Techniques for Applying Pretrained Diffusion Models to Inverse Imaging Problems

Isaac Criddle

Stanford University, EE367

INTRODUCTION

Large-scale image data and computation has enabled rapid advances in image generation over the last decade, especially diffusion models [Ho et al. 20]. As models grow larger and more complex, training becomes more expensive, and as such the appeal of using pretrained models for downstream applications rather than training new models from scratch.

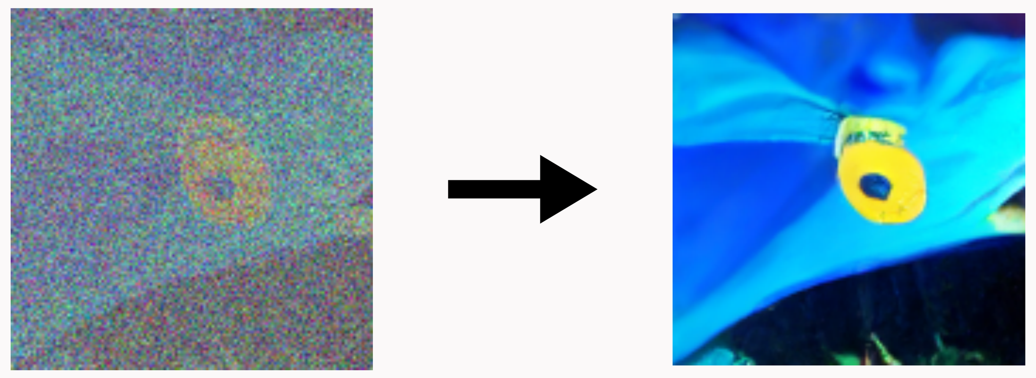
We compare three contemporary techniques for solving inverse imaging problems with generative priors, none of which require additional finetuning: SDEedit [Meng et al. 22], ScoreALD [Jalal et al. 21] and DPS [Chung et al. 23]. We compare both qualitative and quantitative results on inpainting and debarring tasks.

BACKGROUND

Some common inverse imaging problems include inpainting and outpainting, deblurring, deconvolution, among a host of other tasks. Conditioning output of a diffusion model given an input image has proven an effective way to tackle inverse problems.

Denoising diffusion probabilistic models [Ho et al. 20] learns to generate images via learning a reverse stochastic process. By beginning with an image, adding some noise, and learning the reverse step at various levels of noise, a network implicitly learns the distribution of images in the training set, with the ability to generate similar images from pure noise.

SDEedit, ScoreALD, and DPS are each techniques for solving inverse imaging problems such as inpainting and deblurring that can employ retrained diffusion models as priors for the inverse problem.



REFERENCES

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C. (2023). Diffusion posterior sampling for general noisy inverse problems. In ICLR.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. In NeurIPS.

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A. G., and Tamir, J. (2021). Robust compressed sensing mri with deep generative priors.

Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.-Y., and Ermon, S. (2022). Sdedit: Guided image synthesis and editing with stochastic differential equations.

ACKNOWLEDGEMENTS

Wikimedia, for the Mona Lisa image.

Dr. Wetzstein and Sonia Kim for their instruction, mentorship and feedback.

METHODS

We use an unconditioned image diffusion model obtained from Chung et al. as our retrained model in all applications, trained on a dataset of faces. We compare three methods:

SDEedit [Meng et al.]

SDEedit uses an input image as a "prior" by diffusion sampling not from random noise, but from a partially noised version of the input image as though it were a sample from a forward diffusion process. In this way we preserve the broad details from the original image but lose control over fine details as the model performs the backwards denoising process.

We obtain our initial noised image by

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

where y is our initial image and z is noise from a unit gaussian distribution. We can then run the noise image through our reverse diffusion as in DDPM.

ScoreALD [Jalal et al.]

ScoreALD samples from a posterior distribution using our input image as a prior, essentially conditioning the entire diffusion process on our noisy input. In this case, we begin from noise as in DDPM, but at each timestep, after running our model, we add a function of our current diffused image and the original noisy image back to our current diffused image.

$$-\frac{\nabla \left\| y - A(x_t) \right\|^2}{\sigma^2 + \gamma^2}$$

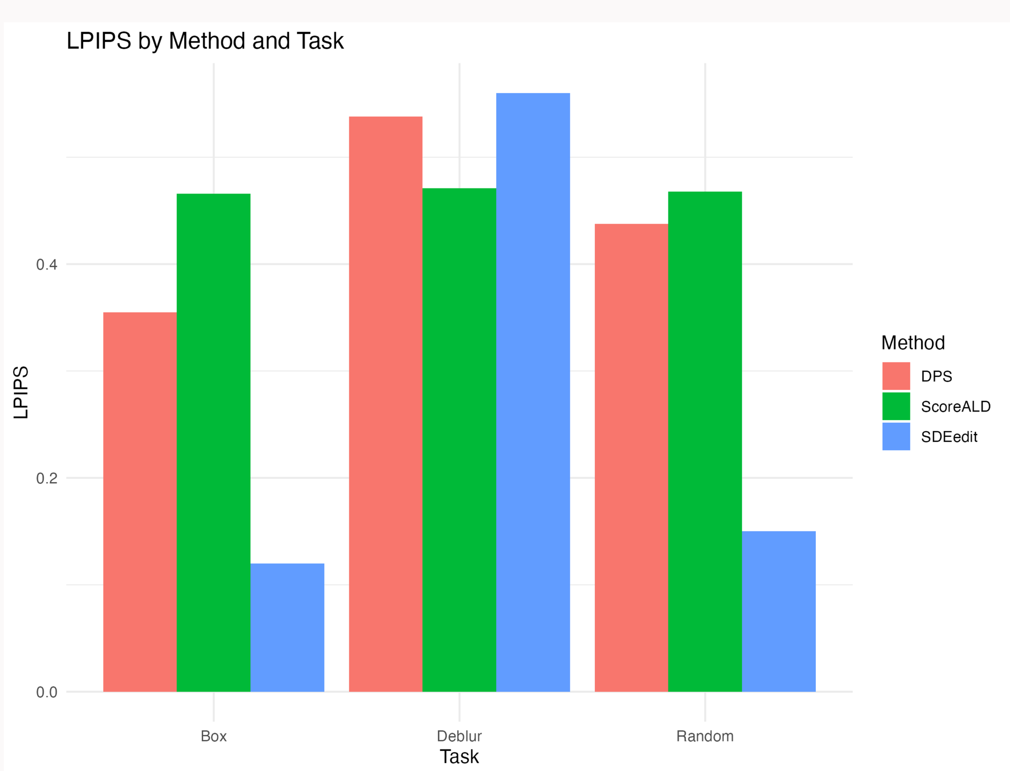
Diffusion Posterior Sampling [Chung et al.]

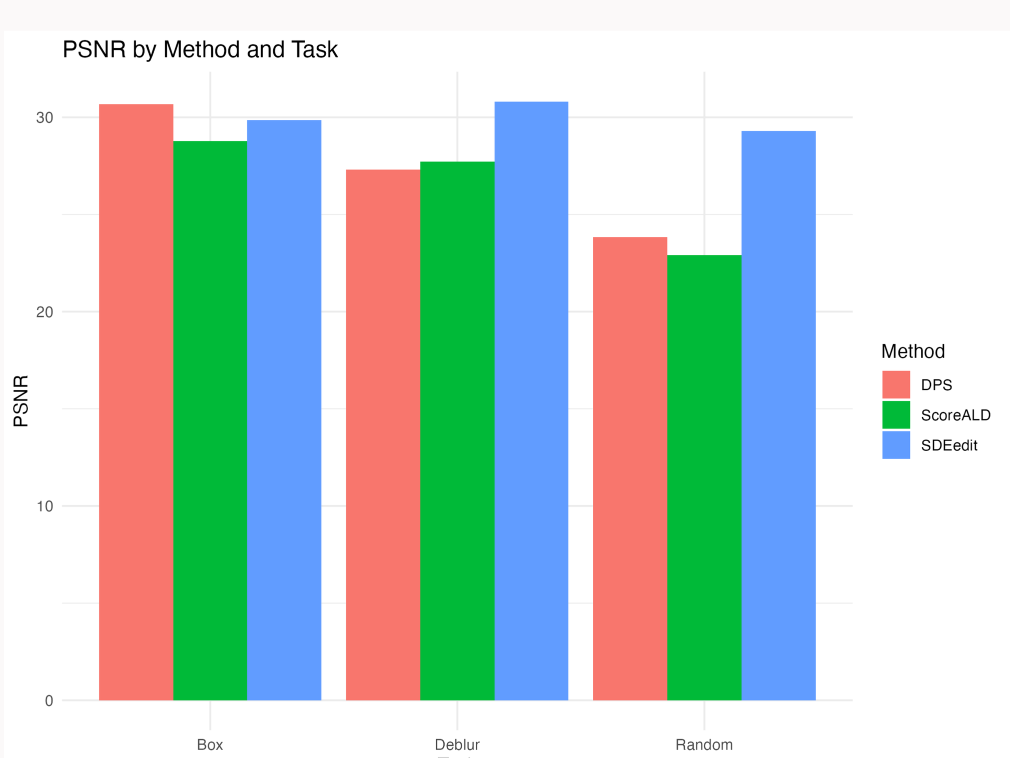
Diffusion Posterior Sampling is similar to ScoreALD in that it is a method of posterior sampling at each step of the diffusion process, though it uses a different expression to estimate the posterior:

$$-\frac{\zeta \nabla \left\| y - A(\hat{x}_0) \right\|^2}{\left\| y - A(\hat{x}_0) \right\|}$$

Here we use an estimate of our completely diffused image and then pass that estimate through our image formation model to obtain a normalized gradient we can tune to guide our diffusion process. Since we use an estimate of the completely diffused image instead of our current image, unlike ScoreALD, DPS can be used in non-linear inverse problem spaces.

RESULTS





Broadly speaking, there was no clear winner either perceptually or quantitatively. On different tasks and different images, different methods will achieve the best results. It seems, though, that SDEedit tends to do very well on inpainting tasks as compared to its performance on the deblur task.

All three methods allow for hyper parameter tuning—for consistency, these hyper parameters were kept at default values.

It is worth noting that none of the techniques allow for generalizing out of the training distribution, as seen in the Mona Lisa cases. In this case, since the model was trained on a dataset of faces, it is only suitable for use on inverse problems involving images of faces.