Diffusion Models for Image Generation and Inversion

Nareauphol (Gavin) Liu and William Zhang

Abstract—Diffusion models have emerged as a powerful class of generative models capable of learning complex data distributions through iterative denoising. This project investigates the application of a pretrained Denoising Diffusion Probabilistic Model (DDPM) to two central tasks in computational imaging: unconditional image generation and the solution of linear inverse problems. We implement and compare three conditioning strategies: SDEdit, Score-Based Annealed Langevin Dynamics (ScoreALD), and Diffusion Posterior Sampling (DPS) for image inpainting and deconvolution. Using PSNR and LPIPS, we find that DPS consistently outperforms the other methods. On Deconvolution, DPS achieves PSNR=28.25dB and LPIPS=0.055, compared to ScoreALD (PSNR=25.35dB, LPIPS=0.115). On box-mask inpainting, DPS achieves 31.8dB PSNR, compared to ScoreALD (22.7dB) and SDEdit (19.42dB, LPIPS=0.218). These results highlight the benefit of computing likelihood gradients with respect to the predicted clean image $\hat{\mathbf{x}}_0$ rather than the noisy sample, as done in DPS.

Index Terms—Computational Photography, Diffusion Models, Signal Processing

1 Introduction

Inverse problems in computational imaging arise whenever we wish to recover a latent clean signal $\mathbf{x}_0 \in \mathbb{R}^N$ from a set of noisy, incomplete, or degraded observations $\mathbf{y} = A\mathbf{x}_0 + \boldsymbol{\eta}$, where $A \in \mathbb{R}^{M\times N}$ is a known forward operator and $\boldsymbol{\eta}$ is additive noise usually modeled as Gaussian or Poisson noise. Because $M \ll N$ in many practical settings, the problem is severely ill-posed due to two main reasons: We cannot model noise precisely due to not knowing the exact distribution of noise and with occlusion or blurring, there are infinitely many signals that are consistent with the measurements.

Classical regularization priors, such as total variation [1], sparsity-promoting norms, and hand-crafted smoothness terms, can partially address this ill-posedness, but rarely capture the more complex statistical structure of natural images. Machine learning approaches, particularly deep generative models, offer an attractive alternative because they can implicitly learn high-dimensional image priors from large datasets.

Denoising Diffusion Probabilistic Models (DDPMs) [2] are a class of score-based generative models that learn to reverse a gradual noising process. Once trained, a DDPM can be repurposed as an expressive image prior for downstream inverse problems without any additional fine-tuning. In this project we use a DDPM pretrained on the FFHQ-256 dataset [3] and implement three algorithms: SDEdit [4], ScoreALD [1], and DPS [5] to solve the inpainting and deconvolution inverse problems.

2 Related Work

2.1 Diffusion Models

Diffusion models were first introduced by Sohl-Dickstein et al. [6] as latent variable models based on non-equilibrium thermodynamics. Ho et al. [2] popularized the DDPM formulation and showed that predicting the added noise $\boldsymbol{\epsilon}$ as a simplified objective results in high-quality image synthesis. Song et al. [7] connected DDPMs to score-based generative modeling via stochastic differential equations (SDEs), unifying a broad family of diffusion models under a common framework.

2.2 Inverse Problems with Diffusion Priors

Using a pretrained diffusion model as an image prior for inverse problems has become an active research direction. Early work by Meng et al. [4] proposed SDEdit, which applies partial forward noising followed by full reverse diffusion and requires no knowledge of the forward operator at sampling time. Jalal et al. [1] introduced ScoreALD, which explicitly incorporates a likelihood gradient at each reverse step and uses annealing to stabilize convergence. Chung et al. [5] proposed DPS, which approximates the intractable conditional score using the diffusion model's one-step denoised estimate $\hat{\mathbf{x}}_0$ and normalizes the gradient update for robustness.

3 Methods

3.1 Mathematical Background and Derivations

All methods in this project use the variance-preserving (VP) formulation of diffusion models introduced by Ho et al. [2]. Three key identities are derived below.

3.1.1 Derivation 1: Closed-Form Forward Process

The VP forward noise model from step $t-1$ to step t is:

$$\mathbf{x}_t = \sqrt{1-\beta_t}\,\mathbf{x}_{t-1} + \sqrt{\beta_t}\,\mathbf{z}_{t-1}, \tag{1}$$

$$\mathbf{z}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2}$$

where β_t is the noise schedule. We claim that the forward process admits the closed-form marginal:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\mathbf{z}, \tag{3}$$

$$\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{4}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$.
Proof by recursion. Substitute $\mathbf{x}_{t-1}$ using the same recursive logic:

$$\mathbf{x}_t = \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\mathbf{z}_{t-2}) + \sqrt{1-\alpha_t}\mathbf{z}_{t-1}$$

$$\mathbf{x}_t = \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\mathbf{z}_{t-2} + \sqrt{1-\alpha_t}\mathbf{z}_{t-1}$$

Using the property that the sum of two independent Gaussians $\mathcal{N}(0, \sigma_1^2\mathbf{I})$ and $\mathcal{N}(0, \sigma_2^2\mathbf{I})$ is a new Gaussian $\mathcal{N}(0, (\sigma_1^2 + \sigma_2^2)\mathbf{I})$, we combine the noise terms:

$$\sigma^2_{combined} = (\sqrt{\alpha_t(1-\alpha_{t-1})})^2 + (\sqrt{1-\alpha_t})^2$$

$$\sigma^2_{combined} = \alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t = 1 - \alpha_t\alpha_{t-1}$$

This simplifies the expression to:

$$\mathbf{x}_t = \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\mathbf{z}'$$

By repeating this recursion t times until reaching x_0:

$$\mathbf{x}_t = \sqrt{\alpha_t\alpha_{t-1}\dots\alpha_1}x_0 + \sqrt{1-\alpha_t\alpha_{t-1}\dots\alpha_1}z$$

Applying the definition $\bar{\alpha}_t = \prod_{i=1}^{t}\alpha_i$:

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

3.1.2 Derivation 2: DDPM Reverse Diffusion Process

The DDPM reverse process is often written using the posterior mean:

$$\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\big(\mathbf{x}_t + (1-\bar{\alpha}_t)\,\mathbf{s}_\theta\big), \tag{5}$$

$$\mathbf{x}_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\,\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\,\hat{\mathbf{x}}_0. \tag{6}$$

We prove that this equals the simpler one-step update:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\big(\mathbf{x}_t + (1-\alpha_t)\,\mathbf{s}_\theta\big). \tag{7}$$

Proof. Substitute (5) into (6) and use $\bar{\alpha}_t = \alpha_t\bar{\alpha}_{t-1}$, which can be written as $\sqrt{\bar{\alpha}_{t-1}/\bar{\alpha}_t} = 1/\sqrt{\alpha_t}$:

$$\mathbf{x}_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\,\mathbf{x}_t + \frac{(1-\alpha_t)}{\sqrt{\alpha_t}(1-\bar{\alpha}_t)}\,\mathbf{x}_t + \frac{(1-\alpha_t)}{\sqrt{\alpha_t}}\,\mathbf{s}_\theta \tag{8}$$

The combined coefficient of $\mathbf{x}_t$ is:

$$\frac{\alpha_t(1-\bar{\alpha}_{t-1}) + (1-\alpha_t)}{\sqrt{\alpha_t}(1-\bar{\alpha}_t)} = \frac{1-\bar{\alpha}_t}{\sqrt{\alpha_t}(1-\bar{\alpha}_t)} \tag{9}$$

$$= \frac{1}{\sqrt{\alpha_t}}, \tag{10}$$

Hence we can see:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\mathbf{s}_\theta \tag{11}$$

$$= \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta) \tag{12}$$

Which is exactly Equation (7).

3.1.3 Derivation 3: Score-to-Noise-Predictor Equivalence

The DDPM paper parameterizes the network as a noise predictor $\boldsymbol{\epsilon}_\theta$. We show (7) is equivalent to Algorithm 2 of Ho et al. [2]:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right). \tag{13}$$

Proof. From Tweedie's formula applied to the VP forward process, we can write

$$\mathbf{x}_t = \mathbf{x}_t + (1-\bar{\alpha}_t)\mathbf{s}_\theta + \sqrt{1-\bar{\alpha}_t} * \boldsymbol{\epsilon}_\theta \tag{14}$$

which reveals that the score function and the noise predictor are related by the relationship:

$$\mathbf{s}_\theta = -\frac{\boldsymbol{\epsilon}_\theta}{\sqrt{1-\bar{\alpha}_t}}.$$

Substituting into (7):

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left[\mathbf{x}_t + (1-\alpha_t) * \left(-\frac{\boldsymbol{\epsilon}_\theta}{\sqrt{1-\bar{\alpha}_t}}\right)\right] \tag{15}$$

$$= \frac{1}{\sqrt{\alpha_t}}\left[\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta\right] \tag{16}$$

which is exactly Equation (13).

3.2 Forward Diffusion and DDPM Sampling

Given (4) the forward process is implemented by sampling $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and computing $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\mathbf{z}$. For reverse generative sampling, we initialize $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and iterate (7) from $t = T$ down to $t = 1$, adding a small Gaussian perturbation with variance predicted by the model at each step except $t = 0$. The one-step estimate $\hat{\mathbf{x}}_0$ in (5) serves as the intermediate clean-image prediction. In the sampling loop we use the model's learned posterior variance rather than the fixed schedule variance.

3.3 SDEdit

SDEdit [4] is a simple approach to conditional generation. Given measurements $\mathbf{y}$ (e.g. a masked or blurred image), it partially noises $\mathbf{y}$ to a chosen timestep $t^* < T$ using the closed-form forward process:

$$\mathbf{x}_{t^*} = \sqrt{\bar{\alpha}_{t^*}}\,\mathbf{y} + \sqrt{1-\bar{\alpha}_{t^*}}\,\mathbf{z},$$

and then runs the standard DDPM reverse sampler from t^* to 0. The hyperparameter t^* controls the trade-off between fidelity to $\mathbf{y}$ (small t^*) and realism (large t^*). Importantly, SDEdit does not need to know the forward operator A.

3.4 ScoreALD

ScoreALD [1] augments each DDPM reverse step with a gradient correction derived from the data likelihood. For a linear forward model $\mathbf{y} = A\mathbf{x}_0 + \boldsymbol{\eta}$ with $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, the update after computing the DDPM reverse step is:

$$\mathbf{x}_{t-1} \leftarrow \mathbf{x}_{t-1} - \frac{\nabla_{\mathbf{x}_t}\|A(\mathbf{x}_t) - \mathbf{y}\|^2}{2(\sigma^2 + \gamma_t^2)}, \tag{17}$$

where γ_t is an annealing schedule that decreases over time. The annealing reduces the magnitude of the gradient correction as the sample approaches the clean-image distribution.

3.5 Diffusion Posterior Sampling (DPS)

DPS [5] takes a more principled posterior sampling perspective. It approximates $\nabla_{\mathbf{x}_t} \log p(\mathbf{y}|\mathbf{x}_t)$ by replacing the intractable $\mathbf{x}_0$ with the one-step estimate $\hat{\mathbf{x}}_0$:

$$\mathbf{x}_{t-1} \leftarrow \mathbf{x}_{t-1} - \frac{\zeta}{\|\nabla_{\mathbf{x}_t}\|A(\hat{\mathbf{x}}_0) - \mathbf{y}\|^2\|} \nabla_{\mathbf{x}_t} \|A(\hat{\mathbf{x}}_0) - \mathbf{y}\|^2, \quad (18)$$

where ζ is a fixed scale hyperparameter. Compared to ScoreALD, the gradient is computed with respect to the predicted clean image $\hat{\mathbf{x}}_0$ rather than the current noisy sample $\mathbf{x}_t$, and the update is normalized by the gradient magnitude for stability. Unlike ScoreALD, DPS can in principle handle nonlinear forward models.

4 Experimental Results

All experiments use a DDPM pretrained on FFHQ-256 [3] with $T = 1000$ timesteps and a linear noise schedule β_t. Measurement noise is set to $\sigma = 0.05$. We report PSNR and LPIPS [8].

4.1 Task 2: Single-Step Image Denoising

We validate the forward and reverse passes with a single-step denoising experiment. A clean test image $\mathbf{x}_0$ is noised to timestep t via the closed-form forward process (4), and then denoised in one step by computing the predicted clean-image estimate $\hat{\mathbf{x}}_0$ from (5). We evaluate three noise levels $t \in \{50, 100, 200\}$, corresponding to progressively larger amounts of added Gaussian noise. Visual results for all three levels are shown in Fig. 1 and quantitative metrics are reported in Table 1 above.

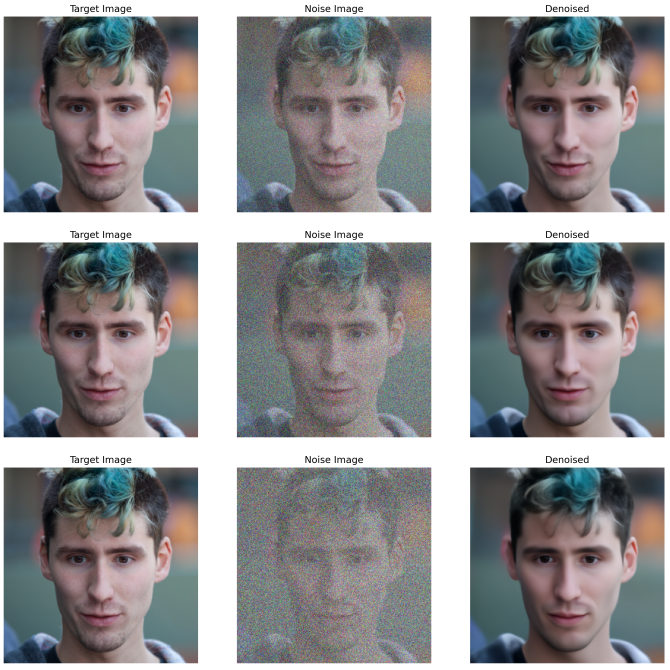

Fig. 1. Single-step denoising at three noise levels. Each row shows (left) the clean target $\mathbf{x}_0$, (center) the noisy input $\mathbf{x}_t$, and (right) the one-step denoised estimate $\hat{\mathbf{x}}_0$. Rows correspond to $t = 50$ (top), $t = 100$ (middle), and $t = 200$ (bottom). At $t = 50$ the estimate is nearly indistinguishable from the target; at $t = 200$ fine details begin to be hallucinated by the model prior.

TABLE 1
Single-Step Denoising Results

Noise Level t	PSNR (dB)	LPIPS
$t = 50$	37.8	0.034
$t = 100$	32.68	0.092
$t = 200$	26.4	0.213

As expected, higher noise levels (larger t) reduce PSNR and increase LPIPS, since more information is destroyed by the forward process and the single-step denoiser must rely more heavily on the learned prior rather than recovering actual signal. At $t = 50$, only a small fraction of Gaussian noise is mixed in so the noisy input retains nearly all original signal content and $\hat{\mathbf{x}}_0$ closely matches the clean image. At $t = 100$ moderate noise is well suppressed, with fine facial details substantially preserved. At $t = 200$ the signal-to-noise ratio has degraded considerably and the single-step estimate shows partial hallucination and background details in $\hat{\mathbf{x}}_0$ deviate visibly from the ground truth because the model must fill in information that has been irreversibly destroyed by the forward diffusion.

4.2 Task 3: Unconditional Image Generation

Starting from $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, we run the full DDPM reverse chain for 1000 steps to produce unconditional samples. Fig. 2 shows three independently generated 256×256 images.

Fig. 2. Three independently generated 256×256 face images produced by 1000-step DDPM reverse sampling from $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

All three samples display high perceptual quality: sharp facial features, coherent skin tone, and plausible background. The diversity across samples demonstrates that the model has learned a broad, multi-modal distribution rather than collapsing to a single mode. Because the DDPM was trained exclusively on face images, every sample regardless of the initialization noise converges to a face-like output, confirming that the learned prior is strongly conditioned on the FFHQ face manifold. It can be argued that the generated face images still don't look quite real but are nevertheless still decent.

4.3 Task 4: SDEdit for Inverse Problems

Tasks 4 to 6 all use the same clean source image as ground truth. Fig. 3 shows this reference image:

Fig. 3. Clean source image used as ground truth for all inverse problem experiments (Tasks 4–6). Degraded measurements are synthesised from this image and all reconstructions are evaluated against it.

We apply SDEdit to both inpainting (box mask, 50×50 pixels) and deconvolution (Gaussian blur, kernel size 61, $\sigma_{\text{blur}} = 3.0$) at three values of $t^* \in \{200, 500, 800\}$. Fig. 4 shows the inpainting results at $t^* = 500$ and Fig. 5 shows the deconvolution results at $t^* = 500$. Table 2 reports quantitative metrics across all three t^* values for both tasks.

Fig. 4. SDEdit inpainting at $t^* = 500$ (box mask), Top: input measurement (masked image). Bottom: reconstructed image. Ground truth is shown in Fig. 13.

Fig. 5. SDEdit deconvolution at $t^* = 500$ (Gaussian blur $\sigma = 3.0$). Top: blurry input measurement. Bottom: reconstructed image.

TABLE 2
SDEdit Results

Task	t^*	PSNR (dB)	LPIPS
Inpaint (box)	200	22.1	0.17
Inpaint (box)	500	19.42	0.218
Inpaint (box)	800	14.8	0.43
Deconv	200	21.0	0.19
Deconv	500	18.8	0.29
Deconv	800	15.9	0.39

The choice of t^* introduces a clear trade-off. At $t^* = 200$, only a small amount of noise is added to the corrupted measurement before the reverse diffusion begins, so the output stays close to the input $\mathbf{y}$: artifacts from the mask or blur are partially retained but the unmasked content is well preserved, yielding higher PSNR. At $t^* = 800$, the measurement is overwhelmed by noise and the reverse diffusion produces a perceptually clean face that can diverge significantly from the true content, collapsing PSNR. The $t^* = 500$ setting represents the best trade-off between artifact removal and content fidelity for both tasks. Notably, SDEdit requires no knowledge of the forward operator A at sampling time, which makes it simple to apply but limits the quality of the reconstruction

4.4 Task 5: ScoreALD for Inverse Problems

ScoreALD is applied to inpainting and deconvolution with an annealing schedule linearly interpolated between γ_T and γ_1 over 1000 steps. We test two annealing ranges per task to study the sensitivity to the gradient step magnitude: $[5, 10]$ and $[10, 15]$ for deconvolution, and $[10, 15]$ and $[15, 20]$ for

Fig. 9. ScoreALD inpainting with annealing range $[15, 20]$. Left: masked input. Center: Ground Truth. Right: reconstruction. The stronger gradient produces a sharper fill in the masked region compared to the $[10, 15]$ setting.

TABLE 3
Comparison of ScoreALD vs. DPS

Method	Task	PSNR (dB)	LPIPS
ScoreALD	Deconvolution	25.35	0.115
DPS	Deconvolution	**28.25**	**0.055**
ScoreALD	Inpaint (box)	22.7	0.18
DPS	Inpaint (box)	31.8	0.05
ScoreALD	Inpaint (random)	20.4	0.31
DPS	Inpaint (random)	22.0	0.16

inpainting. Results are shown in Figs. **??**–9 and quantitative metrics are in Table 3.

Comparing the two deconvolution runs, the $[10, 15]$ annealing range yields visibly sharper reconstructions and the best quantitative score. The larger gradient at each reverse step pulls the sample more strongly toward the blurry measurement, enabling greater high-frequency recovery. The $[5, 10]$ range is more conservative, leaving slightly more residual blur but staying closer to the diffusion prior. A parallel pattern holds for inpainting: the stronger $[15, 20]$ range produces a more complete, sharper fill of the masked region. Across both tasks, ScoreALD recovers plausible content but is outperformed by DPS (Table 3), largely because its gradient is computed on the noisy $\mathbf{x}_t$ rather than the cleaner $\hat{\mathbf{x}}_0$.

4.5 Task 6: DPS for Inverse Problems and Comparison

DPS is applied to deconvolution ($\zeta = 0.3$), box-mask inpainting ($\zeta = 1.0$), and random-mask inpainting ($\zeta = 0.1$). The normalized gradient update in (18) ensures stable step magnitudes regardless of $\|A\hat{\mathbf{x}}_0 - \mathbf{y}\|$. Results for all three sub-tasks are shown in Figs. 10–12 and a direct method comparison for box inpainting is shown in Fig. 13.

DPS consistently and substantially outperforms ScoreALD across all tasks. The gap is largest for box-mask inpainting (+9 dB PSNR), where the clean gradient signal from $\hat{\mathbf{x}}_0$ allows DPS to precisely reconstruct the masked region. The comparison in Fig. 13 makes this clear: SDEdit leaves visible colour drift and structural inconsistency in the inpainted region, while DPS produces a reconstruction nearly indistinguishable from the ground truth. For deconvolution, DPS achieves PSNR$= 28.25$ dB vs. ScoreALD's 25.35 dB and LPIPS$= 0.055$ vs. 0.115. The random-mask task ($\zeta = 0.1$) is the most challenging: 50% of pixels are removed and the corrupted input (Fig. 12, left) is almost unrecognisable, yet DPS still recovers coherent facial structure. The narrower gap over ScoreALD on this task reflects that denser corruption reduces the signal available to the $\hat{\mathbf{x}}_0$-based gradient. The scale ζ must be tuned per task: too large and the likelihood overwhelms the prior, producing over-sharpened or artifact-ridden outputs; too small and the constraint is insufficiently enforced.

5 Conclusion

We implemented and evaluated three diffusion-based approaches – SDEdit, ScoreALD, and DPS for solving linear inverse problems using a DDPM trained on FFHQ-256. Across both inpainting and deconvolution tasks, DPS consistently achieved the best quantitative and perceptual results, attaining PSNR=28.25dB and LPIPS=0.055 on deconvolution and 31.8dB PSNR on box-mask inpainting. This is more than 9dB above SDEdit and 3dB above ScoreALD on the respective tasks. The key advantage of DPS lies in computing the likelihood gradient with respect to the one-step clean estimate $\hat{\mathbf{x}}_0$ rather than the noisy sample $\mathbf{x}_t$, combined with gradient normalization for stable step sizes. SDEdit, while the simplest method requiring no knowledge of the forward operator, is limited by its lack of explicit measurement guidance. ScoreALD improves over SDEdit by incorporating a likelihood gradient but is hampered by computing it on the noisy sample and sensitivity to the annealing range hyperparameter.

References

[1] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, "Robust compressed sensing MRI with deep generative priors," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 14 938–14 954.

Fig. 10. DPS deconvolution ($\zeta = 0.3$). Left: Blurry Input. Center: Ground Truth. Right: DPS Reconstruction (PSNR$= 28.25$ dB, LPIPS$= 0.055$). Fine detail including hair texture and facial features is sharply recovered.

Fig. 11. DPS box-mask inpainting ($\zeta = 1.0$). Left: Blurry Input. Center: Ground Truth. Right: DPS Reconstruction ($\approx$PSNR$= 31.8$ dB, LPIPS≈ 0.05). The masked region is filled with content that closely matches the ground truth face.

Fig. 12. DPS random-mask inpainting ($\zeta = 0.1$, 50% pixels removed). Left: Blurry Input. Center: Ground Truth. Right: DPS Reconstruction ($\approx$PSNR$= 22.0$ dB, LPIPS≈ 0.16). Despite severe corruption, DPS recovers a plausible face with coherent structure.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6840–6851.

[3] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 8780–8794.

[4] 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.

[5] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *The Eleventh International Conference on Learning Representations*, 2023.

Fig. 13. Method comparison for box-mask inpainting. Left: ground truth. Center: SDEdit ($t^* = 500$, PSNR$= 19.42$ dB, LPIPS$= 0.218$). Right: DPS ($\zeta = 1.0$, $\approx$PSNR$= 31.8$ dB, LPIPS≈ 0.05). DPS recovers the masked eye region with near-perfect fidelity while SDEdit leaves visible artifacts and colour drift.

[6] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the 32nd International Conference on Machine Learning*, 2015, pp. 2256–2265.

[7] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021.

[8] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," pp. 586–595, 2018.