Diffusion Models for Solving Inverse Problems

Alaz Cig Department of Mechanical Engineering, Stanford University EE367/CS448I: Computational Imaging alaz@stanford.edu

Abstract

Recovering a signal from incomplete or degraded measurements is an ill-posed problem that requires prior information to resolve. Diffusion models learn expressive image priors through score matching, providing a principled framework for regularizing such inverse problems. We implement and compare three strategies for conditioning the reverse process of a pretrained DDPM (FFHQ-256) on measurements: SDEdit (no gradient), Score-based Annealed Langevin Dynamics (ScoreALD, annealed likelihood gradient), and Diffusion Posterior Sampling (DPS, unit-normalized gradient). On box inpainting and Gaussian deconvolution, DPS achieves the best reconstruction (28.76 dB PSNR, 0.055 LPIPS on deconvolution), outperforming ScoreALD by +3.1 dB. Gradient normalization is the key mechanism: it stabilizes correction magnitudes across timesteps and eliminates task-specific annealing schedules.

1. Introduction

Many problems in computational imaging require recovering a clean signal $x \in \mathbb{R}^n$ from a degraded measurement $y \in \mathbb{R}^m$ related by

$$y = A(x) + n$$

where A is a (possibly non-invertible) forward operator and n is measurement noise. These problems are ill-posed: many signals x are consistent with y, so prior information is needed.

Classical priors such as total variation or sparsity are effective for structured signals but cannot capture the statistics of natural images; GAN-based learned priors improve quality but suffer from mode collapse or require task-specific training.

Diffusion models offer an alternative. A DDPM [1] learns to reverse a gradual noising process, thereby learning the score function $\nabla_x \log p(x)$ of the data distribution. This score serves as an implicit prior: during reverse diffusion, gradient-based corrections can steer samples toward measurement consistency. Several conditioning strategies have been proposed, ranging from gradient-free initialization (SDEdit [2]) to annealed likelihood gradients (ScoreALD [3]) to normalized gradients (DPS [4]).

In this work, we present a unified comparison of all three conditioning strategies on a single pretrained DDPM (FFHQ-256), evaluating on box inpainting and Gaussian deconvolution using both pixel-level (PSNR) and perceptual (LPIPS) metrics. We demonstrate that gradient normalization is the key mechanism behind DPS's advantage, decoupling correction direction from timestep-dependent magnitude and thereby eliminating the need for task-specific annealing schedules.

2. Related Work

Diffusion models for generation. Ho et al. [1] introduced DDPMs, training a network to predict noise at each diffusion step and generating images by iterative denoising. Song and Ermon [5] connected this to score matching, showing the denoiser implicitly learns $\nabla_x \log p_t(x)$. This enables posterior sampling $p(x|y)$ by modifying the learned score with a likelihood term.

Conditioning for inverse problems. SDEdit [2] initializes the reverse process from a noised measurement at timestep t^*, requiring no gradients but offering no data-fidelity guarantee. Jalal et al. [3] introduced ScoreALD, adding a likelihood gradient $\nabla_x \|y - A(\hat{x}_0)\|$ at each step with annealed step sizes. Chung et al. [4] proposed DPS, which normalizes this gradient to unit norm, yielding timestep-independent update magnitudes. Our work compares these three approaches under a unified implementation, focusing on how gradient scaling determines reconstruction quality.

3. Method

3.1 DDPM Background

The forward process adds Gaussian noise over $T = 1000$ steps with a linear schedule ($\beta_1 = 10^{-4}$, $\beta_T = 0.02$). The marginal at timestep t has closed form:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}\, x_0,\ (1 - \bar{\alpha}_t)I)$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1 - \beta_s)$. Equivalently:

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

The reverse process uses a neural network $\epsilon_\theta(x_t, t)$ trained to predict the noise ϵ. This noise prediction is related to the score function by $\nabla_{x_t} \log p_t(x_t) = -\epsilon_\theta(x_t, t)/\sqrt{1 - \bar{\alpha}_t}$, connecting DDPMs to score-based generative modeling [5]. Given the predicted noise, the clean image estimate $\hat{x}_0$ is recovered via Tweedie's formula:

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

The posterior mean for the reverse step is then:

$$\mu_\theta(x_t, t) = \frac{\sqrt{\bar{\alpha}_{t-1}}\, \beta_t}{1 - \bar{\alpha}_t} \hat{x}_0 + \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t$$

A reverse step samples $x_{t-1} \sim \mathcal{N}(\mu_\theta, \sigma_t^2 I)$ where σ_t^2 is the posterior variance, learned via interpolation between β_t and $\tilde{\beta}_t$.

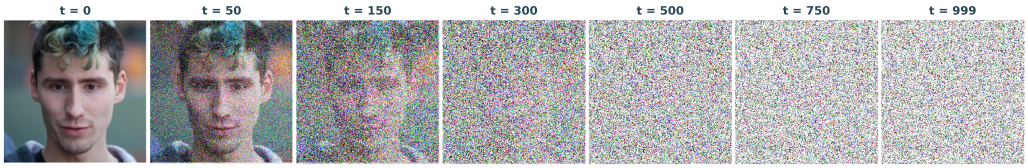

Figure 1: Forward process: progressive noise addition from clean image ($t = 0$) to Gaussian noise ($t = T$).

3.2 Inverse Problem Setup

We consider two linear inverse problems with observation model $y = Ax + n$, $n \sim \mathcal{N}(0, \sigma_n^2 I)$, $\sigma_n = 0.05$:

Box inpainting. A 50×50 pixel region is masked from a 256×256 image. The forward operator $A = M$ is a binary mask (zeros in the masked region).

Gaussian deconvolution. The image is convolved with a 61×61 Gaussian kernel ($\sigma_k = 3.0$). The forward operator $A = K$ is the convolution.

3.3 SDEdit

SDEdit [2] uses no gradient computation. The measurement y is noised to an intermediate timestep t^*:

$$x_{t^*} = \sqrt{\bar{\alpha}_{t^*}}\, y + \sqrt{1 - \bar{\alpha}_{t^*}}\, \epsilon$$

The standard reverse process then runs from t^* to $t = 0$. The parameter t^* controls the tradeoff between measurement fidelity and prior influence: small t^* preserves the measurement but limits the model's ability to fill missing information, while large t^* allows more generation freedom at the cost of consistency. We use $t^* = 500$.

3.4 ScoreALD

ScoreALD [3] adds a measurement-consistency gradient at each reverse step. After computing the standard reverse sample x_t:

$$x_t \leftarrow x_t - \lambda_t \nabla_{x_{t+1}} \|y - A(\hat{x}_0)\|_2$$

where $\hat{x}_0$ is the Tweedie estimate and λ_t is an annealed step size, computed via automatic differentiation. The schedule increases λ_t linearly as $t \to 0$ (stronger data fidelity at lower noise): $\lambda_t \in [10, 15]$ for deconvolution, $[15, 20]$ for inpainting.

3.5 DPS

DPS [4] uses the same likelihood gradient but normalizes it to unit L_2 norm:

$$g = \nabla_{x_{t+1}} \|y - A(\hat{x}_0)\|_2$$

$$x_t \leftarrow x_t - s \cdot \frac{g}{\|g\|_2}$$

where s is a fixed scale ($s = 0.3$ for deconvolution, $s = 1.0$ for inpainting). The normalization ensures constant correction magnitude regardless of timestep; its implications are analyzed in Section 5.1.

3.6 Experimental Details

We use a pretrained DDPM (U-Net, 128 base channels, attention at resolution 16, 4 heads) trained on FFHQ-256. All experiments use a single test image (`00003.png`) at 256×256 resolution.

Metrics: **PSNR** (pixel-level fidelity, dB, higher is better) and **LPIPS** [6] (perceptual distance via AlexNet, lower is better). As a baseline, we include single-pass denoising: an image noised to $t = 100$ and denoised in one step via Tweedie's formula.

4. Results

4.1 Quantitative Results

Method	Task	PSNR (dB) ↑	LPIPS ↓
Single-pass denoise	Denoise ($t = 100$)	32.72	0.092
SDEdit ($t^* = 500$)	Inpaint (box)	20.18	0.185
ScoreALD	Deconv	25.69	0.163
DPS	Deconv	28.76	0.055

Table 1. Quantitative results. ScoreALD and DPS are directly comparable on deconvolution (+3.07 dB for DPS). SDEdit is on inpainting (a different task), so cross-method comparison is qualitative.

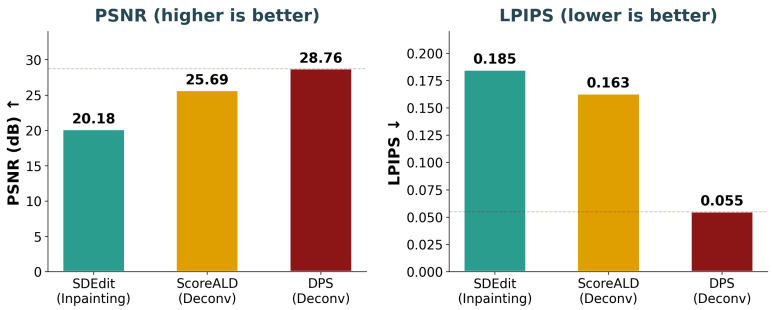

Figure 2: PSNR and LPIPS comparison across methods.

4.2 Single-Pass Denoising

The single-pass baseline achieves 32.72 dB PSNR at $t = 100$, confirming that the pretrained model has learned an accurate score function at moderate noise levels and can serve as a reliable prior for inverse problems.

Figure 3: Single-pass denoising at $t = 100$. Left: original. Middle: noised. Right: denoised.

4.3 Unconditional Generation

Starting from pure noise $x_T \sim \mathcal{N}(0, I)$, the 1000-step reverse process produces realistic face images (Figure 4), confirming both the model's learned prior and our implementation correctness.

Figure 4: Unconditionally generated face from the pretrained DDPM.

4.4 SDEdit Inpainting

SDEdit fills the masked region with plausible content (20.18 dB PSNR, 0.185 LPIPS), generating features consistent with the surrounding face; however, it lacks pixel-level fidelity since no data-consistency term is used. The method runs in roughly half the time of gradient-based approaches (500 steps, no backpropagation), making it the most practical option when approximate consistency suffices.

Figure 5: SDEdit inpainting ($t^* = 500$). Left: masked measurement. Middle: original. Right: reconstruction.

4.5 ScoreALD vs. DPS on Deconvolution

Both methods recover structure from the blurred input, but DPS is substantially sharper. ScoreALD achieves 25.69 dB with residual blur visible in hair and eye regions. DPS reaches 28.76 dB (+3.07 dB) and reduces LPIPS from 0.163 to 0.055 — a 66% perceptual improvement, larger than the PSNR gap alone suggests.

Figure 6: ScoreALD deconvolution. Left: blurred measurement. Middle: original. Right: reconstruction.

Figure 7: DPS deconvolution. Left: blurred measurement. Middle: original. Right: reconstruction.

4.6 Detail Comparison

A zoomed comparison of the eye region (Figure 8) highlights the qualitative gap: ScoreALD retains smoothing artifacts in high-frequency regions, while DPS recovers fine textures including eyelashes and iris detail, consistent with its lower LPIPS score.

Figure 8: Zoomed eye-region crop: ScoreALD (left) vs DPS (right).

5. Discussion

5.1 Why Gradient Normalization Works

Both methods compute the same gradient $g_t = \nabla_x \|y - A(\hat{x}_0)\|_2$ but scale it differently, accounting for a +3.07 dB gap.

The issue is that $\|g_t\|$ varies by orders of magnitude across timesteps. At $t \approx T$, the Tweedie estimate $\hat{x}_0$ is noisy, making $\|g_t\|$ large but its direction unreliable. At $t \approx 0$, the estimate is accurate and the direction informative, but $\|g_t\|$ has shrunk as $\|y - A(\hat{x}_0)\|$ decreases. ScoreALD's linear annealing partially compensates but cannot match the true $\|g_t\|$ profile.

DPS normalizes $g_t/\|g_t\|$, removing magnitude dependence entirely and ensuring that the fixed scale s controls the step size directly in image space. This has two effects: (1) noisy early-timestep gradients are attenuated rather than amplified; (2) small but accurate late-timestep gradients are preserved at full strength s, thereby concentrating effective corrections precisely where gradient direction is most reliable.

5.2 Method Tradeoffs

SDEdit is the simplest and fastest method (no gradients, 500 steps), well-suited for tasks where approximate consistency suffices but unable to enforce explicit data fidelity. ScoreALD introduces measurement consistency via likelihood gradients with an annealed schedule; however, the schedule requires task-specific tuning, and the ranges used here ($[10, 15]$ and $[15, 20]$ for deconvolution and inpainting) were determined through manual search. DPS achieves the best results with the simplest parameterization: a single scale parameter per task replaces ScoreALD's two-endpoint schedule, reducing the tuning burden while improving reconstruction quality.

5.3 Limitations

All quantitative results are reported on a single FFHQ-256 test image, so performance may vary across poses and lighting conditions. The pretrained model is trained exclusively on FFHQ; generalization to other image domains would require retraining or domain adaptation. Gradient-based methods (ScoreALD, DPS) require backpropagation at each of 1000 reverse steps, resulting in 15–20 minutes per reconstruction on a T4 GPU. The current implementation also assumes differentiable linear forward operators; extending to nonlinear models would require approximations to the likelihood gradient.

5.4 Future Work

Replacing the 1000-step DDPM sampler with DDIM [7] could reduce sampling to 50–100 steps while maintaining reconstruction quality. Evaluating on multiple images and datasets would establish statistical robustness, and extending DPS to nonlinear forward models such as JPEG compression or phase retrieval would test generality beyond the linear setting. A systematic sensitivity analysis of the DPS scale parameter and ScoreALD annealing schedule across tasks would provide practical guidance for applying these methods to new inverse problems.

6. Conclusion

We implemented and compared three strategies for solving inverse problems with a pretrained DDPM: SDEdit (no gradient), ScoreALD (annealed gradient), and DPS (normalized gradient). On FFHQ-256 face images, DPS achieves the best reconstruction quality on deconvolution (28.76 dB PSNR, 0.055 LPIPS), outperforming ScoreALD by +3.07 dB. The key insight is that normalizing the likelihood gradient to unit norm decouples the correction direction from its timestep-dependent magnitude, concentrating effective optimization at late timesteps where the gradient direction is most reliable. This single design choice transforms a tuning-sensitive method into a robust one, underscoring that how gradients are scaled matters as much as whether they are used at all.

References

[1] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[2] 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 (ICLR)*, 2022.

[3] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Baraniuk, "Robust compressed sensing MRI with deep generative priors," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[4] 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 (ICLR)*, 2023.

[5] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[6] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[7] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations (ICLR)*, 2021.