Diffusion Models for Image Generation and Inverse Problems

Yue Qi
Stanford Online Program
yueqi@stanford.edu

***Abstract*—**Generative modeling has become an essential component of modern machine learning, offering transformative solutions in data processing and generation across various domains. Among these, diffusion models have gained prominence as a robust approach capable of generating high-quality images and solving challenging inverse problems. In this work, we demonstrate the versatility of diffusion models by employing a pretrained score-predicting function for single-step denoising, and implementing the denoising diffusion probabilistic model (DDPM) framework for unconditional image generation. Furthermore, we solve inverse problems with pretrained diffusion models as priors, showcasing their potential in tasks like image inpainting and deblurring. To achieve this, we implemented score-distillation editing (SDEdit) and two posterior sampling methods, including score annealed Langevin dynamics (ScoreALD) and diffusion posterior sampling (DPS).

Keywords—Computational imaging; Diffusion models; Image generation; Inverse problems

I. Introduction

Generative modeling has emerged as a cornerstone of modern machine learning, finding applications across diverse industries. From enabling chatbots, and powering data generation to advancing signal analysis, generative models have become integral to many aspects of daily life. Over the years, these models have evolved significantly, progressing from early probabilistic approaches to state-of-the-art, deep-learning-based methods. Notable advancements include generative adversarial networks (GANs) [1], variational autoencoders (VAEs)[2], autoregressive models[3], flow-based models[4], and diffusion models[5].

Diffusion models generate data by modeling a progressive transformation of random noise into complex data distributions. This is achieved through a step-by-step reversal of a corruption process, effectively "denoising" the data. Diffusion models are straightforward to define and efficient to train[6], and they have been gaining significant traction as alternatives or improvements to traditional likelihood-based and generative adversarial network-based approaches in tasks such as image and audio synthesis.

Beyond data generation, diffusion models have also shown remarkable potential as generative inverse problem solvers, thanks to their ability to produce high-quality reconstructions and the ease of combining with existing iterative solvers. Inverse problems are prevalent across numerous domains, with applications ranging from signal processing[7] and medical imaging[8] to seismic imaging[9] and astrophysics[10, 11]. By treating them as unsupervised priors, diffusion models have unlocked exciting new possibilities for tackling complex inverse problems, particularly in image restoration and reconstruction.

In this work, we explore the versatility of diffusion modeling in both image generation and three classic inverse problems in computational imaging: denoising, deblurring, and inpainting. Using a pretrained score-predicting function, we implement a single-step method for image denoising. For unconditional image generation, we follow the framework of denoising diffusion probabilistic models (DDPM)[6], using this diffusion model to generate images with an iterative denoising procedure. Furthermore, we used the pretrained diffusion model as priors to address inpainting and deblurring tasks, firstly with the score-distillation editing (SDEdit) method[12], followed by posterior sampling techniques, including score annealed Langevin dynamics (ScoreALD)[13] and diffusion posterior sampling (DPS)[14]. These posterior sampling approaches exemplify modern advancements in solving inverse problems through generative modeling.

II. Related Work

Although this work focuses exclusively on linear inverse problems, diffusion-based approaches have been successfully extended to tackle more general noisy inverse problems, including nonlinear challenges such as super-resolution, non-uniform deblurring, and phase retrieval[14]. Furthermore, the multi-modal extensions of variants of convolutional neural network (CNN)–based diffusion models and large language models (LLMs) have led to the explosive advances in text-to-image synthesis and image-text contrastive learning in recent years[15]. Last but not least, other generative methods, such as GANs[1, 16, 17], VAEs[18] and autoregressive models[19] have demonstrated capabilities of generating high-quality image and audio samples. Interested readers can refer to for a thorough review of other generative methods for multi-modal image synthesis and editing[20].

III. Methods

A. Background: Diffusion Models

A diffusion model is a parameterized Markov chain trained using variational inference to produce samples matching the data after finite time. The forward diffusion process is a Markov chain that gradually adds noise to the data in the opposite direction of sampling until signal is destroyed, and transitions of this chain are learned to reverse a diffusion process[6]. Song et al. [21] defined the stochastic differential equation (SDE) for

the data noising process (i.e. forward SDE) $\boldsymbol{x}(t)$, $t \in [0,T]$, $\boldsymbol{x}(t) \in \mathbb{R}^d \forall t$. The variance preserving (VP) form of the SDE is as the following

$$d\boldsymbol{x} = -\frac{\beta(t)}{2}\boldsymbol{x}dt + \sqrt{\beta(t)}d\boldsymbol{\omega} \tag{1}$$

where $\beta(t) : \mathbb{R} \to \mathbb{R} > 0$ is the noise schedule of the process, and $\boldsymbol{\omega}$ the standard d dimensional Wiener process. Discretizing SDE using the Euler-Maruyama method yields

$$\boldsymbol{x}_t = \sqrt{1-\beta_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_{t-1}, t = 1, \dots, T \tag{2}$$

where $\boldsymbol{z}_{t-1} \sim \mathcal{N}(0, I)$ are i.i.d. Gaussian random variables in each step. Equation (2) can be written as a conditional distribution that does not depend on step $t-1$ but only on the very first step $t = 0$:

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x_0} + \sqrt{1-\bar{\alpha}_t}\mathbf{z} \tag{3}$$

where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, and $\mathbf{z} \sim \mathcal{N}(0, I)$

Poof of equation (3) By math induction, firstly we prove the base case $t = 1$ is true given equation (2):

$$\begin{aligned}\boldsymbol{x}_1 &= \sqrt{1-\beta_1}\boldsymbol{x}_0 + \sqrt{\beta_1}\boldsymbol{z}_0 \\ &= \sqrt{\bar{\alpha}_1}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_1}\boldsymbol{z}_0 \\ &= \sqrt{\bar{\alpha}_1}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_1}\mathbf{z}\end{aligned}$$

Then we prove that if equation (3) holds for any case $t-1$, then it must hold for the next case t as the following.

If $\boldsymbol{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x_0} + \sqrt{1-\bar{\alpha}_{t-1}}\mathbf{z}$ is true,

then equation (2) amounts to:

$$\begin{aligned}\mathbf{x}_t &= \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{z}_{t-1} \\ &= \sqrt{\alpha_t}\left(\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x_0} + \sqrt{1-\bar{\alpha}_{t-1}}\mathbf{z}\right) + \sqrt{\beta_t}\boldsymbol{z}_{t-1} \\ &= \sqrt{\bar{\alpha}_t}\boldsymbol{x_0} + \sqrt{\alpha_t - \bar{\alpha}_t}\mathbf{z} + \sqrt{1-\alpha_t}\boldsymbol{z}_{t-1}\end{aligned}$$

So $$\mathbf{x}_t \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_t}\boldsymbol{x_0}, \left(\sqrt{\alpha_t - \bar{\alpha}_t}\right)^2 + \left(\sqrt{1-\alpha_t}\right)^2\right)$$

$$\mathbf{x}_t \sim \mathcal{N}\left(\sqrt{\bar{\alpha}_t}\boldsymbol{x_0}, 1 - \bar{\alpha}_t\right)$$

which is equivalent to

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x_0} + \sqrt{1-\bar{\alpha}_t}\mathbf{z}$$

This concludes the proof.

B. Unconditional Sampling with Diffusion Models

The image generation problem can be considered as gradually removing noise from a noisy observation $\boldsymbol{x}_t$ to recover $\boldsymbol{x_0}$ through the reverse of the nosing diffusion process (i.e. reverse SDE). The discrete reverse SDE takes the form of the following[6]

$$\boldsymbol{x_{t-1}} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x_t} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon_\theta}(\boldsymbol{x_t}, t)\right) \tag{4}$$

where $\boldsymbol{\epsilon_\theta}$ is a function approximator intended to predict the noise $\boldsymbol{\epsilon}$ from $\boldsymbol{x_t}$. Equivalent process can be achieved with a score-predicting network $\mathbf{s_\theta}$ instead of a noise-predicting network $\boldsymbol{\epsilon_\theta}$, given the relation $\mathbf{s_\theta}(\boldsymbol{x_t}, t) = -\frac{\boldsymbol{\epsilon_\theta}(\boldsymbol{x_t}, t)}{\sqrt{1-\bar{\alpha}_t}}$.

Substituting this into equation (4):

$$\begin{aligned}\boldsymbol{x_{t-1}} &= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x_t} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon_\theta}(\boldsymbol{x_t}, t)\right) \\ &= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x_t} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\left(-\sqrt{1-\bar{\alpha}_t}\right)\mathbf{s_\theta}(\boldsymbol{x_t}, t)\right) \\ \boldsymbol{x_{t-1}} &= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x_t} + (1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t)\right)\end{aligned} \tag{5}$$

Equation (5) can be split into two equations to explicitly include the predicted clean image $\hat{\boldsymbol{x}}_0$, that is

$$\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x_t} + (1-\bar{\alpha}_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t)\right) \tag{6}$$

$$\boldsymbol{x_{t-1}} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 \tag{7}$$

$$\xrightarrow{\textit{is equialent to}}$$

$$\boldsymbol{x_{t-1}} = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x_t} + (1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t)\right) \tag{5}$$

Poof.

$$\begin{aligned}\boldsymbol{x_{t-1}} &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 \\ &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \\ &\qquad \cdot \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x_t} + (1-\bar{\alpha}_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t)\right) \\ &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}\boldsymbol{x_t} \\ &\qquad + \frac{\sqrt{\bar{\alpha}_{t-1}}}{\sqrt{\bar{\alpha}_t}}(1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t) \\ &= \frac{\sqrt{\alpha_t \cdot \bar{\alpha}_t}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_t)}\boldsymbol{x_t} \\ &\qquad + \frac{1}{\sqrt{\alpha_t}}(1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t) \\ &= \frac{\sqrt{\alpha_t \cdot \alpha_t\bar{\alpha}_{t-1}}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_t)}\boldsymbol{x_t} \\ &\qquad + \frac{1}{\sqrt{\alpha_t}}(1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t) \\ &= \frac{\alpha_t\sqrt{\bar{\alpha}_{t-1}}(1-\bar{\alpha}_{t-1}) + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_t)}\boldsymbol{x_t} \\ &\qquad + \frac{1}{\sqrt{\alpha_t}}(1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t) \\ &= \frac{\sqrt{\bar{\alpha}_{t-1}}(\alpha_t - \bar{\alpha}_t + 1 - \alpha_t)}{\sqrt{\bar{\alpha}_t}(1-\bar{\alpha}_t)}\boldsymbol{x_t} + \frac{1}{\sqrt{\alpha_t}}(1-\alpha_t)\mathbf{s_\theta}(\boldsymbol{x_t}, t)\end{aligned}$$

$$= \frac{\sqrt{\bar{\alpha}_{t-1}}}{\sqrt{\bar{\alpha}_t}} \boldsymbol{x_t} + \frac{1}{\sqrt{\alpha_t}}(1-\alpha_t)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{x_t}, t)$$
$$= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x_t} + (1-\alpha_t)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{x_t}, t)\right)$$

This concludes the proof.

With that, an unconditional image generation procedure can be implemented with a score-predicting function by reverse SDE, as summarized by the pseudocode below.

DDPM

$\boldsymbol{x_T} \sim \mathcal{N}(0, I)$

for $t = T, \ldots, 1$ **do**

  $\mathbf{z} \sim \mathcal{N}(0, I)$ **if** $t > 1$, **else** $\mathbf{z} = \mathbf{0}$

  $\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x_t} + (1-\bar{\alpha}_t)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{x_t}, t)\right)$

  $\boldsymbol{x_{t-1}} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \sqrt{1-\alpha_t}\,\mathbf{z}$

end for

return $\boldsymbol{x_0}$

C. Posterior Sampling with Diffusion Models

In inverse problems, the aim is to recover an unknown sample $\boldsymbol{x} \in \mathbb{R}^n$, assuming known measurements $\boldsymbol{y} \in \mathbb{R}^m$ and a forward model $\mathcal{A}$

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{x}) + \boldsymbol{z}, \boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I_m}) \tag{8}$$

Looking through the lens of Bayesian inference, given measurements $\boldsymbol{y}$, the goal can be interpreted as generating plausible reconstructions by sampling from the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$, where $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$ as stated by the Bayes' rule**.** Diffusion models can be utilized as priors for $p(\boldsymbol{x})$ to generate reconstructions by sampling from the posterior distribution. Considering the continuous reverse SDE, the reverse diffusion sampler for sampling from the posterior distribution can be arranged as[21]

$$d\boldsymbol{x} = [-\frac{\beta(t)}{2}\boldsymbol{x} - \beta(t)\left(\nabla_{x_t}\log p_t(\boldsymbol{x_t}) + \nabla_{x_t}\log p_t(\boldsymbol{y}|\boldsymbol{x_t})\right)]dt + \sqrt{\beta(t)}d\bar{\boldsymbol{\omega}} \tag{9}$$

The score function term $\nabla_{x_t}\log p_t(\boldsymbol{x_t})$ can be computed using a pre-trained score-predicting function $\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{x_t}, t)$ **.** However, the likelihood term $\nabla_{x_t}\log p_t(\boldsymbol{y}|\boldsymbol{x_t})$ is analytically intractable due to their dependence on time t. Various approximations have been explored by the community to circumvent the intractability of posterior sampling[22].

As one of the first proposed methods for solving linear inverse problems with diffusion models, ScoreALD method approximates the likelihood gradient as $\nabla_{x_t}\log p(\boldsymbol{y}|\boldsymbol{x_t}) \simeq \frac{\mathcal{A}^H(\boldsymbol{y}-\mathcal{A}\boldsymbol{x_t})}{\sigma^2+\gamma_t^2}$, where $\{\gamma_t\}_{t=1}^T$ are hyperparameters.

Alternatively, the DPS method uses the approximate gradient of the log likelihood: $\nabla_{x_t}\log p(\boldsymbol{y}|\boldsymbol{x_t}) \simeq \nabla_{x_t}\log p(\boldsymbol{y}|\hat{\boldsymbol{x}}_{\boldsymbol{0}})$, where the latter is analytically tractable, as the measurement distribution is given. It is worth noting that, while this paper focuses on solving linear inverse problems, DPS is also capable of tackling more general inverse problems, including nonlinear cases, which ScoreALD and many other early posterior sampling approaches cannot handle[14]. In the experimental section, we implemented ScoreALD and DPS using a pretrained score-predicting function, and the sampling loops are described below.

ScoreALD

$\boldsymbol{x_T} \sim \mathcal{N}(0, I)$

for $t = T, \ldots, 1$ **do**

  $\mathbf{z} \sim \mathcal{N}(0, I)$ **if** $t > 1$, **else** $\mathbf{z} = \mathbf{0}$

  $\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x_t} + (1-\bar{\alpha}_t)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{x_t}, t)\right)$

  $\boldsymbol{x_{t-1}}' = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \sqrt{1-\alpha_t}\,\mathbf{z}$

  $\boldsymbol{x_{t-1}} = \boldsymbol{x_{t-1}}' - \frac{1}{\sigma^2+\gamma_t^2}\nabla_{x_t}\|\boldsymbol{\mathcal{A}}(\boldsymbol{x_t}) - \boldsymbol{y}\|^2$

end for

return $\boldsymbol{x_0}$

DPS

$\boldsymbol{x_T} \sim \mathcal{N}(0, I)$

for $t = T, \ldots, 1$ **do**

  $\mathbf{z} \sim \mathcal{N}(0, I)$ **if** $t > 1$, **else** $\mathbf{z} = \mathbf{0}$

  $\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\boldsymbol{x_t} + (1-\bar{\alpha}_t)\mathbf{s}_{\boldsymbol{\theta}}(\boldsymbol{x_t}, t)\right)$

  $\boldsymbol{x_{t-1}}' = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \sqrt{1-\alpha_t}\,\mathbf{z}$

  $\boldsymbol{x_{t-1}} = \boldsymbol{x_{t-1}}' - \frac{\zeta_t}{2\sigma^2}\nabla_{x_t}\|\boldsymbol{\mathcal{A}}(\hat{\mathbf{x}}_0) - \boldsymbol{y}\|^2$

end for

return $\boldsymbol{x_0}$

IV. Experimental Results and Discussion

A. Single-Step Denoising

In this section, the diffusion model pretrained on the FFHQ-256 dataset was utilized to perform single-step denoising. Noisy images ($\mathbf{x}_t$) were generated by adding i.i.d. Gaussian noise $\mathbf{z} \sim \mathcal{N}(0, I)$ to the clean test image ($\mathbf{x}_0$), following the forward process described by equation (3) with the noise schedule β_t. Images at different noise levels were obtained by selecting various diffusion steps t in the forward noising process. The predicted denoised image ($\hat{\mathbf{x}}_0$) was obtained by applying the reverse process to the noisy image ($\mathbf{x}_t$) using the single-step

model prediction method described in equation (6), with the pretrained score predictor $\mathbf{s}_{\boldsymbol{\theta}}$.

Fig. 1 illustrates the results of denoising a test human face image at three progressively increasing noise levels ($t = 50$, $t = 100$, and $t = 200$). PSNR and LPIPS were computed as quantitative metrics to assess the quality of the denoised images. Although PSNR gradually decreases as the noise level increases, it retains a reasonable value of 29.6 dB even at the highest noise level. Conversely, LPIPS increases from 0.0549 to 0.1531 as the noise level rises from $t = 50$ to $t = 200$, reflecting a decline in perceptual similarity relative to the ground truth as the input starts with more noise and the denoising process involves more reverse diffusion steps.

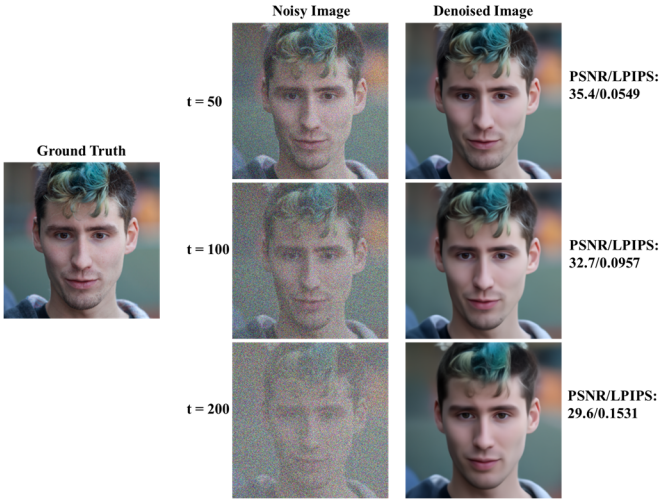

Fig. 1 Single-step denoising at different noise levels

Furthermore, this denoising method was evaluated on a non-human face input image (Fig. 2). The model exhibited plausible performance on this test image of a red panda, achieving a relatively high PSNR of 29.2 dB and a low LPIPS of 0.2123. This outcome is likely due to the red panda face shares plenty structural similarities to the human faces on which the diffusion model was trained. However, when compared to the denoised human face image at the same input noise level (t=100) the denoised red panda image demonstrates a lower PSNR and a higher LPIPS, suggesting the model indeed performs worse as the input image deviates from the distribution of the model's training data.

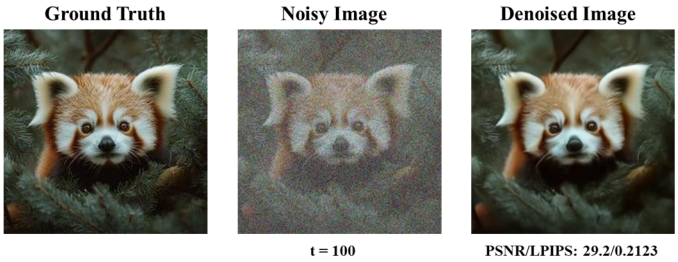

Fig. 2 Single-step denoising on a non-human face noisy image

B. *Unconditional Image Generation*

Unconditional imaging generation with the pretrained diffusion model was demonstrated in this section, using the DDPM procedure. The process began by initializing the image as pure Gaussian random noise, which was then iteratively denoised according to a parameter schedule over 1,000 steps. Three images ($\mathbf{x}_{t=0}$) were generated using this approach and are presented in Fig. 3 alongside partially denoised intermediate results at 700 ($\mathbf{x}_{t=300}$) and 950 ($\mathbf{x}_{t=50}$) steps. The generated images are all identifiable as photorealistic human portraits; however, certain imperfections are evident. For instance, the first image contains an arbitrary object in the foreground. Additionally, the third image is oversaturated, resulting in the loss of fine details, particularly the hairs.

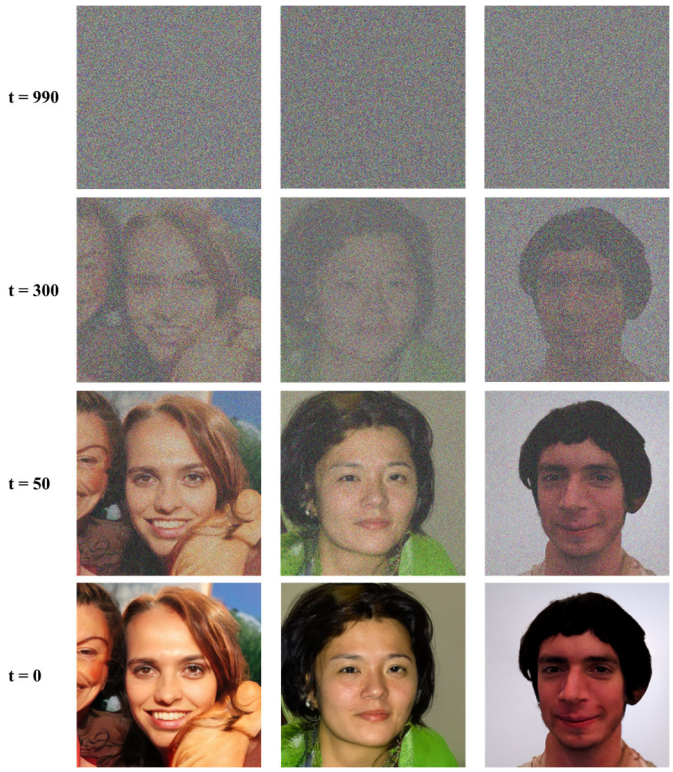

Fig. 3 Unconditional image generation

C. *Solving Inverse Problems*

Here we explored the applications of the diffusion model in solving two inverse problems: image inpainting and deconvolution with SDEdit, ScoreALD and DPS.

1) SDEdit

SDEdit is an image synthesis and editing framework based on stochastic differential equations developed by Meng et al.[12]. In this implementation of SDEdit, we first add noise to the input, then subsequently denoises the resulting partially noised version of the input image through the DDPM denoising procedure to enhance its realism.

Fig. 4, 5 exhibit the results of applying SDEdit to two tasks: inpainting an image partially obscured by a box mask and deblurring an image with by Gaussian blur. For each task, three noise levels ($t = 250$, $t = 500$, and $t = 750$) were evaluated to

analyze the impact of the noise magnitude on the overall quality of the reconstructed images.

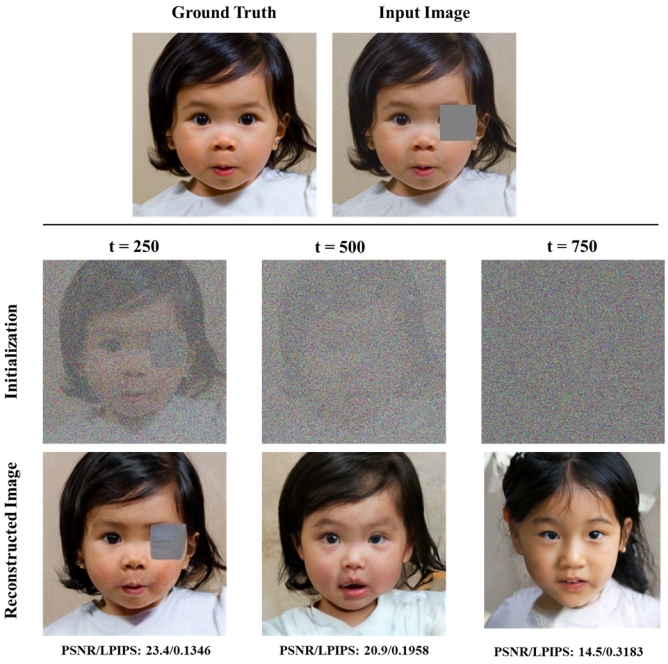

Fig. 4 Box-type inpainting with SDEdit at three noise levels

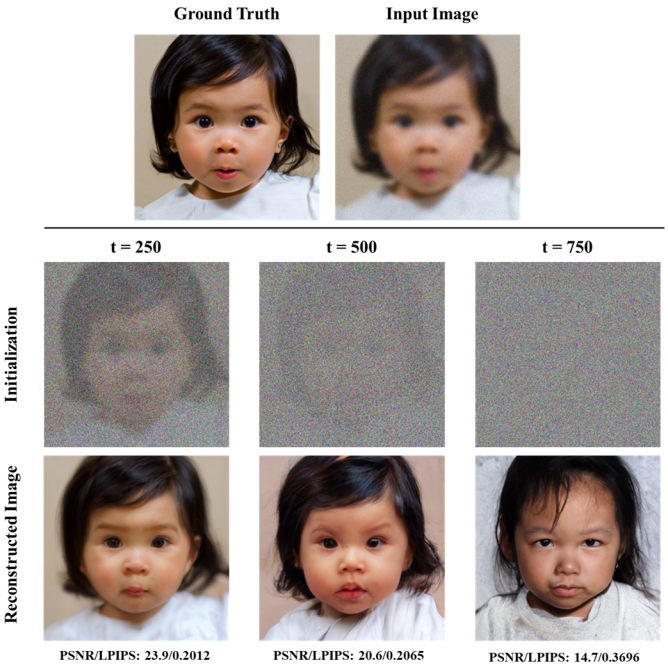

Fig. 5 Gaussian deblurring with SDEdit at three noise levels

The results reveal a clear trend: as more noise is added to the input image and additional reverse diffusion steps are taken during denoising, more artifacts from the original input are removed. However, this also leads to the erosion of other features, resulting in outcomes that are less faithful to the input image. As shown in Fig. 4, 5, when the noise level is insufficient ($t = 250$) to obscure the artifacts in the input, the inpainting or deblurring task essentially "fails", as the output images retain significant traces of the original mask or blur. On the other extreme of the spectrum, when an excessive amount of noise is added during initialization ($t = 750$), the noisy input lacks sufficient information about the original image, rendering the denoising process effectively "unconditional" and producing outputs that no longer resemble the input images.

Among the three noise levels evaluated, $t = 500$ yields the best results for both inpainting and devolution tasks, as judged by parametric metrics such as PSNR and LPIPS, as well as qualitative assessments. Nonetheless, even the results at $t = 500$ are suboptimal, highlighting the inherent difficulty of balancing restoration and faithfulness in a controllable manner in this approach. As demonstrated in the following section, posterior sampling approaches offer more effective solutions to these inverse problems.

2) ScoreALD and DPS

In this section, we experiment with two posterior sampling approaches for solving the inpainting and linear deblurring inverse problems, namely ScoreALD and DPS.

Fig. 6 summarizes the results of applying ScoreALD and DPS to box-type inpainting and Gaussian deblurring tasks. For comparison, the results from the SDEdit method at noise level $t = 500$, discussed in the previous section, are also included. Additionally, we applied these three diffusion-based approaches to the random-type inpainting task. Detailed anneal and scale parameters used for each task can be found in the code.

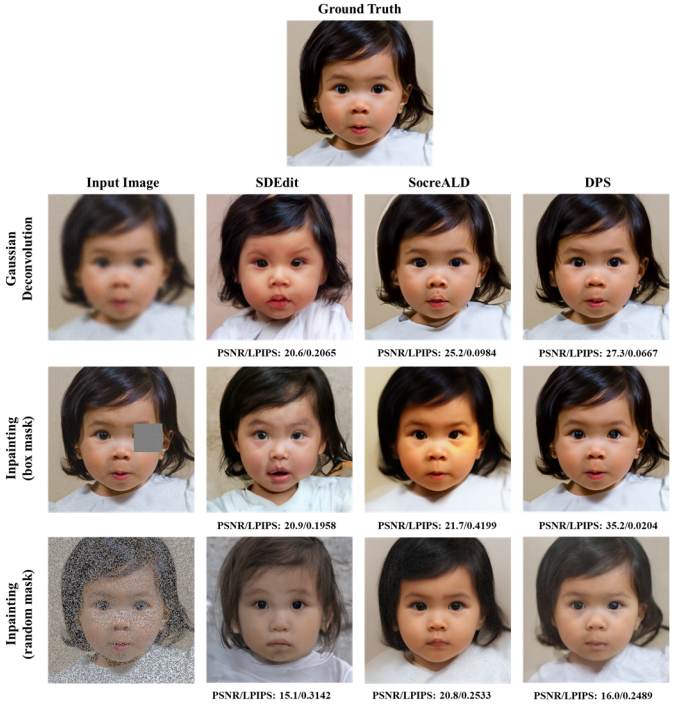

Fig. 6 Box-/random-type inpainting and Gaussian deblurring with SDEdit, ScoreALD, and DPS

As shown in Fig. 6, both posterior sampling methods significantly outperform SDEdit across all tasks, achieving consistently higher PSNR and lower LPIPS values. Between theses posterior sampling approaches, DPS demonstrates superior performance in terms of parametric metrics and qualitative evaluation, with the most notable improvements observed in the box inpainting task.

While both ScoreALD and DPS produce high-quality outputs for box-type inpainting and Gaussian deblurring tasks, neither method achieves satisfactory results for random-type inpainting using the current anneal and scale parameters. Specifically, they fail to achieve both sufficiently high PSNR and low LPIPS simultaneously. Improved outcomes are anticipated with further fine-tuning of these parameters.

V. Conclusion

In this project, we have implemented a single-step method for image denoising using a pretrained score predictor, achieving high-quality denoised outputs for moderate to high noise levels. However, as expected, the performance of the diffusion-based denoiser deteriorated when applied to images outside the model's training distribution. Images generated using the DDPM framework were generally of good quality, although occasional imperfections in composition or saturation were observed. For image inpainting and deblurring tasks, posterior sampling methods such as ScoreALD and DPS significantly outperformed the more basic SDEdit approach, consistently achieving higher PSNR and lower LPIPS values. Notably, DPS demonstrated superior performance, particularly in box-type inpainting scenarios. While SDEdit requires careful selection of noise levels to effectively perturb the input images while maintaining a certain level of faithfulness to other general features, the posterior sampling methods provided more robust and effective solutions to these inverse problems, exemplifying the modern advancements enabled by generative modeling.

Acknowledgment

The author extends gratitude to the teaching team of EE367, winter 2025, for outlining this project and providing the starter code.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: 'Generative adversarial nets', Adv Neural Inf Process Syst, 2014, 27

[2] Kingma, D.P., and Welling, M.: 'An introduction to variational autoencoders', Foundations and Trends® in Machine Learning, 2019, 12, (4), pp. 307-392

[3] Dalal, M., Li, A.C., and Taori, R.: 'Autoregressive models: What are they good for?', arXiv preprint arXiv:1910.07737, 2019

[4] Chen, R.T., Rubanova, Y., Bettencourt, J., and Duvenaud, D.K.: 'Neural ordinary differential equations', Adv Neural Inf Process Syst, 2018, 31

[5] Song, Y., and Ermon, S.: 'Generative modeling by estimating gradients of the data distribution', Adv Neural Inf Process Syst, 2019, 32

[6] Ho, J., Jain, A., and Abbeel, P.: 'Denoising diffusion probabilistic models', Adv Neural Inf Process Syst, 2020, 33, pp. 6840-6851

[7] Lemercier, J.-M., Richter, J., Welker, S., Moliner, E., Välimäki, V., and Gerkmann, T.: 'Diffusion Models for Audio Restoration: A review [Special Issue On Model-Based and Data-Driven Audio Signal Processing]', IEEE Signal Process Mag, 2025, 41, (6), pp. 72-84

[8] Song, Y., Shen, L., Xing, L., and Ermon, S.: 'Solving inverse problems in medical imaging with score-based generative models', arXiv preprint arXiv:2111.08005, 2021

[9] Virieux, J., and Operto, S.: 'An overview of full-waveform inversion in exploration geophysics', Geophysics, 2009, 74, (6), pp. WCC1-WCC26

[10] Zhao, X., Ting, Y.-S., Diao, K., and Mao, Y.: 'Can diffusion model conditionally generate astrophysical images?', Mon Not R Astron Soc, 2023, 526, (2), pp. 1699-1712

[11] Reddy, P., Toomey, M.W., Parul, H., and Gleyzer, S.: 'DiffLense: a conditional diffusion model for super-resolution of gravitational lensing data', Machine Learning: Science and Technology, 2024, 5, (3), pp. 035076

[12] Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.-Y., and Ermon, S.: 'Sdedit: Guided image synthesis and editing with stochastic differential equations', arXiv preprint arXiv:2108.01073, 2021

[13] Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., and Tamir, J.: 'Robust compressed sensing mri with deep generative priors', Adv Neural Inf Process Syst, 2021, 34, pp. 14938-14954

[14] Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., and Ye, J.C.: 'Diffusion posterior sampling for general noisy inverse problems', arXiv preprint arXiv:2209.14687, 2022

[15] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., and Salimans, T.: 'Photorealistic text-to-image diffusion models with deep language understanding', Adv Neural Inf Process Syst, 2022, 35, pp. 36479-36494

[16] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A.A.: 'Generative adversarial networks: An overview', IEEE Signal Process Mag, 2018, 35, (1), pp. 53-65

[17] Brock, A., Donahue, J., and Simonyan, K.: 'Large scale GAN training for high fidelity natural image synthesis', arXiv preprint arXiv:1809.11096, 2018

[18] Razavi, A., Van den Oord, A., and Vinyals, O.: 'Generating diverse high-fidelity images with vq-vae-2', Adv Neural Inf Process Syst, 2019, 32

[19] Kingma, D.P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M.: 'Improved variational inference with inverse autoregressive flow', Adv Neural Inf Process Syst, 2016, 29

[20] Zhan, F., Yu, Y., Wu, R., Zhang, J., Lu, S., Liu, L., Kortylewski, A., Theobalt, C., and Xing, E.: 'Multimodal image synthesis and editing: The generative AI era', IEEE Trans Pattern Anal Mach Intell, 2023, 45, (12), pp. 15098-15119

[21] Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., and Poole, B.: 'Score-based generative modeling through stochastic differential equations', arXiv preprint arXiv:2011.13456, 2020

[22] Daras, G., Chung, H., Lai, C.-H., Mitsufuji, Y., Ye, J.C., Milanfar, P., Dimakis, A.G., and Delbracio, M.: 'A Survey on Diffusion Models for Inverse Problems', CoRR, 2024