

Fast & Learnable Measurement-conditioned Diffusion Probabilistic Model for Under-sampled MRI Reconstruction

Khushbu Pahwa
Electrical & Computer Engineering Department
UID: 805498854
University of California, Los Angeles

Abstract

We propose a fast and learnable version of the measurement-conditioned denoising diffusion probabilistic model, MC-DDPM [Yutong et al. 2022], for under-sampled medical image reconstruction based on DDPM. This work can be considered as a novel extension of the MC-DDPM. MC-DDPM is defined in measurement domain (e.g. k-space in MRI reconstruction) and conditioned on under-sampling mask. We apply this method to accelerate MRI reconstruction and the experimental results show excellent performance, outperforming full supervision baseline and the state-of-the-art score-based reconstruction method, and achieve the performance of MCDDPM with around 200 steps by replacing the DDPM module of MC-DDPM with fast-DPM architecture proposed in [Zhifeng et al. 2021]. In addition to faster sampling speed, another contribution of this project is the learning of noise schedule which has been kept constant in the MC-DDPM paper. This improvement is adapted from the Variational Diffusion Model (VDM) proposed in [Kingma et al. 2021]. We show that our results are comparable and in some cases better than the MC-DDPM paper for the PD and PDFs knee single coil MRI data from the fastMRI dataset [Zbontar et al. 2018]. However, due to lack of compute, the experiments have been conducted on just 1 volume of pd4x, pd8x, pdfs4x, and pdfs8x. The evaluation metrics are kept same as the MC-DDPM paper : PSNR and SSIM, however, we report results for only 1 volume compared to the results for 6 randomly selected volumes used for testing in the MC-DDPM paper.

1 Introduction

The problem of reconstruction from the under-sampled medical imaging data including magnetic resonance imaging (MRI) has been studied in depth over many years. [Aggarwal et al. 2018, Hammernik et al., 2018, Eo et al., 2018, Han et al., 2019], sparse view or limited angles computed tomography (CT)

[Han and Ye, 2018, Zhang et al., 2019, WWang et al., 2019] and digital breast tomosynthesis (DBT). Most of works aim to obtain one sample of the posterior distribution $p(\mathbf{x} | \mathbf{y})$ where \mathbf{x} is the reconstructed target image and \mathbf{y} is the under-sampled measurements.

Diffusion probabilistic models are a class of deep generative models that use Markov chains to gradually transform between a simple distribution (e.g., isotropic Gaussian) and the complex data distribution [Sohl-Dickstein et al., 2015; Ho et al., 2020]. Most recently, these models have obtained the state-of-the-art results in several important domains, including image synthesis [Ho et al., 2020; Song et al., 2020b; Dhariwal and Nichol, 2021], audio synthesis [Kong et al., 2020b; Chen et al., 2020], and 3-D point cloud generation [Luo and Hu, 2021; Zhou et al., 2021]. They have demonstrated superior performance and have been widely used in various image processing tasks DDPM utilizes a latent variable model to reverse a diffusion process, where the data distribution is perturbed to the noise distribution by gradually adding Gaussian noise. Similar to DDPM, score-based generative models [Hyvärinen and Dayan, 2005, Song and Ermon, 2019] also generate data samples by reversing a diffusion process. Both DDPM and score-based models are proved to be discretizations of different continuous stochastic differential equations by [Song et al. 2020]. The difference between them lies in the specific setting of diffusion process and sampling algorithms. They have been applied to the

1

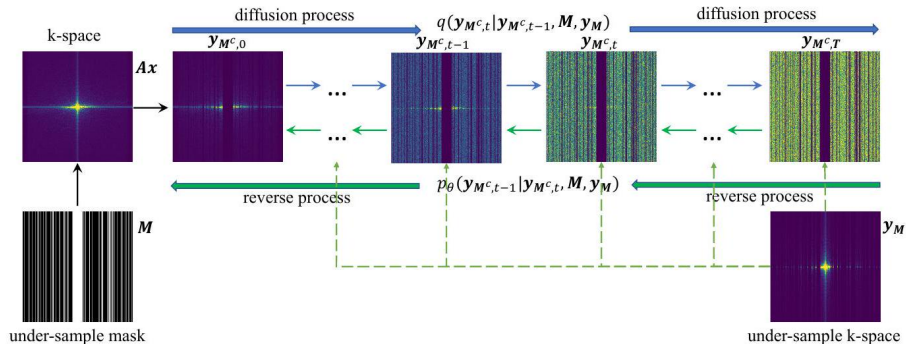


Figure 1: MC-DDPM approach illustrated by the example of under-sampled MRI reconstruction. Diffusion process: starting from the non-sampled k-space $\mathbf{y}_{M^c,0}$, Gaussian noise is gradually added until time T . Reverse process: starting from total noise, $\mathbf{y}_{M^c,0}$ is generated step by step. The details of notations is presented in Sect. 3

generation of image [Song et al. 2020, Nichol and Dhariwal, 2021, Dhariwal and Nichol, 2021], audio [Kong et al., 2020] or graph [Niu et al., [2020], and to conditional generation tasks such as in in-painting [Song and Ermon, 2019] Song et al., 2020], super-resolution [Choi et al., 2021, Saharia et al. 2021] and image editing [Meng et al.. 2021]. In these applications, the diffusion process of DDPM or score-based generative model is defined in data domain, and is

unconditioned although the reverse process could be conditioned given certain downstream task. Particularly, the score-based generative model has been used for under-sampled medical image reconstruction [Jalal et al., 2021 . Song et al., 2021. Chung et al. 2021], where the diffusion process is defined in the domain of image \mathbf{x} and is irrelevant to under-sampled measurements \mathbf{y} .

In this project, we attempt to explore the diffusion models efficacy in undersampled MR reconstruction. In particular our study is based on the (MC-DDPM) Measured Conditioned Denoising Diffusion Probabilistic Model proposed in [Yutong et al. 2022]. The details about the MC-DDPM method are discussed in Sect. 3 and 4. The MC-DDPM paper adopted the DDPM model rather than score-based generative model because DDPM is more flexible to control the noise distribution. In this study, we made enormous efforts to reproduce the results of the MC-DDPM paper and in addition , we propose two variations of the MC-DDPM architecture, namely FAST MCD-PPM and VAR MC-DDPM. The FAST MC-DDPM method is introduced to speed up the traditional DDPM model by taking inspiration from the Denoising Diffusion Implicit Model proposed by [Song et al., 2021]. We demonstrate that FAST MC-DDPM can achieve almost similar performance as that of the MC-DDPM with just 200 steps. However, it should be noted, that due to lack of compute, the comparison has been done on a single volume as opposed to 6 volumes tested in the MC-DDPM paper. So, we leave the comprehensive evaluation to ensure that the proposed variations are indeed useful to the scientific community, as future work. VAR MC-DDPM is another proposed variation of the MC-DDPM method. Here, instead of the constant noise schedule used in the MC-DDPM paper, the noise schedule is learned. This modification in the MC-DDPM architecture is adapted by the VDM model proposed in [Karras et al., 2022]. The details of FAST MC-DDPM and VAR MC-DDPM are discussed in the Background section : Sect. 2. The measurement-conditioned DDPM (MC-DDPM) for under-sampled medical image reconstruction is based on DDPM (Fig 1 illustrates the method by the example of under-sampled MRI reconstruction), where the under-sampling is in the measurement space (e.g. k -space in MRI reconstruction) and thus the conditional diffusion process is also defined in the measurement sapce. In MC-DDPM, the diffusion and sampling process are defined in measurement domain rather than image domain; and the diffusion process is conditioned on under-sampling mask so that data consistency is contained in the model naturally and inherently, and there is no need to execute extra data consistency when sampling. Thus MC-DDPM model can be used to sample multiple reconstruction results from the same measurements \mathbf{y} . This is useful for the uncertainty quantification for $q(\mathbf{x} | \mathbf{y})$, such as pixel-variance. Our experiments on the the accelerated MRI reconstruction show that FAST MC-DDPM can achieve comparable results to that of MC-DDPM with much less no of steps (200), compared to 1000 in the original paper. In addition, the VAR MC-DDPM variant does show improvemnt over the MC-DDPM, which indicates that noise schedule is an important hyperparameter which was not paid much attention in the MC-DDPM paper. We also compare the results with U-NET. Our evaluation is based on single volume average metrics : Peak

Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Normalized Mean Squared Error (NMSE).

This project is organized as follows: relevant background on DDPM, DDIM (Denosing Diffusion Implicit Models), and VDM (Variational Diffusion Model), and the under-sampled medical image reconstruction task is in Sect. 2, details of the competing method: MC-DDPM is presented in Sect. 3, specifications about the reproduction of the MCDDPM paper are given in Sect. 4, FAST MC-DDPM and VAR MC-DDPM are discussed in Sect. 5, and finally the results of the two proposed improvements: FAST MC-DDPM and VAR MC-DDPM are shown compared to the Zero Filled, MC-DDPM reconstruction, and U-NET, for a single volume of the pd4x, pd8x, pdfs4x and pdfs8x data obtained from the fastMRI dataset, are presented in Sect. 6. The conclusion and future steps are presented in Sect. 7.

2 Background

2.1 Denoising Diffusion Probabilistic Model

DDPM [Ho et al. 2020] is a certain parameterization of diffusion models [Sohl-Dickstein et al. 2015], which is a class of latent variable models using a Markov chain to convert the noise distribution to the data distribution. It has the form of $p_\theta(\mathbf{x}_0) := \int p_\theta(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T}$, where \mathbf{x}_0 follows the data distribution $q(\mathbf{x}_0)$ and $\mathbf{x}_1, \dots, \mathbf{x}_T$ are latent variables of the same dimensionality as \mathbf{x}_0 . The joint distribution $p_\theta(\mathbf{x}_{0:T})$ is defined as a Markov chain with learned Gaussian transitions starting from $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{0}, \mathbf{I})$

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t), p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) := \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}).$$

The sampling process of $p_\theta(\mathbf{x}_0)$ is: to sample \mathbf{x}_T from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ firstly; then, to sample \mathbf{x}_{t-1} from $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$ until \mathbf{x}_0 is obtained. It can be regarded as a reverse process of the diffusion process, which converts the data distribution to the noise distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$. In DDPM the diffusion process is fixed to a Markov chain that gradually adds Gaussian noise to the data according to a variance schedule β_1, \dots, β_T :

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}), \quad q(\mathbf{x}_t | \mathbf{x}_{t-1}) := \mathcal{N}(\alpha_t \mathbf{x}_{t-1}, \beta_t^2 \mathbf{I}),$$

where $\alpha_t^2 + \beta_t^2 = 1$ for all t and β_1, \dots, β_T are fixed to constants and their value are set specially so that $q(\mathbf{x}_T | \mathbf{x}_0) \approx \mathcal{N}(\mathbf{0}, \mathbf{I})$.

2.2 Under-sampled Medical Image Reconstruction

Suppose $\mathbf{x} \in \mathbb{R}^n$ represents a medical image and $\mathbf{y} \in \mathbb{R}^m, m < n$ is the under-sampled measurements which is obtained by the following forward model:

$$\mathbf{y} = \mathbf{P}_\Omega \mathbf{A} \mathbf{x} + \boldsymbol{\epsilon},$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is the measuring matrix and usually is invertible, $\mathbf{P}_\Omega \in \mathbb{R}^{m \times n}$ is the undersampling matrix with the given sampling pattern ΩL^2 and $\boldsymbol{\epsilon}$ is the noise. For example, \mathbf{x} is a CT image, \mathbf{A} is the Radon transform matrix and \mathbf{y} is the sinogram of limited angles. Under-sampled medical image reconstruction is to reconstruct \mathbf{x} from \mathbf{y} as possible. Assuming \mathbf{x} follows a distribution of $q(\mathbf{x})$ and given \mathbf{P}_Ω , according to Bayesian Formula, we can derive the posterior distribution as follows (usually \mathbf{P}_Ω is neglected):

$$q(\mathbf{x} | \mathbf{y}, \mathbf{P}_\Omega) = \frac{q(\mathbf{x}, \mathbf{y} | \mathbf{P}_\Omega)}{q(\mathbf{y})} = \frac{q(\mathbf{y} | \mathbf{x}, \mathbf{P}_\Omega) q(\mathbf{x})}{q(\mathbf{y})}.$$

Therefore, the task of under-sampled medical image reconstruction to reconstruct the posterior distribution.

2.3 Denoising Diffusion Implicit Models

Diffusion models usually comprise: *i*) a parameter-free T -step Markov chain named the diffusion process, which gradually adds random noise into the data, and *ii*) a parameterized T -step Markov chain called the reverse or denoising process, which removes the added noise as a denoising function. The likelihood in diffusion models is intractable, but they can be efficiently trained by optimizing a variant of the variational lower bound. In particular, [Ho et al. 2020] propose a certain parameterization called the denoising diffusion probabilistic model (DDPM) and show its connection with denoising score matching [Song and Ermon, 2019], so the reverse process can be viewed as sampling from a scorebased model using Langevin dynamics. DDPM can produce high-fidelity samples reliably with large model capacity and outperforms the state-of-the-art models in image and audio domains [Dhariwal and Nichol, 2021; Kong et al., 2020b]. However, a noticeable limitation of diffusion models is their expensive denoising or sampling process. For example, DDPM requires a Markov chain with $T = 1000$ steps to generate high quality image samples [Ho et al., 2020], and DiffWave requires $T = 200$ to obtain high-fidelity audio synthesis [Kong et al., 2020b]. In other words, one has to run the forward-pass of the neural network T times to generate a sample, which is much slower than the state-of-the-art GANs or flow-based models for image and audio synthesis [e.g., Karras et al., 2020; Kingma and Dhariwal, 2018; Kong et al., 2020a; Ping et al., 2020]. The authors of [Song et al., 2022] highlighted a critical drawback of the DDPM models, stating that they require many iterations to produce a high quality sample. For DDPMs, this is because that the generative process (from noise to data) approximates the reverse of the forward diffusion process (from data to noise), which could have thousands of steps; iterating over all the

steps is required to produce a single sample, which is much slower compared to GANs, which only needs one pass through a network. [Song et al., 2021] demonstrated that it takes around 20 hours to sample 50k images of size 32×32 from a DDPM, but less than a minute to do so from a GAN on a Nvidia 2080 Ti GPU. This becomes more problematic for larger images as sampling 50k images of size 256×256 could take nearly 1000 hours on the same GPU.

To overcome this limitation of the DDPM’s and to reduce the computational gap between the DDPM and GANs, the denoising diffusion implicit models (DDIMs) were proposed by [Song et al. 2021]. DDIMs are implicit probabilistic models [Mohamed & Lakshminarayanan, 2016] and are closely related to DDPMs, in the sense that they are trained with the same objective function.

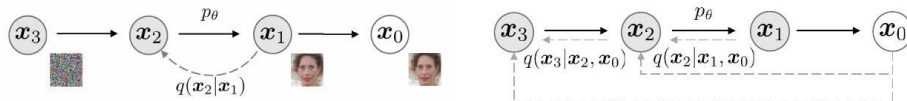


Figure 2: Graphical models for diffusion (left) and non-Markovian (right) inference models.

Given samples from a data distribution $q(\mathbf{x}_0)$, we are interested in learning a model distribution $p_\theta(\mathbf{x}_0)$ that approximates $q(\mathbf{x}_0)$ and is easy to sample from. Denoising diffusion probabilistic models (DDPMs, Sohl-Dickstein et al. (2015); Ho et al. (2020)) are latent variable models of the form

$$p_\theta(\mathbf{x}_0) = \int p_\theta(\mathbf{x}_{0:T}) d\mathbf{x}_{1:T}, \quad \text{where} \quad p_\theta(\mathbf{x}_{0:T}) := p_\theta(\mathbf{x}_T) \prod_{t=1}^T p_\theta^{(t)}(\mathbf{x}_{t-1} | \mathbf{x}_t)$$

where $\mathbf{x}_1, \dots, \mathbf{x}_T$ are latent variables in the same sample space as \mathbf{x}_0 (denoted as \mathcal{X}). The parameters θ are learned to fit the data distribution $q(\mathbf{x}_0)$ by maximizing a variational lower bound:

$$\max_{\theta} \mathbb{E}_{q(\mathbf{x}_0)} [\log p_\theta(\mathbf{x}_0)] \leq \max_{\theta} \mathbb{E}_{q(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)} [\log p_\theta(\mathbf{x}_{0:T}) - \log q(\mathbf{x}_{1:T} | \mathbf{x}_0)]$$

where $q(\mathbf{x}_{1:T} | \mathbf{x}_0)$ is some inference distribution over the latent variables. Unlike typical latent variable models (such as the variational autoencoder (Rezende et al., 2014)), DDPMs are learned with a fixed (rather than trainable) inference procedure $q(\mathbf{x}_{1:T} | \mathbf{x}_0)$, and latent variables are relatively high dimensional. For example, Ho et al. (2020) considered the following Markov chain with Gaussian transitions parameterized by a decreasing sequence $\alpha_{1:T} \in (0, 1]^T$:

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}), \quad \text{where} \quad q(\mathbf{x}_t | \mathbf{x}_{t-1}) := \mathcal{N}\left(\sqrt{\frac{\alpha_t}{\alpha_{t-1}}} \mathbf{x}_{t-1}, \left(1 - \frac{\alpha_t}{\alpha_{t-1}}\right) \mathbf{I}\right)$$

where the covariance matrix is ensured to have positive terms on its diagonal. This is called the forward process due to the autoregressive nature of the sampling procedure (from \mathbf{x}_0 to \mathbf{x}_T). We call the latent variable model $p_\theta(\mathbf{x}_{0:T})$,

which is a Markov chain that samples from \mathbf{x}_T to \mathbf{x}_0 , the generative process, since it approximates the intractable reverse process $q(\mathbf{x}_{t-1} | \mathbf{x}_t)$. Intuitively, the forward process progressively adds noise to the observation \mathbf{x}_0 , whereas the generative process progressively denoises a noisy observation (Figure 1, left).

A special property of the forward process is that

$$q(\mathbf{x}_t | \mathbf{x}_0) := \int q(\mathbf{x}_{1:t} | \mathbf{x}_0) d\mathbf{x}_{1:(t-1)} = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_0, (1 - \alpha_t)\mathbf{I})$$

so we can express \mathbf{x}_t as a linear combination of \mathbf{x}_0 and a noise variable ϵ :

$$\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1 - \alpha_t}\epsilon, \text{ where } \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

When we set α_T sufficiently close to 0, $q(\mathbf{x}_T | \mathbf{x}_0)$ converges to a standard Gaussian for all \mathbf{x}_0 , so it is natural to set $p_\theta(\mathbf{x}_T) := \mathcal{N}(\mathbf{0}, \mathbf{I})$. If all the conditionals are modeled as Gaussians with trainable mean functions and fixed variances, the objective in Eq. (2) can be simplified to :

$$L_\gamma(\epsilon_\theta) := \sum_{t=1}^T \gamma_t \mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0), \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\left\| \epsilon_\theta^{(t)} (\sqrt{\alpha_t}\mathbf{x}_0 + \sqrt{1 - \alpha_t}\epsilon_t) - \epsilon_t \right\|_2^2 \right]$$

where $\epsilon_\theta := \left\{ \epsilon_\theta^{(t)} \right\}_{t=1}^T$ is a set of T functions, each $\epsilon_\theta^{(t)} : \mathcal{X} \rightarrow \mathcal{X}$ (indexed by t) is a function with trainable parameters $\theta^{(t)}$, and $\gamma := [\gamma_1, \dots, \gamma_T]$ is a vector of positive coefficients in the objective that depends on $\alpha_{1:T}$. In Ho et al. (2020), the objective with $\gamma = \mathbf{1}$ is optimized instead to maximize generation performance of the trained model; this is also the same objective used in noise conditional score networks [Song & Ermon, 2019] based on score matching [Hyvärinen, 2005; Vincent, 2011]. From a trained model, \mathbf{x}_0 is sampled by first sampling \mathbf{x}_T from the prior $p_\theta(\mathbf{x}_T)$, and then sampling \mathbf{x}_{t-1} from the generative processes iteratively.

The length T of the forward process is an important hyperparameter in DDPMs. From a variational perspective, a large T allows the reverse process to be close to a Gaussian [Sohl-Dickstein et al., 2015], so that the generative process modeled with Gaussian conditional distributions becomes a good approximation; this motivates the choice of large T values, such as $T = 1000$ in Ho et al. (2020). However, as all T iterations have to be performed sequentially, instead of in parallel, to obtain a sample \mathbf{x}_0 , sampling from DDPMs is much slower than sampling from other deep generative models, which makes them impractical for tasks where compute is limited and latency is critical.

In a nutshell, DDIM proposes a change to the recently popular diffusion models, motivated by increasing the speed of sampling. This is accomplished by changing the “forward” process which adds noise to the data. In the original diffusion models, this forward process is a Markov process whose marginals and conditionals can be computed efficiently in closed form. This paper proposes to replace this Markov forward process with a non-markovian process that is designed to have the same marginals. The generative model, in this case, changes

such that to predict the next step in the process, the model must first predict the “clean” sample at the end of the chain which is then used to give an estimate for the next step in the chain. Intriguingly, the objective for training this new generative model is identical to training a standard diffusion model. Thus, the models differ only at sampling time. Under the new interpretation, we can sample from a family of models after training. This family can be tuned to increase the speed of sampling, at the cost of some sample quality. For details, the DDIM paper, [Song et al., 2021] can be referred.

2.4 Variational Diffusion Model

The paper by [Karras et al., 2021] presents a novel perspective on denoising diffusion models. The main contributions of their work includes a (i) new formulation of ELBO using $SNR(t)$ and showing invariance to the noise schedule under this formulation, (ii) training noise schedule, and (iii) several new architecture improvements. The authors show that the proposed method’s : Variational Diffusion Model’s discrete-time versions include previous finite-length Markov chain models as special cases. The paper then demonstrates that the ELBOs of the proposed continuous-time models exist. In here, the resulting ELBOs include an integration (wrt time) of a reconstruction loss weighted by the time-derivative of $SNR(t)$. Based on this, the paper shows that the losses of the previous continuous-time diffusion-based models follow the same integration form but with different weighting functions. The authors also present interesting properties of the proposed method due to $SNR(t)$. In particular, remind that $SNR(t)$ is invertible because of its monotonicity, and the continuous-time loss’s integration contains the time-derivative of $SNR(t)$. The paper shows that the integral can be re-written wrt $SNR(t)$ ’s output by change-of-variable. In this form, the integral only depends on the range of the integration but doesn’t depend on the shape of $SNR(t)$ within the range. As a result, acknowledging the invariance of the loss on the $SNR(t)$ ’s output space, the paper proposes to parameterize the score models conditioning on log-scaled $SNR(t)$ instead of $SNR(t)$. The paper proposes to train for the discrete and continuous-time models. For the discrete-time models, ϵ_{θ} is defined as a monotonic network wrt $SNR(t)$. For continuous-time models, the paper proposes to train networks by maximizing the ELBO together with the diffusion models. For continuous-time models, ϵ_{θ} is trained by minimizing the variance of the Monte-Carlo estimation of the continuous-time ELBOs.

3 Competing Method: Measurement-conditioned DDPM

We base our improvements, tasking MC-DDPM as the baseline. Thus, it is important to understand the nuts and bolts of MC-DDPM. [Yutong et al., 2022] proposed measurement-conditioned DDPM (MC-DDPM) which is designed for under-sampled medical image reconstruction. In this section, the formulate of MC-DDPM, is discussed, including the diffusion process and its reverse process,

training objective and sampling algorithm. The under-sampled forward model can be represented as:

$$\mathbf{y}_M = \mathbf{M}\mathbf{A}\mathbf{x} + \epsilon_M,$$

where $\mathbf{M} \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements are either 1 or 0 depending on the sampling pattern $\Omega \mathbb{3}^3 \mathbf{y}_M$ and ϵ_M are both n -dimension vectors and their components at nonsampled positions are 0. The merit of the new notations is that we can further define $\mathbf{M}^c = \mathbf{I} - \mathbf{M}$ (the superscript c means complement) and $\mathbf{y}_{M^c} = \mathbf{M}^c \mathbf{A} \mathbf{x}$ which represents the non-sampled measurements. In this paper, we assume $\epsilon_M = \mathbf{0}$. Then, we have $\mathbf{y}_M + \mathbf{y}_{M^c} = \mathbf{A} \mathbf{x}$, i.e. $\mathbf{y}_M + \mathbf{y}_{M^c}$ is the full-sampled measurements. In addition, the posterior distribution of reconstruction can be rewritten as $q(\mathbf{x} | \mathbf{M}, \mathbf{y}_M)$. Through this paper, the subscript \mathbf{M} or \mathbf{M}^c in notations indicates that only components at under-sampled or non-sampled positions are not 0.

The purpose of reconstruction task is to estimate $q(\mathbf{x} | \mathbf{M}, \mathbf{y}_M)$. Since \mathbf{y}_M is known and $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y}_M + \mathbf{y}_{M^c})$, the problem is transformed to estimate $q(\mathbf{y}_{M^c} | \mathbf{M}, \mathbf{y}_M)$. Because \mathbf{M} and \mathbf{M}^c are equivalent as the condition, we can replace $q(\mathbf{y}_{M^c} | \mathbf{M}, \mathbf{y}_M)$ by $q(\mathbf{y}_{M^c} | \mathbf{M}^c, \mathbf{y}_M)$. Based on this observation, we propose MC-DDPM which solves the reconstruction problem by generating samples of $q(\mathbf{y}_{M^c} | \mathbf{M}^c, \mathbf{y}_M)$. MC-DDPM is defined in measurement domain, instead of image domain as usual DDPM, and is conditioned on the non-sampling matrix \mathbf{M}^c and sampled measurements \mathbf{y}_M . It has the following form:

$$p_\theta(\mathbf{y}_{M^c,0} | \mathbf{M}^c, \mathbf{y}_M) := \int p_\theta(\mathbf{y}_{M^c,0:T} | \mathbf{M}^c, \mathbf{y}_M) d\mathbf{y}_{M^c,1:T},$$

where $\mathbf{y}_{M^c,0} = \mathbf{y}_{M^c} \cdot p_\theta(\mathbf{y}_{M^c,0:T} | \mathbf{M}^c, \mathbf{y}_M)$ is defined as follows:

$$p_\theta(\mathbf{y}_{M^c,0:T} | \mathbf{M}^c, \mathbf{y}_M) := p(\mathbf{y}_{M^c,T} | \mathbf{M}^c, \mathbf{y}_M) \prod_{t=1}^T p_\theta(\mathbf{y}_{M^c,t-1} | \mathbf{y}_{M^c,t}, \mathbf{M}^c, \mathbf{y}_M),$$

$$p_\theta(\mathbf{y}_{M^c,t-1} | \mathbf{y}_{M^c,t}, \mathbf{M}^c, \mathbf{y}_M) := \mathcal{N}(\boldsymbol{\mu}_\theta(\mathbf{y}_{M^c,t}, t, \mathbf{M}^c, \mathbf{y}_M), \sigma_t^2 \mathbf{M}^c),$$

where $\sigma_t^2 \mathbf{M}^c$ is the covariance matrix and it means the noise is only added at non-sampled positions because for all t the components of $\mathbf{y}_{M^c,t}$ at under-sampled positions are always 0. If the conditions $(\mathbf{M}^c, \mathbf{y}_M)$ in equations above is removed, they degrade to the form of Eq. 1.

Similar to DDPM, the sampling process of $p_\theta(\mathbf{y}_{M^c,0} | \mathbf{M}^c, \mathbf{y}_M)$ is a reverse process of the diffusion process which is also defined in measurement domain. Specifically, the Gaussian noise is gradually added to the non-sampled measurements $\mathbf{y}_{M^c,0}$. The diffusion process has the following form:

$$q(\mathbf{y}_{M^c,1:T} | \mathbf{y}_{M^c,0}, \mathbf{M}^c, \mathbf{y}_M) := \prod_{t=1}^T q(\mathbf{y}_{M^c,t} | \mathbf{y}_{M^c,t-1}, \mathbf{M}^c, \mathbf{y}_M),$$

$$q(\mathbf{y}_{M^c,t} | \mathbf{y}_{M^c,t-1}, \mathbf{M}^c, \mathbf{y}_M) := \mathcal{N}(\alpha_t \mathbf{y}_{M^c,t-1}, \beta_t^2 \mathbf{M}^c),$$

There are two points worthy of noting: (1) α_t, β_t are not restricted to satisfy $\alpha_t^2 + \beta_t^2 = 1$; (2) formally, we add \mathbf{y}_M as one of the conditions, but it has no

effect on the diffusion process in fact. Let $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, $\bar{\beta}_t^2 = \sum_{i=1}^t \frac{\alpha_i^2}{\alpha_i^2} \beta_i^2$, and we additionally define $\bar{\alpha}_0 = 1$, $\bar{\beta}_0 = 0$. Then, we can derive that:

$$q(\mathbf{y}_{\mathbf{M}^c, t} | \mathbf{y}_{\mathbf{M}^c, 0}, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}}) = \mathcal{N}(\bar{\alpha}_t \mathbf{y}_{\mathbf{M}^c, 0}, \bar{\beta}_t^2 \mathbf{M}^c),$$

$$q(\mathbf{y}_{\mathbf{M}^c, t-1} | \mathbf{y}_{\mathbf{M}^c, t}, \mathbf{y}_{\mathbf{M}^c, 0}, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}}) = \mathcal{N}(\tilde{\boldsymbol{\mu}}_t, \tilde{\beta}_t^2 \mathbf{M}^c),$$

where $\tilde{\boldsymbol{\mu}}_t = \frac{\alpha_t \bar{\beta}_{t-1}^2}{\bar{\beta}_t^2} \mathbf{y}_{\mathbf{M}^c, t} + \frac{\bar{\alpha}_{t-1} \beta_t^2}{\bar{\beta}_t^2} \mathbf{y}_{\mathbf{M}^c, 0}$, $\tilde{\beta}_t = \frac{\beta_t \bar{\beta}_{t-1}}{\bar{\beta}_t}$. In MC-DDPM, we assume that α_t is set specially so that $\bar{\alpha}_T \approx 0$, i.e. $q(\mathbf{y}_{\mathbf{M}^c, T} | \mathbf{y}_{\mathbf{M}^c, 0}) \approx \mathcal{N}(\mathbf{0}, \bar{\beta}_T^2 \mathbf{M}^c)$ is a noise distribution independent of $\mathbf{y}_{\mathbf{M}^c, 0}$.

Next, we discuss how to train MC-DDPM $p_\theta(\mathbf{y}_{\mathbf{M}^c, 0} | \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})$. Firstly, let $p(\mathbf{y}_{\mathbf{M}^c, T} | \mathbf{M}^c, \mathbf{y}_{\mathbf{M}}) = \mathcal{N}(\mathbf{0}, \bar{\beta}_T^2 \mathbf{M}^c)$ so that it is nearly equal to $q(\mathbf{y}_{\mathbf{M}^c, T} | \mathbf{y}_{\mathbf{M}^c, 0})$. Training of $p_\theta(\mathbf{y}_{\mathbf{M}^c, 0} | \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})$ is performed by optimizing the variational bound on negative log likelihood:

$$\begin{aligned} \mathbb{E}[-\log p_\theta(\mathbf{y}_{\mathbf{M}^c, 0} | \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})] &\leq \mathbb{E}_q \left[-\log \frac{p_\theta(\mathbf{y}_{\mathbf{M}^c, 0:T} | \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})}{q(\mathbf{y}_{\mathbf{M}^c, 1:T} | \mathbf{y}_{\mathbf{M}^c, 0}, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})} \right] \\ &= \mathbb{E}_q \left[-\log p(\mathbf{y}_{\mathbf{M}^c, T} | \mathbf{M}^c, \mathbf{y}_{\mathbf{M}}) - \sum_{t \geq 1} \log \frac{p_\theta(\mathbf{y}_{\mathbf{M}^c, t-1} | \mathbf{y}_{\mathbf{M}^c, t}, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})}{q(\mathbf{y}_{\mathbf{M}^c, t} | \mathbf{y}_{\mathbf{M}^c, t-1}, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})} \right] =: L. \end{aligned}$$

Assuming that

$$\boldsymbol{\mu}_\theta(\mathbf{y}_{\mathbf{M}^c, t}, t, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}}) = \frac{1}{\alpha_t} \left(\mathbf{y}_{\mathbf{M}^c, t} - \frac{\beta_t^2}{\bar{\beta}_t} \varepsilon_\theta(\mathbf{y}_{\mathbf{M}^c, t}, t, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}}) \right),$$

and supposing $\mathbf{y}_{\mathbf{M}^c, t} = \bar{\alpha}_t \mathbf{y}_{\mathbf{M}^c, 0} + \varepsilon$, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \bar{\beta}_t^2 \mathbf{M}^c)$ (Eq. 8, after reweighting L can be simplified as follows:

$$L_{\text{simple}} = \mathbb{E}_{\mathbf{y}_{\mathbf{M}^c, 0}, t, \varepsilon} \|\varepsilon - \varepsilon_\theta(\bar{\alpha}_t \mathbf{y}_{\mathbf{M}^c, 0} + \bar{\beta}_t \varepsilon, t, \mathbf{M}^c, \mathbf{y}_{\mathbf{M}})\|_2^2, \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{M}^c),$$

Algorithm 1 Training	Algorithm 2 Sampling
1: repeat	1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$	2: for $t = T, \dots, 1$ do
3: $t \sim \text{Uniform}(\{1, \dots, T\})$	3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$	4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: Take gradient descent step on $\nabla_\theta \ \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t)\ ^2$	5: end for
6: until converged	6: return \mathbf{x}_0

Algorithm 1 MC-DDPM Training	Algorithm 2 MC-DDPM Sampling
1: repeat	1: Given \mathbf{M}^c and \mathbf{y}_M
2: $\mathbf{x} \sim q(\mathbf{x})$, obtain \mathbf{M} and \mathbf{M}^c	2: $\mathbf{y}_{M^c, T} \sim \mathcal{N}(\mathbf{0}, \bar{\beta}_T^2 \mathbf{M}^c)$
3: $\mathbf{y}_M = \mathbf{M}\mathbf{A}\mathbf{x}$, $\mathbf{y}_{M^c} = \mathbf{M}^c\mathbf{A}\mathbf{x}$	3: for $t = T, \dots, 1$ do
4: $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}^c)$	4: $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{M}^c)$ if $t > 1$, else $\mathbf{z}_t = \mathbf{0}$
5: $t \sim \text{Uniform}(\{1, \dots, T\})$	5: $\boldsymbol{\mu}_t = \boldsymbol{\mu}_\theta(\mathbf{y}_{M^c, t}, t, \mathbf{M}^c, \mathbf{y}_M)$
6: $\mathbf{y}_{M, t}^c = \bar{\alpha}_t \mathbf{y}_{M, 0}^c + \bar{\beta}_t \boldsymbol{\varepsilon}$	6: $\mathbf{y}_{M^c, t-1} = \boldsymbol{\mu}_t + \sigma_t \mathbf{z}_t$
7: Take gradient descent step on	7: end for
$\ \nabla_\theta \ \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_\theta(\mathbf{y}_{M, t}^c, t, \mathbf{M}^c, \mathbf{y}_M)\ _2^2$	8: return $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y}_M + \mathbf{y}_{M^c, 0})$
8: until converged	

where t is uniform between 1 and T .

Algorithm 1: Training displays the complete training procedure of the Denoising Diffusion Probabilistic Models with this simplified objective and Algorithm 2: Sampling shows the sampling process of the Denoising Diffusion Probabilistic Models.

Algorithm 1: MC-DDPM training displays the complete training procedure of MC-DDPM with this simplified objective and Algorithm 2: MC-DDPM Sampling shows the sampling process.

Since MC-DDPM can produce multiple samples of the posterior distribution $q(\mathbf{x} | \mathbf{y}_M, \mathbf{M})$, the pixel-variance can be computed by Monte Carlo approach which is used to quantify uncertainty of reconstruction.

4 MC-DDPM Experiments

We apply MC-DDPM to accelerated MRI reconstruction where \mathbf{A} is 2 d Fourier transform and \mathbf{y}_M is the under-sampled k-space data. The specific design for $\boldsymbol{\varepsilon}_\theta(\mathbf{y}_{M^c, t}, t, \mathbf{M}^c, \mathbf{y}_M)$ in our experiments is given as follows:

$$\boldsymbol{\varepsilon}_\theta(\mathbf{y}_{M^c, t}, t, \mathbf{M}^c, \mathbf{y}_M) = \mathbf{M}^c f(g(\mathbf{A}^{-1}(\mathbf{y}_{M^c, t} + \mathbf{y}_M), \mathbf{A}^{-1}\mathbf{y}_M), t; \theta),$$

where f is a deep neural network and $g(\cdot, \cdot)$ is the concatenation operation. Because MR image \mathbf{x} is in complex field, we use $|\mathbf{x}|$, the magnitude of it, as the final image. Pixel-wise variance is also computed using magnitude images.

5 Proposed Methods: FAST MC-DDPM & VAR MC-DDPM

5.1 FAST MC-DDPM

We replace the DDPM with DDIM model in the MC-DDPM architecture. Denoising diffusion probabilistic models (DDPMs) have achieved high quality image generation without adversarial training, yet they require simulating a Markov chain for many steps to produce a sample. To accelerate sampling, DDIM model has been proposed by [Song et al. 2021] which is a more efficient

class of iterative implicit probabilistic models with the same training procedure as DDPMs, as discussed in Sect. 2. DDIM constructs a class of non-Markovian diffusion processes that lead to the same training objective, but whose reverse process can be much faster to sample from. Thus, faster sampling has motivated us to replace the DDPM with DDIM for sampling in the MC-DDPM stage.

5.2 VAR MC-DDPM

In this variant of MC-DDPM, we replace the DDPM with Variational Diffusion Model so that the noise schedule can be learned instead of the fixed noise schedule used in the MC-DDPM paper. The details regarding the Variational Diffusion Model are outlined in Sect 2., and can also be referred from Karras et al., 2022

6 Experimental Setting

All experiments are performed with fastMRI single-coil knee dataset [Zbontar et al. 2018], which is publicly available ⁴ and is divided into two parts, proton-density with (PDFS) and without fat suppression (PD). We trained the network with k-space data which were computed from 320×320 size complex images. In order to reproduce the results of MC-DDPM, the same experimental procedure has been adopted, using the guided-DDPM [Dhariwal and Nichol, 2021] and aand the diffusion process in [Dhariwal and Nichol [2021] but multiply β_t by 0.5 so that $\bar{\beta}_T \approx 0.5$. All networks were trained with learning rate of 0.0001 using AdamW optimizer. We were able to reproduce the results of the MC-DDPM paper, however, for the scope of the project and due to the compute and storage limitations, we compare the proposed variations of the MC-DDPM - FAST MC-DDPM and VAR MC-DDPM on a single volume. We also show comparison with Zero Filled and U-Net architecture used in the MC-DDPM paper, to keep consistent settings. For all volumes of training data, we drop the first and last five slices to avoid training the model with noise-only data as [Chung et al. [2021] did. For testing, we randomly select single volume from the validation set and dropped the first and last 5 slices from each volume for both PD and PDFS. The model architectures used in experiments stems from U-Net [Ronneberger et al., 2015] and is added by time embedding modules and self-attention layers. The model is trained with batch size of 48 for 35k steps as in MC-DDPM experiments. To reproduce the exact pipeline of MC-DDPM, the same settings are used : about α_t and β_t , the "cosine" scheduled which is used similar to that in [Dhariwal and Nichol [2021], and then multiplied the β_t by 0.5. All code has been implemented in PyTorch.

To verify superiority, we perform comparison studies with baseline methods (U-Net [Ronneberger et al. 2015]) used in [Zbontar et al. 2018]. The evaluation metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), of score-based reconstruction method proposed in [Chung et al. 2021] are also used for comparison since their experiments are conducted on the

same dataset.

6.1 Experimental Results

We show the results of PD with $4\times$ (the first row) and $8\times$ (the second row) acceleration in Fig. 2. We compare our method to zero-filled reconstruction (ZF), U-Net, and MC-DDPM. Since MC-DDPM can produce multiple reconstruction samples, we use the mean of 20 samples as the object for comparison. We observe that that the proposed method - FAST MC-DDPM attains comparable performance to that of MC-DDPM with much fewer steps (around 200), compared to 1000 used in the MC-DDPM paper. We also observe that the Variational MC-DDPM outperforms the MC-DDPM on all experiemntal settings. Moreover, we are able to reproduce the trend demonstrated in the MC-DDPM paper where they show that their method outperforms U-Net baseline. For lack of time, we don't compare our approach with the score based diffusion model. However, since the MC-DDPM showed superior performance over the score based diffusion model, we compare our results, instead with MC-DDPM.



Figure 3: An example of reverse diffusion process for reconstruction

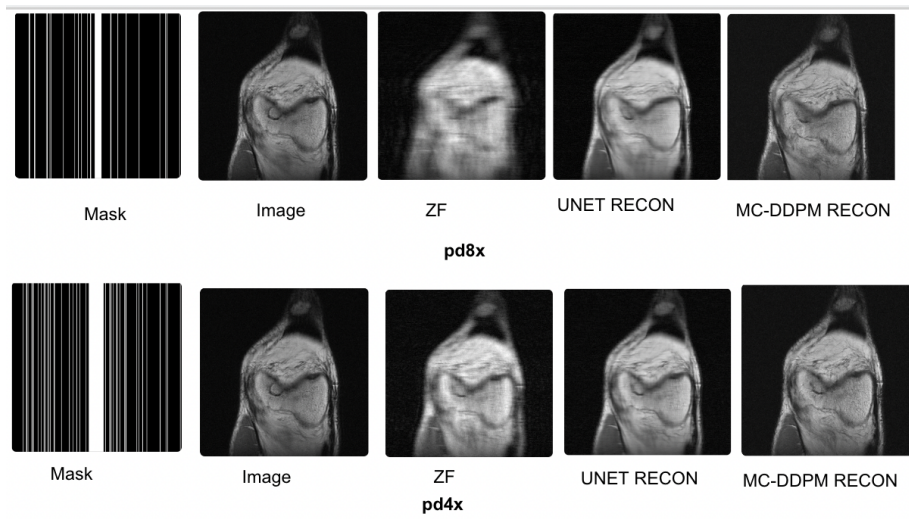


Figure 4: Sample Reconstruction results of ZF, U-NET, MC-DDPM for pd4x and pd8x

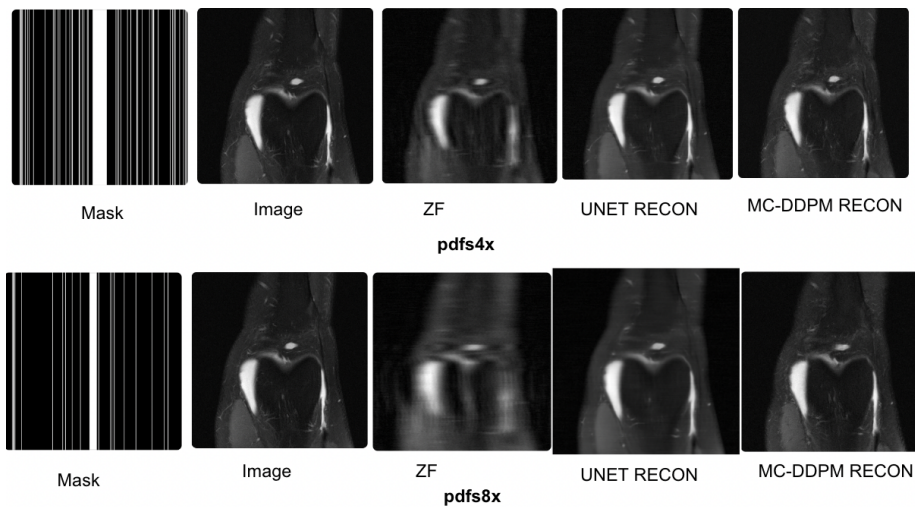
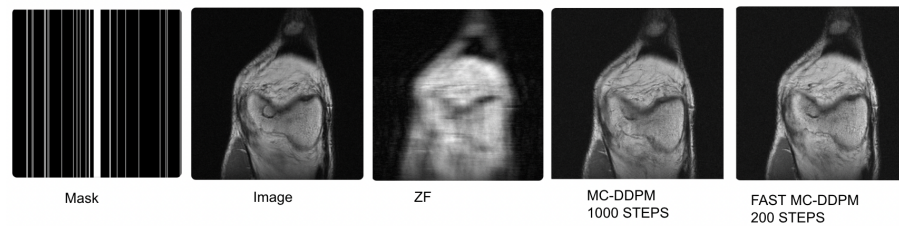


Figure 5: Sample Reconstruction results of ZF, U-NET, MC-DDPM for pdfs4x and pdfs8x

FAST MC-DDPM



VAR MC-DDPM

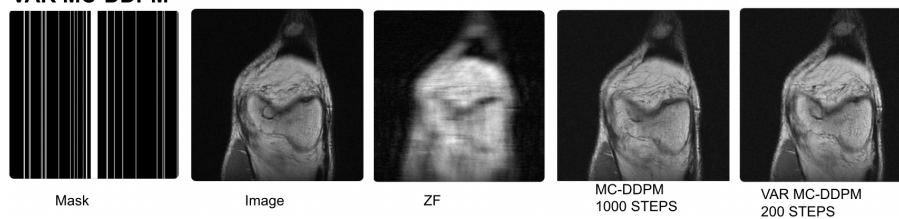


Figure 6: Comparison of the reconstruction results of ZF, MC-DDPM , FAST MC-DDPM, and VAR MC-DDPM for pd4x and pd8x.

pd						
		ZF	U-NET	MCDDPM	FAST MC-DDPM (200 STEPS)	VAR MC-DDPM
	PSNR	24.92	31.25	31.57	31.4	32.19
x 4	SSIM	0.683	0.771	0.776	0.773	0.792
	NMSE	0.034	0.019	0.0178	0.0184	0.0169
	PSNR	26.59	28.59	29.2	28.91	30.12
x 8	SSIM	0.586	0.669	0.716	0.709	0.749
	NMSE	0.0523	0.0353	0.0365	0.042	0.0398
pdfs						
		ZF	U-NET	MCDDPM	FAST MC-DDPM (200 STEPS)	VAR MC-DDPM
	PSNR	24.63	30.75	31.506	30.932	31.638
x 4	SSIM	0.62	0.711	0.744	0.745	0.755
	NMSE	0.041	0.026	0.019	0.022	0.192
	PSNR	25.24	26.179	29.83	28.73	29.91
x 8	SSIM	0.059	0.624	0.665	0.648	0.672
	NMSE	0.076	0.0617	0.028	0.0419	0.032

Table 1: Quantitative metrics. The first half of the table reports the results for the pd case, and the second half for pdfs. "x4", and "x8" in the first column denote the acceleration factor. Numbers in red indicate the best metric out of all the methods. Numbers in bold face indicate the second best metric out of all the methods i.e. the MC-DDPM, however, the FAST MC-DDPM performance is comparable with just 200 steps.

6.2 Discussion

It is very common in medical imaging that the measurement is under sampled to reduce the cost or dosage. Therefore, it is important to define the conditional diffusion process in the measurement space for a reconstruction task. In this, project, we reproduce the pipeline of MC-DDPM, and propose two novel variants, specifically FAST MC-DDPM and VAR MC-DDPM. However, due to lack of compute and storage, we report our experimental results on a single volume of pd4x, pd8x, pdfs4x and pdfs8x of knee single-coil mri data obtained from fstMRI.

7 Conclusion

In this project we present fast and learnable noise-schedule variants of the MC-DDPM or medical image reconstruction using under-sampled measurements. Our method applies diffusion process in measurement domain with conditioned under-sampling mask with faster sampling speed, improved training, and enhanced quantity of synthesis. We obtain comparable performance to that attained by MC-DDPM using our much faster FAST MC-DDPM variant with the only difference being replacing DDPM with Denoising Diffusion Implic Model

proposed in [Song et al. 2021]. Moreover, we are able to attain superior performance with out VAR MC-DDPM variant which learns the the variance schedule $\{\beta_t\}$ effectively - which is an important hyperparameter related to the sampling quality and efficiency, as shown in [Kerras et al., 2022]. We wish to follow up this study with the following future steps: (i) verifying the improvements of FAST MC-DDPM & VAR MC-DDPM on all the 6 volumes tested in the MC-DDPM paper, (ii) checking if integration of FAST and VAR is possible. (iii)checking if adaptive diffusion without conditioned on image prior yields better results than conditional, and finally (iv) exploring scheduling between conditional and unconditional DDPM.

8 Acknowledgements

Thanks Dr. Dan for your guidance and immense support throughout the project.

9 References

Xie, Yutong, and Quanzheng Li. "Measurement-conditioned Denoising Diffusion Probabilistic Model for Under-sampled Medical Image Reconstruction." arXiv preprint arXiv:2203.03623 (2022).

Kong, Zhifeng, and Wei Ping. "On fast sampling of diffusion probabilistic models." arXiv preprint arXiv:2106.00132 (2021).

Kingma, Diederik, et al. "Variational diffusion models." *Advances in neural information processing systems* 34 (2021): 21696-21707.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. arXiv preprint arXiv:2006.11239, 2020.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256-2265. PMLR, 2015.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456, 2020.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. arXiv preprint arXiv:2105.05233, 2021.

Yutong Xie, Dufan Wu, Bin Dong, and Quanzheng Li. Trained model in supervised deep learning is a conditional risk minimizer. arXiv preprint arXiv:2202.03674, 2022.

Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. arXiv preprint

arXiv:2111.08005, 2021.

Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alex Dimakis, and Jonathan Tamir. Robust compressed sensing mri with deep generative priors. In Thirty-Fifth Conference on Neural Information Processing Systems, 2021.

Hyungjin Chung et al. Score-based diffusion models for accelerated mri. arXiv preprint arXiv:2110.05243, 2021.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, et al. fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839, 2018.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computerassisted intervention*, pages 234-241. Springer, 2015.

Jiaxi Wang, Li Zeng, Chengxiang Wang, and Yumeng Guo. Admm-based deep reconstruction for limited-angle ct. *Physics in Medicine & Biology*, 64(11):115011, 2019.

Haimiao Zhang, Bin Dong, and Baodong Liu. Jsr-net: a deep network for joint spatial-radon domain ct reconstruction from incomplete data. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3657-3661. IEEE, 2019.

Yoseob Han and Jong Chul Ye. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE transactions on medical imaging*, 37(6):1418-1429, 2018.

Yoseob Han, Leonard Sunwoo, and Jong Chul Ye. k-space deep learning for accelerated mri. *IEEE transactions on medical imaging*, 39(2):377-386, 2019.

Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2):394-405, 2018.

Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho-Joon Lee, and Dosik Hwang. Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic resonance in medicine*, 80(5) : 2188 – 2201, 2018.

Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055-3071, 2018.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162-8171. PMLR, 2021.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. arXiv preprint arXiv:2009.09761,

2020.

Chenhao Niu, Yang Song, Jiaming Song, Shengjia Zhao, Aditya Grover, and Stefano Ermon. Permutation invariant graph generation via score-based generative modeling. In International Conference on Artificial Intelligence and Statistics, pages 4474-4484. PMLR, 2020.

Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. arXiv preprint arXiv:2108.02938. 2021. Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. arXiv preprint arXiv:2104.07636, 2021.

Chenlin Meng, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sedit: Image synthesis and editing with stochastic differential equations. arXiv preprint arXiv:2108.01073. 2021.