNN [3] and DU-GAN [13]. In terms of computational efficiency, Fast-DDPM reduced

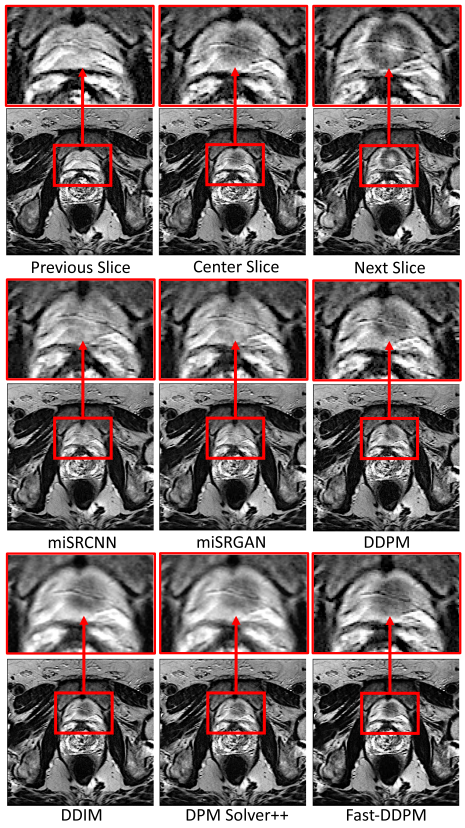


Fig. 4. Qualitative results of the MR multi-image super-resolution task.

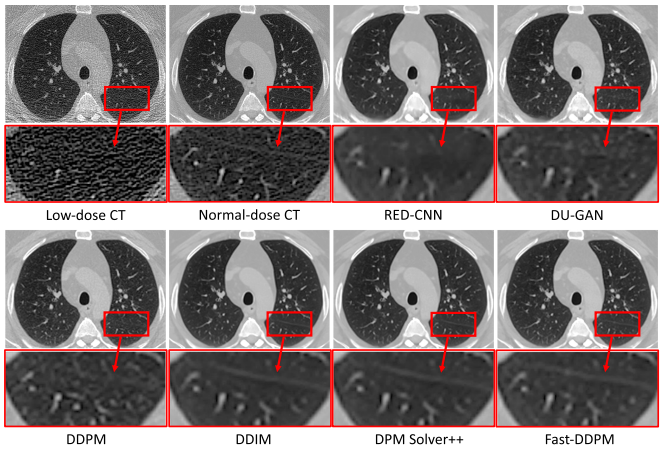


Fig. 5. Qualitative results of the CT image denoising task.

inference time by approximately 100-fold and training time by 5-fold compared to DDPM. Furthermore, compared to prior state-of-the-art fast-sampling methods, DPM-Solver++ and DDIM, Fast-DDPM achieved higher PSNR and SSIM scores, setting a new benchmark.

Fig. 5 shows the low-dose, normal-dose, and predicted normal-dose CT images generated by various models for a representative subject. The zoomed-in region, highlighted by red rectangular boxes, demonstrates the ability of our Fast-DDPM model to preserve fine structural details in the denoised images compared to other methods. Notably, the lung fissure appears clearest and sharpest in the image produced by Fast-DDPM,

TABLE III
COMPARISON OF THE PERFORMANCE AND AVERAGED INFERENCE TIME PER BATCH (360 SLICES) FOR VARIOUS IMAGE-TO-IMAGE TRANSLATION METHODS

Method	PSNR	SSIM	Training time	Inference Time
Pix2Pix [15]	25.6	0.85	6 h	3.3 s
RegGAN [20]	26.0	0.86	9 h	3.1 s
DDPM [10]	**26.3**	**0.89**	135 h	22.2 m
DDIM [47]	26.2	0.88	27 h	13.2 s
DPM-Solver++ [27]	26.2	0.84	27 h	13.2 s
Fast-DDPM	**26.3**	**0.89**	27 h	13.2 s

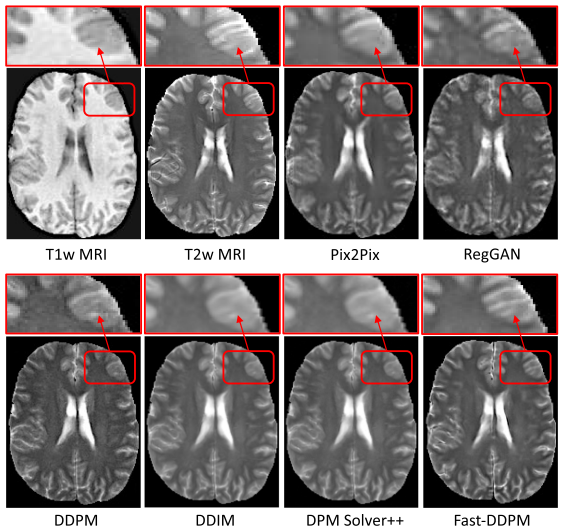


Fig. 6. Qualitative results of the T1w MRI to T2w MRI translation task.

while other methods (except for DDIM and DPM-Solver++) exhibit noticeable blurring and loss of detail in this region. A pulmonary fissure is the boundary between lobes of the lung. Preserving fissure details enables automated segmentation of the lung lobes, facilitating more detailed analysis of lung diseases at the lobar level.

C. Image-to-Image Translation

As shown in Table III, Fast-DDPM outperformed all other leading methods on the image-to-image translation task, including Pix2Pix [15], RegGAN [20], DDPM, DDIM, and DPM-Solver++. While DDPM achieved comparable performance, it required a total training time of 135 hours and 22.2 minutes per batch for sampling. In contrast, Fast-DDPM significantly reduced the training time to 27 hours and the sampling time to 13.2 seconds per batch. These improvements highlight the practical advantages of Fast-DDPM in real-world applications.

Fig. 6 shows the T2w MR images generated by various image-to-image translation methods for a representative subject. Notably, the predictions from Pix2Pix and RegGAN appear blurry and fail to reconstruct the intricate brain structures highlighted by the arrows. DDIM and DPM-Solver++ also produce blurry predictions and struggle to recover fine details effectively. Although DDPM restores some structural details, the reconstructed T2w MRI exhibits a relatively low signal-to-noise ratio. In contrast, Fast-DDPM produces high-quality translations with

TABLE IV
THE IMPACT OF THE NUMBER OF TIME STEPS ON DIFFERENT IMAGE-TO-IMAGE TASKS

Number of Time Steps	Super-Resolution		Denoising		Translation	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3	24.8	0.79	32.7	0.83	25.3	0.86
5	26.9	0.88	37.4	0.92	25.8	0.91
7	27.0	0.88	37.4	0.92	25.9	0.91
10	**27.1**	**0.89**	**37.5**	**0.92**	26.3	0.91
20	26.7	0.88	37.4	0.91	26.5	0.91
50	26.2	0.87	36.9	0.91	26.5	0.92
100	25.8	0.86	36.5	0.90	**26.6**	**0.92**
200	25.6	0.85	36.0	0.89	26.6	0.91
500	25.4	0.84	35.7	0.88	26.4	0.91
1000	25.3	0.83	35.4	0.87	26.3	0.89

the sharpest structural fidelity and minimal noise, demonstrating its superior performance.

D. Ablation Study

Impact of the Number of Time Steps We investigated the impact of the number of time steps on the performance of Fast-DDPM. Specifically, we conducted experiments using 3, 5, 7, 10, 20, 50, 100, 200, 500 and 1000 time steps across all three datasets. A uniform noise scheduler (Fig. 3(a)) was used for Denoising and Image-to-Image Translation, while a non-uniform noise scheduler (Fig. 3(b)) was used for Super-Resolution.

Table IV shows that for Super-Resolution and Denoising, model performance improves with the number of time steps up to 10, but declines beyond that. Both tasks involve progressively refining a random noise image with structural guidance from a condition image containing strong anatomical priors (e.g., in low-dose CT denoising, the low-dose CT image serves as the condition), meaning that excessive diffusion steps may introduce unnecessary noise or blurring, reducing perceptual quality. Since these tasks primarily enhance image fidelity rather than synthesizing entirely new structures, fewer steps enable efficient denoising while preserving fine details.

In contrast, Image-to-Image Translation follows a different trend, achieving peak performance at 100 time steps before declining. Unlike Super-Resolution and Denoising, this task involves generating new image appearances, often requiring more extensive transformations between the condition image and the final output. A higher number of diffusion steps likely allows the model to refine these transformations more effectively, leading to improved synthesis quality. However, beyond 100 steps, additional diffusion may introduce unnecessary noise refinements or reduce structural consistency, leading to a performance drop.

These results highlight that the optimal number of time steps depends on the nature of the task. Tasks focused on enhancing image fidelity (Super-Resolution, Denoising) benefit from fewer steps, whereas tasks requiring structural transformation (Image-to-Image Translation) require more steps to fully reconstruct new details. Future research could explore adaptive strategies to dynamically adjust the number of time steps based on task-specific requirements.

Impact of Noise Scheduler Table V illustrates the impact of the two proposed noise schedulers on the performance of Fast-DDPM. The uniform noise scheduler is a better choice for image denoising and image-to-image translation tasks, while the non-uniform noise scheduler performed better in the super-resolution task. These findings highlight the importance of selecting an appropriate noise scheduler tailored to the specific image-to-image task. This adaptability enhances Fast-DDPM's performance across various medical image processing applications, demonstrating its versatility and effectiveness in handling diverse imaging challenges.

TABLE V
THE IMPACT OF UNIFORM AND NON-UNIFORM NOISE SCHEDULERS ON DIFFERENT IMAGE-TO-IMAGE TASKS

Noise Scheduler	Super-Resolution		Denoising		Translation	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Uniform	26.6	0.88	**37.5**	**0.92**	**26.3**	**0.89**
Non-uniform	**27.1**	**0.89**	37.3	0.91	26.1	0.88

VI. DISCUSSION

A. Alignment With Clinical Standards

In the medical field, image generation and manipulation must adhere to clinical standards and protocols to ensure practical relevance and patient safety. Fast-DDPM addresses this critical need by focusing on clinically relevant tasks such as multi-image super-resolution, image denoising, and image-to-image translation, which directly impact diagnostic accuracy and streamline clinical workflows, thereby enhancing patient care. Unlike general-purpose image generators that prioritize visual diversity and creativity, Fast-DDPM is designed to ensure alignment with ground truth data, evaluated using clinically relevant metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). This alignment guarantees that the generated outputs preserve the intensity and structural fidelity essential for clinical decision-making. Furthermore, our approach significantly accelerates both the training and inference of diffusion models, paving the way for real-time clinical applications.

B. Speed-Accuracy Trade-Off

In unconditional image generation, the model must generate realistic samples entirely from noise, requiring it to learn the full data distribution across all possible variations. This typically necessitates a large number of time steps to incrementally refine details. Reducing the number of steps can disrupt this process, leading to a mismatch between the learned and true distribution and ultimately degrading sample quality. As a result, there is often a trade-off between speed and generation quality.

However, in conditional image generation, the model is guided by an external signal (e.g., a low-resolution version of the target image), significantly reducing the complexity of the learning task. Instead of synthesizing content from scratch, the model learns to refine existing structures. Consequently, using fewer, well-optimized time steps can improve performance by allowing the model to focus its capacity on relevant noise levels rather than being diluted across unnecessary diffusion steps.

We hypothesize that training a diffusion model across a wide range of noise levels without prioritization may introduce excessive redundancy, making it harder to learn distinctive features at different noise levels. Instead, a more selective, step-aware training approach—where priority is given to the noise levels that directly influence the sampling path—can foster more effective representation learning and ultimately lead to superior sample quality.

C. Diffusion Models Vs. GANs

While generative adversarial networks (GANs) have proven effective in various image processing tasks, diffusion models—particularly Fast-DDPM—offer several distinct advantages over GANs. First, Fast-DDPM consistently generates higher-quality samples than GANs, avoiding artifacts that can obscure critical details in medical images. It achieves this without the extended training and sampling times typical of other diffusion models. Second, Fast-DDPM ensures superior training stability, effectively avoiding the mode collapse issue often encountered in GANs due to their adversarial framework. Additionally, Fast-DDPM is more robust to hyperparameter tuning, as its performance does not depend on balancing generator and discriminator losses—a common challenge with GANs. This efficiency, combined with adaptability to applications such as text and audio synthesis, positions Fast-DDPM as a highly promising solution for a wide range of use cases.

D. Limitations and Future Work

This study has a few limitations that could be addressed in future work. First, the noise schedulers used were manually selected and may not be optimal for the specific tasks. Future work will focus on developing adaptive noise schedulers that can be automatically learned and dynamically adjusted based on the input data and the specific task requirements. Second, this paper focuses on 2D diffusion models, which may not yield optimal performance compared to applying 3D diffusion models. However, model development in 3D requires significant computational resources, which is beyond our current capacity. Nevertheless, the concept of Fast-DDPM can be directly applied to 3D. Future work will focus on generalizing Fast-DDPM to 3D by running the diffusion models in a smaller 3D latent space using autoencoders.

E. Significance and Implications

The advancements introduced by Fast-DDPM have significant implications for the field of medical imaging. Our results demonstrate that Fast-DDPM achieves state-of-the-art performance in enhancing medical images across tasks such as super-resolution, denoising, and image-to-image translation. These capabilities directly impact clinical workflows by improving image quality, which can enhance diagnostic accuracy, reduce the need for repeat imaging, and support better-informed treatment planning. At the same time, Fast-DDPM significantly reduces training and inference time while preserving image quality, addressing a key limitation of traditional diffusion models, where high computational cost and time requirements have limited their practical applications. The dramatic reduction in resource demand makes diffusion models more accessible and practical for real-world clinical applications, particularly in time-sensitive scenarios. The improved efficiency of Fast-DDPM enables real-time image analysis, leading to quicker diagnostic decisions and more timely interventions. Furthermore, Fast-DDPM offers significant flexibility for fine-tuning through adjustments in time steps and noise schedulers. Its design encourages exploration of adaptive noise scheduling methods and task-specific configurations, potentially inspiring future advancements in the field.

VII. Conclusion

This paper introduces Fast-DDPM, a simple yet effective approach that accelerates both the training and sampling of diffusion models by reducing the number of time steps from 1,000 to 10. Our evaluation on three imaging datasets demonstrates the state-of-the-art performance of Fast-DDPM for medical image-to-image generation tasks. This advancement has the potential to further research in applying diffusion models for real-time medical imaging applications.

VII. Conflict of interest statement

The authors have no conflict of interest to declare.

Acknowledgment

The authors express their sincere gratitude to the NVIDIA AI Technology Center at the University of Florida for their invaluable feedback, technical guidance, and support throughout this project.

References

[1] K. Armanious et al., "MedGAN: Medical image translation using GANs," *Computerized Med. Imag. Graph.*, vol. 79, 2020, Art. no. 101684.
[2] C. Cao et al., "High-frequency space diffusion model for accelerated MRI," *IEEE Trans. Med. Imag.*, vol. 43, no. 5, pp. 1853–1865, May 2024.
[3] H. Chen et al., "Low-dose CT with a residual encoderdecoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.
[4] T. Chen, "On the importance of noise scheduling for diffusion models," 2023, *arXiv:2301.10972*.
[5] H. Chung, E. S. Lee, and J. Chul Ye, "MR image denoising and super-resolution using regularized reverse diffusion," *IEEE Trans. Med. Imag.*, 42, no. 4, pp. 922–934, Apr. 2023.
[6] 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.
[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[8] K. Gong, K. Johnson, G. El Fakhri, Q. Li, and T. Pan, "PET image denoising based on denoising diffusion probabilistic model," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 51, pp. 358–368, 2023.
[9] A. Güngör UH et al., "Adaptive diffusion priors for accelerated MRI reconstruction," *Med. Image Anal.*, vol. 88, 2023, Art. no. 102872.
[10] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 6840–6851.
[11] E. Hoogeboom, J. Heek, and T. Salimans, "Simple diffusion: End-to-end diffusion for high resolution images," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 13213–13232.

[12] X. Hu, Y.-J. Chen, T.-Y. Ho, and Y. Shi, "Conditional diffusion models for weakly supervised medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2023, pp. 756–765.

[13] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, "DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[14] H. Iqbal, U. C. Khalid Chen, and J. Hua, "Unsupervised anomaly detection in medical images using masked diffusion model," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2024, pp. 372–381.

[15] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.

[16] B. Kim, I. Han, and J. Chul Ye, "Diffusemorph: Unsupervised deformable image registration using diffusion model," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 347–364.

[17] B. Kim, Y. Oh, and J. Chul Ye, "Diffusion adversarial representation learning for self-supervised vessel segmentation," in *Proc. 11th Int. Conf. Learn. Representations*, 2023.

[18] J. Kim and H. Park, "Adaptive latent diffusion model for 3D medical image to image translation: Multi-modal magnetic resonance imaging study," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jan. 2024, pp. 7604–7613.

[19] D. Kingma, T. Salimans, B. Poole, and J. Ho, "Variational diffusion models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 21696–21707.

[20] L. Kong, C. Lian, D. Huang, Z.J. Li, Y. Hu, and Q. Zhou, "Breaking the dilemma of medical image-to-image translation," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, 2021, pp. 1964–1978.

[21] Z. Kong and W. Ping, "On fast sampling of diffusion probabilistic models," in *Proc. INNF+ Conf.*, 2021.

[22] H. Li et al., "SRDIFF: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.

[23] X. Li et al., "Q-diffusion: Quantizing diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2023, pp. 17535–17545.

[24] J. Liu et al., "Dolce: A model-based probabilistic diffusion framework for limited-angle CT reconstruction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2023, pp. 10498–10508.

[25] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," in *Proc. Int. Conf. Learn. Representations*, 2023.

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

[27] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "DPM-solver : Fast solver for guided sampling of diffusion probabilistic models," 2022, *arXiv:2211.01095*.

[28] Q. Lyu and G. Wang, "Conversion between CT and MRI images using diffusion and score-matching models," 2022, *arXiv:2209.12104*.

[29] J. Ma, Y. Zhu, C. You, and B. Wang, "Pre-trained diffusion models for plug-and-play medical image enhancement," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2023, pp. 3–13.

[30] X. Ma, G. Fang, and X. Wang, "DeepCache: Accelerating diffusion models for free," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 15762–15772.

[31] C. McCollough et al., "Low dose CT image and projection data (LDCT-and-projection data)(Version 4)," *Med. Phys.*, 48, pp. 902–911, 2021.

[32] C. Meng et al., "On distillation of guided diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 14297–14306.

[33] B.H. Menze et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[34] M. I. Miller, A. Trouvé, and L. Younes, "Geodesic shooting for computational anatomy," *J. Math. Imag. Vis.*, vol. 24, pp. 209–228, 2006.

[35] A. Nichol, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 16784–16804.

[36] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8162–8171.

[37] M. Özbey et al., "Unsupervised medical image translation with adversarial diffusion models," *IEEE Trans. Med. Imag.*, vol. 42, no. 12, pp. 3524–3539, Dec. 2023.

[38] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," 2022, *arXiv:2204.06125*.

[39] 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.*, 2022, pp. 10684–10695.

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

[41] C. Saharia et al., "Photorealistic text-to-image diffusion models with deep language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 36479–36494.

[42] C. Saharia, J. Ho, W. Chan, T. Salimans, D.J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2022.

[43] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," 2022, *arXiv:2202.00512*.

[44] A. Sauer, D. Lorenz, A. Blattmann, and R. Rombach, "Adversarial diffusion distillation," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2025, pp. 87–103.

[45] Y. Shang, Z. Yuan, B. Xie, B. Wu, and Y. Yan, "Post-training quantization on diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 1972–1981.

[46] W. Shao et al., "Geodesic density regression for correcting 4DCT pulmonary respiratory motion artifacts," *Med. Image Anal.*, vol. 72, 2021, Art. no. 102140.

[47] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learn. Representations*, 2021.

[48] Y. Song, P. Dhariwal, M. Chen, and I. Sutskever, "Consistency models," 2023, *arXiv:2303.01469*.

[49] Y. Song et al., "Score-based generative modeling through stochastic differential equations," in *Proc. Int. Conf. Learn. Representations*, 2021.

[50] G.A. Sonn et al., "Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device," *J. Urol.*, vol. 189, no. 1, pp. 86–92, 2013.

[51] R.R. Sood et al., "3D registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction," *Med. Image Anal.*, vol. 69, 2021, Art. no. 101957.

[52] J. Wang et al., "Inversesr: 3d Brain MRI super-resolution using a latent diffusion model," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, H. Greenspan, A. P. Madabhushi, S. Mousavi, J. Salcudean, T. Duncan Syeda-Mahmood, and R. Taylor, Eds. Cham, Springer Nature Switzerland, 2023, pp. 438–447.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[54] D. Watson, W. Chan, J. Ho, and M. Norouzi, "Learning fast samplers for diffusion models by differentiating through sample quality," in *Proc. Int. Conf. Learn. Representations*, 2022.

[55] J. Wolleb, F. Bieder, R. Sandkühler, and P. C. Cattin, "Diffusion models for medical anomaly detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2022, pp. 35–45.

[56] Z. Wu, X. Chen, S. Xie, J. Shen, and Y. Zeng, "Super-resolution of brain MRI images based on denoising diffusion probabilistic model," *Biomed. Signal Process. Control*, vol. 85, 2023, Art. no. 104901.

[57] Z. Xing, L. Wan, H. Fu, G. Yang, and L. Zhu, "Diff-UNet: A diffusion embedded network for volumetric segmentation," 2023, *arXiv:2303.10326*.

[58] R. Xu, Y. Wang, and B. Du, "MAEDiff: Masked autoencoder-enhanced diffusion models for unsupervised anomaly detection in brain images," 2024, *arXiv:2401.10561*.

[59] T. Zhang, H. Jiang, K. Gong, and W. Shao, "Geodesic diffusion models for medical image-to-image generation," 2025, *arXiv:2503.00745*.

[60] H. Zheng, W. Nie, A. Vahdat, K. Azizzadenesheli, and A. Anandkumar, "Fast sampling of diffusion models via operator learning," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 42390–42402.