

# It Takes Two to Better Denoise

Yongqian Peng\*  
2100017731  
Yuanpei College

Yujie Hou  
2300013108  
School of EECS

Yue Tan  
2200012743  
School of EECS

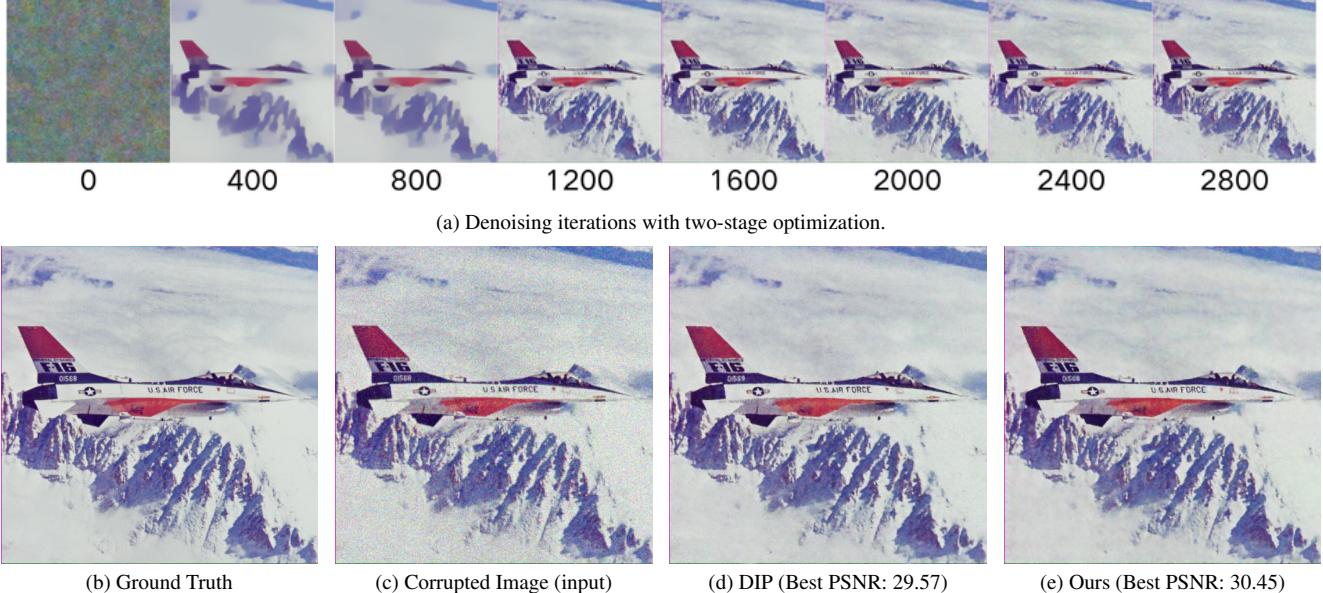


Figure 1. **Qualitative Results: DIP vs. Two-Stage Optimization on the F16 Image.** Fig. 1a illustrates the iterative progression of the two-stage optimization process. Comparisons of key results are presented in Figs. 1b to 1e, which include: the ground truth image, the corrupted input image used for denoising, the output generated by DIP, and the output produced by our proposed two-stage optimization process.

## Abstract

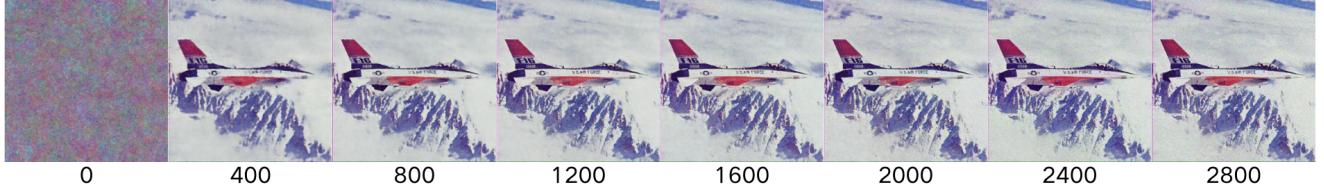
Inverse image reconstruction tasks have been long studied in computer vision. It shows its particular advantages in domains where data is noisy or insufficient, such as medicine and astronomy. Deep Image Prior (DIP) is one of the methods for these tasks, leveraging the structure of convolutional neural networks as hand-crafted prior without requiring sufficient training. This paper delves into replicating and extending the foundational work on DIP. We replicate results across key tasks and explore the challenges of balancing noise removal with the reconstruction of high-frequency details. To address these challenges, we propose a two-stage optimization process combining DIP with Total Variation (TV) regularization. Our method disentangles the denoising and detailing processes, improving denoising while preserving fine details. Results show significant improvement in image restoration tasks, achieving high-quality outcomes on both standard datasets and real-world images.

*These findings highlight the potential of enhanced DIP techniques for broader applications in image restoration.*

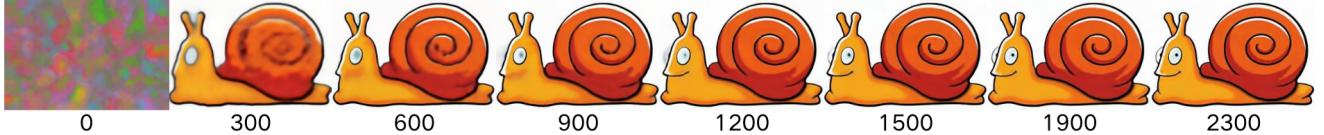
## 1. Introduction

Inverse image reconstruction problems have long been a focal point of research in computer vision, addressing tasks such as denoising, inpainting, and super-resolution [1, 3, 8, 10, 11]. Traditional approaches have evolved significantly, with learning-based methods demonstrating remarkable capabilities in restoring degraded images [2, 6]. However, in fields such as medicine and astronomy, image data often suffer from noise and scarcity.

To address these challenges, researchers have explored alternative methods that do not require large training datasets [4, 9, 10, 12]. One promising approach is the Deep Image Prior (DIP) [10], which utilizes the inherent structure of convolutional neural networks as a prior for image restoration. This method effectively eliminates the ne-

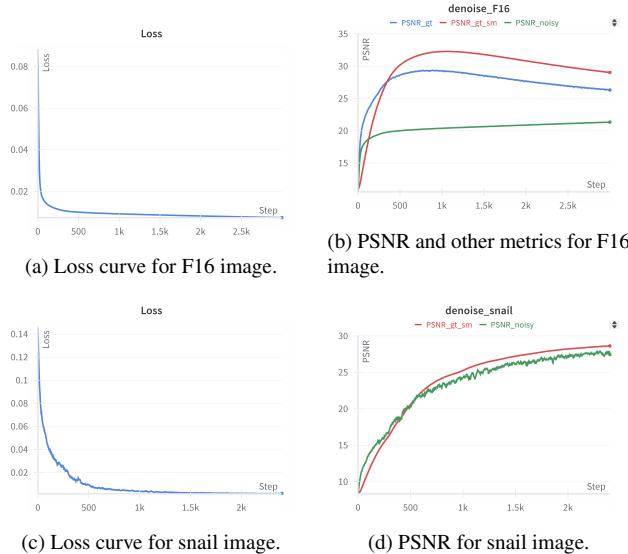


(a) **Visual and quantitative results of the denoising process for the F16 image.** The images show progressive noise reduction, while the PSNR graph indicates a significant initial increase, stabilizing as the model converges. The loss curve demonstrates rapid initial learning, tapering off as the process continues.



(b) **Snail denoising results.** This image sequence illustrates the effectiveness of the denoising algorithm on a different subject. Similar trends in PSNR and loss curves are observed, highlighting the model's consistent performance across varied inputs.

**Figure 2. Denoising results for different images.** The F16 (top) and snail (bottom) images demonstrate the denoising process's effectiveness and consistency, with both visual improvements and quantitative metrics indicating successful noise reduction.



**Figure 3. Visual and quantitative results of the denoising process for the F16 and snail images.** The F16 (top) and snail (bottom) images show different convergence behaviors in terms of PSNR and loss curves, indicating the effectiveness of the denoising process and the nature of the noise in each image.

cessity for extensive training on external data, providing a novel solution for enhancing image quality in these domains.

Although DIP can handle various types of image restoration, its performance still lags behind state-of-the-art training-based methods. One notable limitation is the difficulty in balancing noise removal with the reconstruction of high-frequency details. This issue is inherent to the nature of DIP, which relies on early stopping during training on a single image for denoising. With fewer iterations, the model effectively produces a denoised but blurred image.

However, as the number of iterations increases, the reconstruction of fine textures is accompanied by the reintroduction of noise, compromising the overall image quality (see Fig. A5).

Inspired by the work of Mataev et al. [7], we incorporate an additional regularization stage into the image denoising process. The key insight is that the current DIP method iteratively restores an image from Gaussian noise, simultaneously denoising and reconstructing details in a unified process. To disentangle these two processes, we design a two-stage optimization for image restoration: denoising and detailing, which aim to further compress the noise and enhance image quality.

In sum, we delve into the foundational work presented in the original DIP paper, meticulously replicating their results across key tasks: denoising, inpainting, and super-resolution. Furthermore, we comprehensively analyze the results in the original paper and extend these methods to apply them to our images, exploring the potential and limitations of DIP in diverse real-world scenarios. In addition, we propose a two-stage optimization process, equipping DIP with an additional regularization stage, which improves the results of denoising.

## 2. Related Work

### 2.1. Deep Image Prior

Image restoration tasks require models capable of reconstructing images with lost or corrupted data back to the original images with high quality. Traditional methods, such as conventional learning-based methods, often struggle with complex and unknown degradation processes as they rely on extensive training datasets, which may not always be feasible [4, 9, 12].

## 2.2. Regularization Denoising (RED)

To improve the lagging performance of DIP compared to the state-of-the-art methods on several image reconstruction problems, reconsidering the explicit regularization in the optimization process to enrich the implicit regularization is one of the feasible methods. This method uses Total-Variation (TV) regularization to denoise the images [5].

## 3. Data

We use the same data set provided in the paper and we also experiment on our own images taken in real-world scenarios or found on the Internet. Our work has successfully replicated the results of DIP, achieving similar performance for denoising, super-resolution, and inpainting on both the provided dataset and our images, as measured by Peak Signal-to-Noise Ratio (PSNR) and qualitative comparisons.

## 4. Methods

DIP proposes a unique method that uses untrained ConvNets for image restoration and fits a generator network to a single corrupted image, leveraging their intrinsic structures as priors. DIP uses the networks' structure as a handcrafted image prior, which can effectively capture low-level image statistics without the need for extensive prior learning, thus bridging the gap between learning-based and learning-free methods.

DIP employs a deep convolutional network with a U-Net type “hourglas” architecture with skip-connections. Formally, this architecture is designed to map a random code vector  $Z$  to an image  $X$  without any pre-training on data. Then we formulate the image restoration tasks as energy minimization problems, where the goal is to minimize a data term  $E(x; x_0)$  and a regularizer  $R(x)$ . The implicit prior from a single input image captured by the network is then defined as an indicator function that allows the network to produce images from the random code vector. The network parameters are updated using task-specific loss functions (*e.g.*, MSE for denoising). For the optimization process, the parameters of the network are optimized by using gradient descent, starting from a random initialization. The optimization aims to find parameters  $\theta^*$  that reproduce the target image  $X_0$ .

Beyond DIP, we combine DIP with Denoising Regularization (RED) in order to improve the results of denoising. We divide the optimization process of DIP into two stages. In the first stage, we mainly focus on removing the noise in the corrupted images more effectively. We implement the TV regularization in the denoising process and compute the TV Loss of the images. The optimization process turns into:

$$\min_u \frac{1}{2} \|u - u_0\|_2^2 + \lambda TV(u)$$

, where  $u_0$  is the image with noise,  $u$  is the denoised image and  $\lambda$  is the regularization parameter controlling the trade-off between the fidelity and smoothness of the data. The TV Loss is implemented as:

$$TV(u) = \sum_{i,j} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|$$

In the second stage, we reuse the DIP process to restore a clean image from the denoised image beforehand. We hope to generate cleaner images by pre-processing the noise in the input images to maximize the reconstruction performance of DIP.



(a) Portrait inpainting results after 6000 iterations.



(b) Vase inpainting results after 3000 iterations.



(c) Library inpainting results after 5000 iterations.

**Figure 4. Inpainting results using Deep Image Prior across three different scenes.** Portrait (top), vase (middle), and library (bottom). Each row shows the original image, the corrupted image, and the result of inpainting, demonstrating the method's ability to fill the mask with fine details.

## 5. Replicate Results

We have replicated the results of the paper using the provided data and obtained the following images and curves. Our results reveal that the DIP method indeed holds promise in various image restoration tasks. Specifically, we focused on three key areas: denoising, inpainting, and super-resolution. Below, we present a summary of our findings for each task.

### 5.1. Denoising

In the denoising task, we tested the DIP approach on several noisy images from the provided dataset. Our experiments

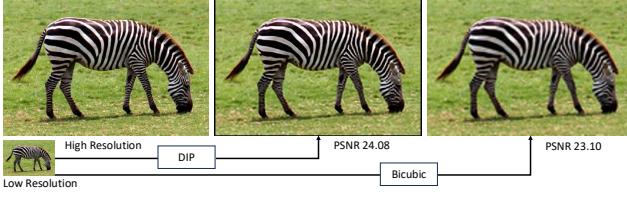


Figure 5. Super-resolution results of the Zebra image.

showed that DIP effectively reduced noise while preserving important image details (see Fig. 2). The reconstructed images demonstrated comparable quality to those reported in the original paper, achieving high PSNR values (see Fig. 3). The qualitative results confirmed that DIP could capture the underlying structure of the images.

## 5.2. Inpainting

For the inpainting task, we evaluated DIP’s ability to fill in missing parts of images. The method performed well, reconstructing missing regions and maintaining coherence with the surrounding image content. Our results were consistent with the original paper, showing that DIP could handle various inpainting scenarios, including large missing areas (see Fig. A1, Fig. 4).

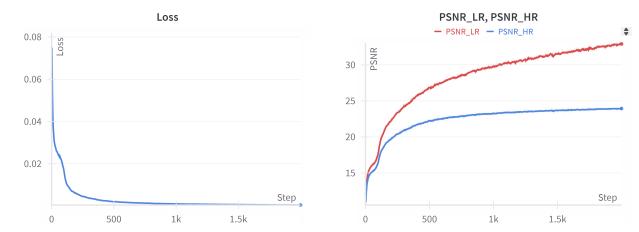
## 5.3. Super-Resolution

In the super-resolution task, DIP was applied to enhance the resolution of low-resolution images. Our experiments confirmed that DIP could effectively increase image resolution while maintaining visual fidelity. The results were in line with those of the original study, demonstrating that DIP could recover fine details and produce sharp, high-resolution images.

# 6. Applying for our images

## 6.1. Denoising

To thoroughly assess DIP, we manually constructed a complex scenario, as depicted in Fig. A4a. This setup features a predominantly dark-light contrast with a small highlighted area. Additionally, it includes numerous characters in both English and Chinese, allowing us to evaluate the effectiveness of the denoising algorithm in handling such elements (see Fig. 7a). Notably, we captured these photos at different ISO settings to introduce varying noise levels, while adjusting the shutter speed to maintain consistent brightness. We also conduct our experiments on images generated by the text-to-image model. The results reveal that: (i) The greater the noise present in the original image, the more effective the denoising process. (ii) There is a significant difference between artificial noise and the noise inherent in camera-captured images. DIP performs better when dealing with artificial noise, indicating that natural noise is more complex. (iii) Restoring text is particularly challenging for DIP.



(a) Loss curve of the super-resolution process. (b) PSNR curve of the super-resolution process.

Figure 6. **Curves of the super-resolution process.** The left sub-figure shows the loss curve, while the right subfigure presents the PSNR curve, illustrating the performance and effectiveness of the super-resolution process.

## 6.2. Inpainting

For our inpainting experiments, we selected an image we captured ourselves. This image features a person on a road with a mask covering him. This image is notable for its diverse frequency composition. The mountains in the background represent high-frequency areas due to their intricate details, while the road and the sky are characterized by low-frequency areas with smoother textures. Additionally, the red and blue color areas contribute to the image’s complexity, making it ideal for testing our inpainting techniques. Our results indicate that the DIP model faces challenges in balancing the inpainting of the mask with minimal artifacts while also preserving the details in high-frequency areas. This suggests that the model struggles to balance avoiding inpainting artifacts and restoring details of the original image, highlighting an area for potential improvement in inpainting algorithms.

## 6.3. Super-resolution

In our super-resolution experiments, we selected an image of a flying bird that we filmed ourselves. Images of wildlife, particularly those of fast-moving and distant animals, usually require using super-resolution algorithms due to their small size in the image plane. Our results indicate that the DIP method performs notably well, surpassing the results of traditional bicubic interpolation.

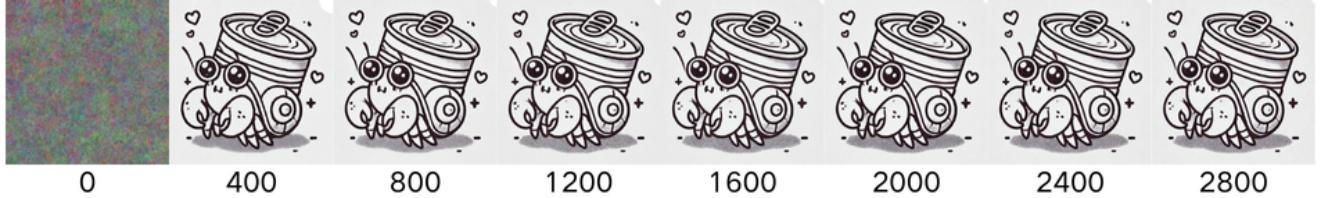
# 7. Extension and Proposed Improvements

## 7.1. Revisiting Limitations of Deep Image Prior

While the DIP method shows impressive performance across multiple image restoration tasks, it suffers from several notable limitations. First, the reliance on early stopping is one of the drawbacks. This approach introduces subjectivity and variability in determining the optimal stopping point. Moreover, the single-image optimization process can lead to the reintroduction of noise in the reconstructed image as the number of iterations increases. Another limitation is that DIP struggles to handle complex degradation

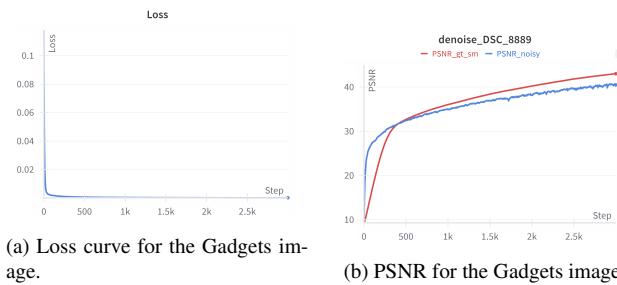


(a) **Denoising results of the Gadgets image.** The images show progressive noise reduction, showing that DIP works well in a real-world situation.

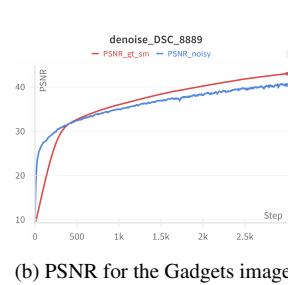


(b) **Denoising results of the Herbitcrap image.** This image sequence illustrates the effectiveness of the denoising algorithm on a more complicated animation subject than the Snail image.

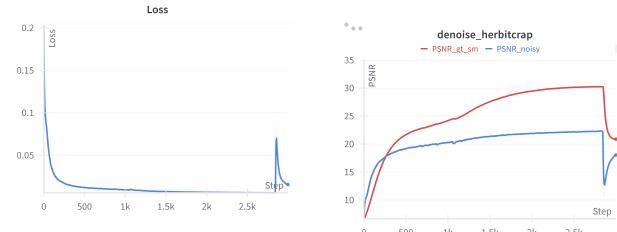
**Figure 7. Denoising results for different images of our own dataset.** The Gadgets Fig. 7a and Herbitcrap Fig. 7b images demonstrate the denoising process's effectiveness and consistency, with both visual improvements and quantitative metrics indicating successful noise reduction.



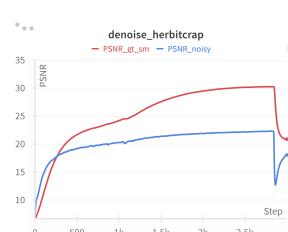
(a) Loss curve for the Gadgets image.



(b) PSNR for the Gadgets image.



(c) Loss curve for the Herbitcrap image.



(d) PSNR for the Herbitcrap image.

**Figure 8. Visual and quantitative results of the denoising process on our dataset.** While the Gadgets Fig. 8a and Herbitcrap Fig. 8c images show different convergence behaviors in terms of PSNR and loss curves, the curves still indicate the effectiveness of the denoising process.

scenarios or images with intricate textures, where it fails to balance noise removal with detail preservation effectively.

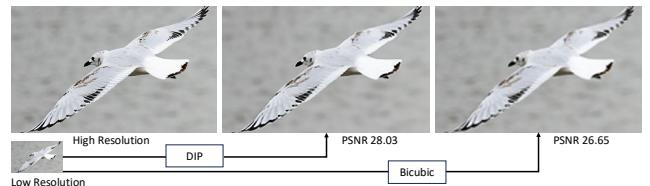
## 7.2. Two-Stage Optimization Process

To address the limitations of traditional DIP methods, we propose a **two-stage optimization framework** that separates the tasks of denoising and detailing.

In the first stage, we focus on effective noise suppression by integrating **Total Variation (TV) regularization** into



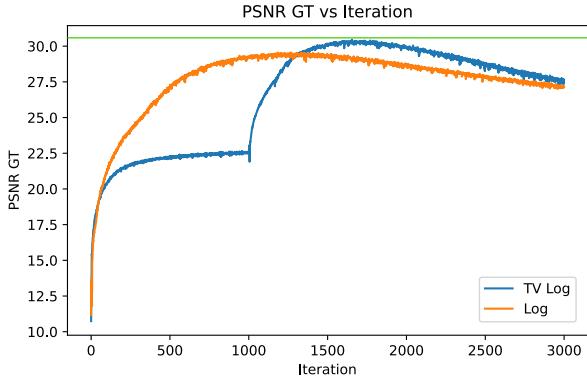
**Figure 9. Inpainting results of the Scenario image after 3000 iterations.** This row shows the original image, the corrupted image, and the result of inpainting, demonstrating the method's ability to fill the mask with fine details while showing that DIP lacks the ability to adjust the restoration between high and low-frequency regions effectively.



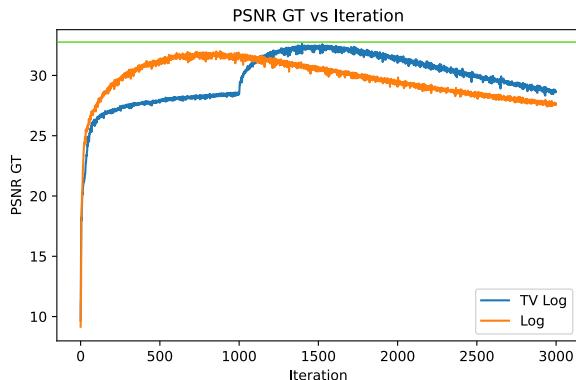
**Figure 10. Super-resolution results of the bird image.**

the DIP framework. TV regularization enhances smoothness while preserving sharp edges, effectively reducing the risk of overfitting to noise. During the initial phase of the optimization process (first one-third of the iterations), TV regularization is applied to prioritize noise removal. This ensures the elimination of artifacts without compromising significant structural details in the image.

After completing the denoising stage, we remove the regularization and pass the denoised image into the standard



(a) The PSNR curve of denoising on the F16 image.



(b) The PSNR curve of denoising on the Gadgets image.

**Figure 11. Comparison of the PSNR for DIP and DIP with TV regularization of the denoising process.** The results of both the F16 and the Gadgets images show that DIP with TV regularization has significant differences in the pattern of PSNR. The PSNR of the original DIP reaches a peak within 1000 iterations while DIP with TV regularization has a boost in performance after 1000 iterations and achieves higher PSNR in the end. This indicates the effective improvement using TV regularization on DIP.

DIP optimization process. This second stage emphasizes refining high-frequency details and restoring textures, enabling the model to recover intricate patterns and features.

By disentangling the tasks of denoising and detailing, this targeted approach mitigates the risk of reintroducing noise and ensures the preservation of fine details in the reconstruction process.

### 7.3. Results and Observations

Our experiments demonstrate that the proposed two-stage optimization process significantly enhances the performance of DIP in denoising tasks. Quantitatively, as shown in Fig. 11, the PSNR values for images processed using the two-stage method stabilize during the first third of the iterations without any drop. After this point, the PSNR values increase rapidly, surpassing the highest values achieved by the original DIP method. Qualitatively, as illustrated

in Fig. 1 and Fig. A4d, the first stage effectively denoises the images, while the second stage further restores details. These results highlight the effectiveness of integrating explicit regularization with DIP, paving the way for its application in more complex and high-demand scenarios.

### 7.4. Future Directions

Building on these improvements, future work could explore adaptive regularization techniques that dynamically adjust the regularization parameters based on image characteristics. Additionally, integrating learning-based priors with DIP could enhance its ability to handle diverse types of noise and degradation.

### 8. Limitation

Our two-stage optimization method has several limitations. First, although we find it useful in the denoising process, we find it limited in tasks of inpainting. In the inpainting task, it hurt the inpainting process by over-blurring the areas of the mask, making the detailing process useless in restoring the details of these areas. Second, we did not recognize significant improvement when denoising natural noise, indicating its limitations in dealing with more complex types of noise.

### 9. Conclusion

In this work, we critically examined the DIP framework and its application to image restoration tasks, highlighting its strengths and limitations. While DIP has demonstrated remarkable results, it still struggles to balance denoising and detailing.

To address these issues, we proposed a two-stage optimization process that separates the tasks of denoising and detailing. By incorporating TV regularization in the initial stage, we enhanced the noise suppression capability of DIP. The detailing stage leveraged the standard DIP framework to refine high-frequency details.

Our experiments confirmed the effectiveness of the proposed method, demonstrating significant improvements in both quantitative and qualitative visual results. The disentanglement of noise suppression and detail enhancement not only improved the stability of the DIP framework but also expanded its applicability to more complex image restoration scenarios.

This study underscores the potential of combining explicit regularization techniques with DIP to overcome its inherent limitations. By providing a possible way to enhance the ability of DIP, we hope to contribute to the broader field of image restoration, offering insights and methodologies that can inform the development of more effective solutions.

## References

- [1] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1
- [2] Du Chen, Jie Liang, Xindong Zhang, Ming Liu, Hui Zeng, and Lei Zhang. Human guided ground-truth generation for realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14082–14091, 2023. 1
- [3] Guy Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2024–2036, 1989. 1
- [4] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. Pet image reconstruction using deep image prior. *IEEE transactions on medical imaging*, 38(7):1655–1665, 2018. 1, 2
- [5] Jiaming Liu, Yu Sun, Xiaojian Xu, and Ulugbek S. Kamilov. Image restoration using total variation regularized deep image prior. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 7715–7719. IEEE, 2019. 3
- [6] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 3(8), 2023. 1
- [7] Gary Mataev, Peyman Milanfar, and Michael Elad. Deepred: Deep image prior powered by red. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2
- [8] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003. 1
- [9] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1):86–109, 2019. 1, 2
- [10] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1
- [11] Jin Wang, Yanwen Guo, Yiting Ying, Yanli Liu, and Qunsheng Peng. Fast non-local algorithm for image denoising. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 1
- [12] Hai-Miao Zhang and Bin Dong. A review on deep learning in medical image reconstruction. *Journal of the Operations Research Society of China*, 8(2):311–340, 2020. 1, 2



(a) **Portrait inpainting results.** Starting from random noise, the algorithm gradually reconstructs the portrait, with noticeable improvements at each iteration. By iteration 5600, the image is nearly complete, showcasing the model's capacity to capture complex facial features and textures.

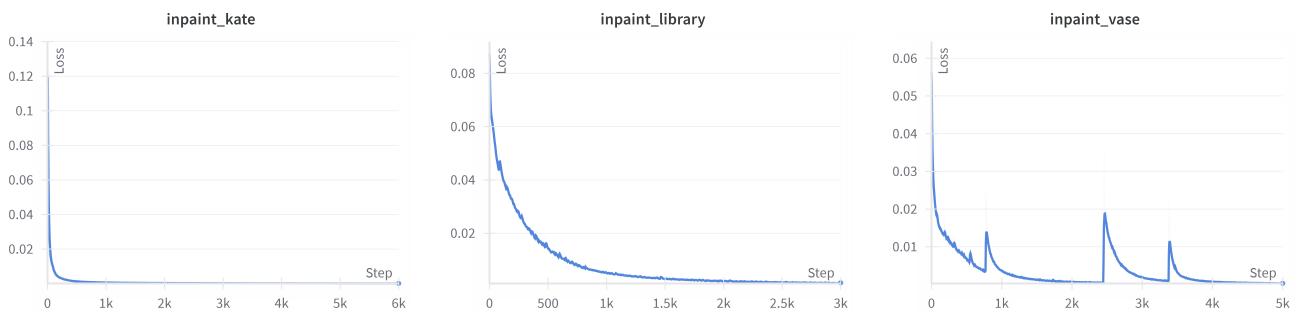


(b) **Library inpainting results.** The restoration of the library scene highlights the method's effectiveness in recreating structured environments. From iteration 0 to 2800, the intricate details of books and architectural elements become increasingly defined, indicating the model's proficiency in handling complex patterns and repetitive structures. However, in the upper-right corner, where a large area of the mask is present in the original image, some artifacts are still noticeable, similar to the results in the original paper.



(c) **Vase inpainting results.** The vase scene further illustrates the Deep Image Prior's ability to reconstruct symmetrical and ornate features. In the early interactions, some artifacts can still be seen at the positions of the masks. Over 4900 iterations, the model adeptly restores the reflective surfaces and intricate designs, emphasizing its strength in managing both texture and symmetry.

**Figure A1. Inpainting results using Deep Image Prior across three different scenes.** Portrait (top), library (middle), and vase (bottom). Each row shows the progressive improvement of the inpainted images over several iterations, demonstrating the method's ability to recover fine details from initial noise.



(a) Loss curve for portrait image.

(b) Loss curve for library image.

(c) Loss curve for vase image.

**Figure A2. Loss curves of the inpainting process for different images.** Comparing all three curves, the portrait image is the easiest to reconstruct, likely due to the small area of the masks in the original inputs. The reconstruction of the library scenario converges more stably than that of the vase.

## inpaint\_vase

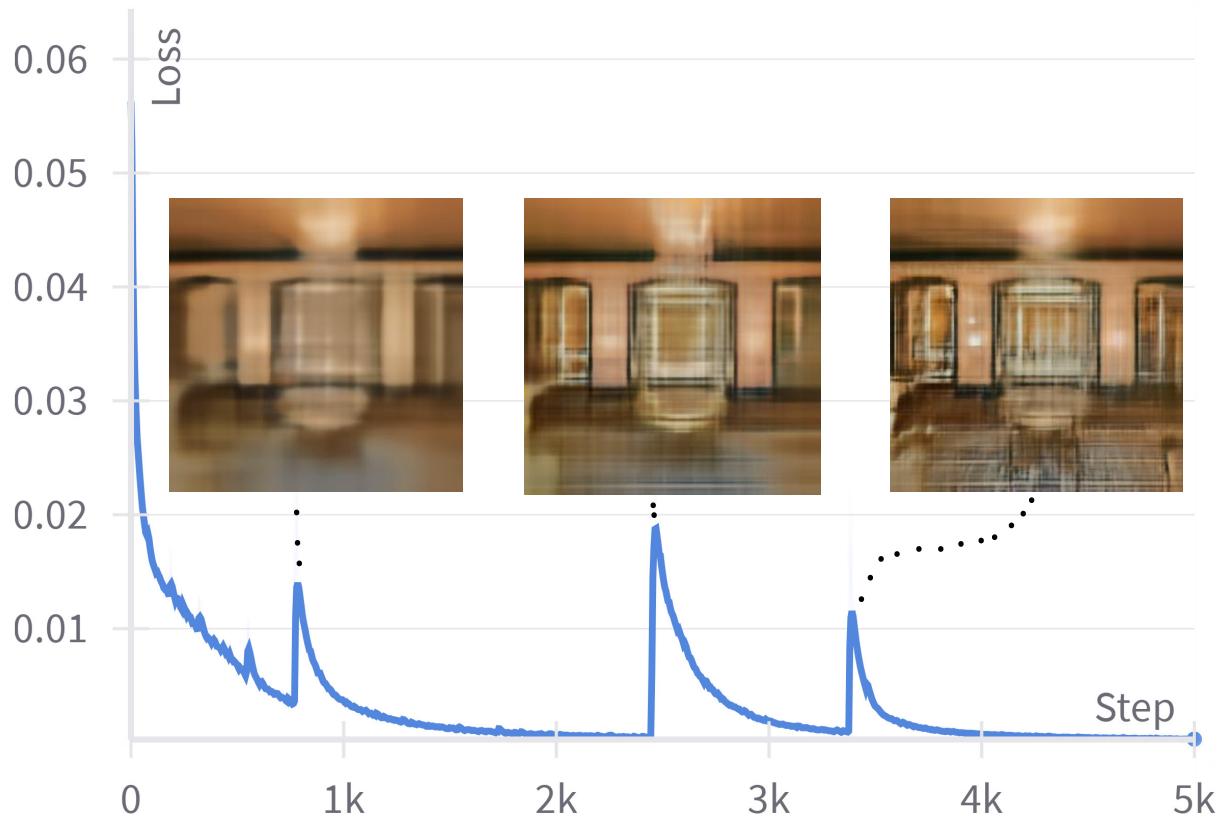


Figure A3. The visualization of the images at spikes of the loss curve.

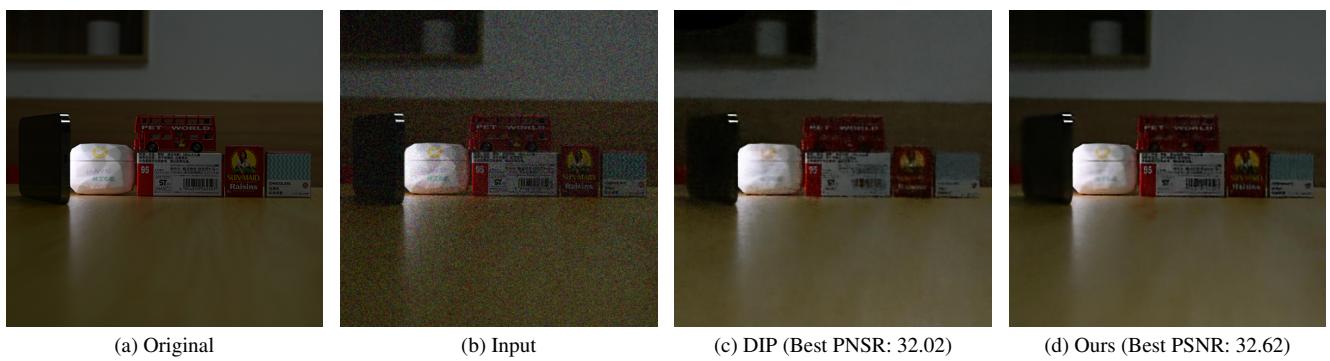


Figure A4. Comparisons of key results are presented in Figs. A4a to A4d, which include: the ground truth image, the corrupted input image used for denoising, the output generated by DIP, and the output produced by our proposed two-stage optimization process.

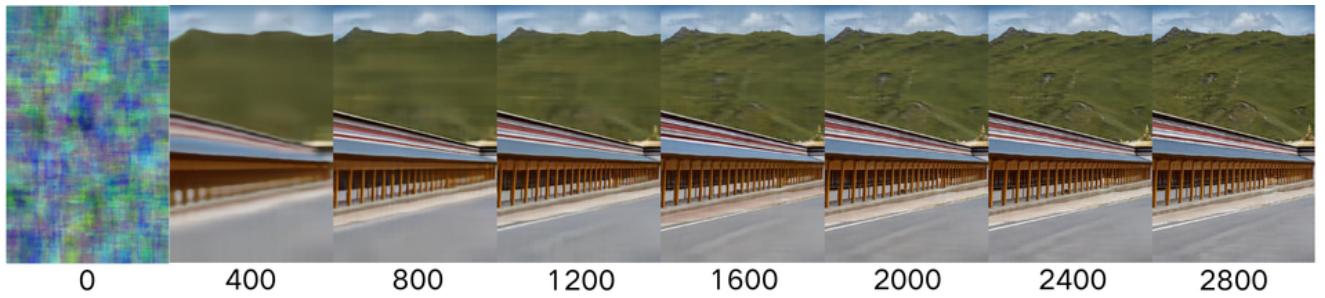


Figure A5. **Inpainting results of the Scenery image.** Starting from random noise, the algorithm gradually reconstructs the scene, with noticeable improvements at each iteration. By iteration 2000, the image is nearly complete, showcasing the model's capacity to capture complex object features and textures in the natural world and on human constructions.