Diffusion Image Prior

Hamadi Chihaoui* and Paolo Favaro
University of Bern, Switzerland
hamadi.chihaoui@unibe.ch, paolo.favaro@unibe.ch

Abstract

Zero-shot image restoration (IR) methods based on pre-trained diffusion models have recently achieved significant success. These methods typically require at least a parametric form of the degradation model. However, in real-world scenarios, the degradation may be too complex to define explicitly without relying on crude approximations. To handle this general case, we introduce the DIffusion Image Prior (DIIP). We take inspiration from the Deep Image Prior (DIP) [17], since it can be used to remove artifacts without the need for an explicit degradation model. However, in contrast to DIP, we find that pretrained diffusion models offer a much stronger prior, despite being trained without knowledge from corrupted data. We show that, the optimization process in DIIP first reconstructs a clean version of the image before eventually overfitting to the degraded input, but it does so for a broader range of degradations than DIP. In light of this result, we propose a blind image restoration (IR) method based on early stopping, which does not require prior knowledge of the degradation model. We validate DIIP on various degradation-blind IR tasks, including JPEG artifact removal, waterdrop removal, denoising and super-resolution with state-of-the-art results.

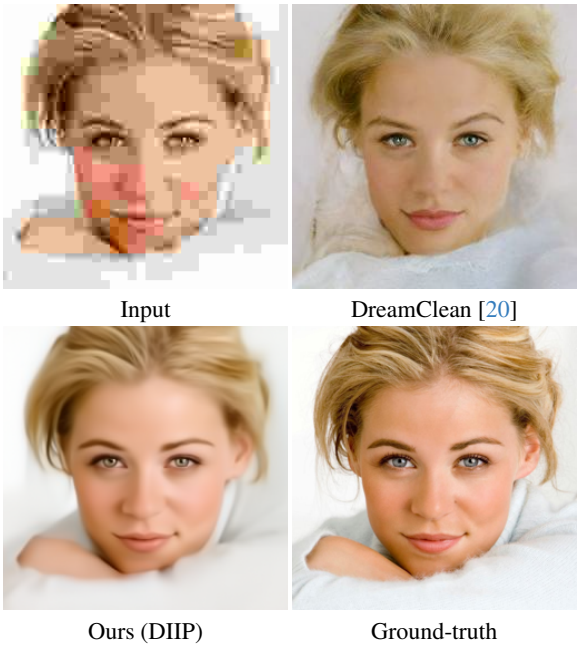


Figure 1. In many applications, such as JPEG artifacts removal, image degradation is often unknown and challenging to model. In comparison to the recent state-of-the-art DreamClean [20], DIffusion Image Prior (DIIP) consistently produces results that more effectively preserve the original nature of the image.

1. Introduction

One of the most well-known dataset training-free and degradation blind methods for image restoration is Deep Image Prior (DIP) [17]. A notable feature of DIP is its applicability in cases where the degradation model is too complex to be accurately modeled, such as with JPEG compression. DIP leverages the implicit prior embedded within a convolutional neural network (CNN). It demonstrates that through an iterative reconstruction process of a corrupt image with an untrained CNN (by using direct matching, since the degradation model is unknown), one can recover a clean image. Inspired by this approach, we explore a similar investigation, where we also exploit the reconstruction process of a degraded image, but use instead a frozen pre-trained (diffusion) model (see Figure 1). Although our reconstruction optimizes only the input noise to the model, while in DIP the CNN parameters were optimized, we find that the two processes share the same property of reconstructing a clean image before converging to the corresponding degraded input. Surprisingly, we find that while in DIP this property holds primarily for high frequency degradations such as noise, with pre-trained diffusion models this phenomenon is observed across a wider range of degradations, including low-frequency ones such as blurring. This happens despite the fact that the diffusion model was never exposed to such degraded data during training to either generate it or to be adversarial to it. We call this property the *DIffusion Image*

*Corresponding author.

Prior (DIIP).

DIIP has immediate applicability in image restoration (IR), where the aim is to reconstruct a clean image from its degraded version. While DIP can handle only image corruption processes that inject high-frequency artifacts in an image without using the degradation operator, DIIP can handle a broader range of artifacts. Recently, DreamClean [20] has also demonstrated the remarkable capability of restoring images without knowledge of the degradation model for a wide range of cases. However, as shown in Figure 1, DIIP yields reconstructions that are more faithful to the original image. The capability of the above three methods drastically simplifies the solution of IR tasks, as they do not require any collection of specialized datasets for supervised training [12, 18], which can be costly and time-consuming, nor the use of additional knowledge about the artifacts [3–6, 9, 11, 14], which is often not available in practice.

We demonstrate DIIP on several IR problems, non-uniform deformation, waterdrop removal, different noise distributions (mixtures of Gaussian and speckle noise), JPEG compression and superresolution. Our findings indicate that in all these degradation cases the reconstruction of the initial noise through a pre-trained and frozen diffusion model initially generates clean reconstructions before overfitting to the degraded images. This suggests that DIIP can be used in a similar fashion to DIP via *early stopping*. In this paper, we present some self-supervised early stopping criteria that result in state of the art performance on all the above IR tasks despite their simplicity. In summary, our contributions are

- We conduct a study showing the implicit priors of a frozen pretrained diffusion model when used for image reconstruction in the case of a wide range of degradations that either remove high-frequency details or introduce high-frequency artifacts; we show that trained diffusion models exhibit a similar implicit prior behavior as observed with DIP, but on a broader range of cases;
- Based on our findings, we propose a new dataset training-free and fully blind image restoration method, DIIP, which does not assume any prior knowledge of the degradation model, making it broadly applicable to a wide range of complex image restoration tasks;
- We demonstrate state-of-the-art performance on CelebA and ImageNet benchmarks for several blind image restoration tasks, including denoising, waterdrop removal, super-resolution, and JPEG artifact removal.

2. Related Work

Image restoration has a long and rich history in the literature. We focus on image restoration methods that are similar to our approach in terms of their training paradigm (test-time training methods) and blindness (fully blind).

2.1. Test-Time Image Restoration

Recently, test-time training methods have gained popularity. These methods eliminate the need for a training dataset and are applied directly to the degraded image at test time. We can further classify these methods based on their level of blindness of the degradation model.

Non-blind and partially-blind image restoration. Non-blind methods assume that the degradation model is fully known; for example, in image deblurring, they assume the degradation takes the form of a blur kernel and that the entries of this kernel are known. Some recent non-blind methods include [4, 11, 19]. DDRM [11] introduces a variational inference objective and proposes an inverse problem solver based on posterior sampling to learn the posterior distribution of the inverse problem. DDNM [19] presents a zero-shot framework for image restoration tasks using range-null space decomposition, refining only the null-space contents during the reverse diffusion process to both ensure data fidelity and realism. DPS [4] offers a more general framework that addresses both non-linear and noisy cases. However, these methods assume complete knowledge of the degradation model, which is rarely available in real-world scenarios. On the other hand, partially blind methods relax this assumption by only requiring knowledge of the parametric form of the degradation model. These methods include [3, 6, 9, 15]. GibbsDDRM [15] extends DDRM to scenarios where only partial information about the degradation is available. BlindDPS [6] builds on the work of [4], addressing blind deblurring by jointly optimizing the blur operator and the sharp image during the reverse diffusion process. BIRD [3] proposes a fast diffusion inversion method that simultaneously inverts the diffusion model and infers the degradation model. In contrast, DIIP handles the fully degradation-blind case, *i.e.*, when we do not assume knowledge of the degradation's parametric form. Moreover, BIRD uses a fixed stopping criterion (*i.e.*, a fixed number of iterations), whereas DIIP relies on a self-supervised stopping criterion.

Fully-blind image restoration methods. There have been a few attempts to solve image restoration in a zero-shot manner when the explicit parametric form of the degradation is unknown. One such method is Deep Image Prior (DIP), which can handle complex degradation scenarios despite its reliance on just the implicit prior of untrained neural networks. Recently, DreamClean [20], a fully blind image restoration method based on a pre-trained diffusion model, was introduced. DreamClean works by inverting the input image and iteratively moving the intermediate latents through an additional variance-preserving step during the reverse diffusion sampling process. DIIP instead uses a pre-trained diffusion model by reconstructing the initial noise (and by making the diffusion process deterministic).

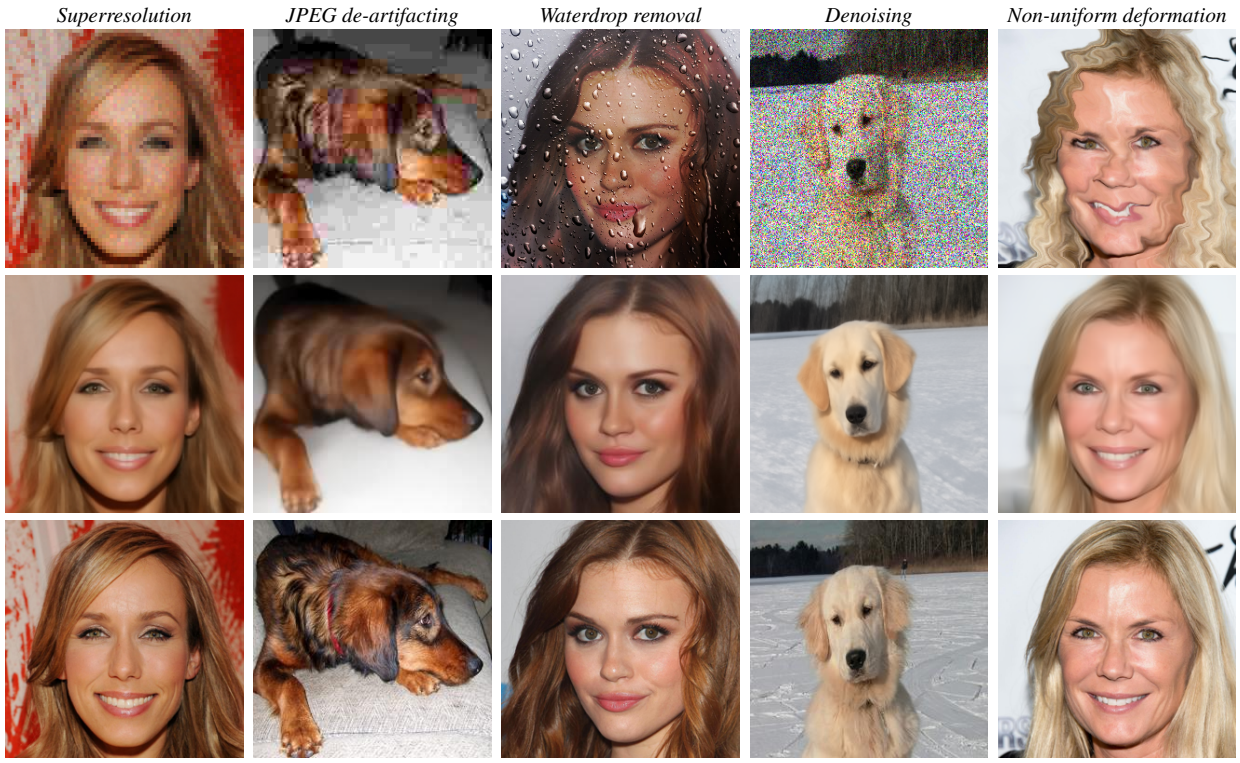


Figure 2. DIIP on several **degradation-blind** image restoration tasks. Top: Input image. Middle: Our prediction. Bottom: Ground truth.

3. The DIffusion Image Prior

3.1. Image Restoration: Problem Statement

We aim to solve the task of image restoration in a test-time training and fully blind manner. Given a degraded input image y, our goal is to recover the degradation-free (clean) image x. We leverage a pre-trained diffusion model, specifically DDIM [16], a class of diffusion models that enables deterministic mapping from noise space to data space. g denotes the **deterministic** version of DDIM mapping a Gaussian vector to a data sample from the target distribution.

We focus on degradations that are rarely addressed by existing test-time and blind methods, as these degradations are either too complex to model in a test-time setting or simply unknown. Examples include noise with unknown distributions (*e.g.*, correlated noise or non-Gaussian/Poisson distributions), JPEG compression artifacts, and non-uniform blur (see also Figure 2). In such cases, the IR task can be formulated as an energy minimization problem of the form

$$x^* = \arg\min_x E(x; y) + R(x), \tag{1}$$

where $E(x; y)$ is a data fidelity term that measures how well x matches the observed image y, and $R(x)$ is a regularization term that enforces prior knowledge about the distribution of x.

3.2. Revisiting the Deep Image Prior

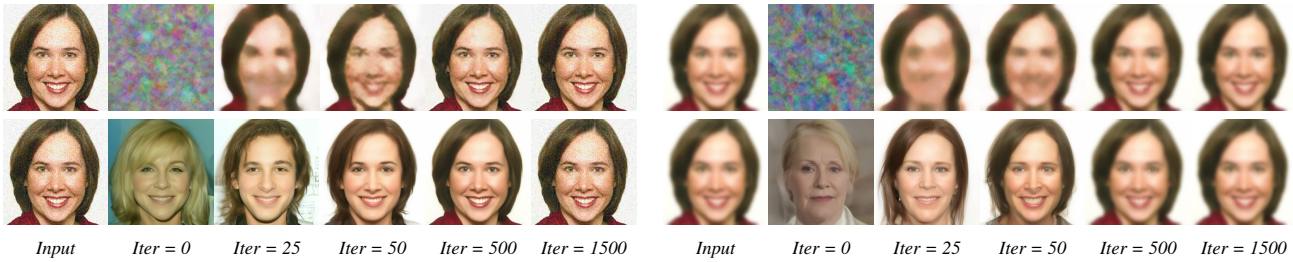
Deep Image Prior (DIP) addresses the optimization problem in Eq. (1) by introducing an implicit prior captured by a class of neural networks, specifically, deep Convolutional Neural Networks (CNNs), and by employing *early stopping*, as follows

$$\theta^* = \arg\min_\theta \|f_\theta(z) - y\|^2, \quad \text{with} \quad x^* = f_{\theta^*}(z), \tag{2}$$

where we imply $E(f_\theta(z); y) = \|f_\theta(z) - y\|^2$. The minimizer θ^* is obtained using an iterative optimizer such as gradient descent, and starts from a random initialization of the parameters of a convolutional neural network f_θ fed a fixed input noise vector z (which is not optimized). [17] applies early stopping to avoid overfitting the degraded image y. This is implemented by defining a fixed and reasonable number of iterative steps T_{max}. The early stopped solution θ^* is then used to compute the restored image $x^* = f_{\theta^*}(z)$.

3.3. Implicit Priors of a Pretrained Diffusion Model

Motivated by the findings of [17] on the implicit prior of an untrained CNN, a relatively weak model due to its lack of training, we aim to explore a much stronger alternative: a diffusion model pre-trained on clean data. More specifically, we aim to address two key questions:



(a) Gaussian noise (b) Gaussian blur

Figure 3. Intermediate outputs of the iterative reconstruction of a degraded image in the case of DIP (top) and our proposed optimization scheme using a frozen pretrained diffusion model (bottom). In the case of a pre-trained diffusion model, the optimization consistently produces clean images at an intermediate stage *irrespective of the type of degradation* ((a) noise or (b) blur).

Q1: Does a pre-trained diffusion model exhibit similar implicit priors to those observed in DIP?

Q2: If such implicit priors exist, how do they differ from those in DIP?

To investigate these questions, we explore the task of reconstructing a degraded image using a pre-trained diffusion model, where g represents its induced mapping from noise to image space. Moreover, while DIP fixes the network input z and trains the network parameters, here we keep the pre-trained diffusion model frozen and optimize with respect to the input z alone. The optimization objective is then equivalent to

$$z^* = \arg\min_z \|g(z) - y\|^2, \quad \text{with} \quad x^* = g(z^*), \tag{3}$$

where we also let $E(g(z); y) = \|g(z) - y\|^2$. In general, the degradation can either *remove* high-frequency details (resulting in smoothing) or *add* high-frequency artifacts (*e.g.*, noise). Here, we empirically investigate the behavior of the optimization described in Eq. (3) under synthetic degradations that either add high-frequency artifacts (Gaussian noise) or remove high-frequency details (Gaussian blur). For this study, we randomly select images from the FFHQ [10] dataset and apply: (1) Gaussian noise, and (2) Gaussian blur. We frame this as *a pure reconstruction task* and examine how the optimization in Eq. (3) behaves when reconstructing the degraded images versus the original reference (clean) images. Our ultimate goal is to generalize these findings to more complex degradations, such as JPEG artifacts and non-uniform blur. We run the optimization until it converges (this occurs with $N = 1500$ iterations). In the case of a diffusion model, the mapping $g(z)$ could be computationally demanding, and this could make the iterative reconstruction in Eq. (3) unfeasible. To enable this optimization, we adopt the efficient inversion strategy used in [3], which was originally proposed in [16]. However, in contrast to [3] our final objective is not to use the inversion to reconstruct the clean image, but rather the degraded one. Also, unlike [3] we do not use any degradation model. The detailed algorithm for our optimization is provided in Section 1 of our supplementary material. Our findings are that the answer to Q1 is positive and the answer to Q2 can be split into the following two observations

1. *There are two distinct regimes observed during the reconstruction, regardless of degradation type.* Interestingly, a key difference when using a pre-trained diffusion model compared to vanilla DIP is that, irrespective of the type of degradation (noise or blur), the optimization consistently produces clean images at an intermediate stage, even though it is purely a reconstruction process with *no integrated degradation model*. Specifically, there are two distinct regimes observed: (I) an initial regime where the generated images are clean, realistic, and progressively approach the clean reference images, followed by (II) a regime where the generated images start to reflect the degradation, becoming either blurred or noisy. This behavior is illustrated in Figure 3, by showing an example of intermediate reconstructions obtained during our optimization and that of DIP. In the case of DIP, in contrast, the reconstruction path is not consistent across the degradations. DIP fails to produce sharp and clean images during the iterative reconstruction in the case of blurry images.
2. *There is a high inertia to reconstructing high-frequency artifacts (*e.g.*, from noise).* One commonality observed between DIP and our proposed optimization scheme is the tendency to resist overfitting to high-frequency artifacts (such as noise). This behavior is illustrated in Figure 3, where the intermediate image remains clean until a late stage of the optimization (iteration = 500) when reconstructing an image corrupted with Gaussian noise.

A crucial requirement for leveraging the findings from the previous section is an automated method to determine when to stop the optimization. In the next paragraphs, we outline two simple and effective procedures to do so.

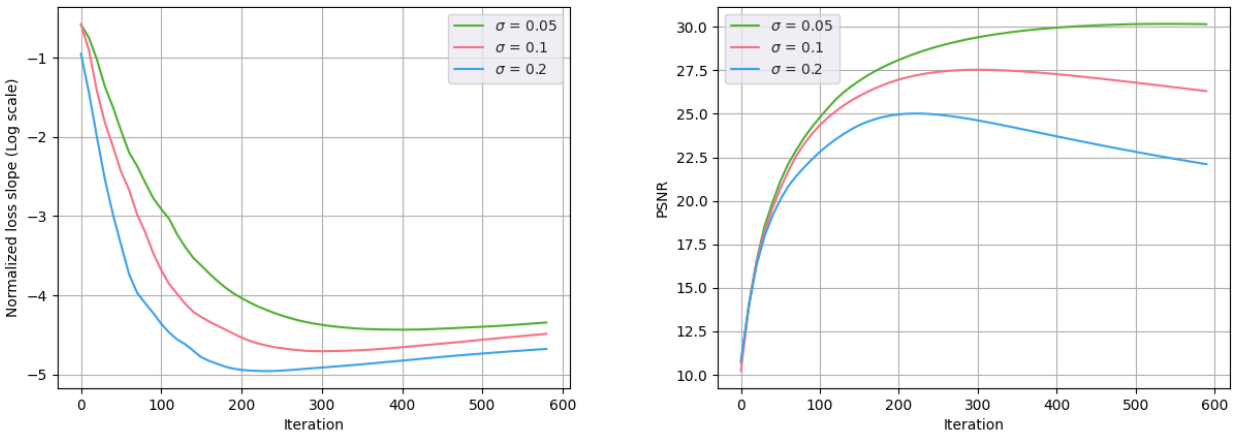


Figure 4. **Stopping criterion in the case of degradations that introduce high-frequency artifacts.** We study the optimization trend across varying levels of degradation that introduce high-frequency artifacts, focusing specifically on input images with different noise levels.(a) shows the normalized slopes of the loss functions for input images with increasing noise levels. (b) shows the Peak Signal-to-Noise Ratio (PSNR) relative to the clean reference image as optimization progresses for various noise intensities. We observe that the slopes in (a) reach their minima at iterations corresponding to the points where the PSNR in (b) is maximal. For noise-based degradations, which introduce high-frequency artifacts, the slope in (a) can serve as an indicator of peak performance, providing guidance on the optimal stopping point for the optimization process.

3.4. Self-Supervised Stopping Criteria

3.4.1 Degradations removing high-frequency content

Identifying an absolute measurement to characterize image sharpness is challenging. However, it is more manageable to assess *the relative change in the sharpness* of intermediate reconstructions. Inspired by Chao et al. [2], who demonstrated that the variance of the Laplacian (LV) provides a robust measure of image sharpness, we adopt LV to detect blurriness. Notably, we track the trend of LV over optimization iterations, rather than its absolute value, to assess whether the image is losing sharpness. In our first finding, we observed that during regime (I) the generated images are realistic and sharp. Therefore, as the optimization progresses, we can detect when intermediate reconstructions begin to lose sharpness by using the variance of the Laplacian of the reconstructed image as a sharpness score. In Figure 6, we illustrate the trend of the variance of the Laplacian (LV) operator throughout our optimization process. In the case of a blurry image, aside from a transient phase in the initial iterations, the variance steadily decreases as the image becomes less sharp, indicating a transition to overfitting. We define a minimum number of iterations k_{min} to avoid the transitory regime and we keep track of the iteration number that corresponds to the last peak of sharpness. If the iteration $k > k_{min}$ and $\sigma^2[k+1] < \sigma^2[k]$, where $\sigma^2[k] \doteq LV(g(z^k))$, we stop the optimization. In this case, the optimal reconstruction is not necessarily at the current iteration (k), rather the one with the last highest sharpness (*i.e.*, the highest LV).

3.4.2 Degradations adding high-frequency content

In the case of degradations that introduce high-frequency artifacts, such as noise (as shown in Figure 3), the reconstructed image quickly captures the correct structure and the loss drops significantly during the initial iterations. For example, in Figure 3 the reconstructed face rapidly converges to the correct identity. However, due to the model's slow convergence to high-frequency artifacts, the loss takes a long time to approach zero. This behavior motivates the use of the normalized loss slope, *i.e.*, the change in loss normalized by the loss itself as a criterion for detecting resistance to high-frequency artifacts like noise. In Figure 4, we display the *normalized slope*

$$\Delta_k = \frac{E(z^k; y) - E(z^{k-1}; y)}{E(z^{k-1}; y)}, \tag{4}$$

where k is the iteration number, in log scale (a) and the PSNR trend (b) relative to the clean reference image for input images with varying noise levels. We can immediately notice that the locations at which the slope Δ in (a) reach a minimum correspond to the maxima of the reconstruction in (b). Thus, a self-supervised stopping criterion based on

Algorithm 1 DIIP

Require: Degraded image y, g,η, k_{min}, ϵ
Ensure: Restored image $\hat{x}$
$k = 0$, $LastSharpnessPeakIter = 0$, $detected_H$ = **false**, $detected_L$ = **false**
while NOT $detected_H$ **and** NOT $detected_L$ **do**
  $z^{k+1} = z^k - \eta \nabla_z \|g(z^k) - y\|^2$
  Compute the Laplacian variance $\sigma^2[k]$ of $g(z^k)$
  //keep track of the iteration with last sharpness peak
  if $\sigma^2[k] < \sigma^2[k-1]$ and $\sigma^2[k-2] < \sigma^2[k-1]$ **then**
    $LastSharpnessPeakIter = k - 1$
  end if
  //Low frequency degradation stopping criterion
  if $k > k_{\text{min}}$ and $\sigma^2[k] < \sigma^2[k-1]$ **then**
    $detected_L$ = **true**
    $n^* = LastSharpnessPeakIter$
  end if
  //High frequency degradation stopping criterion
  if $k > 0$ and $\Delta_k < \epsilon$ **then**
    $detected_H$ = **true**
    $n^* = k$
  end if
end while
Return the restored image $\hat{x} = g(z^{n^*})$

the normalized slope of the loss can effectively detect when to halt optimization, maximizing reconstruction quality before overfitting occurs. For simplicity and robustness, our stopping criterion $detected_H$ is triggered (*true*) when the normalized loss decrease is below a threshold ϵ. Formally, when $\Delta_k < \epsilon$, we stop and return the reconstruction $g(z^k)$.

3.5. Restoration via the DIffusion Image Prior

Based on previous work, we propose a simple algorithm for image restoration. Given a degraded image y, we perform the optimization defined in Eq. (3) using gradient descent with a randomly initialized latent code z. At each iteration, we take a gradient step and check whether either of the two stopping criteria is satisfied. One criterion targets high-frequency artifacts, and the other targets low-frequency artifacts. To handle low-frequency artifacts, we monitor the Laplacian variance over time. If a minimum number of iterations has been reached and the Laplacian variance is still decreasing, we stop the optimization and return the intermediate prediction corresponding to the last local peak in the Laplacian variance. To address high-frequency artifacts, we track the loss reduction. Specifically, if the normalized slope of the loss falls below a predefined threshold ϵ, we terminate the optimization. The full procedure is summarized in Algorithm 1.

Table 1. Effect of k_{min} on DIIP performance. PSNR (dB) is reported when using different k_{min} values.

k_{min}	Non-uniform deformation	Waterdrop removal
50	22.18	22.48
100	23.45	23.78
150	23.52	23.82

Table 2. Effect of ϵ on DIIP performance. PSNR (dB) is reported when using different ϵ values.

ϵ	Denoising	JPEG-deartifacting
0.005	27.25	22.38
0.001	28.37	25.29
0.0005	28.14	25.02

Table 3. Performance gap when using optimal stopping. PSNR (dB) is reported for our stopping criterion and an optimal stopping criterion that has access to ground-truth data.

	Denoising	Water-drop removal
Optimal stopping	28.63	23.73
Ours	28.37	23.45

4. Experiments

Experimental settings. We evaluate DIIP across various image restoration (IR) tasks. For quantitative evaluation, we conduct experiments on both structured degradations (*e.g.*, denoising, super-resolution) and complex unstructured degradations (*e.g.*, non-uniform deformation, water drop removal). We evaluate our method on images from ImageNet 1K [7] and CelebA 1K [13] at a resolution of 256×256 pixels. For CelebA 1K, we use the model pre-trained on CelebA by [14]. For ImageNet 1K , we use the model pre-trained on ImageNet by [8]. To ensure a fair comparison, the same pre-trained models are used across all methods. Performance is measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) for fidelity, and Learned Perceptual Image Patch Similarity (LPIPS) for perceptual quality. DIIP is compared against state-of-the-art zero-shot and partially-blind methods, including BlindDPS [6], GDP [9], BIRD [3], and GibbsDDRM [15]. Additionally, we compare against fully blind zero-shot methods such as DIP [17] and the recent state-of-the-art DreamClean [20]. We emphasize that only a few methods have been proposed for unstructured degradations (*e.g.*, water drop removal), which explains the smaller number of baselines in Table 5 compared to Table 4. For super-resolution, we apply an 8×8 Gaussian blur followed by $8\times$ downsampling. For denoising, we use a mixture of Gaussian and speckle noise with $\sigma \approx 0.3$. JPEG artifact removal is performed using the

Figure 5. Results of blind restoration methods applied to structured degradations. Top row: Denoising. Bottom row: 4× super-resolution.

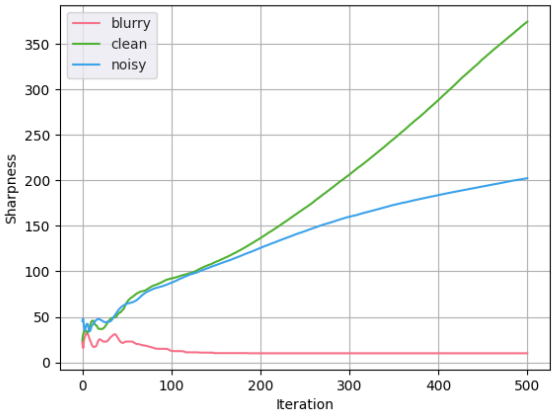


Figure 6. **Stopping criterion for degradations that remove high-frequency details from the image**. Here we show the trend of the variance of the Laplacian operator throughout our optimization scheme. After the first regime, in case of blurry image the variance of the Laplacian keeps decreasing while in the case of a sharp image the optimization adds more high frequency details and the variance of the Laplacian does not decrease. A similar trend to the clean image is observed for a noisy image, which shows that a separate stopping criterion is needed.

ImageIO [1] Python library with a quality factor $q = 5$. All tasks include additive Gaussian noise with $\sigma = 0.02$. For generating unstructured degradations, we use an online tool[1] to create degraded inputs. We set $k_{min} = 100$ and $\epsilon = 0.001$. Optimization is performed using the Adam optimizer with a learning rate of 0.0015.

Quantitative and qualitative comparison. Tables 4 and 5 show the performance of DIIP compared to other methods in both the partially blind and fully blind settings. DIIP consistently outperforms or matches the state-of-the-art approaches across different image restoration tasks. It is important to note that methods such as [3, 6, 9, 15] are not applicable in the fully blind scenario, as they require knowledge of the parametric form of the degradation model, and therefore are not included in Table 5. Figure 5 presents a visual comparison of CelebA images restored for image denoising and super-resolution tasks. Figure 7 showcases a visual comparison for unstructured restoration tasks. Despite lacking prior knowledge of the degradation, DIIP produces realistic reconstructions and generally preserves higher fidelity compared to other competing methods.

[1] https://online.visual-paradigm.com/

5. Ablations

We conduct different ablations to analyze the impact of each component of DIIP. Specifically, we examine the influence of the two hyperparameters, k_{min} and ϵ, in Algorithm 1.

Effect of the minimum number of iterations k_{min}. In Table 1, we present the effect of the minimum iteration k_{min} on DIIP. After approximately $k_{min} = 100$, the performance stabilizes, making it a good trade-off between image quality and efficiency.

Effect of the threshold ϵ. In Table 2, we show the effect of the threshold ϵ. A higher ϵ causes earlier stopping than optimal, resulting in lower PSNR. Conversely, a very small ϵ allows the process to continue longer, leading to high-frequency artifacts. We found that $\epsilon = 0.001$ yields the best performance.

Gap to the optimal stopping. In Table 3, we quantify the gap between our stopping criterion and the optimal one, assuming access to the clean target image. To determine the optimal stopping point, we run our optimization and stop when PSNR with respect to the ground truth stops improv-

Table 4. Quantitative comparison with training free and zero-shot blind zero-shot methods on structured IR tasks on the CelebA validation dataset. The best method is indicated in bold.

Method	Denoising			Superresolution (×4)			Superresolution (×8)		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
GDP [9]	27.73	0.817	0.232	24.21	0.708	0.337	21.66	0.618	0.374
Gibbsddrm [15]	27.38	0.809	0.255	24.38	0.689	0.330	21.45	0.605	0.364
BIRD [3]	27.92	0.821	0.238	**25.26**	0.751	**0.294**	22.63	0.626	0.352
BlindDPS [6]	27.56	0.813	0.246	24.51	0.722	0.324	21.73	0.620	0.360
DIP [17]	25.81	0.606	0.345	21.33	0.566	0.426	20.34	0.488	0.471
DreamClean [17]	27.05	0.771	0.236	23.44	0.663	0.322	21.33	0.586	0.344
Ours	**28.37**	**0.842**	**0.224**	25.14	**0.764**	0.301	**22.86**	**0.651**	**0.336**

Table 5. Quantitative comparison with training free and zero-shot blind methods on unstructured IR tasks. We note that[3, 6, 9, 15] could not be applied to unstructured degradations. The best method is indicated in bold.

Method	JPEG De-artifacting			Non-uniform Deformation			Water-drop Removal		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
DIP [17]	20.43	0.593	0.622	18.83	0.437	0.643	20.37	0.517	0.642
DreamClean [20]	23.92	0.691	0.342	22.16	0.612	0.398	22.94	0.643	**0.361**
Ours	**25.29**	**0.783**	**0.325**	**23.45**	**0.689**	**0.392**	**23.78**	**0.702**	0.377

Table 6. Runtime (in seconds) and Memory consumption (in Gigabytes) comparison of training-free methods on CelebA. The input image is of size 256×256.

Method	Runtime [s]	Memory [GB]
GDP [9]	168	1.1
BIRD [3]	234	1.2
BlindDPS [6]	270	6.1
DreamClean [20]	125	1.3
Ours	138	1.2

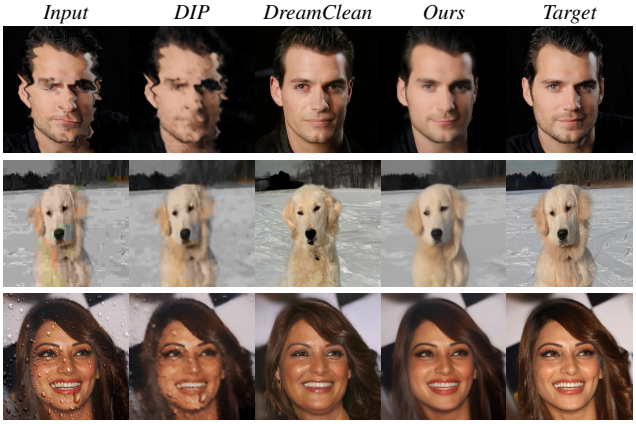


Figure 7. Visual comparison of blind restoration methods on tasks with unstructured degradations: non-uniform deformation (top), JPEG de-artifacting (middle), and water-drop removal (bottom).

ing. DIIP lags behind this optimal point by approximately 0.3 dB in the denoising and water drop removal tasks.

Efficiency comparison. In Table 6, we compare the runtime and memory consumption of training-free methods. We run DIIP on 100 degraded images with different degradation types and report the average runtime. Our method achieves a good balance between image quality and computational efficiency while offering a broader range of applicability than most existing methods.

6. Conclusion

IIn this work, we presented DIffusion Image Prior (DIIP), a novel blind image restoration method that leverages pretrained diffusion models to handle a wide range of degradation types without requiring explicit knowledge of the degradation process. Drawing inspiration from Deep Image Prior (DIP), we showed that pretrained diffusion models offer a much stronger prior for restoration tasks, enabling the reconstruction of clean images even under complex and unknown degradations. Our experiments demonstrated that the optimization process in DIIP consistently produces high-fidelity restorations across various degradation types, including JPEG artifact removal, waterdrop removal, non-uniform deformation, and super-resolution. By incorporating an early stopping strategy, we effectively prevented overfitting to the degraded input, further enhancing restoration quality. Overall, DIIP achieves competitive performance in both restoration quality and robustness, offering a practical solution for real-world image restoration scenarios where degradation models are unknown or difficult to specify.

Acknowledgments We acknowledge the support of the SNF project number 200020 200304.

References

[1] Parag Chandra and Luis Ibanez. Imageio: design of an extensible image input/output library. *XRDS: Crossroads, The ACM Magazine for Students*, 7(4):10–15, 2001. 7

[2] Jesse T Chao, Calvin D Roskelley, and Christopher JR Loewen. Maps: machine-assisted phenotype scoring enables rapid functional assessment of genetic variants by high-content microscopy. *BMC bioinformatics*, 22(1):202, 2021. 5

[3] Hamadi Chihaoui, Abdelhak Lemkhenter, and Paolo Favaro. Blind image restoration via fast diffusion inversion. *arXiv preprint arXiv:2405.19572*, 2024. 2, 4, 6, 7, 8

[4] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 2

[5] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. In *Advances in Neural Information Processing Systems*, 2022.

[6] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6059–6069, 2023. 2, 6, 7, 8

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[8] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 6

[9] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9935–9946, 2023. 2, 6, 7, 8

[10] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 4

[11] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022. 2

[12] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 2

[13] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August*, 15(2018):11, 2018. 6

[14] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022. 2, 6

[15] Naoki Murata, Koichi Saito, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano Ermon. Gibbsddrm: A partially collapsed gibbs sampler for solving blind inverse problems with denoising diffusion restoration. In *International conference on machine learning*, pages 25501–25522. PMLR, 2023. 2, 6, 7, 8

[16] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 3, 4

[17] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. 1, 3, 6, 8

[18] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 2

[19] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. *arXiv preprint arXiv:2212.00490*, 2022. 2

[20] Jie Xiao, Ruili Feng, Han Zhang, Zhiheng Liu, Zhantao Yang, Yurui Zhu, Xueyang Fu, Kai Zhu, Yu Liu, and Zheng-Jun Zha. Dreamclean: Restoring clean image using deep diffusion prior. In *The Twelfth International Conference on Learning Representations*, 2024. 1, 2, 6, 8