Plug and Play Diffusion Priors for Image Deconvolution

Ryan Po

Abstract—This paper introduces an novel method for image deconvolution using diffusion models as plug-and-play priors. We explore the potential of diffusion models trained on large datasets to solve challenging, ill-posed inverse problems like image deconvolution under high Gaussian noise. Unlike previous methods that require ground truth knowledge of noise characteristics and rely on the denoising ability of diffusion models, our approach leverages their generative capabilities without constraints on the noise variance. We demonstrate the effectiveness of our method over traditional techniques, particularly in the presence of high noise levels, showcasing the ability of diffusion models to guide images towards the natural image manifold while maintaining high perceptual quality.

Index Terms—Computational Photography

1 Introduction

Recent advances in diffusion models for image synthesis has proved to be incredibly effective not just for generation, but also solving challenging, ill-posed inverse problems such as inpainting and 3D reconstruction by utilizing these powerful models as a general prior on the distribution of natural images. While diffusion models can be viewed as a generator, through its intuitive modeling of a data distribution's gradient, it can also be interpreted as a native denoiser of Gaussian noise at different variance levels due to the way these models are trained.

Utilizing these powerful diffusion models as plug-and-play priors for solving inverse problems through optimization-based methods such as ADMM and HQS is an under-explored area. In this project, we aim to introduce a new framework that allows us to utilize the generative capabilities of diffusion models trained on large scale data for solving challenging inverse problems such as image deconvolution in the presence of high Gaussian noise corruption.

2 Related Work

2.1 Diffusion Models.

Foundation models like Stable Diffusion, Imagen, Midjourney, DALL-E 2, and DALL-E 3 have revolutionized visual computing through generative AI. These models, trained on text-image pairs ranging from hundreds of millions to billions in number, possess an expansive knowledge base and operate with billions of learnable parameters, making them exceptionally large. They achieve their capabilities after extensive training on a vast array of high-performance GPUs, laying the groundwork for cutting-edge generative AI applications. The core technology behind image, video, and 3D object generation in these tools primarily involves variants of convolutional neural networks (CNNs) infused with diffusion models. These are further enhanced by integrating multimodal interactions with text, using transformer-based frameworks like CLIP.

2.2 Image Editing and SDEdit

Due to their progressive and attention-focused design, diffusion models naturally lends itself to the task of image editing. They enable precise adjustments throughout various stages and aspects of the network, allowing for modifications in both the spatial arrangement and visual appeal of images. Current research aims to improve the control and versatility of editing processes, ensuring they are user-friendly. Some common applications include the modification of an image's visual characteristics while maintaining its original layout and guiding an incomplete image to the natural image manifold. In this area, SDEdit [1] offers a simple yet effective method by introducing a controlled amount of noise to create a partially noised image. This is then refined through a reverse diffusion process, guided by a new conditioning signal.

2.3 Diffusion Models as Plug-and-play Denoiser

Prior work [2], has shown the potential of using diffusion models as a general prior for solving inverse problems such as image deconvolution. However, this method is predicated on the fact that the image formation models involves an additive noise term that is Gaussian in nature. Patel et al. assumes that the measurement image lies on the diffusion trajectory, and can therefore treat the trained diffusion model as a denoiser, rather than a generator. Treating the diffusion model as a Gaussian denoiser also requires knowledge of the ground truth variance of the injected noise in the image formation model, another constraint that is not easily satisfied. In our method, we aim to leverage the *generative* capabilities of the diffusion model rather than its denoising capabilities. Our method requires no constraints on the image formation model, and does not need ground truth knowledge of the noise characteristics.

3 Preliminaries

3.1 Diffusion Model Preliminaries

Recent work has shown that diffusion models can achieve state-of-the-art quality for image generation tasks [3]. Specif-

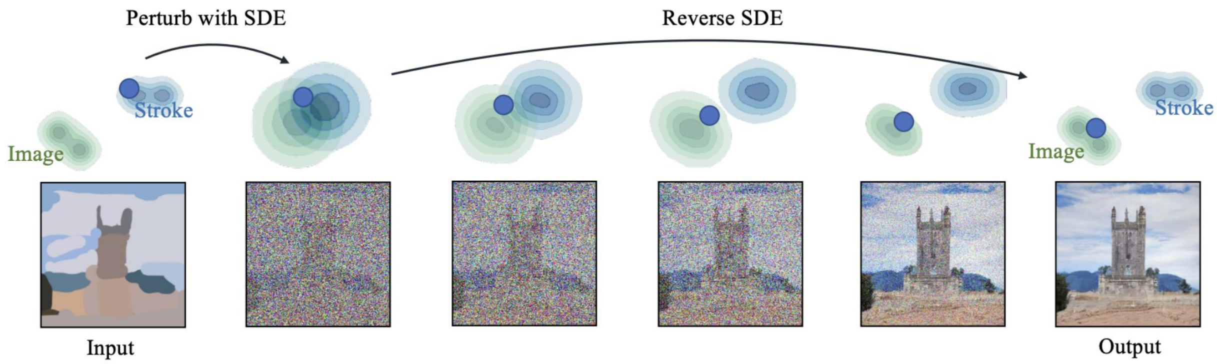

Fig. 1. **SDEdit Visualization.** SDEdit involves introducing a calibrated level of noise to an already existing image and then guiding the reverse diffusion process with a new conditioning signal.

ically, Denoising Diffusion Probabilistic Models (DDPMs) implement image synthesis as a denoising process. DDPMs begin from sampled Gaussian noise x_T and apply T denoising steps to create a final image x_0. The forward diffusion process q is modelled as a Markov chain that gradually adds Gaussian noise to a ground truth image according to a predetermined variance schedule $\beta_1, \beta_2, \ldots, \beta_T$

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}\right) \tag{1}$$

The goal of DDPMs is to train a diffusion model to revert the forward process. Specifically, a function approximator $\boldsymbol{\epsilon}_\phi$ is trained to predict the noise $\boldsymbol{\epsilon}$ contained in a noisy image x_t at step t. $\boldsymbol{\epsilon}_\phi$ is typically represented as a convolutional neural network characterised by its parameters ϕ. Most successful models [3], [4], [5] train their models using a simplified variant of the variational lower bound on the data distribution:

$$\mathcal{L}_{\text{DDPM}} = \mathbb{E}_{t,x,\boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\phi\left(x_t, t\right)\right\|^2\right] \tag{2}$$

with t uniformly sampled from $\{1, \ldots, T\}$. The resulting update step for obtaining a sample for x_{t-1} from x_t is then

$$x_{t-1} = x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\phi(x_t, t) + \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t\mathcal{N}(0, \mathbf{I}) \tag{3}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$, $\alpha_t = 1-\beta_t$.

Text-to-image diffusion models build upon the above theory to introduce conditional diffusion processes using classifier-free guidance [6]. Given a condition y, usually represented as a text prompt, a diffusion model $\boldsymbol{\epsilon}_\phi(x_t, t, y)$ is trained to predict noise in an image as shown in Eq. 2. During training, conditioning y is randomly dropped out, leaving the diffusion model to predict noise without it. At inference, noise prediction is instead represented by:

$$\hat{\boldsymbol{\epsilon}}_\phi(x_t, t, y) = \boldsymbol{\epsilon}_\phi(x_t, t, \emptyset) + s\Big(\boldsymbol{\epsilon}_\phi(x_t, t, y) - \boldsymbol{\epsilon}_\phi(x_t, t, \emptyset)\Big) \tag{4}$$

Where s is a user-defined constant controlling the degree of guidance and $\boldsymbol{\epsilon}(x_t, t, \emptyset)$ represents the noise prediction without conditioning.

3.2 Image Deconvolution

The task of deconvolution is often ill-posed. Given a convolution operator $\mathbf{A}$, a vectorized image $\mathbf{x}$ and measurements $\mathbf{b}$, then the image formation model can be posed as,

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{5}$$

where n is some signal-independent additive noise. The problem of recovering $\mathbf{x}$ from the observed measurements $\mathbf{b}$ is an ill-posed problems, with many possible solutions. In order to constrain the solution space, it is often useful to introduce some prior $\Psi(\mathbf{x})$ into the system. In this case, the reconstruction problem can be posed as,

$$\underset{x}{\text{minimize}}\ \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\Psi(\mathbf{x}), \tag{6}$$

where $\Psi\mathbf{x}$ is the custom prior evaluated on the recovered image $\mathbf{x}$ and λ is some scalar weight determining the strength of the prior. We further rephrase the above optimization problem to the following form, separating the reconstruction and prior terms,

$$\begin{aligned} &\underset{\{x\}}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2}_{f(\mathbf{x})} + \underbrace{\lambda\Psi(z)}_{g(\mathbf{z})} \\ &\text{subject to} \quad D\mathbf{x} - z = 0. \end{aligned} \tag{7}$$

The above formulation lends itself nicely to be solved by common optimization methods such as ADMM and HQS.

3.3 Alternating Direction Method of Multipliers

In this project, we aim to employ an optimization-based method for solving the problem of image deconvolution. We focus on the Alternating Direction Method of Multipliers (ADMM), as it lends well to the plug-and-play fashion in which we wish to apply diffusion priors on. ADMM is a powerful optimization algorithm that combines ideas from dual decomposition and augmented Lagrangian methods to solve a wide range of complex optimization problems efficiently. ADMM is particularly well-suited for distributed computing and large-scale optimization problems that can be separated into smaller subproblems.

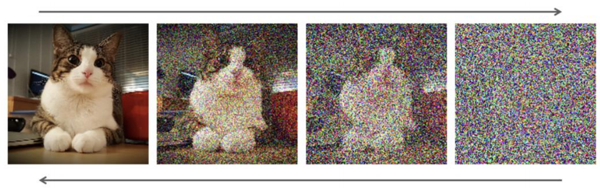

Fig. 2. **Diffusion Models.** Diffusion models simulate the forward and reverse process of iteratively adding Gaussian noise to data samples. Samples are generated by traversing backwards from randomly sampled Gaussian noise.

Following the formulation of the deconvolution problem presented in Equation 7, we can take the Augmented Lagrangian giving,

$$L_\rho^{(ADMM)}(x,z,u) = f(x)+g(z)+\frac{\rho}{2}\|Dx-z+u\|_2^2-\frac{\rho}{2}\|u\|_2^2. \quad (8)$$

Using the above formulation, the following iterative update rules can be derived,

$$x \leftarrow \text{prox}_{\|\cdot\|_2,\rho}(v) = \underset{\{x\}}{\arg\min}\ \frac{1}{2}\|Ax-b\|_2^2+\frac{\rho}{2}\|Dx-z+u\|_2^2, \quad (9)$$

$$z \leftarrow \text{prox}_{\Psi,\rho}(v) = \underset{\{z\}}{\arg\min}\ \lambda\Psi(z)+\frac{\rho}{2}\|Dx-z+u\|_2^2, \quad (10)$$

$$u \leftarrow u + Dx - z. \quad (11)$$

In practice, the following is implemented by the following operations.

$$x \leftarrow (A^TA+\rho D^TD)^{-1}A^Tb+\rho D^Tv \quad (12)$$

$$z \leftarrow \arg\min_z g(z)+\frac{\rho}{2}\|v-z\|_2^2 \quad (13)$$

$$u \leftarrow u+Dx-z \quad (14)$$

In the case where we're working with a Gaussian denoiser, the proximal operator for the z update can be derived as,

$$\text{prox}_{D,\rho}(v) = D\left(v, \sigma^2=\frac{\lambda}{\rho}\right) \quad (15)$$

Where D is any Gaussian denoiser of our choosing, including a diffusion model.

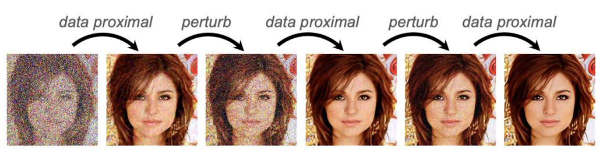

Fig. 3. **Proposed Method.** By artificially injecting Gaussian noise into measured data, we place our measured data onto the diffusion trajectory. Alternating between injecting noise and solving the data proximal problem allows us to leverage the diffusion prior without knowledge of existing noise in the measurement.

4 Proposed Method

Inspired by SDEdit, we propose to leverage the generative capabilities of diffusion models by manually placing our measured data near a known point on the diffusion trajectory.

Given an observed measurement $\mathbf{b}$, we manually add Gaussian noise to our measurement at some known scale. Specifically, we sample

$$\mathbf{b}(t) \sim \mathcal{N}(\mathbf{b}; \sigma(t)\boldsymbol{I}) \quad (16)$$

, where t corresponds to some timestep in the diffusion trajectory, and $\sigma(t)$ is the variance defined at timestep t. We then solve the data proximal problem using ADMM to obtain some estimated solution $\hat{\mathbf{x}}_t$. Given our updated measurement $\mathbf{b}(t)$, we use the pre-trained diffusion model as a native denoiser, as we reverse the diffusion process starting from timestep t, producing a clean, denoised estimate of $\mathbf{b}(t)$. This clean estimate serves as the prior during the optimization process. The above process of injecting noise into the observed measurement and solving the data proximal problem using the diffusion model as a native denoising prior is then repeated for another timestep.

Our method, removes the need for ground truth knowledge of the variance of the Gaussian noise in the measured image, as the process of injecting noise into the measurement image places the measurement image onto a known spot along the diffusion trajectory. While we are treating the diffusion model as a native denoiser when substituting it for D in Equation 15, we are still leveraging the generative capabilities of the diffusion model through the use of an SDEdit-like editing process.

5 Experiments

5.1 Setup

In order to test the capability of our setup, we start off with a clean image, convolving it with some random kernel, then add a known amount of Gaussian noise to the convolved image. We test our set up on two different noise levels, $\sigma^2 = 0.1, 0.5$. The higher noise level is expected to leverage the generative capabilities of the diffusion model more, as the inverse problem becomes more ill-posed.

5.2 Metrics

We measure the performance of our method in the task of image deconvolution using PSNR. We also include a perceptual metric in the form of LPIPS in our quantitative evaluations. Since we are experimenting with high noise levels, the ill-posed nature of the problem will likely lead to reconstructed image to stray away from the natural image manifold. Diffusion models, on the other hand, have the capability to hallucinate details when facing such ill-posed settings. In this case, we believe perceptual metrics (e.g. LPIPS) is a better measurement for a high quality reconstruction of the original image.

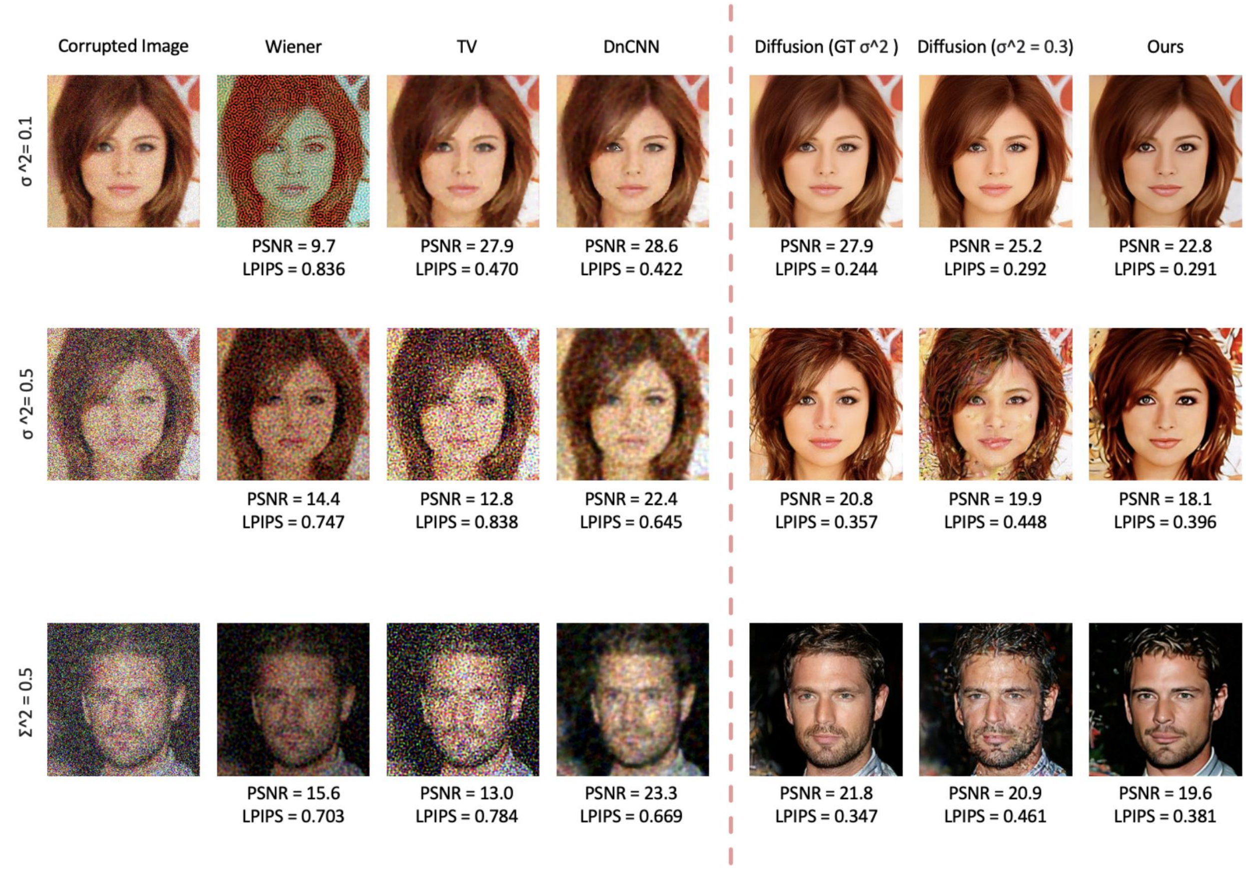

Fig. 4. **Main Results.** We evaluate or method and other baselines under two different noise levels $\sigma^2 \in \{0.1, 0.5\}$. While non-diffusion results tend to lead to higher PSNR, diffusion-based methods tend to outperform in perceptual metrics. Our method also outperforms other diffusion methods when the ground truth injected noise is unknown.

5.3 Baselines

We compare our method with a variety of others priors used in conjunction with ADMM. Namely, Wiener, total variation (TV), and DnCNN.

For diffusion prior-based methods, we also test out three different settings.

- In the first, we assume that we know the ground truth variance of the noise injected into the measurement image. This allows us to fully leverage the diffusion model as a denoiser, as we replace D with the diffusion model starting from the timestep corresponding to the ground truth noise level.
- In the second setting, we assume no knowledge of the noise injected into the measurements, but we still treat the diffusion model as a denoiser, but starting from a fixed timestep corresponding to $\sigma^2 = 0.3$.
- Finally, we run the diffusion model using our method, where we alternate between injecting noise into the measurement image at a known noise level and recovering the image using ADMM. We use a linearly decreasing time-schedule of $0.5, 0.3, 0.1$ for the higher noise setting, and $0.15, 0.1, 0.05$ for the lower noise setting.

6 Results

6.1 Comparisons

Figure 4 shows reconstructed results using the baseline methods and the three diffusion-based methods outlined in the section above. For $\sigma^2 = 0.1$, when noise levels are low, conventional methods such as TV prior already work surprisingly well, with DnCNN performing the best leading to the highest PSNR. However, looking at the perceptual metrics, diffusion methods lead to much better results as their generative capabilities lead to the generation of reconstructed images that lie on the natural image manifold. This effect is further illustrated at higher noise levels ($\sigma^2 = 0.5$), where conventional priors perform much worse. The generative capabilities still manage to generate images that look more natural. However, although these images look more ”clean”, they lead to higher PSNR values than baseline methods due to the generative nature of the diffusion priors used.

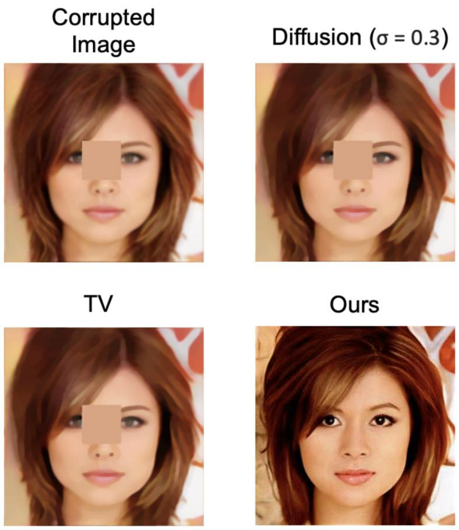

Fig. 5. **Diffusion Priors for Other Inverse Problems.** Preliminary results from applying our method to a more challenging setting, where the convolved image is corrupted with a mask.

The advantage of our method is also illustrated at high noise levels. Although the diffusion prior can be used with great performance given ground truth knowledge of the injected noise level, it can be observed that reconstruction quality drops drastically when the noise used for the denoising prior during ADMM is not equal to the actual injected noise level. Our method, which makes no assumptions about the noise level leads to better perceptual metrics when compared to using the diffusion prior without ground truth noise level.

However, there are also key limitations to our method. Firstly, the generative nature of the diffusion prior leads to results that don't necessarily align well with the original ground truth image. This is especially apparent in the results shown in the bottom row, where the recovered face has a different complexion than the other recovered results. Our method also leads to a certain degree of saturation, a known limitation when traversing back and forth along the diffusion trajectory too many times [7].

6.2 Other Inverse Problems

We experiment with other forms of image formation models where the measurements are corrupted through an averaging mask instead of additive Gaussian noise. We show preliminary results in Figure 5. Our method is able to reconstruct and hallucinate details in the woman's face by leveraging the generative capabilities of the diffusion model.

7 Conclusion and Discussion

We present a novel method for image deconvolution by leveraging the generative capabilities of diffusion models. Using a pre-trained diffusion model as denoising prior for ADMM-based image deconvolution, we show that our method outperforms existing baselines in perceptual metrics in a less constrained setting. We also show preliminary results of where diffusion priors can be applied to solve even more challenging inverse problems where information seems to be completely loss, such as in the case of masking/inpainting.

While diffusion models are shown to be powerful priors, they are also prone to hallucinate due to their generative capabilities. This combined with unsolved issues of saturation resulting from traversing back and forth in the diffusion trajectory leaves room for improvement in future work.

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," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:245704504

[2] K. K. Zakaria Patel, Akash Haridas, "Deconvolution using admm with diffusion prior," https://www.cs.toronto.edu/~lindell/teaching/2529/past_projects/2022/report/akash_haridas-kartikaeya_kumar-zakaria_patel.pdf, 2023.

[3] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *ArXiv*, vol. abs/2105.05233, 2021.

[4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *ArXiv*, vol. abs/2006.11239, 2020.

[5] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE transactions on pattern analysis and machine intelligence*, vol. PP, 2021.

[6] J. Ho, "Classifier-free diffusion guidance," *ArXiv*, vol. abs/2207.12598, 2022.

[7] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," *ArXiv*, vol. abs/2209.14988, 2022.