Analysis of Diffusion Models in Inverse Problems

Munir Bshara

Abstract—Diffusion models are increasingly used as learned priors for inverse problems, but it is still unclear how different conditioning strategies shape reconstruction behavior. This poster studies SDEdit, ScoreALD, and diffusion posterior sampling (DPS) on image deconvolution and inpainting. In addition to PSNR and LPIPS comparisons, the internals of the diffusion model are analyzed via various techniques borrowed from standard ODE analysis. The experiments show that DPS gives the strongest reconstructions, while the proposed analyses provide intuition for where the methods could differ. Overall, the results support the view that diffusion models can serve as implicit regularizers and that their inverse-problem behavior can be interpreted through local dynamics of the reverse process.

Index Terms—Diffusion models, inverse problems, image restoration, DPS, SDEdit, ScoreALD

1 Introduction

The act of sensing the world is fundamental to nearly all living systems. Humans perceive their surroundings through vision, hearing, touch, taste, and smell; many animals possess even richer sensory capabilities; and even plants exhibit measurable responses to external stimuli. The desire to replicate and augment this ability has driven the development of increasingly sophisticated sensing technologies.

Modern engineering systems measure the physical world using sensors such as radar, LiDAR, and cameras. These devices provide powerful and scalable ways to acquire information about an environment, but they are inherently imperfect. Measurements are often corrupted by noise, blur, limited resolution, or incomplete sampling, all of which degrade the fidelity of the recorded data. As a result, many sensing systems give rise not to direct access of the underlying scene, but rather to indirect and imperfect observations.

Recovering the latent signal from such degraded measurements is the central goal of an inverse problem. In general, inverse problems seek to infer an unknown quantity of interest from observations that have been transformed by a forward measurement process and corrupted by imperfections. These problems are frequently ill-posed in the sense that a solution may fail to exist, may not be unique, or may not depend continuously on the observed data. Consequently, even small perturbations in the measurements can lead to large changes in the reconstructed solution.

In computational imaging, two canonical examples of inverse problems are deconvolution and denoising. Deconvolution arises naturally when an image is blurred by defocus, motion, or an optical point spread function, while denoising is motivated by the stochastic and electronic noise introduced by the image acquisition pipeline. In both cases, the objective is to recover an estimate of the original image from a corrupted measurement. Although this objective is intuitively simple, it is challenging in practice because the corruption process destroys information and because many candidate reconstructions can be consistent with the same observation.

For this reason, successful inverse methods must incorporate additional structure beyond the measurement itself. Classical approaches impose hand-crafted priors such as smoothness, sparsity, or total variation, while more recent learning-based methods rely on data-driven priors that capture richer image statistics. The key challenge is therefore not only to invert the forward model, but also to identify what assumptions or prior knowledge most effectively distinguish a plausible underlying image from an arbitrary one.

This work focuses on understanding inverse problems from that perspective. Rather than viewing reconstruction purely as a numerical procedure, we emphasize the role of priors and regularization in shaping what solutions are preferred. This viewpoint is especially relevant in modern computational imaging, where increasingly expressive learned models offer strong empirical performance but can be difficult to interpret. In the following sections, we review the relevant prior-based approaches and then introduce diffusion-model-based methods as a powerful modern framework for solving such inverse problems.

2 Related Work

Inverse problems in imaging have traditionally been approached through explicit forward models combined with carefully designed priors. Classical restoration methods for deconvolution and denoising include Wiener filtering, and variational formulations based on smoothness or sparsity assumptions. In particular, total variation regularization became a foundational tool because it suppresses noise while preserving edges, making it especially effective for piecewise-smooth images [1], [2]. More broadly, these methods cast reconstruction as an optimization problem in which a data-fidelity term enforces consistency with the measurement process and a regularizer encodes prior assumptions on the latent image.

As inverse problems grew more challenging, research increasingly shifted toward richer and more adaptive priors. Sparse coding and learned denoisers provided more

• *M. Bshara is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305.*
E-mail: bshara[at]stanford.edu

expressive alternatives to purely hand-crafted regularization. This line of work led to plug-and-play methods, in which a powerful denoiser is inserted into an iterative optimization procedure and effectively acts as an implicit prior without requiring an explicit regularization functional [3]. These approaches helped bridge classical model-based reconstruction and modern learning-based restoration, and they demonstrated that high-quality image priors could be incorporated into inverse solvers in a modular way.

More recently, score-based and diffusion-based generative models have emerged as powerful learned priors for imaging. Denoising diffusion probabilistic models (DDPMs) and score-based generative models define a gradual noising process together with a learned reverse process that maps noise back to the data distribution [4], [5]. A key advantage of these models is that the learned score function provides local information about the data distribution, which makes diffusion models particularly attractive for inverse problems where one seeks to combine measurement consistency with a strong generative prior. The stochastic differential equation formulation of score-based modeling further provides a unified view of sampling, denoising, and conditional generation.

Several recent works have adapted diffusion models to inverse problems directly. Early formulations showed that pretrained score models could be combined with measurement operators to solve tasks such as inpainting, colorization, and other linear restoration problems [5]. SDEdit demonstrated that partial corruption followed by reverse-time denoising can effectively trade off realism and fidelity, making diffusion priors useful even when the observation already contains substantial structure [6]. More general posterior-oriented approaches, such as diffusion posterior sampling (DPS), incorporate the likelihood into the reverse trajectory and extend diffusion solvers to noisy and nonlinear inverse problems [7]. These methods suggest that diffusion models can function not merely as generators, but as flexible priors for posterior inference under degraded measurements.

Despite their strong empirical performance, the behavior of diffusion-based inverse solvers remains less interpretable than that of many classical reconstruction methods. Much of the literature evaluates these methods primarily through perceptual or distortion-based metrics, while comparatively less attention has been given to understanding how different conditioning strategies alter the reverse dynamics, stability, and sensitivity of the learned prior. This motivates the present work: rather than studying diffusion methods only as black-box reconstruction tools, we analyze how approaches how to understand the model's local behavior.

3 METHOD

3.1 Diffusion Model Identities

In this subsection, we collect three equivalent identities used throughout the remainder of the paper.

3.1.1 Closed-Form Forward Marginal

We begin with the variance-preserving forward process

$$\mathbf{x}_t = \sqrt{1-\beta_t}\,\mathbf{x}_{t-1} + \sqrt{\beta_t}\,\mathbf{z}_{t-1}, \qquad \mathbf{z}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (1)$$

and define

$$\alpha_t = 1 - \beta_t, \qquad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i. \quad (2)$$

We claim that this implies the closed-form marginal

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\mathbf{z}, \qquad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (3)$$

We prove this by induction. For $t = 1$,

$$\mathbf{x}_1 = \sqrt{1-\beta_1}\,\mathbf{x}_0 + \sqrt{\beta_1}\,\mathbf{z}_0 = \sqrt{\alpha_1}\,\mathbf{x}_0 + \sqrt{1-\alpha_1}\,\mathbf{z}_0, \quad (4)$$

which agrees with (3) since $\bar{\alpha}_1 = \alpha_1$.

Now assume that

$$\mathbf{x}_n = \sqrt{\bar{\alpha}_n}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_n}\,\mathbf{z}, \quad (5)$$

for some Gaussian $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then

$$\mathbf{x}_{n+1} = \sqrt{1-\beta_{n+1}}\,\mathbf{x}_n + \sqrt{\beta_{n+1}}\,\mathbf{z}_n \quad (6)$$

$$= \sqrt{\alpha_{n+1}}\left(\sqrt{\bar{\alpha}_n}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_n}\,\mathbf{z}\right) + \sqrt{1-\alpha_{n+1}}\,\mathbf{z}_n \quad (7)$$

$$= \sqrt{\bar{\alpha}_{n+1}}\,\mathbf{x}_0 + \sqrt{\alpha_{n+1}(1-\bar{\alpha}_n)}\,\mathbf{z} + \sqrt{1-\alpha_{n+1}}\,\mathbf{z}_n. \quad (8)$$

Since a linear combination of independent Gaussian variables is again Gaussian, the last two terms may be written as a single Gaussian with variance

$$\alpha_{n+1}(1-\bar{\alpha}_n) + (1-\alpha_{n+1}) = 1 - \alpha_{n+1}\bar{\alpha}_n = 1 - \bar{\alpha}_{n+1}. \quad (9)$$

Therefore,

$$\mathbf{x}_{n+1} = \sqrt{\bar{\alpha}_{n+1}}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_{n+1}}\,\mathbf{z}, \quad (10)$$

which completes the induction.

3.1.2 Equivalence of Two Reverse-Step Forms

We next show that the two common expressions for the reverse diffusion update are equivalent. Suppose

$$\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t + (1-\bar{\alpha}_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right). \quad (11)$$

Substituting (11) into the standard DDPM reverse mean

$$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t, \quad (12)$$

gives

$$\mathbf{x}_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \cdot \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t + (1-\bar{\alpha}_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right) + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t \quad (13)$$

$$= \frac{1-\alpha_t}{\sqrt{\alpha_t}(1-\bar{\alpha}_t)}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\mathbf{s}_\theta(\mathbf{x}_t, t) + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t. \quad (14)$$

Collecting the $\mathbf{x}_t$ terms yields

$$\mathbf{x}_{t-1} = \left[\frac{1-\alpha_t}{\sqrt{\alpha_t}(1-\bar{\alpha}_t)} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\right]\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\mathbf{s}_\theta(\mathbf{x}_t, t) \quad (15)$$

$$= \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t + \frac{1-\alpha_t}{\sqrt{\alpha_t}}\mathbf{s}_\theta(\mathbf{x}_t, t), \quad (16)$$

where we used $\bar{\alpha}_t = \alpha_t\bar{\alpha}_{t-1}$. Hence,

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right). \quad (17)$$

Thus, (12) and (17) are equivalent.

3.1.3 Equivalence of Score Prediction and Noise Prediction

Finally, we show that a score-predicting model may be written equivalently as a noise-predicting model. Using Tweedie's formula,

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) = \frac{\sqrt{\bar{\alpha}_t}\,\mathbb{E}[\mathbf{x}_0 \mid \mathbf{x}_t] - \mathbf{x}_t}{1-\bar{\alpha}_t}, \tag{18}$$

and approximating the posterior mean with a denoiser $D_\theta(\mathbf{x}_t, t) \approx \hat{\mathbf{x}}_0$, we write

$$\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)}{\sqrt{\bar{\alpha}_t}}. \tag{19}$$

Substituting (19) into Tweedie's formula yields

$$\mathbf{s}_\theta(\mathbf{x}_t, t) = \frac{\sqrt{\bar{\alpha}_t}\hat{\mathbf{x}}_0 - \mathbf{x}_t}{1-\bar{\alpha}_t} \tag{20}$$

$$= \frac{\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) - \mathbf{x}_t}{1-\bar{\alpha}_t} \tag{21}$$

$$= -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t). \tag{22}$$

Plugging (22) into (17) gives

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right), \tag{23}$$

which is exactly the noise-prediction form of the reverse diffusion process. Therefore, score prediction and noise prediction are equivalent parameterizations of the same reverse dynamics.

3.2 Image Denoising with a Pretrained Diffusion Model

Given a noisy image $\mathbf{x}_t$, Task 2 uses the pretrained score model $\mathbf{s}_\theta(\mathbf{x}_t, t) \approx \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ to denoise a single corrupted sample. Mathematically, this is just one application of the reverse diffusion prior without any measurement operator:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right). \tag{24}$$

Thus, the denoising task differs from the later inverse problems in that it uses only the learned prior and does not impose any explicit data-consistency constraint.

3.3 Unconditional Sampling with DDPM

Task 3 extends denoising to full image generation by starting from pure Gaussian noise, $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and iteratively applying the reverse diffusion update for all timesteps:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}_t. \tag{25}$$

Compared with Task 2, the mathematical difference is that denoising starts from a given noisy observation, whereas unconditional sampling starts from random noise and generates an image entirely from the learned prior.

3.4 Inverse Problems with SDEdit

Task 4 introduces conditional generation for inverse problems such as inpainting and deconvolution. Let $\mathbf{y} = A(\mathbf{x})$ denote the measurement. SDEdit first constructs a partially noised version of the measurement-consistent initialization, say $\tilde{\mathbf{x}}_{t_0}$, and then runs reverse diffusion from time t_0 back to 0:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right), \quad t = t_0, \ldots, 1. \tag{26}$$

The key mathematical point is that SDEdit does *not* add an explicit likelihood gradient during sampling; conditioning enters only through the choice of the starting state $\tilde{\mathbf{x}}_{t_0}$. The assignment describes SDEdit precisely in this way: start from a partially noised observation and apply iterative DDPM denoising.

3.5 Inverse Problems with ScoreALD

Task 5 turns the inverse problem into posterior sampling by combining the diffusion score with a measurement likelihood term. If $\ell(\mathbf{x}; \mathbf{y})$ denotes the data-fidelity objective, then ScoreALD modifies the reverse step by adding a likelihood correction:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right) + \zeta_t \mathbf{g}(\mathbf{x}_t, \mathbf{y}), \tag{27}$$

where

$$\mathbf{g}(\mathbf{x}_t, \mathbf{y}) = -\nabla_{\mathbf{x}_t} \ell(\mathbf{x}_t; \mathbf{y}), \quad \zeta_t \text{ is an annealing factor.} \tag{28}$$

Thus, unlike SDEdit, ScoreALD enforces measurement consistency at every reverse step through likelihood gradients.

3.6 Inverse Problems with DPS

Task 6 uses the same posterior-sampling viewpoint as ScoreALD, but normalizes the likelihood gradient to improve stability. A generic DPS-style update can be written as

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t + (1-\alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right) + \lambda_t \frac{\mathbf{g}(\mathbf{x}_t, \mathbf{y})}{\|\mathbf{g}(\mathbf{x}_t, \mathbf{y})\|_2}, \tag{29}$$

where λ_t is a scale parameter. Therefore, the mathematical distinction between ScoreALD and DPS is not the presence of a likelihood correction itself, but the fact that DPS uses a normalized likelihood-gradient update. This is essentially an adaptive reweighting of how much one trusts the model output.

3.7 Methods of Analysis

To better understand the local behavior of the diffusion prior, we studied how small perturbations in the input image propagate through the score network and through one reverse-diffusion step. This perspective was motivated by the role of Green's functions in linear PDEs, where one probes a system by applying localized perturbations and observing the resulting spatial response. In our setting, the score network approximates the time-dependent score

$$\mathbf{s}_\theta(\mathbf{x}_t, t) \approx \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t), \tag{30}$$

which is the central object governing reverse-time generation in score-based diffusion models [8].

We therefore define the score Jacobian

$$J_s(\mathbf{x}_t, t) = \frac{\partial \mathbf{s}_\theta(\mathbf{x}_t, t)}{\partial \mathbf{x}_t}, \tag{31}$$

and estimate selected columns of this operator with finite differences by perturbing individual pixels and measuring the resulting change in the score output. Concretely, for the ith pixel basis vector $\mathbf{e}_i$ and a small perturbation size ε, we approximate

$$J_s(\mathbf{x}_t, t)\mathbf{e}_i \approx \frac{\mathbf{s}_\theta(\mathbf{x}_t + \varepsilon \mathbf{e}_i, t) - \mathbf{s}_\theta(\mathbf{x}_t, t)}{\varepsilon}. \tag{32}$$

This provides a local sensitivity map that quantifies how a perturbation at one pixel influences the score field across the image.

To understand how this local sensitivity affects generation, we propagate the same perturbation through one deterministic reverse step. Using the score-form reverse update,

$$\mathbf{x}_{t-1} = f_t(\mathbf{x}_t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t + (1 - \alpha_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\right), \tag{33}$$

the Jacobian of the reverse map is

$$J_f(\mathbf{x}_t, t) = \frac{\partial f_t(\mathbf{x}_t)}{\partial \mathbf{x}_t} = \frac{1}{\sqrt{\alpha_t}} \left(I + (1 - \alpha_t) J_s(\mathbf{x}_t, t)\right). \tag{34}$$

Thus, the reverse-step response can also be estimated by finite differences and interpreted as the spatial footprint of a pixel perturbation after one inner iteration of deterministic generation.

3.8 Compression Investigation

Because diffusion sampling requires repeated evaluations of a large neural network, we also investigated whether the reverse dynamics exhibit low-dimensional or slowly varying structure that could be exploited computationally. Neural networks are often substantially compressible in practice, motivating the question of whether all score evaluations are equally necessary at every timestep [9].

Our first diagnostic was spectral. Inspired by the eigenfaces viewpoint, in which principal components provide a compact representation of high-dimensional image variation, we computed singular vectors of the linearized operators above to identify dominant response directions [10]. In particular, the top singular vectors of J_s and J_f reveal whether the score network responds primarily along a small number of structured modes, while the associated singular values quantify the strength of those modes. A rapid decay in the singular-value spectrum would suggest that the local action of the model is effectively low rank, which in turn supports the possibility of compressed or intermittent evaluation strategies.

Our second diagnostic was motivated by numerical ODE stability. Explicit time-stepping methods are conditionally stable, with admissible step sizes governed by the spectrum of the linearized update operator; for explicit Euler, stability depends on the location of $h\lambda$ relative to the method's stability region [11]. Although reverse diffusion is not literally an explicit Euler discretization of a fixed linear system, this analogy suggests that the spectral radius

$$\rho(J_f) \tag{35}$$

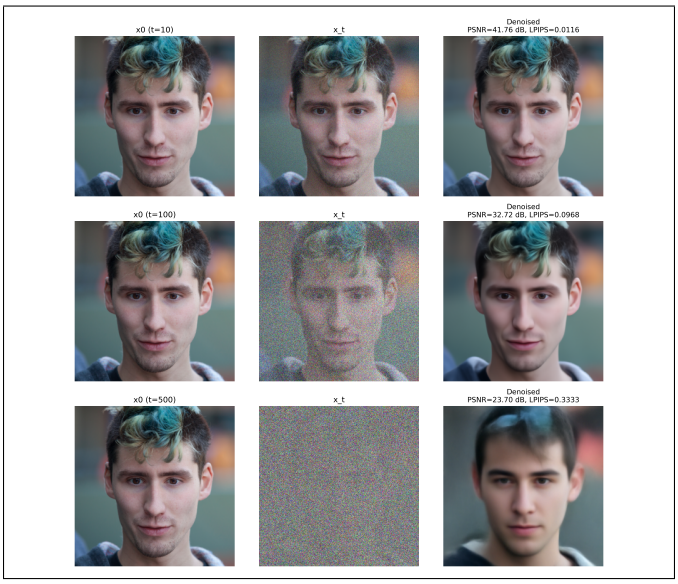

Fig. 1. Task 2: Evaluating Diffusion Denoising

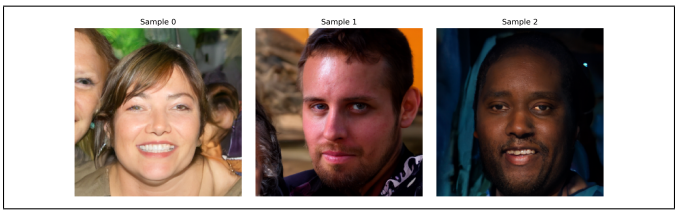

Fig. 2. Task 3: Unconditional Generation

can serve as a useful local proxy for how aggressively one may subsample or skip score evaluations in time before the dynamics become inaccurate or unstable. We therefore use the maximum singular value and spectral radius of the local reverse map as exploratory indicators of timestep sensitivity.

4 Experimental Results

4.1 Task 2: Pure Diffusion Model Image Denoising

By iteratively applying the DDPM reverse updates to a noisy image, the pretrained diffusion model can remove moderate noise and recover meaningful visual structure. At small noise levels, the reconstruction remains faithful to the original image. At larger noise levels, however, too much information is lost, and the model increasingly relies on the learned prior rather than the observation. This is quantitatively backed up by the LPIPS metric which decreases when the starting image is less noisy.

4.2 Task 3: Unconditional Generation

Starting from pure Gaussian noise, the reverse diffusion process generates samples from the learned image distribution. The resulting images exhibit clear human-like structure and texture, showing that the model has learned a strong prior over faces, though some samples remain imperfect or ambiguous. Attached in Figure 2are three examples of these unconditionally generated images.

4.3 Task 4: SDEdit

SDEdit initializes the reverse process from a corrupted observation rather than pure noise. This allows it to refine partially degraded inputs, but because it does not explicitly enforce measurement consistency at every step, it may drift toward visually plausible but less faithful reconstructions when the corruption is severe. Visually this means that the less noise added initially, the closer you are staying to the measurement, and the more noise one adds the worse the data matches the true measurement. This is shown in Figure 3 by the decreasing PSNR values as initial noise added increases in both the deconvolution and inpainting case.

4.4 Task 5: ScoreALD

ScoreALD improves on SDEdit by incorporating a likelihood-based correction during reverse diffusion. This additional measurement guidance yields reconstructions that are more consistent with the observation and improves performance on inverse problems such as inpainting and deconvolution. These gains are visible qualitatively in Figure 4.4 and are also reflected quantitatively through higher PSNR values relative to the initial degraded measurement.

4.5 Task 6: DPS

DPS provides the strongest reconstructions by combining the diffusion prior with a normalized likelihood correction. This adaptive weighting stabilizes the reverse dynamics while preserving fidelity to the measurement, leading to the best overall tradeoff between realism and data consistency. For the examples considered here, the resulting reconstructions are often visually very close to the underlying ground truth, which is further supported by their consistently high PSNR values in Figure 5,

4.6 Sensitivity Analysis

To better understand these behaviors, we performed perturbation-based sensitivity analysis of the score function and deterministic reverse step. By perturbing individual pixels and measuring the resulting response, we visualize how local changes propagate through the learned dynamics. A sample has been attached in Figure 6 where the first column shows understanding of facial structure of the noise and cheek bones and the third shows understanding of lighting where the perturbation on the right eyelid corresponds with the right cheek bone

This model is not just a 1 time step it also has a 3rd dimension of depth so we decided to sensitivity probe at many different times and we get these nice curves of how much of the effect(I call this energy) is enclosed in a specific receptive field. As t increases it becomes more local as shown by the energy metric in Figure 7.

4.7 Compression Investigation

We also studied whether the reverse dynamics exhibit low-dimensional structure that could support compressed or less frequent score evaluations. Singular value and singular vector analysis of the local linearized operators provide an initial view into the dominant modes of the diffusion process and its possible computational simplifications. Figure 8 illustrates how dominant modes only start appearing when later on into the diffusion process.

Another analysis done was whether there exists any decrease in size of the neural network that can be done and as the spectral figure hints at, only low noise images in Figure 9 have understandable dominant singular vectors. These are singular vectors of the Jacobian of the reverse map and thus the computation was iterative via power iteration.

5 Conclusion

Overall, while the two methods share the same posterior-sampling philosophy, DPS achieves the stronger empirical performance for the same measurement data.

We also observe that relying too heavily on the learned diffusion prior alone is not always sufficient for accurate reconstruction. As seen in the denoising and SDEdit results, the diffusion model can produce images that are visually plausible and lie on the learned data manifold, yet still deviate from the true underlying image when measurement consistency is weakly enforced. This highlights an important point: realism under the prior does not necessarily imply fidelity to the observation. The strongest results therefore come from balancing the expressive prior learned by the diffusion model with explicit measurement-based corrections, which is precisely what allows DPS to outperform the other approaches considered here.

We also found that the reverse dynamics of DDPM sampling exhibit a clear time dependence in both spatial sensitivity and local stability. In particular, the diffusion model appears to have a time-dependent receptive field: perturbations at different timesteps propagate over different spatial extents, indicating that the score network responds more locally at some noise levels and more globally at others. At the same time, our spectral analysis suggests that as the sample approaches the clean-image regime, the spectral radius of the linearized reverse map tends to grow. This indicates that the late stages of sampling may be more sensitive to perturbations and potentially less stable, even though they are also the stages where fine image details are being reconstructed.

6 Future Work

Taking both results from the conclusion, these observations suggest that the reverse process is not uniformly difficult across timesteps. Early noisy stages may primarily require coarse, global structure formation, while later low-noise stages demand finer corrections and must contend with increasingly delicate dynamics. This opens the possibility of designing a stacked or hierarchical architecture in which different models are specialized for different noise regimes. For example, one model could operate over high-noise timesteps to capture large-scale structure efficiently, while another model, activated only at low noise levels, could take more timesteps or use a more robust update rule to better handle the larger spectral radius and the instability associated with explicit time-stepping. Such a design could improve both computational efficiency and reconstruction

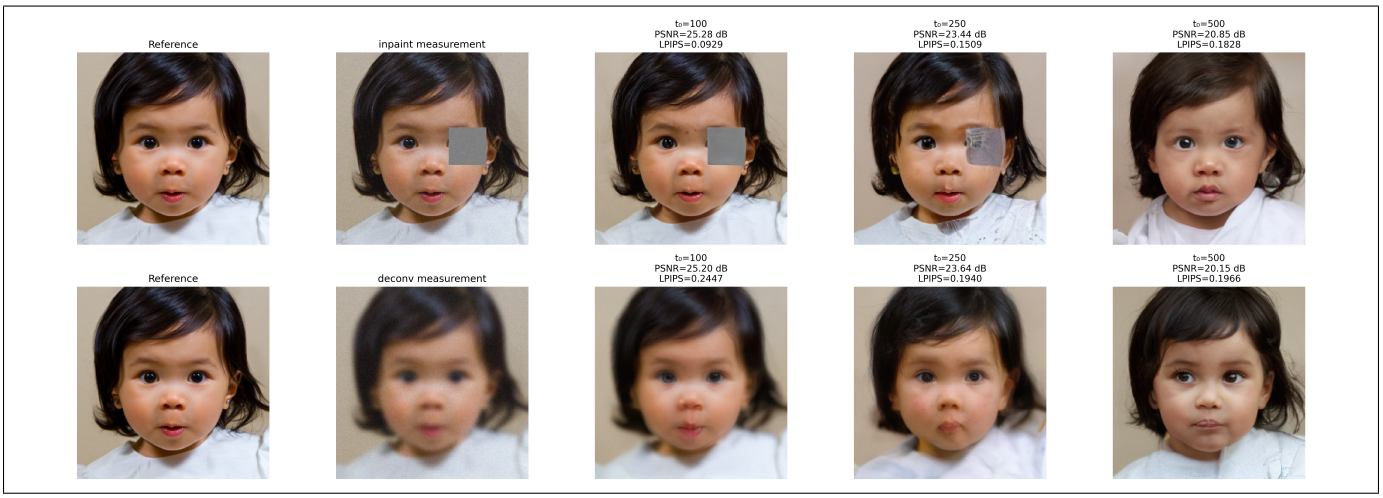

Fig. 3. Task 4: SDEdit Quantitative Assessment

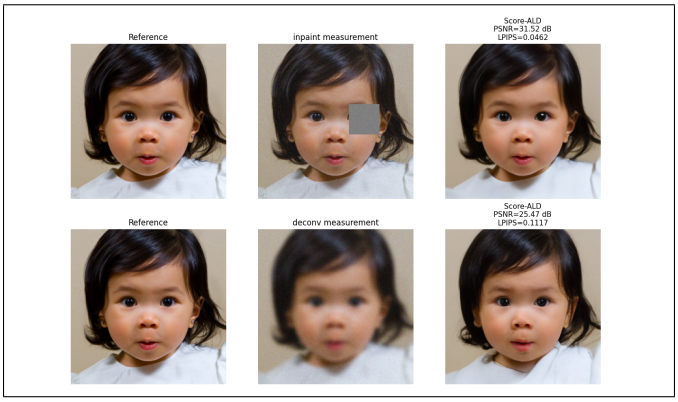

Fig. 4. Task 5: ScoreALD Quantitative Assessment

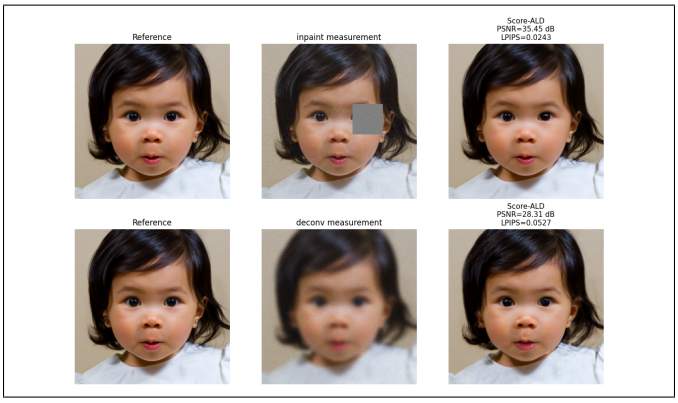

Fig. 5. Task 6: DPS Quantitative Assessment

quality by allocating model capacity where the reverse dynamics are most challenging.

More broadly, this perspective suggests that diffusion sampling should not necessarily be viewed as a homogeneous iterative process. Instead, the reverse trajectory may be better understood as a sequence of qualitatively different regimes, each with its own spatial scale, sensitivity, and numerical behavior. Exploiting this structure could lead to new architectures and samplers that are better matched to the underlying dynamics of diffusion generation.

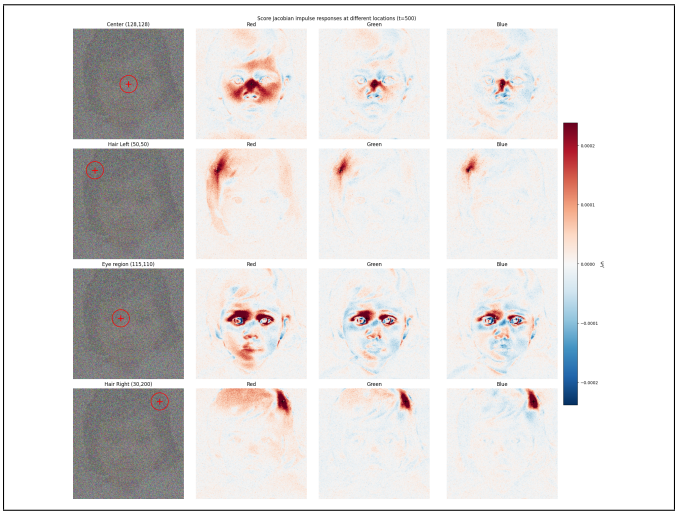

Fig. 6. Sensitivity Analysis

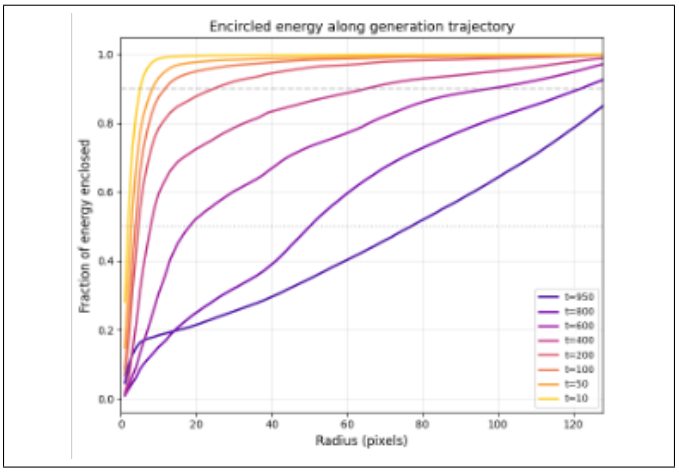

Fig. 7. Proxy for radius of effect of one pixel

References

[1] W. H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, 1972.

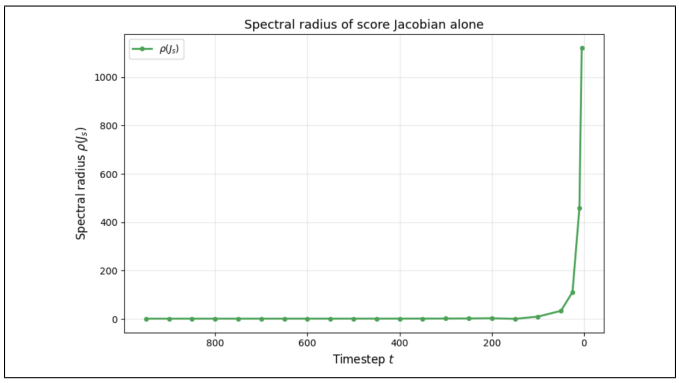

Fig. 8. Eigenvalue proxy showing less stability later on in the iterative process. Less stability via the metric defined by ODE analysis.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.

[3] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.

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

[5] 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.

[6] C. Meng, Y. He, J. Song, Y. 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.

[7] 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.

[8] 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.

[9] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding," in *International Conference on Learning Representations*, 2016.

[10] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[11] MIT Mathematics, "Introduction to numerical analysis: Stability of euler methods," 2012, course notes.

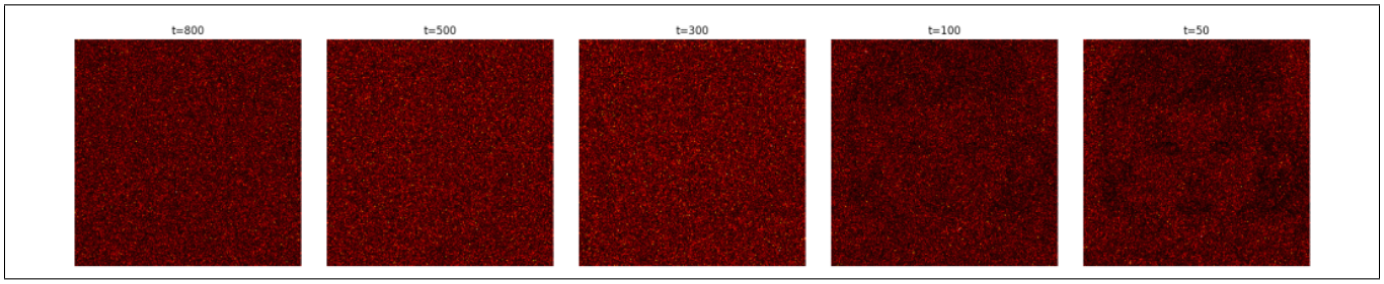

Fig. 9. Dominant singular vectors over time.