Inverse Imaging Methods with Diffusion Model Prior

Jinhyo Huh

Abstract—Inverse imaging problems aim to recover clean images from incomplete or corrupted observations. In this work, we investigate the use of a pre-trained diffusion model as a learned prior for solving inverse imaging problems. Using a score-based diffusion model trained on the Flickr-Faces-HQ dataset, we implement several inference strategies including unconditional DDPM sampling, SDEdit, Score-based Annealed Langevin Dynamics (ScoreALD), and Diffusion Posterior Sampling (DPS). We evaluate these methods on image inpainting and deconvolution. Our experiments demonstrate how diffusion models can reconstruct realistic images while enforcing consistency with observed measurements. Quantitative evaluation using PSNR and LPIPS, along with qualitative comparisons, shows that the posterior sampling approach DPS produces higher-quality reconstructions with greater data fidelity than simpler diffusion-based editing methods. These results highlight the effectiveness of diffusion models as powerful learned priors and the importance of incorporating measurement information for solving inverse problems in computational imaging.

Index Terms—Diffusion Models, Inverse Imaging, Computational Imaging, Image Inpainting, Image Deconvolution, Diffusion Posterior Sampling, Score-based Generative Models

1 Introduction

Inverse imaging problems arise when recovering a clean image from incomplete, obscured, or corrupted measurements. These problems appear in many areas of computational imaging, including image generation, inpainting, and deconvolution (Fig. 1). Mathematically, inverse imaging is commonly modeled as

$$y = Ax + \mu \tag{1}$$

where x represents the unknown clean ground truth image, A is the image formation model, and μ represents measurement noise. A fundamental challenge is that these problems are typically ill-posed: multiple possible images x may produce the same observed measurement y. As a result, the accuracy of image reconstruction is greatly enhanced by additional assumptions, such as prior knowledge about what constitutes a valid natural image or the reliance on the observed measurement to a controlled degree.

Diffusion models are highly effective generative models capable of learning the distribution of natural images, representing our belief about what a valid clean image should look like. The diffusion model learns a score function

$$s_\theta(x_t, t) = \nabla_{x_t} \log p_t(x_t) \tag{2}$$

which corresponds to the gradient of the log probability density of the data distribution—a vector field pointing toward regions of higher probability in the image distribution. During inference, this learned score provides directions that progressively transform noisy or corrupted images toward realistic images. In this work, we use a diffusion model [1] pre-trained on the Flickr-Faces-HQ Dataset, following the variance-preserving formulation of Denoising Diffusion Probabilistic Models [2], as a learned image prior to solve inverse imaging problems.

- *J. Huh is with the Department of Computer Science, Stanford University, Stanford, CA, 94305. Email: jinhyo@stanford.edu.*

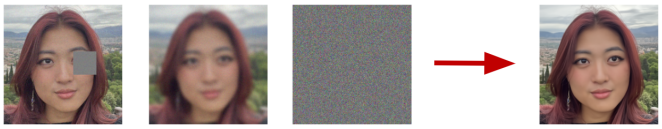

Fig. 1. The goal of inverse imaging problems is clean image reconstruction from corrupted measurements.

Using this model, we explore how diffusion priors can be integrated into reconstruction procedures that guide the solution toward the learned distribution of natural images, while balancing consistency with the observed data for inverse imaging tasks of box inpainting, random inpainting, and deconvolution. We provide a quantitative and qualitative evaluation of several inverse imaging methods with the diffusion model prior, such as SDEdit [3], ScoreALD [4], and Diffusion Posterior Sampling (DPS) [1], .

2 Related Work

Classical approaches to inverse problems include Maximum Likelihood (ML) or Maximum a Posteriori (MAP) estimation. Maximum Likelihood methods estimate the solution that best explains the observed measurements while ignoring any prior knowledge about the underlying image distribution. In contrast, MAP methods incorporate a prior and estimate the parameters that both explain the data and are plausible under the prior. However, MAP produces a single deterministic solution, providing only a point estimate rather than capturing the full distribution of feasible reconstructions. This limitation motivates the use of generative models that approximate the full data distribution and enable sampling-based reconstruction.

A foundational framework for diffusion models is the Denoising Diffusion Probabilistic Model (DDPM) intro-

Fig. 2. The DDPM forward diffusion process gradually corrupts a clean image by adding Gaussian noise over timesteps $t = 0 \rightarrow T$. A learned reverse process then iteratively denoises the image to recover a realistic sample as $t \rightarrow 0$.

duced by Ho et al. [2], which formulates image generation as discretizations of stochastic differential equations (SDEs), where the reverse-time dynamics gradually transform noise into realistic images (Fig. 2). In the forward diffusion process, Gaussian noise is progressively added to an image through a noise schedule controlled by β, defined by

$$x_t = \sqrt{1-\beta_t}\, x_{t-1} + \sqrt{\beta_t}\, z_{t-1}, \quad z_{t-1} \sim \mathcal{N}(0, I) \tag{3}$$

which can be reformulated to depend on only the first timestep $t = 0$

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, z \tag{4}$$

where

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

As we integrate methods from DDPM to posterior sampling, we can reformulate previous state prediction

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

to

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

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

that allows DPS to operate through an intermediate estimate of the clean image $\hat{x}_0$. These formulations are mathematically equivalent through the relationship between the noise schedule parameters α_t and $\bar{\alpha}_t$, but provide different perspectives that simplify implementation of diffusion-based sampling and posterior conditioning.

The DDPM model uses a noise-predicting network ϵ_θ in its one-step reverse diffusion step

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right), \tag{8}$$

but we reformulate this to Equation (5) to use the score-predicting network s_θ of the data distribution, enabling reverse diffusion to reconstruct clean images from noisy samples, leveraged for inpainting and deconvolution processes.

To evaluate reconstruction quality, we use two standard image metrics. Peak Signal-to-Noise Ratio (PSNR) measures pixel-wise fidelity between the reconstructed and ground truth images, where higher values indicate better reconstruction accuracy. In contrast, Learned Perceptual Image Patch Similarity (LPIPS) measures perceptual similarity using deep neural network features, with lower values indicating closer perceptual alignment. Together, these metrics capture both low-level accuracy and perceptual image quality.

3 Methods

In this section, we describe the diffusion-based approaches used for image generation and inverse imaging reconstruction. We begin with the standard unconditional sampling procedure for diffusion models, where images are generated from pure Gaussian noise through iterative reverse diffusion. We then extend this framework to conditional reconstruction for inverse problems by incorporating measurement information into the sampling process. Specifically, we investigate three diffusion-based methods: SDEdit, ScoreALD, and DPS. These methods leverage a pre-trained diffusion model as a prior while the latter two enforce consistency with observed measurements, enabling more accurate reconstruction of ground truth images from corrupted or incomplete observations.

3.1 Image Denoising & Unconditional Generation

As a baseline, we implement the standard sampling procedure from DDPM [2]. The diffusion model is first used as a denoiser that maps a noisy image x_t to an estimate of the underlying clean image $\hat{x}_0$ using the learned score function. This denoising step leverages the pretrained diffusion model's knowledge of the natural image distribution. Given the current noisy sample x_t and the predicted clean image $\hat{x}_0$, the reverse diffusion step estimates the most likely previous state x_{t-1}. This is done by computing the mean of the diffusion posterior distribution $p(x_{t-1} \mid x_t, \hat{x}_0)$, which represents the expected value of the previous timestep conditioned on the current state and the denoised estimate. Unconditional generation is a specific case in which the process is initialized with Gaussian noise $x_T \sim \mathcal{N}(0, I)$ and iteratively applying the reverse diffusion updates. Repeating this process across all timesteps gradually transforms noise into a realistic image sample x_0 drawn from the learned data distribution.

3.2 SDEdit

The first method implemented is SDEdit [3], a simple modification of the DDPM sampling procedure. Instead of starting from pure Gaussian noise, SDEdit begins with a masked or blurry measurement y and simulates the forward diffusion process to obtain a noisy sample at timestep t (Fig. 3).

$$x_t = \sqrt{\bar{\alpha}_t}\, y + \sqrt{1-\bar{\alpha}_t} \tag{9}$$

Here, t is a user-defined hyperparameter that controls the amount of noise added to the measurement. Starting from this partially noised state x_t, we apply the standard DDPM reverse diffusion process to iteratively denoise the image.

Fig. 3. SDEdit image reconstruction from masked or blurred measurement y, with noised y at example timestep $t = 500$.

3.3 ScoreALD

Score-based Annealed Langevin Dynamics (ScoreALD) [4] extends the diffusion sampling process by incorporating measurement consistency during the reverse diffusion steps. Unlike SDEdit, which conditions the reconstruction only through initialization, ScoreALD explicitly modifies each reverse diffusion step using a likelihood-guided gradient.

The method begins with the standard DDPM reverse diffusion procedure. Starting from Gaussian noise $x_T \sim \mathcal{N}(0, I)$, the model iteratively denoises the image. At each timestep, the diffusion model first predicts a clean image estimate (Eq. 6) which is then used to compute the posterior mean of the diffusion transition (Eq. 7).

To enforce measurement consistency, ScoreALD introduces an additional gradient step derived from the likelihood of the measurement model. Assuming measurements $y = \mathcal{A}(x) + \epsilon$, the reverse diffusion update is modified as

$$x_{t-1} = x_{t-1} - \frac{1}{2(\sigma^2 + \gamma_t^2)} \nabla_{x_t} \left\| \mathcal{A}(x_t) - y \right\|^2 . \tag{10}$$

Here, γ_t is an annealing factor that controls the strength of the measurement-guided update across timesteps t. Larger values of γ_t reduce the influence of the likelihood gradient, while smaller values increase measurement consistency of the reconstruction. This update relies on the naive approximation

$$\nabla_x \log p(b|x_0) \approx \nabla_x \log p_t(b|x_t), \tag{11}$$

which allows the likelihood gradient to be evaluated using the current noisy sample x_t. By combining the diffusion prior with measurement-driven gradients, ScoreALD performs posterior sampling that balances realism from the learned image distribution with fidelity to the observed measurements.

3.4 Diffusion Posterior Sampling (DPS)

Diffusion Posterior Sampling (DPS) [1] improves upon ScoreALD by incorporating measurement consistency directly into the reverse diffusion process using the diffusion model's estimate of the clean image. Instead of computing likelihood gradients using the noisy state x_t, DPS uses the predicted clean image $\hat{x}_0$ (Eq. 6) as guidance for the measurement update x'_{t-1} (Eq. 7).

DPS then modifies this update using a posterior-guided gradient step that enforces measurement consistency

$$x_{t-1} = x'_{t-1} - \zeta_t \nabla_{x_t} \left\| \mathcal{A}(\hat{x}_0) - y \right\|^2 . \tag{12}$$

Here $\mathcal{A}$ represents the forward measurement operator and ζ_t controls the strength of the likelihood-guided update. In practice, the step size is normalized as

$$\zeta_t = \frac{\zeta}{\|\nabla_{x_t} \|\mathcal{A}(\hat{x}_0) - y\|^2\|}. \tag{13}$$

DPS provides a better approximation to the posterior gradient by evaluating the likelihood using the clean image estimate $\hat{x}_0$ rather than the noisy sample x_t. This corresponds to the approximation

$$\nabla_x \log p(b|x_0) \approx \nabla_x \log p_t(b|x_0 = \mathbb{E}[x_0|x_t]). \tag{14}$$

By combining the diffusion prior with posterior-guided updates, DPS performs stable posterior sampling while maintaining consistency with the observed measurements.

4 Experimental Results

We evaluate the performance of diffusion-based reconstruction methods on three inverse imaging tasks: box inpainting, random inpainting, and image deconvolution. These problems represent different types of measurement corruption, from localized missing regions to widespread noise and blur. We compare three diffusion-based approaches—SDEdit, ScoreALD, and DPS—which integrate the diffusion prior with measurement information in a different way. SDEdit reconstructs images by partially noising the input measurement before applying the reverse diffusion process. ScoreALD and DPS incorporate likelihood-guided reverse diffusion to more directly enforce measurement consistency. We analyze the reconstructed images' qualitative traits and quantitative metrics of PSNR and LPIPS of across tasks.

4.1 Image Denoising and Unconditional Generation

In the results shown in Figure 4, as the timestep t increases, the amount of noise added during the forward diffusion process increases, making the denoising task more challenging. This trend is reflected in both quantitative metrics and qualitative results. At lower noise level $t = 30$, the model successfully reconstructs fine image details, achieving high reconstruction quality with PSNR of 34.96 and LPIPS of 0.033. The reconstructed image is visually nearly indistinguishable from the ground truth.

As the noise level increases to $t = 100$, the reconstruction begins to lose some high-frequency details, resulting in a decrease in PSNR to 30.18 and an increase in LPIPS to 0.094. While the overall facial structure is preserved, subtle texture information becomes less accurate. At larger timestep $t = 300$, the input image becomes heavily corrupted, and the denoising process relies more strongly on the learned diffusion prior rather than the observed image content. Consequently, reconstruction quality decreases significantly with PSNR 25.80 and LPIPS 0.208, and the output image appears smoother and less faithful to the original image.

This method operates under a trade-off between measurement fidelity and reliance on the learned image prior.

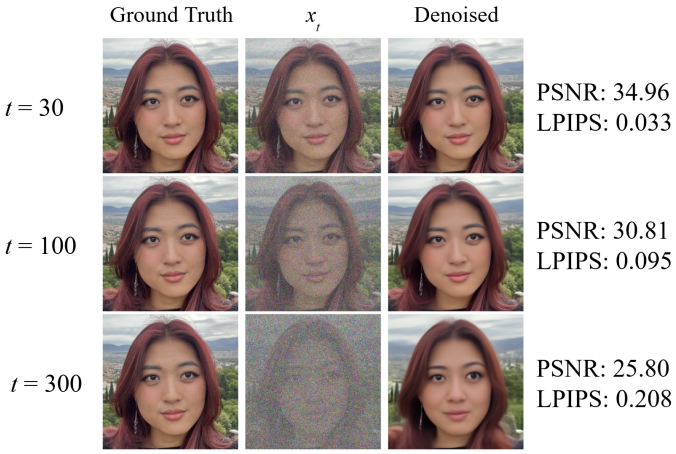

Fig. 4. Results of DDPM denoising with noise added up to different timesteps t.

Fig. 5. Results of unconditional generation. Left to right: Hui bin, Rudolph, Brent.

When the noise level is small, the model primarily recovers the original image content. At higher noise levels, however, the model increasingly generates plausible images from the learned distribution rather than strictly reconstructing the ground truth.

Unconditional generation synthesizes images by starting from pure Gaussian noise and iteratively applying the reverse diffusion process. In this case, the model relies entirely on the learned prior over natural images without conditioning on any external measurement. The three sample generated images (Fig. 5) demonstrate that the model is able to produce realistic human faces with coherent structure, consistent lighting, and plausible textures. Facial features such as eyes, hair, and skin tone appear well-formed, indicating that the learned score function successfully captures the underlying distribution of natural images.

Qualitatively, the generated faces skew toward masculine, Caucasian-appearing faces, suggesting that the model's learned distribution is biased toward these populations overrepresented in the dataset. These relatively realistic images highlight the capability of diffusion models as reasonable generative priors for natural images of human faces, forming the foundation for the inverse imaging methods that follow.

4.2 SDEdit

We evaluate SDEdit's performance for inverse imaging tasks while varying the noise level parameter t. As shown in Figure 6, for all noise levels and tasks, SDEdit produces

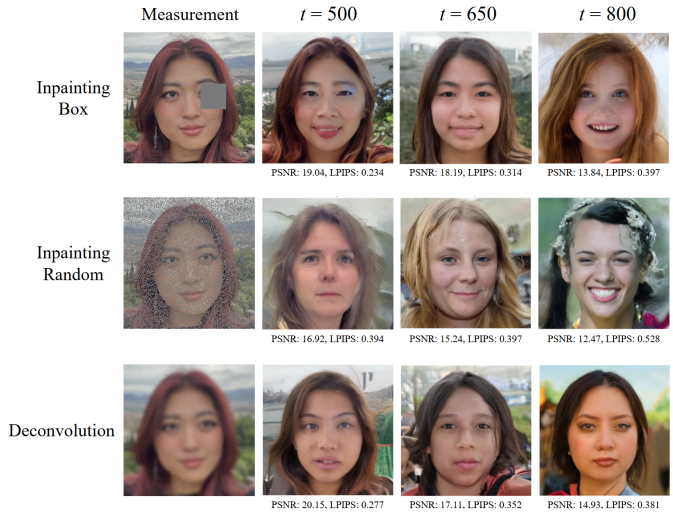

Fig. 6. Results of SDEdit in box inpainting, random inpainting, and deconvolution tasks across different noise levels t.

visually plausible reconstructions, but there is minimal coherence to the ground truth image. The most qualitatively similar result is with $t = 500$ for the box inpainting and deconvolution tasks, which also yielded the highest PSNR and lowest LPIPS scores. The smallest noise level $t = 500$ had the best quantitative scores for all three tasks, performance worsening as noise level increases—at $t = 800$, the generated images look very far from the original image. Higher additive noise levels introduce greater stochasticity into the reconstruction and the identity begins to drift from the ground truth.

Across different types of tasks, the inpainting problem with random masking led to results with the lowest PSNR and higher LPIPS scores, likely due to the highest severity of masking with stark high-frequency changes across the entire image that make the overall result more grey and less consistent with the original colors. SDEdit performs comparably in box inpainting and deconvolution tasks both quantitatively and in maintaining color vibrancy, but is unable to control measurement fidelity as the noise level t gets higher. SDEdit reconstructions illustrate the trade-off controlled by the noise level parameter t—smaller values of t preserve more information from the input measurement but may limit the model's ability to correct severe corruption. Larger values of t allow the diffusion prior to play a stronger role, producing more realistic images but potentially sacrificing fidelity to the original measurement.

4.3 ScoreALD

ScoreALD introduces a likelihood-guided correction that enforces consistency with the measurement y, controlled by the annealing factor γ. We evaluate its performance on inverse imaging tasks, with results shown in Figure 7. For box inpainting, ScoreALD is somewhat able to reconstruct the missing region while maintaining global facial structure, although there are noticeable artifacts and differences in color near the masked region. Among the tested schedules, $\gamma = [15, 15]$ achieves the highest PSNR (22.98) and lowest LPIPS (0.194), indicating the best balance between realism and measurement consistency. Increasing the annealing

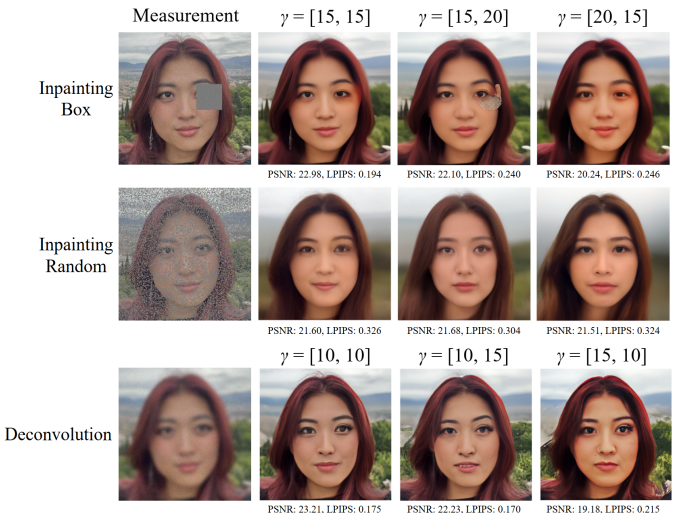

Fig. 7. Results of ScoreALD in various inverse imaging problems across different values of the annealing factor γ scheduled across timesteps t.

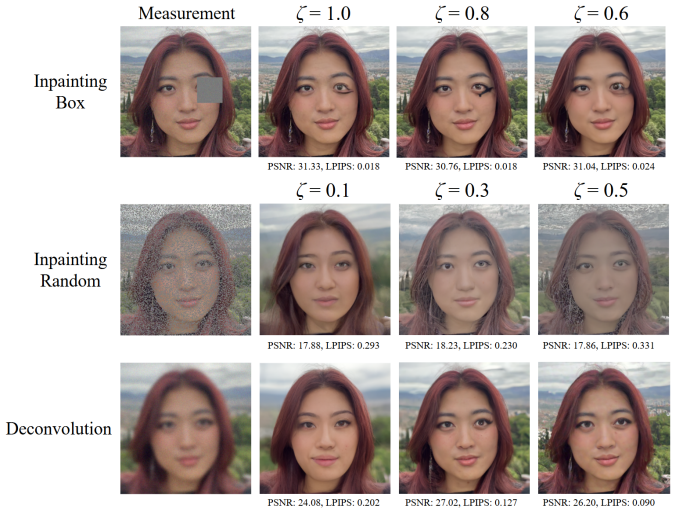

Fig. 8. Results of DPS in various inverse imaging problems with varying weight ζ of the data fidelity term.

parameter weakens the likelihood correction and leads to slightly lower fidelity to the ground truth.

Random inpainting is a more challenging task because a large fraction of pixels are corrupted. While ScoreALD still produces realistic faces, the reconstructed identity deviates more from the ground truth as the model relies more heavily on the learned diffusion prior. The PSNR values are smaller and the LPIPS values are bigger than box inpainting, indicating worse performance. For random inpainting, an annealing factor schedule $\gamma = [15, 20]$ that increases (and decreases weight of data fidelity) as the reconstruction approaches the clean image has the highest PSNR and lowest LPIPS, performing the best. For image deconvolution, ScoreALD effectively sharpens the blurred measurement and recovers high-frequency facial features while aligning relatively well to the observed measurement. The schedule $\gamma = [10, 15]$ achieves the lowest LPIPS (0.170), indicating better perceptual similarity, while $\gamma = [10, 10]$ achieves the highest PSNR (23.21).

When γ decreases over time (e.g., $[15, 10]$), the likelihood correction, pushing the image toward the measurement, becomes stronger in later diffusion steps. This allows the diffusion process to drift toward a plausible but measurement-inconsistent solution before the correction is applied, which results in reduced reconstruction fidelity and led to lower PSNR and higher LPIPS across all tasks. For ScoreALD, it appears that either a constant or increasing annealing factor yields the best results—enforcing data fidelity in the beginning of the denoising process to ensure alignment of the general shape is more effective than later on, which just refines smaller details. Overall, these results highlight the importance of tuning the annealing parameter γ. Smaller values strengthen the measurement correction but can introduce artifacts, while larger values allow the diffusion prior to dominate. A balanced schedule provides the best trade-off between reconstruction fidelity and perceptual realism.

4.4 Diffusion Posterior Sampling (DPS)

We evaluate DPS on its results in Figure 8 for box inpainting, random inpainting, and image deconvolution while varying the scaling parameter ζ, which is directly correlated with the strength of the likelihood-guided update, how strongly the reconstruction is pulled toward measurement consistency. For box inpainting, DPS achieves high-quality reconstructions across a range of ζ values, as indicated by the high PSNR of above 30 and low LPIPS of around 0.02. Qualitatively, we see that DPS, compared with the other methods like SDEdit and ScoreALD, maintains almost exact fidelity to the measurement in the unmasked regions. However, the box inpainting task doesn't create very plausible results, perhaps due to the ζ value being too high and suppressing the diffusion model prior.

For random inpainting, the task is more challenging due to the large amount of missing pixel information and the severe loss of colors of the observed measurement. Lower ζ values produce smoother, more vibrant, but less accurate reconstructions, reflected in lower PSNR and higher LPIPS. Increasing ζ improves measurement consistency up to a point, but overly large values reintroduces noise similar to the measurement and degrade perceptual quality. For deconvolution, increasing ζ generally improves sharpness and perceptual quality, as reflected by decreasing LPIPS (from 0.202 to 0.090). However, excessively large ζ can lead to oversharpening or amplified artifacts like noise from the observed measurement.

Overall, ζ controls the trade-off between adherence to the measurement and reliance on the diffusion prior. Smaller values favor realism but may ignore the measurement, while larger values enforce consistency but risk instability. Moderate values provide the best balance across tasks. DPS qualitatively produces reconstructions that are most similar in shape and color to the ground truth image compared to prior methods.

4.5 Comparison of Diffusion Prior Methods

For box inpainting and deconvolution inverse imaging problems, DPS achieves the strongest quantitative performance, with significantly higher PSNR and lower LPIPS, indicating both better pixel fidelity and perceptual quality.

TABLE 1
Quantitative comparison of SDEdit, ScoreALD, and DPS across inverse imaging tasks.

	SDEdit			ScoreALD			DPS		
	t	PSNR	LPIPS	γ	PSNR	LPIPS	ζ	PSNR	LPIPS
Box Inpainting	**500**	**19.04**	**0.234**	**[15, 15]**	**22.98**	**0.194**	**1.0**	**31.33**	**0.018**
	650	18.19	0.314	$[15, 20]$	22.10	0.240	0.8	30.76	0.018
	800	13.84	0.397	$[20, 15]$	20.24	0.246	0.6	31.04	0.024
Random Inpainting	**500**	**16.92**	**0.394**	$[15, 15]$	21.60	0.326	0.1	17.88	0.293
	650	15.24	0.397	**[15, 20]**	**21.68**	**0.304**	**0.3**	**18.23**	**0.230**
	800	12.47	0.528	$[20, 15]$	21.51	0.324	0.5	17.86	0.331
Deconvolution	**500**	**20.15**	**0.277**	$[10, 10]$	**23.21**	0.175	0.1	24.08	0.202
	650	17.11	0.352	$[10, 15]$	22.23	**0.170**	0.3	**27.02**	0.127
	800	14.93	0.381	$[15, 10]$	19.18	0.215	0.5	26.20	**0.090**

This is especially pronounced in structured tasks such as box inpainting, where strong measurement guidance with $\zeta = 1.0$ leads to quantitatively accurate reconstruction. ScoreALD consistently improves over the baseline DDPM (SDEdit) by incorporating likelihood gradients, yielding better alignment with the measurement while maintaining reasonable perceptual quality. However, its performance remains sensitive to the choice of γ, and suboptimal scheduling can lead to either oversmoothing or artifacts. ScoreALD also achieves the highest PSNR for random inpainting tasks, while DPS produces the lowest LPIPS scores—differing results across the two metrics.

SDEdit performs worst quantitatively, particularly as t increases, where reconstructions drift further from the measurement. While it produces plausible images due to the strong diffusion prior, it lacks explicit measurement enforcement, resulting in lower PSNR and higher LPIPS. The results of these methods with a diffusion model prior highlight the importance of explicitly incorporating measurement information: DPS and ScoreALD methods that balance diffusion priors with likelihood guidance significantly outperform the purely prior-driven approach of SDEdit.

5 Conclusion

In this work, we explored the use of diffusion models as learned priors for solving inverse imaging problems, including box inpainting, random inpainting, and image deconvolution. Building from the baseline DDPM forward and reverse diffusion process, we compared three approaches—SDEdit, ScoreALD, and Diffusion Posterior Sampling (DPS)—that incorporate the diffusion prior with measurement information to varying degrees.

Our results demonstrate that explicitly enforcing measurement consistency is critical for accurate reconstruction. While SDEdit produces visually plausible images, it lacks strong data fidelity and degrades as noise increases. ScoreALD improves reconstruction by incorporating likelihood-based corrections, but its performance is sensitive to the annealing schedule and relies on a naive approximation that evaluates gradients using the noisy sample. In contrast, DPS achieves the best overall performance, consistently producing reconstructions with higher PSNR and lower LPIPS by leveraging a more accurate posterior-guided update based on the diffusion model's predicted clean image.

These findings highlight the effectiveness of diffusion models as powerful priors for inverse problems, particularly when combined with principled posterior sampling techniques. Future work could explore improved scheduling strategies, more robust likelihood formulations, and applications to more complex or real-world imaging systems.

Acknowledgments

The author would like to thank Professor Gordon Wetzstein and Sonia Kim for their instruction and support in the class CS 448I: Computational Imaging.

References

[1] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion Posterior Sampling for General Noisy Inverse Problems," in *International Conference on Learning Representations*, 2023.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[3] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations," in *International Conference on Learning Representations*, 2022.

[4] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, "Robust compressed sensing MRI with deep generative priors," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

Appendix

Equivalence of Eq. 3 and Eq. 4

Let us recursively substitute previous states to express x_t directly in terms of the original image x_0.

For $t = 2$,

$$\begin{aligned} x_2 &= \sqrt{\alpha_2} x_1 + \sqrt{1-\alpha_2} z_1 \\ &= \sqrt{\alpha_2}\left(\sqrt{\alpha_1} x_0 + \sqrt{1-\alpha_1} z_0\right) + \sqrt{1-\alpha_2} z_1 \\ &= \sqrt{\alpha_1 \alpha_2}\, x_0 + \sqrt{\alpha_2(1-\alpha_1)}\, z_0 + \sqrt{1-\alpha_2}\, z_1. \end{aligned}$$

Continuing this expansion recursively yields

$$x_t = \sqrt{\alpha_t \alpha_{t-1} \cdots \alpha_1}\, x_0 + \sum_{i=1}^{t} \left(\sqrt{(1-\alpha_i) \prod_{j=i+1}^{t} \alpha_j}\, z_{i-1} \right).$$

Define the cumulative product

$$\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i.$$

The coefficient of x_0 becomes $\sqrt{\bar{\alpha}_t}$. The remaining terms are a linear combination of independent Gaussian variables z_i. Since a weighted sum of independent Gaussian variables is also Gaussian, these terms can be combined into a single Gaussian variable $z \sim \mathcal{N}(0, I)$ and the total variance of the noise terms equals $1 - \bar{\alpha}_t$, which yields the closed-form expression

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, z, \quad z \sim \mathcal{N}(0, I).$$

Thus, the forward diffusion process can be written directly as a conditional distribution depending only on the original image x_0.

Equivalence of Eq. 5 and Eq. 6, 7

Substitute the expression for $\hat{x}_0$ into the formula for x_{t-1}:

$$x_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \cdot \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t + (1-\bar{\alpha}_t) s_\theta(x_t, t) \right).$$

Using

$$\bar{\alpha}_t = \alpha_t \bar{\alpha}_{t-1},$$

we get

$$\frac{\sqrt{\bar{\alpha}_{t-1}}}{\sqrt{\bar{\alpha}_t}} = \frac{1}{\sqrt{\alpha_t}}.$$

Therefore,

$$\begin{aligned} x_{t-1} &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} \left(x_t + (1-\bar{\alpha}_t) s_\theta(x_t, t) \right) \\ &= \left[\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} + \frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} \right] x_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}} s_\theta(x_t, t). \end{aligned}$$

Now simplify the coefficient of x_t:

$$\begin{aligned} \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} + \frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} &= \frac{\alpha_t(1-\bar{\alpha}_{t-1}) + (1-\alpha_t)}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} \\ &= \frac{1-\alpha_t\bar{\alpha}_{t-1}}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} \\ &= \frac{1-\bar{\alpha}_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} \\ &= \frac{1}{\sqrt{\alpha_t}}. \end{aligned}$$

Hence,

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

Thus, the two forms of the DDPM reverse diffusion step are equivalent.

Equivalence of Eq. 5 and Eq. 8

Using Tweedie's formula for the variance-preserving diffusion process,

$$\hat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t + (1-\bar{\alpha}_t) \nabla_{x_t} \log p_t(x_t) \right),$$

and identifying the score function as

$$s_\theta(x_t, t) = \nabla_{x_t} \log p_t(x_t),$$

we have

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

From the forward diffusion equation,

$$x_t = \sqrt{\bar{\alpha}_t}\, \hat{x}_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t),$$

so solving for $\hat{x}_0$ gives

$$\hat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t) \right).$$

Equating the two expressions for $\hat{x}_0$,

$$\frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t + (1-\bar{\alpha}_t) s_\theta(x_t, t) \right) = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t) \right).$$

Multiplying both sides by $\sqrt{\bar{\alpha}_t}$ and canceling x_t yields

$$(1-\bar{\alpha}_t) s_\theta(x_t, t) = -\sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(x_t, t),$$

and therefore

$$s_\theta(x_t, t) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(x_t, t).$$

Substituting this into the reverse diffusion update,

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

Thus, the score-prediction and noise-prediction forms are equivalent.