Diffusion Models for Inverse Problems

Nick Ping

Abstract—Diffusion models have emerged as powerful generative priors for image synthesis and restoration. In this work, we evaluate diffusion-based approaches for solving inverse problems in computational imaging. We compare three methods: SDEdit, Score Annealed Langevin Dynamics (Score ALD), and Diffusion Posterior Sampling (DPS). These three methods are employed on image denoising, inpainting, and deconvolution tasks using a pretrained model trained on the Flickr-Faces-HQ dataset. Performance is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS). Results show that diffusion models can effectively reconstruct corrupted images, though performance degrades with increasing noise and on out-of-distribution inputs. Among the evaluated methods, DPS achieves the best reconstruction quality across tasks, while SDEdit provides lower-quality results but with reduced computational cost. These results demonstrate the effectiveness of diffusion models as general-purpose priors for inverse imaging problems and highlight the benefits of incorporating measurement-guided sampling.

Index Terms—Computational Photography, Diffusion Model

1 Introduction

Diffusion models have emerged as a powerful tool for image synthesis and editing with widespread usage in academia and industry. Image restoration tasks such as inpainting and deconvolution, recovering an image from a measured image with missing patches or global blur, are fundamental inverse problems in computational imaging. Diffusion models prove to be effective alternatives to traditional approaches that rely on deep neural nets and handcrafted priors since diffusion models can generalize well across various reconstruction tasks and can be used as general-purpose priors for a variety of inverse problems.

Diffusion models learn to generate images by iterative denoising according to stochastic differential equations. The forward diffusion process gradually adds noise while the reverse process removes this noise. For image restoration tasks, the corrupted measurement can be incorporated into this iterative denoising procedure, enabling reconstruction methods that combine measurement consistency with learned image priors. SDEdit [1] is one proposed strategy that injects noise according to the forward diffusion process and then applies the reverse diffusion process to remove the noise and, in the process, restores the image. While this method is simple, guidance strategies help direct the trajectory towards a desired image prior by using the gradient of the score function. Methods such as Score Annealed Langevin Dynamics (Score ALD [2]) and Diffusion Posterior Sampling (DPS [3]) use this principle for guiding the reverse diffusion process for a better restoration.

In this paper, we investigate the effectiveness of diffusion-based approaches for fundamental image restoration tasks and conduct a comparative study of SDEdit, Score ALD, and DPS. We evaluate their performance on two representative inverse problems: image inpainting and image deconvolution. Furthermore, to assess the robustness and generalization capabilities of these methods, we include both in-distribution and out-of-distribution samples in the evaluation set.

- *N. Ping is with the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305.*
  E-mail: nping2@stanford.edu

2 Related Work

Diffusion models have recently achieved strong performance in conditional image generation by guiding the reverse diffusion process with auxiliary information such as class labels or natural language descriptions. Large-scale systems such as Imagen [4] demonstrate that diffusion-based models can produce high-quality images conditioned on text prompts, enabling a wide range of generative and editing applications.

2.1 Related Image Restoration Tasks

Beyond image inpainting and deconvolution tasks, diffusion models have proven to be effective for a variety of image editing and restoration tasks. Tasks such as super-resolution, which is image upsampling, and image customization, which augments images with custom features.

2.2 Video and Cross-Modal Generation

The diffusion framework can generalize to video synthesis by employing temporal consistency. Recent models such as Imagen Video and Sora [5] demonstrate consistent video sequences based on natural language or image prompts. Cross-modal generation also extends the diffusion framework by utilizing multiple data modalities such as text, image, and audio representations as both source data or generated data. Such applications include multimodal translation, audio generation, and image captioning tasks.

2.3 Imitation Learning and Robotics

More recently, diffusion models have been explored in robotics and control, particularly for imitation learning and policy generation. Approaches such as Diffusion Policy [6] use diffusion processes to learn complex manipulation behaviors from demonstration data. These methods illustrate the broader applicability of diffusion-based modeling beyond visual generation tasks.

3 METHODOLOGY

We will utilize the variance preserving (VP) formulation of the diffusion models introduced by the Denoising Diffusion Probabilistic Models (DDPM [7]) paper.

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}z_{t-1}, \qquad t = 1, 2, \ldots, T,$$

where $z_{t-1} \sim \mathcal{N}(0, I)$ are i.i.d Gaussian random variables for each step and β_t is the noise schedule parameter for each step. This formulation is equivalent to:

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

, where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, and $z \sim \mathcal{N}(0, I)$. A derivation of this equivalence is given in the Appendix.

With this formulation together with Tweedie's formula, the reverse diffusion process is then given as

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

$$x_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t}\hat{x}_0 \tag{2}$$

, where $s_\theta(x_t, t)$ is a score-predicting model. By iteratively applying this reverse process starting from random noise, images can be generated from the learned distribution.

In this work, we will utilize a pretrained diffusion model that is trained on Flickr-Faces-HQ Dataset which contains images of human faces.

3.1 Image denoising and generalizability

We evaluate model generalizability through image denoising on in distribution images compared to out of distribution images. To achieve this, we add noise to both images using the forward diffusion process with different number of forward steps, and then perform image denoising with an appropriate number of reverse diffusion steps. We will then consider the Peak Signal-to-Noise Ratio (PSNR) and the Learned Perceptual Image Patch Similarity (LPIPS) as the comparison metrics.

3.2 Inverse Problems

We further evaluate the capability of diffusion-based methods to solve two classes of inverse problems: image inpainting and image deconvolution. In the image inpainting task, a region of the image is masked and must be reconstructed. In the image deconvolution task, the observed image is blurred using a known convolution kernel, and the objective is to recover the underlying sharp image.

For both tasks, reconstruction performance is evaluated using PSNR and LPIPS. We compare three diffusion-based approaches for solving these inverse problems: SDEdit, Score ALD, and DPS.

3.2.1 SDEdit

SDEdit reconstructs a corrupted image by applying the forward diffusion process for a limited number of steps to add noise, then removes the noise by applying the reverse diffusion process.

TABLE 1
PSNR and LPIPS of the recovered images shown in Figure (2).

	In Distribution		Out of Distribution	
Forward Steps	PSNR	LPIPS	PSNR	LPIPS
10	41.6	0.012	38.8	0.008
100	32.5	0.101	27.6	0.167
500	23.4	0.323	20.3	0.701

3.2.2 Score ALD

Score ALD reconstructs a corrupted image by augmenting the reverse diffusion process with an additional step. After taking a single reverse step given in Equation (1), we take a conditioning step:

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

, where γ is an annealing factor which controls the strength of conditioning, and $\mathcal{A}(x_t)$ is the forward image formation process on the current noisy image.

We should expect Score ALD to perform better than SDEdit because we can guide the diffusion trajectory with conditioning.

3.2.3 DPS

DPS is similar to Score ALD by augmenting the reverse diffusion step given in Equation (1) but with another conditioning function:

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

, where ζ is a scaling factor, and $\mathcal{A}(\hat{x}_0)$ is the forward image formation process on the estimated starting image.

We should expect DPS to perform better than Score ALD because the conditioning function should better reflect the forward image formation model by applying it to the estimated starting image rather than a noisy current step.

4 EXPERIMENTAL RESULTS

By iteratively applying 1000 reverse diffusion steps on random noise, we obtain three faces that are unconditionally generated shown in Figure (1). These faces are reflective of the training distribution of human facial images. We observe quality facial features, but backgrounds or shoulders are warped and distorted.

4.1 Image denoising and generalizability

We evaluate denoising performance by applying $(10, 100, 500)$ forward diffusion steps to generate different noise levels for both in-distribution and out-of-distribution target images. The resulting reconstruction quality is measured using PSNR and LPIPS. Quantitative results for both in-distribution and out-of-distribution reconstructions are reported in Table (1), while representative target, noisy, and reconstructed image triplets are shown in Figure (2).

From these results, we observe that the model can successfully denoise images at moderate noise levels for both in-distribution and out-of-distribution examples. However, reconstruction quality is consistently higher for the in-distribution images. As expected, denoising performance

Fig. 1. Examples of unconditionally generated images using the diffusion model.

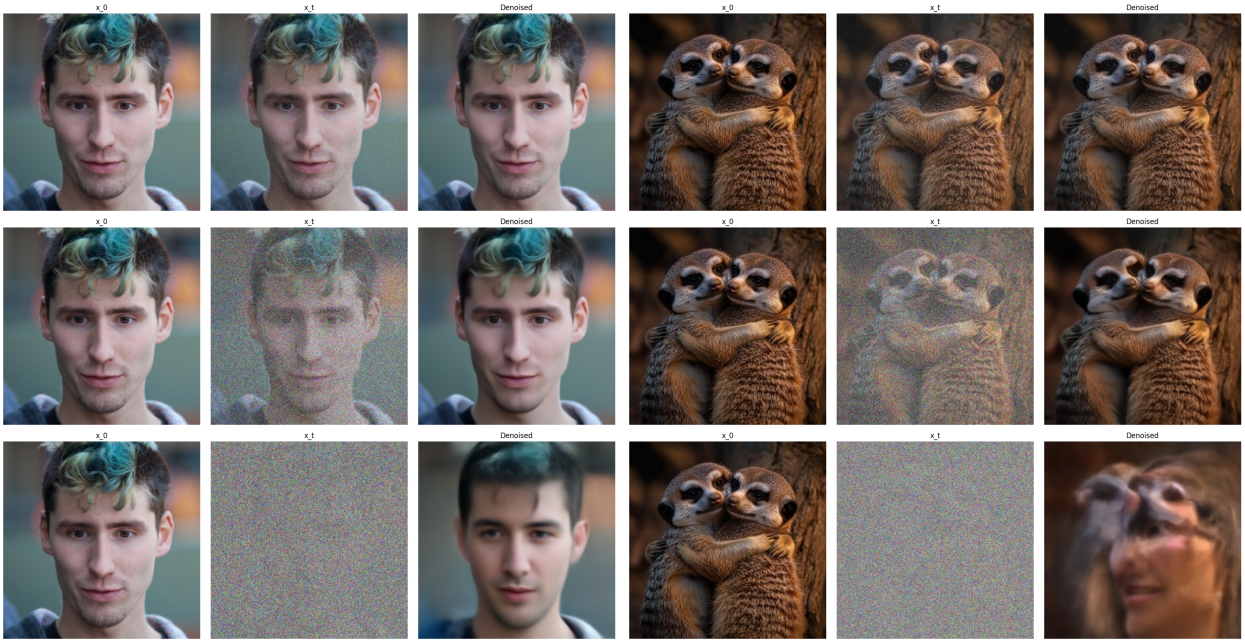

Fig. 2. Ground truth, noisy, and reconstructed image triplets for 3 noise levels for an in-distribution and out-of-distribution examples. From top to bottom, (10, 100, 500) noise steps are applied. The left triplets are of faces which are in the training distribution. The right triplets are of meerkats which are not in the training distribution. PSNR and LPIPS for the recovered images are given in Table (1).

improves when fewer forward diffusion steps are applied, corresponding to lower injected noise levels.

At higher noise levels, reconstruction quality degrades. For the in-distribution example, some fine image details are lost but the overall structure of the face is preserved. In contrast, the out-of-distribution example deteriorates more significantly, with the reconstruction drifting toward the model's learned face prior. In extreme cases, the reconstructed image resembles a face even when the ground-truth image does not contain facial structure.

4.2 Image Reconstruction Comparison

We evaluate image reconstruction performance on the inverse problems of image deconvolution and image inpainting using three diffusion-based reconstruction methods: SDEdit, Score ALD, and DPS. For each experiment, we vary the measurement noise level with $\sigma = (0.01, 0.05, 0.1)$.

TABLE 2
PSNR and LPIPS of the recovered images using the three different methods SDEdit, Score ALD, and DPS. The images for SDEdit are shown in Figure (3). The images for Score ALD are shown in Figure (4). The images for DPS are shown in Figure (5).

Method	Noise Level	Deconvolution		Inpainting	
		PSNR	LPIPS	PSNR	LPIPS
SDEdit	0.01	20.052	0.230	20.322	0.203
	0.05	20.542	0.187	20.186	0.204
	0.1	20.108	0.230	20.579	0.195
Score ALD	0.01	22.404	0.141	27.031	0.086
	0.05	22.100	0.162	28.148	0.061
	0.1	21.953	0.153	28.188	0.057
DPS	0.01	28.213	0.052	35.032	0.023
	0.05	28.828	0.058	34.730	0.024
	0.1	28.066	0.070	34.034	0.022

Reconstruction quality is evaluated using PSNR and LPIPS, with quantitative results reported in Table (2).

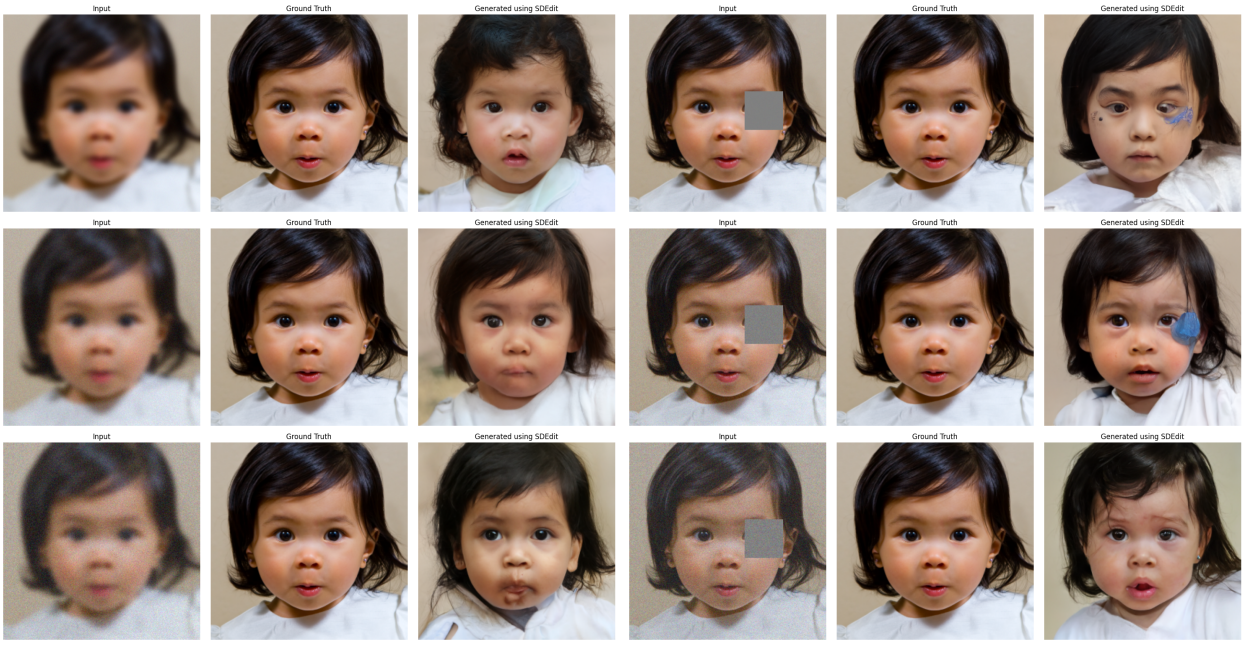

Fig. 3. Noisy, ground truth, and reconstructed image triplets for 3 noise levels for the deconvolution and inpainting tasks using SDEdit. From top to bottom, $\sigma = (0.01, 0.05, 0.1)$. The left triplets are of the deconvolution task. The right triplets are of the inpainting task. PSNR and LPIPS for the recovered images are given in Table (2).

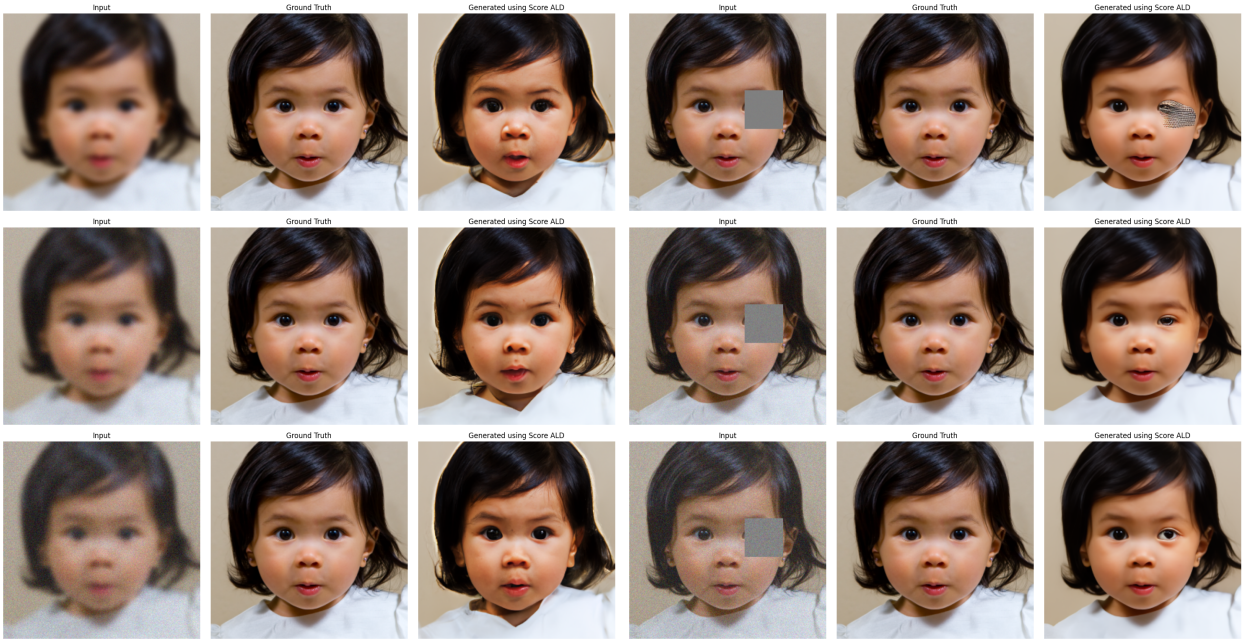

Fig. 4. Noisy, ground truth, and reconstructed image triplets for 3 noise levels for the deconvolution and inpainting tasks using Score ALD. From top to bottom, $\sigma = (0.01, 0.05, 0.1)$. The left triplets are of the deconvolution task. The right triplets are of the inpainting task. PSNR and LPIPS for the recovered images are given in Table (2).

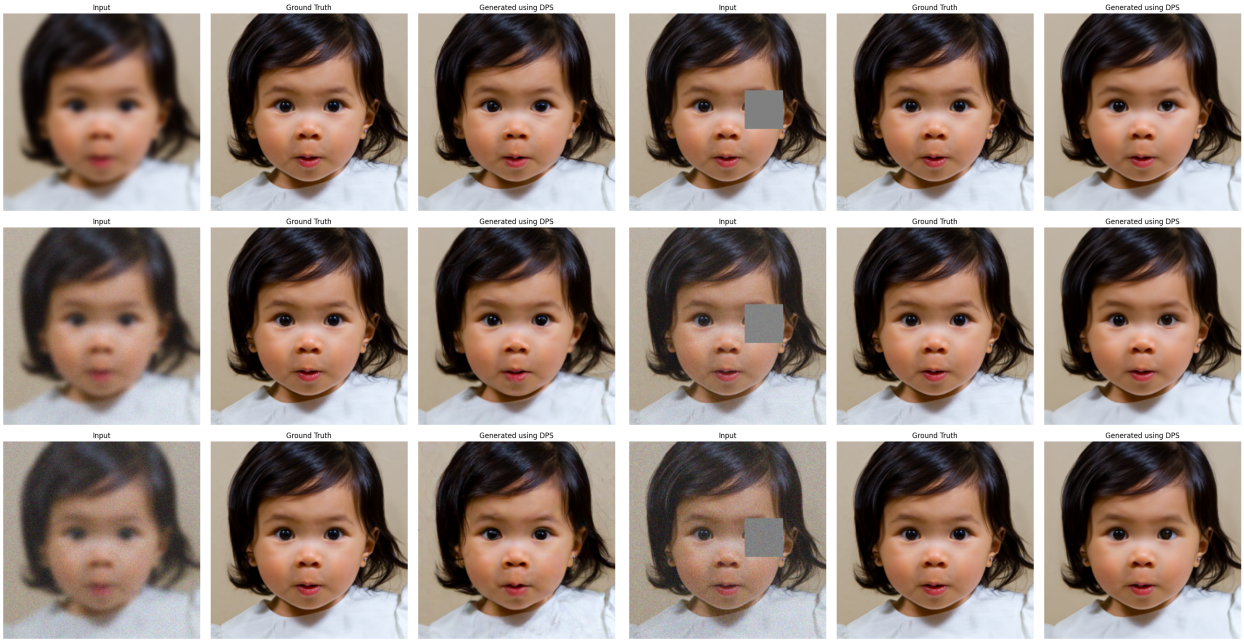

Fig. 5. Noisy, ground truth, and reconstructed image triplets for 3 noise levels for the deconvolution and inpainting tasks using DPS. From top to bottom, $\sigma = (0.01, 0.05, 0.1)$. The left triplets are of the deconvolution task. The right triplets are of the inpainting task. PSNR and LPIPS for the recovered images are given in Table (2).

Representative measured, ground-truth, and reconstructed image triplets are shown in Figures (3)-(5). Figure (3) presents results obtained using SDEdit, Figure (4) shows reconstructions from Score ALD, and Figure (5) illustrates results produced by DPS.

From the results, we observe that DPS achieves the best performance for both image deconvolution and image inpainting across all tested noise levels. In contrast, SDEdit consistently yields the lowest reconstruction quality among the three methods. We also observe that image inpainting is an easier task than image deconvolution.

These results align with our hypothesis that methods incorporating a more faithful image formation model can achieve improved reconstruction performance. In particular, DPS includes the forward measurement model on a predicted starting image into the diffusion sampling process, which helps enforce measurement consistency and leads to higher-quality reconstructions.

However, the improvements in reconstruction quality come at the cost of increased runtime. DPS consistently requires approximately 2-3 times the runtime of Score ALD, while Score ALD requires roughly twice the runtime of SDEdit.

4.2.1 *Method and Noise Level*

We observe that the measurement noise level has no clear effect on reconstruction quality. The variations in PSNR and LPIPS across different noise levels show no consistent trend and appear to fall within the range of normal algorithmic variability.

For all three reconstruction methods, the measurement noise level has a weak or inconsistent effect on the reconstruction PSNR and LPIPS. This suggests that the reconstruction performance is primarily determined by the diffusion prior and guidance strategy rather than the measurement noise level.

5 Conclusion

We investigated diffusion-based methods for solving inverse problems in image restoration, focusing on denoising, inpainting, and deconvolution. Using a pretrained diffusion model, we compared SDEdit, Score ALD, and Diffusion Posterior Sampling across varying noise levels. Results show that diffusion models can successfully reconstruct corrupted images, with higher performance on in-distribution data and degradation for out-of-distribution inputs. Among the evaluated methods, DPS consistently achieves the best reconstruction quality, while SDEdit offers lower computational cost but reduced performance. These results demonstrate the effectiveness of diffusion models as priors for inverse problems and highlight the importance of incorporating measurement-consistent guidance in the sampling process.

References

[1] 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," 2022. [Online]. Available: https://arxiv.org/abs/2108.01073

[2] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. I. Tamir, "Robust compressed sensing mri with deep generative priors," 2021. [Online]. Available: https://arxiv.org/abs/2108.01368

[3] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," 2024. [Online]. Available: https://arxiv.org/abs/2209.14687

[4] Imagen-Team-Google, :, J. Baldridge, J. Bauer, M. Bhutani, N. Brichtova, A. Bunner, L. Castrejon, K. Chan, Y. Chen, S. Dieleman, Y. Du, Z. Eaton-Rosen, H. Fei, N. de Freitas, Y. Gao, E. Gladchenko, S. G. Colmenarejo, M. Guo, A. Haig, W. Hawkins, H. Hu, H. Huang, T. P. Igwe, C. Kaplanis, S. Khodadadeh, Y. Kim, K. Konyushkova, K. Langner, E. Lau, R. Lawton, S. Luo, S. Mokrá, H. Nandwani, Y. Onoe, A. van den Oord, Z. Parekh, J. Pont-Tuset, H. Qi, R. Qian, D. Ramachandran, P. Rane, A. Rashwan, A. Razavi, R. Riachi, H. Srinivasan, S. Srinivasan, R. Strudel, B. Uria, O. Wang, S. Wang, A. Waters, C. Wolff, A. Wright, Z. Xiao, H. Xiong, K. Xu, M. van Zee, J. Zhang, K. Zhang, W. Zhou, K. Zolna, O. Aboubakar, C. Akbulut, O. Akerlund, I. Albuquerque, N. Anderson, M. Andreetto, L. Aroyo, B. Bariach, D. Barker, S. Ben, D. Berman, C. Biles, I. Blok, P. Botadra, J. Brennan, K. Brown, J. Buckley, R. Bunel, E. Bursztein, C. Butterfield, B. Caine, V. Carpenter, N. Casagrande, M.-W. Chang, S. Chang, S. Chaudhuri, T. Chen, J. Choi, D. Churbanau, N. Clement, M. Cohen, F. Cole, M. Dektiarev, V. Du, P. Dutta, T. Eccles, N. Elue, A. Feden, S. Fruchter, F. Garcia, R. Garg, W. Ge, A. Ghazy, B. Gipson, A. Goodman, D. Górny, S. Gowal, K. Gupta, Y. Halpern, Y. Han, S. Hao, J. Hayes, J. Heek, A. Hertz, E. Hirst, E. Hoogeboom, T. Hou, H. Howard, M. Ibrahim, D. Ike-Njoku, J. Iljazi, V. Ionescu, W. Isaac, R. Jana, G. Jennings, D. Jenson, X. Jia, K. Jones, X. Ju, I. Kajic, C. Kaplanis, B. K. Ayan, J. Kelly, S. Kothawade, C. Kouridi, I. Ktena, J. Kumakaw, D. Kurniawan, D. Lagun, L. Lavitas, J. Lee, T. Li, M. Liang, M. Li-Calis, Y. Liu, J. L. Alberca, M. K. Lorrain, P. Lu, K. Lum, Y. Ma, C. Malik, J. Mellor, T. Mensink, I. Mosseri, T. Murray, A. Nematzadeh, P. Nicholas, S. Nørly, J. G. Oliveira, G. Ortiz-Jimenez, M. Paganini, T. L. Paine, R. Paiss, A. Parrish, A. Peckham, V. Peswani, I. Petrovski, T. Pfaff, A. Pirozhenko, R. Poplin, U. Prabhu, Y. Qi, M. Rahtz, C. Rashtchian, C. Rastogi, A. Raul, A. Razavi, S.-A. Rebuffi, S. Ricco, F. Riedel, D. Robinson, P. Rohatgi, B. Rosgen, S. Rumbley, M. Ryu, A. Salgado, T. Salimans, S. Singla, F. Schroff, C. Schumann, T. Shah, E. Shaw, G. Shaw, B. Shillingford, K. Shivakumar, D. Shtatnov, Z. Singer, E. Sluzhaev, V. Sokolov, T. Sottiaux, F. Stimberg, B. Stone, D. Stutz, Y.-C. Su, E. Tabellion, S. Tang, D. Tao, K. Thomas, G. Thornton, A. Toor, C. Udrescu, A. Upadhyay, C. Vasconcelos, A. Vasiloff, A. Voynov, A. Walker, L. Wang, M. Wang, S. Wang, S. Wang, Q. Wang, Y. Wang, Ágoston Weisz, O. Wiles, C. Wu, X. F. Xu, A. Xue, J. Yang, L. Yu, M. Yurtoglu, A. Zand, H. Zhang, J. Zhang, C. Zhao, A. Zhaxybay, M. Zhou, S. Zhu, Z. Zhu, D. Bloxwich, M. Bordbar, L. C. Cobo, E. Collins, S. Dai, T. Doshi, A. Dragan, D. Eck, D. Hassabis, S. Hsiao, T. Hume, K. Kavukcuoglu, H. King, J. Krawczyk, Y. Li, K. Meier-Hellstern, A. Orban, Y. Pinsky, A. Subramanya, O. Vinyals, T. Yu, and Y. Zwols, "Imagen 3," 2024. [Online]. Available: https://arxiv.org/abs/2408.07009

[5] Y. Liu, K. Zhang, Y. Li, Z. Yan, C. Gao, R. Chen, Z. Yuan, Y. Huang, H. Sun, J. Gao, L. He, and L. Sun, "Sora: A review on background, technology, limitations, and opportunities of large vision models," 2024. [Online]. Available: https://arxiv.org/abs/2402.17177

[6] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, "3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations," 2024. [Online]. Available: https://arxiv.org/abs/2403.03954

[7] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020. [Online]. Available: https://arxiv.org/abs/2006.11239

APPENDIX

Given

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}z_{t-1}, \qquad t = 1, 2, \ldots, T,$$

where $z_{t-1} \sim \mathcal{N}(0, I)$ are i.i.d Gaussian random variables for each step and β_t is the noise schedule parameter for each step. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, and the inductive hypothesis is

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

Consider $t = 1$:

$$\begin{aligned} x_1 &= \sqrt{1-\beta_1}x_0 + \sqrt{\beta_1}z_0 \\ &= \sqrt{\alpha_1}x_0 + \sqrt{1-\alpha_1}z \\ &= \sqrt{\bar{\alpha}_1}x_0 + \sqrt{1-\bar{\alpha}_1}z \end{aligned}$$

Then Equation (3) holds.

Consider $t = k + 1$:

$$\begin{aligned} x_{k+1} &= \sqrt{1-\beta_{k+1}}x_k + \sqrt{\beta_{k+1}}z_k \\ &= \sqrt{\alpha_{k+1}}x_k + \sqrt{1-\alpha_{k+1}}z \\ &= \sqrt{\alpha_{k+1}}\left(\sqrt{\bar{\alpha}_k}x_0 + \sqrt{1-\bar{\alpha}_k}z\right) + \sqrt{1-\alpha_{k+1}}z \\ &= \sqrt{\alpha_{k+1}\bar{\alpha}_k}x_0 + \sqrt{\alpha_{k+1}(1-\bar{\alpha}_k)}z + \sqrt{1-\alpha_{k+1}}z \\ &= \sqrt{\alpha_{k+1}\prod_{i=1}^{k}\alpha_i}x_0 + \left(\sqrt{\alpha_{k+1}\left(1-\prod_{i=1}^{k}\alpha_i\right)} + \sqrt{1-\alpha_{k+1}}\right)z \\ &= \sqrt{\prod_{i=1}^{k+1}\alpha_i}x_0 + \left(\sqrt{\alpha_{k+1}-\prod_{i=1}^{k+1}\alpha_i} + \sqrt{1-\alpha_{k+1}}\right)z \\ &= \sqrt{\bar{\alpha}_{k+1}}x_0 + \left(\sqrt{\alpha_{k+1}-\bar{\alpha}_{k+1}} + \sqrt{1-\alpha_{k+1}}\right)z \\ &= \sqrt{\bar{\alpha}_{k+1}}x_0 + \sqrt{\alpha_{k+1}-\bar{\alpha}_{k+1}+1-\alpha_{k+1}}z \\ &= \sqrt{\bar{\alpha}_{k+1}}x_0 + \sqrt{1-\bar{\alpha}_{k+1}}z \end{aligned}$$

So, we have shown that Equation (3) holds for $t = 1$ and $t = k + 1$, so by induction we have proved that it holds for all $t \geq 1$

Nick Ping Nick Ping is studying his Master of Science in Electrical Engineering at Stanford University.