

# Mixed Degradation Image Restoration via Deep Image Prior Empowered by Deep Denoising Engine

Weimin Yuan, Yinuo Wang, Ning Li, Cai Meng\*, Xiangzhi Bai

**Abstract**—Deep Image Prior (DIP) is a powerful unsupervised learning image restoration technique. However, DIP struggles when handling complex degradation scenarios involving mixed image artifacts. To address this limitation, we propose a novel technique to enhance DIP’s performance in handling mixed image degradation. Our method leverages additional deep denoiser, which is deployed as a denoising engine in the regularization by denoising (RED) framework. A new objective function is constructed by combining DIP with RED, and solved by the alternating direction method of multiplier (ADMM) algorithm. Our method explicitly learns a more comprehensive representation of the underlying image structure and being robust to different types of degradation. Experimental results demonstrate the effectiveness of our method, showing effective improvements in restoring images corrupted by mixed degradation on several image restoration tasks, such as image inpainting, super-resolution and deblurring.

**Index Terms**—Mixed degradation, Deep image prior, Regularization by denoising.

## I. INTRODUCTION

This paper begins with the general image degraded model:

$$Y = HX + n, \quad (1)$$

here,  $Y$ ,  $X$  indicate the observed degraded and latent clean images, respectively.  $n$  is additive Gaussian noise.  $H$  is the degradation matrix relating to degradation system. Depending on the specific form of  $H$ , Eq. (1) can be expressed as different restoration tasks. For example, Eq. (1) becomes the image inpainting when  $H$  is a masking matrix; Eq. (1) represents the image super-resolution when  $H$  is a sub-sampling operator; Eq. (1) becomes the image deblurring when  $H$  is convolution operator, and image denoising when  $H$  is an identity matrix.

To cope with the ill-posed issue of Eq. (1), image prior  $\rho(X)$  is adopted as the regularization to constrain the solution space by solving a minimization problem:

$$\hat{X} = \arg \min_X \frac{1}{2} \|Y - HX\|_2^2 + \lambda \rho(X). \quad (2)$$

here  $\lambda$  is regularization factor. The quest for an effective regularization in the past has played an important role in the vast progress of this field. A variety of prior models have been

developed, including non-local self-similarity [8] and sparsity [9].

In recent years, deep learning methods have shifted the attention in this field towards new imaging formulations. Supervised networks learn how to execute image restoration based on labeled training data, and then optimizing degraded image through training procedures. Being task specific, these networks rely heavy on large amounts of training data. However, the range of tasks they are able to handle can be constrained by their reliance on training data, which may difficult or expensive to obtain [10].

Apart from using training sets to perform supervised learning, unsupervised learning has the ability to learn via the neural networks itself, thus extracting information resembling to the input context. Deep Image Prior (DIP) [1], [11] leverages the intrinsic structure of deep convolutional neural networks (CNNs) to restore images without the need for a specific dataset of clean images. Unlike supervised learning methods that rely on large datasets to learn the mapping, DIP relies instead on the internal structure of CNN architecture itself to impose a strong prior on the space of natural images. Rather than directly solving the optimization in Eq. (2) with explicit regularization  $\rho(X)$ , DIP offers an alternative way to implicitly regularize the process of restoration. The core idea behind DIP is to minimize

$$\hat{\Theta} = \arg \min_{\Theta} \|HT_{\Theta}(Z) - Y\|_2^2 \quad s.t. \hat{X} = T_{\hat{\Theta}}(Z) \quad (3)$$

here,  $\Theta$  represents the parameters of CNN,  $Z$  is a randomly initialized noise vector. The optimization process is performed through gradient descent method, iteratively updating  $\Theta$  to minimize  $\|HT_{\Theta}(Z) - Y\|_2^2$ . Given the optimal  $\hat{\Theta}$ , final recovering result  $\hat{X}$  is the output of deep network  $T_{\hat{\Theta}}(Z)$ . DIP has shown impressive results in various image restoration tasks [11]–[17]. The key advantage lies in without relying on specific training datasets, making DIP to be a valuable tool for image restoration with limited or no access to clean data.

However, DIP struggles to effectively handle mixed degradation scenarios where multiple types of degradation are present simultaneously existing. Specifically, image restoration tasks in real scenes, such as image inpainting, super-resolution and deblurring, in addition to their own degradation factors (pixels missing, low resolution and blur, respectively), are inevitably affected by additional noise, which actually belongs to the category of image mixed degradation.

For example, in Fig. 1 (a), after a long period of air oxidation and flow, murals inevitably suffered from information loss and noise interference simultaneously. This means that the murals need to be denoised and inpainting at the same

Corresponding author: Cai Meng.

W. Yuan, Y. Wang, N. Li, C. Meng and X. Bai are with the Image Processing Center, Beihang University, Beijing 100191, China (e-mail: yuanweimin@buaa.edu.cn; wangyinuo@buaa.edu.cn; zy1915215@buaa.edu.cn; Tsai@buaa.edu.cn; jackybxz@buaa.edu.cn).

Acknowledgments: This work has been supported in part by the National Natural Science Foundation of China (Grants Nos. 92148206, 61873010), and Beijing NSFC project (Grant No. 7202103).

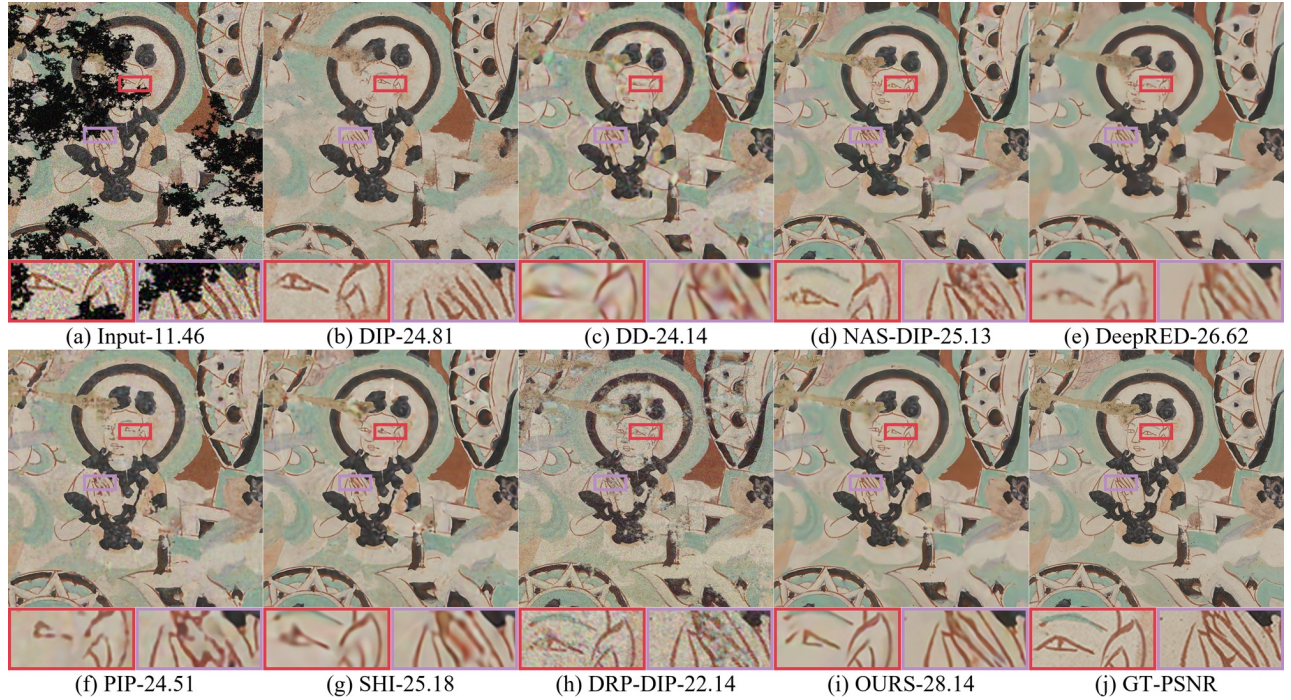


Fig. 1. Visual comparisons of mixed degradation image restoration results produced by different unsupervised learning methods, take Dunhuang historical mural as an example (image inpainting). Aging and influenced by factors like air oxidation, historical mural undergoes a mixed degradation, encompassing information loss and noise contamination simultaneously as shown in (a). (b) DIP [1]-(2018-CVPR), (c) DD [2]-(2018-ICLR), (d) NAS-DIP [3]-(2020-ECCV), (e) DeepRED [4]-(2019-ICCV), (f) PIP [5]-(2022-Arxiv), (g) SHI [6]-(2022-IJCV), (h) DRP-DIP [7]-(2023-CVPR), (i) OURS, (j) Truth. Compared to DIP and its six variants (c-h), our method produces a more naturally looking reconstructed result, with sharp and clean edges, achieving the highest PSNR. For performance on other mixed degraded image restoration tasks, please refer to Section V.

time, thus becoming a challenging mixed degraded image restoration problem. It can be further inferred that during the reconstruction process of image super-resolution or deblurring, it should also not overlook the noise factor. However, both DIP [1] and its variants [2], [3], [5]–[7], [18] ignore the influence of noise in these restoration tasks, and only regard these tasks as a single degradation category. Randomly initialized neural network might lack the specific architectural complexities required to model and represent the intricate relationships between different types of degradation. Continuously optimizing the structural complexity and parameter quantity of the network may not achieve better results. As a result, the network’s capacity may not be sufficient to simultaneously capture and address multiple degradation factors, leading to sub-optimal restoration results or the introduction of visual artifacts, as shown in Figs. 1(b)-(h).

To fill this gap, in this paper, regularization by denoising (RED) framework [19] is introduced as an explicitly regularizer in the optimization process of DIP. The RED terms explicitly encourages the network to learn a more general representation of the underlying image structure. By introducing the pre-trained denoiser FFDNet [20], as denoising engine of RED, this denoising step encourages the network to focus on learning the essential image features while suppressing the impact of noise and artifacts caused by mixed degradations. Our proposed method inherit the unsupervised/data-free property of DIP and is therefore able to adapt to image restoration tasks where training data is not available. Moreover, compared to [4], [18], [21] that directly introducing shallow regularization

terms (such as BM3D [9], TV [22], NLM [8]) to restore a single degraded image, our method is more adaptable and capable of handling the types of mixed degradation driven by pre-trained deep denoiser. As result, our restored images exhibit improved quality and fidelity in situations where mixed degradation has occurred, as shown in Fig. 1 (i).

**Our Contributions.** In summary, the main contributions of this paper are:

- To the best of our knowledge, we take the first step forward in addressing the image restoration task of DIP dealing with mixed degradation. Specifically, we introduce the RED framework, powered by FFDNet, as an extension of DIP to establish our objective function. In order to tackle the resulting optimization problem, we employ the Alternating Direction Method of Multipliers (ADMM), capitalizing on its significant advantages in decomposing the objective function into three simpler sub-problems (see Section III).
- We provide analysis of the feasibility of our method. We verify in detail the three assumptions, that is, the pre-trained deep denoiser FFDNet  $f(\mathbf{X})$  deployed in our RED framework should simultaneously satisfy differentiability, homogeneity and the Jacobian matrix of  $f(\mathbf{X})$  satisfies symmetry. And how these three conditions lead to the RED gradient holds (see Section IV).
- Experimental results have demonstrated the effectiveness of our method by comparing DIP and its various variants on three image restoration tasks (see Section V).

The paper is organized as follows. Section II introduces

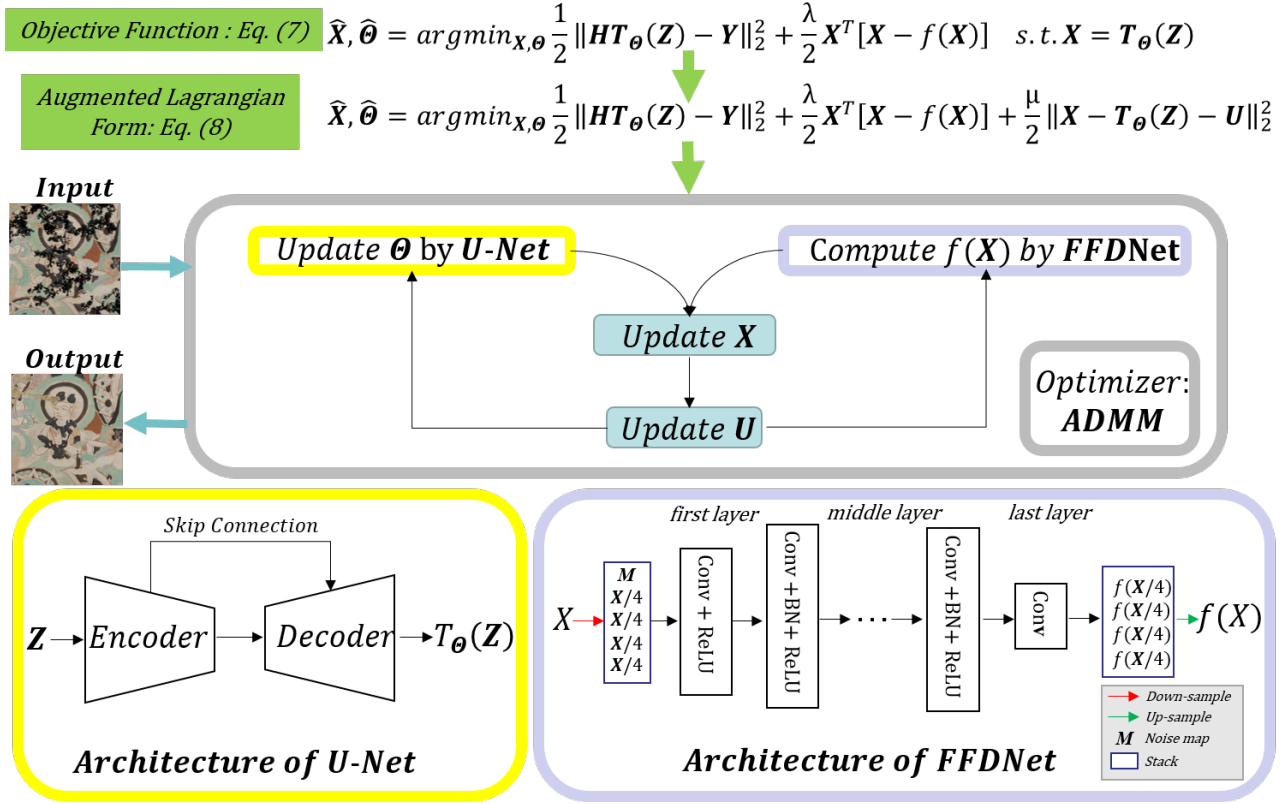


Fig. 2. The flowchart of our method. By incorporating RED, powered by deep denoising engine FFDNet, into DIP, we establish the objective function Eq. (7). The ADMM algorithm transform the problem of augmented Lagrangian form Eq. (8) into updating three sub-problems:  $\Theta$ ,  $X$  and  $U$  sequentially. See Section III for details.

the related work. Section III presents the proposed method. Section IV provides the analysis of the feasibility of our method, followed by experimental validation in Section V. Finally, Section VI concludes this paper.

## II. RELATED WORK

By taking the structure of a randomly initialized network as an image prior, DIP [1], [11] shows that a deep CNN has an intrinsic ability to learn the uncorrupted image. DIP has an edge in real-world scenes because it needs no training data. And the result achieved is on par with some sophisticated technique that depends on a large amount of well-curated data [10]. Such advantages of DIP have attracted widespread attention, and motivated a large number of studies.

For example, Deep decoder (DD) [2] learns to decode noise tensor with no convolutions. By leveraging the CNN architecture search, neural architecture search deep image prior (NAS-DIP) [3] and image-specific neural architecture search deep image prior (ISNAS-DIP) [16] search for neural architectures that capture stronger structured priors than DIP. Frequency-band correspondence measure is introduced in [6] to characterize the spectral bias of DIP, revealing that low-frequency image signals are learned more efficiently than high-frequency ones. From neural implicit representation perspective, positional-encoding prior (PIP) [5] introduces fourier features, allows for the substitution of convolutions with simple multi-layer perception. DRP-DIP [7] freezes randomly-initialized network weights and reducing the network depth to speed up DIP. To

mitigate the negative impact caused by over-fitting, a series of plug-and-play prior schemes, such as TV [21], BM3D [18], NLM [4], IRCNN [23], are incorporated to extend the framework of DIP for better performance.

Different from the above mentioned methods targeted on single degradation image restoration, in this paper, we go one step further and first explore the situation where DIP copes with mixed degradation of images.

## III. PROPOSED METHOD

As mentioned earlier, both DIP [1], [11] and its variants [2], [3], [5]–[7] tend to overlook the impact of noise in the context of image restoration tasks. They often treat these tasks as a single degradation category. Randomly initialized neural networks may lack the specific architectural intricacies necessary to effectively model and represent the intricate relationships among different degradation types. Merely optimizing the network’s structural complexity and parameter count continuously may not lead to improved performance. Consequently, DIP might be inadequate for simultaneously capturing and addressing multiple degradation factors. This limitation will result in unsatisfactory results, see Fig. 1.

In this paper, the performance of DIP under mixed degradation is improved by introducing RED framework as an explicit regularization term. The flowchart of our method is depicted in Fig. 2. The subsequent sections elaborate the details.



### A. Proposed objective function

Compared to forthright employ denoiser into a regularization in plug-and-play [24] (PNP), regularization by denoising [19] (RED) is a radically new way to exploit an image denoising engine. RED embeds an arbitrarily specified denoising engine  $f(\mathbf{X})$  into the regularization for image restoration. The mathematical framework of RED is the inner product between image  $\mathbf{X}$  and its denoising residual  $\mathbf{X} - f(\mathbf{X})$ :

$$\rho(\mathbf{X}) = \frac{1}{2} \mathbf{X}^T [\mathbf{X} - f(\mathbf{X})] \quad (4)$$

here,  $f(\mathbf{X})$  is a chosen denoising engine. Under the following three assumptions of  $f(\mathbf{X})$ :

- **Assumption 1: Differentiability.**  $f(\mathbf{X})$  is differentiable.
- **Assumption 2: Homogeneity.**  $f[(1 + \epsilon)\mathbf{X}] = (1 + \epsilon)f(\mathbf{X})$ ,  $\epsilon$  is a small constant.
- **Assumption 3: Symmetry.** The Jacobian matrix of  $f(\mathbf{X})$  meets symmetry:  $Jf(\mathbf{X}) = [Jf(\mathbf{X})]^T$ .

a key and highly beneficial property is obtained, namely  $\rho(\mathbf{X})$  is a convex function and its gradient can be efficiently computed as:

$$\nabla \rho(\mathbf{X}) = \mathbf{X} - f(\mathbf{X}) \quad (5)$$

we choose FFDNet [20], depicted in Fig. 2, as denoising engine  $f(\mathbf{X})$  in our framework. With a tunable noise map as input, FFDNet works on downsampled sub-images, achieving a good trade-off between speed and denoising performance. FFDNet is implicitly formulated as:

$$\hat{\mathbf{X}}_F = f(\mathbf{X}, \mathbf{M}; \Theta_F) \quad (6)$$

$\mathbf{M}$  is noise level map.  $\Theta_F$  indicates pre-trained parameters.

DIP follows the generic template of Eq. (2), but drop the explicit regularizer  $\rho(\mathbf{X})$  and use instead the implicit prior captured by CNN. In our method, RED framework Eq. (4) is implemented by adding a regularization term to the loss function Eq. (3) of DIP. More specifically, the objective function of our method is:

$$\hat{\mathbf{X}}, \hat{\Theta} = \arg \min_{\mathbf{X}, \Theta} \frac{1}{2} \|\mathbf{H}\mathbf{T}_\Theta(\mathbf{Z}) - \mathbf{Y}\|_2^2 + \frac{\lambda}{2} \mathbf{X}^T [\mathbf{X} - f(\mathbf{X})] \quad (7)$$

*s.t.*  $\mathbf{X} = \mathbf{T}_\Theta(\mathbf{Z})$

$\lambda$  controls the RED regularization strength. Compared to Eq. (3), our objective function Eq. (7) explicitly promotes the learning of a more encompassing representation of the underlying image structure and demonstrates robustness against mixed types of degradation.

### B. Optimizer

To solve the problem Eq. (7), ADMM [25] is implemented with the great advantages of decomposing the Eq. (7) into three sub-problems. We turn the constraint in Eq. (7) into a penalty using the Augmented Lagrangian:

$$\hat{\mathbf{X}}, \hat{\Theta} = \arg \min_{\mathbf{X}, \Theta} \frac{1}{2} \|\mathbf{H}\mathbf{T}_\Theta(\mathbf{Z}) - \mathbf{Y}\|_2^2 + \frac{\lambda}{2} \mathbf{X}^T [\mathbf{X} - f(\mathbf{X})] + \frac{\mu}{2} \|\mathbf{X} - \mathbf{T}_\Theta(\mathbf{Z}) - \mathbf{U}\|_2^2 \quad (8)$$

$\mu$  is ADMM factor,  $\mathbf{U}$  is corresponding Lagrangian multiplier.

The ADMM algorithm is used to addresses Eq. (8), and it transforms the problem into solving the following three sub-problems:  $\Theta$ ,  $\mathbf{X}$  and  $\mathbf{U}$  sequentially:

1) *Update of  $\Theta$  sub-problem:* When  $\mathbf{X}$  and  $\mathbf{U}$  are fixed,  $\mathbf{X}^T [\mathbf{X} - f(\mathbf{X})]$  in Eq. (8) is omitted, and we have:

$$\hat{\Theta} = \arg \min_{\Theta} \|\mathbf{H}\mathbf{T}_\Theta(\mathbf{Z}) - \mathbf{Y}\|_2^2 + \mu \|\mathbf{X} - \mathbf{U} - \mathbf{T}_\Theta(\mathbf{Z})\|_2^2 \quad (9)$$

Eq. (9) shows a resemblance with the optimization in DIP, which uses back-propagation. The second term forces  $\mathbf{T}_\Theta(\mathbf{Z})$  to approach  $(\mathbf{X} - \mathbf{U})$ , which provides a stabilizing and robustifying effect. We optimize the Eq. (9) using gradient descent. U-Net [26] with skip-connections is employed in our method as  $\mathbf{T}_\Theta(\cdot)$ .

2) *Update of  $\mathbf{X}$  sub-problem:* We freeze  $\Theta$  and  $\mathbf{U}$  in Eq. (8) and we have:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \frac{\mu}{2} \|\mathbf{X} - \mathbf{C}\|_2^2 + \frac{\lambda}{2} \mathbf{X}^T [\mathbf{X} - f(\mathbf{X})] \quad (10)$$

here,  $\mathbf{C} = [\mathbf{T}_\Theta(\mathbf{Z}) + \mathbf{U}]$ . Eq. (10) is a classic RED objective, representing a denoising of the image  $\mathbf{C}$ .

As note in [19], when the gradient expression of  $\rho(\mathbf{X})$  satisfies Eq. (5), Eq. (10) can be solved through fixed-point method by zeroing the derivative of Eq. (10). Here, we assume that the denoising engine  $f(\mathbf{X})$  FFDNet used in this method satisfies the above three assumptions sated in III-A, that is **differentiability**, **homogeneity** and **symmetry**, so that the gradient expression defined in Eq. (5) holds. We give detailed proof of this part in Section IV.

Based on the above, by setting the gradient of Eq. (10) to zero, we have:

$$0 = \mu(\mathbf{X} - \mathbf{C}) + \lambda[\mathbf{X} - f(\mathbf{X})] \quad (11)$$

Eq. (11) can be solved iteratively by fixed-point method:

$$\mathbf{X}^{k+1} = \frac{1}{\lambda + \mu} [\lambda f(\mathbf{X}^k) + \mu \mathbf{C}] \quad (12)$$

$k$  is the iteration number.

3) *Update of  $\mathbf{U}$  sub-problem:* As last, Lagrangian multiplier  $\mathbf{U}$  is updated as:

$$\mathbf{U}^{k+1} = \mathbf{U}^k - \mathbf{X} + \mathbf{T}_\Theta(\mathbf{Z}) \quad (13)$$

Till now, we have solved the proposed method by ADMM in detail. **Algorithm 1** summarizes the pseudo-code of our method. Fig. 1 shows the case of image inpainting, where our method can achieve better objective and visual quality.

---

#### Algorithm 1 The Proposed Method

---

**Input:**  $\mathbf{Y}$ ,  $\mathbf{H}$ ,  $\lambda$  and  $\mu$ .

**Initialization:** Set  $\mathbf{U}_0 = \mathbf{0}$ ,  $\mathbf{X}_0 = \mathbf{Y}$ , noise vector  $\mathbf{Z}$ .

**for**  $k = 0, 1, \dots, Iter$  **do**

Update  $\Theta^{k+1}$  by solving Eq. (9).

Update  $\mathbf{X}^{k+1}$  by solving Eq. (12).

Update  $\mathbf{U}^{k+1}$  by solving Eq. (13).

**end for**

**Output:** The recovered image  $\hat{\mathbf{X}}$ .

---

#### IV. MEET THREE ASSUMPTIONS

In III-B for updating  $\mathbf{X}$  sub-problem, we assume that denoising engine  $f(\mathbf{X})$  FFDNet satisfies three assumptions so that the gradient expression Eq. (5) is established. Thus the fixed-point method can be employed in solving Eq. (12).

Now, we verify that denoising engine  $f(\mathbf{X})$  FFDNet satisfies the three required assumptions: **1. differentiability**, **2. homogeneity**, and **3. symmetry**. And how these three conditions lead to the gradient expression Eq. (5) holds.

##### A. Verify Assumption 1

**Differentiability: Denoising engine  $f(\mathbf{X})$  FFDNet is differentiable.**

**Analysis:** As shown in Fig. 2, FFDNet down-samples the input  $\mathbf{X}$  to four halved sub-images  $\mathbf{X}/4$ , and stacks noise map  $\mathbf{M}$  to form a tensor. Then the tensor is fed into three different types of CNN in turn. The first layer consists **Conv + ReLU** [27], the middle layers consist **Conv + BN** [28] + **ReLU**, and the last layer consists **Conv**. An up-sample operation is performed at last, which generates the estimated result with same dimensions of input. Common differentiable operations include sampling (bilinear or bicubic), convolution, ReLU and BN. Regarding their respective differentiability, when these operations are used sequentially within FFDNet, the linear combination  $f(\mathbf{X})$  is differentiable.

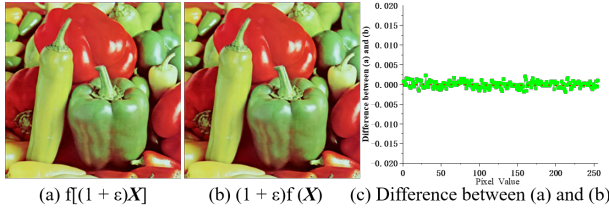


Fig. 3. The evaluation of homogeneity. (a)  $f[(1 + \epsilon)\mathbf{X}]$ , (b)  $(1 + \epsilon)f(\mathbf{X})$ . SSIM and MSE between (a) and (b) is 0.999,  $1.16\text{e-}6$ .

##### B. Verify Assumption 2

**Homogeneity:**  $f[(1 + \epsilon)\mathbf{X}] = (1 + \epsilon)f(\mathbf{X})$ , where  $\epsilon$  is a small constant.

**Analysis:** Due to the implicit nature of  $f(\mathbf{X})$  in Eq. (6) for FFDNet, we provide experimental evidence to proof the homogeneity. Testing image *pepper* is corrupted by gaussian noise with noise level  $\sigma_n = 5$  to be  $\mathbf{X}$ .  $\epsilon$  is set to 0.01. Fig. 3 shows the evaluation of homogeneity, SSIM [29] and MSE are used to measure the similarity between  $f[(1 + \epsilon)\mathbf{X}]$  and  $(1 + \epsilon)f(\mathbf{X})$ , SSIM and MSE between  $f[(1 + \epsilon)\mathbf{X}]$  (Fig. 3 (a)) and  $(1 + \epsilon)f(\mathbf{X})$  (Fig. 3 (b)) are very close to 1 and 0, respectively. Moreover, in Fig. 3 (c), it can be seen that the difference between Fig. 3 (a) and Fig. 3 (b) fluctuates very slightly around 0. The above verify the homogeneity property of  $f(\mathbf{X})$ .

Based on the verification of homogeneity, we introduce a conclusion from [30] for facilitate subsequent proof:

$$[Jf(\mathbf{X})]\mathbf{X} = f(\mathbf{X}) \quad (14)$$

##### C. Verify Assumption 3

**The Jacobian matrix of  $f(\mathbf{X})$  satisfies symmetry:**  $Jf(\mathbf{X}) = [Jf(\mathbf{X})]^T$ .

**Analysis:** Normalized Error Metric (NEM) [30] is employed to verify the symmetry property. NEM is defined as:

$$e_f^J(\mathbf{X}) = \frac{\|Jf(\mathbf{X}) - [Jf(\mathbf{X})]^T\|_F^2}{\|Jf(\mathbf{X})\|_F^2} \quad (15)$$

here,  $[Jf(\mathbf{X})]_{i,n} = [f_i(\mathbf{X} + \varrho e_n) - f_i(\mathbf{X} - \varrho e_n)]/2\varrho$ .  $\varrho$  is small constant and  $e_n$  is the  $n$ -th column of identity matrix. The smaller the NEM value of  $f(\mathbf{X})$ , the closer it is to symmetry. Table I lists the averaged NEM of five common denoisers: NLM [8], BM3D [9], TNRD [31], DnCNN [32], FFDNet [20]) under  $\sigma_n = 50$  and  $\varrho = 0.001$ . It can be observed that  $e_f^J(\mathbf{X})$  of FFDNet is the smallest, closest to 0. This verifies the symmetry of the Jacobian matrix of FFDNet.

TABLE I  
AVERAGED NEM OF FIVE COMMON DENOISERS ON SET12.

Denoiser	NLM	BM3D	TNRD	DnCNN	FFDNet
$e_f^J(\mathbf{X})$	0.7638	0.7039	0.0259	0.0116	0.0073

##### D. Verify gradient expression Eq. (5)

We have verified that the denoising engine  $f(\mathbf{X})$  FFDNet satisfies differentiability (IV-A), homogeneity (IV-B), and symmetry (IV-C). Now we prove how these three conditions lead to the gradient expression Eq. (5) holds.

**Analysis:** It is hard to directly prove that the gradient expression of  $\rho(\mathbf{X})$  is Eq. (5). Instead, we verify that Eq. (5) is equivalent to the true gradient expression of  $\rho(\mathbf{X})$  [30] 1:

$$\nabla \rho(\mathbf{X}) = \mathbf{X} - \frac{1}{2}f(\mathbf{X}) - \frac{1}{2}[Jf(\mathbf{X})]^T \mathbf{X} \quad (16)$$

Eq. (16) not requires that  $f(\mathbf{X})$  is homogeneous and symmetric Jacobian.

Now, our goal is converted to prove that the two gradient expressions of RED: Eq. (5) and Eq. (16) are equivalent. For this, we assume that Eq. (5) and Eq. (16) are equivalent, thus we have:

$$\mathbf{X} - f(\mathbf{X}) = \mathbf{X} - \frac{1}{2}f(\mathbf{X}) - \frac{1}{2}[Jf(\mathbf{X})]^T \mathbf{X} \quad (17)$$

so we can obtain:

$$f(\mathbf{X}) = [Jf(\mathbf{X})]^T \mathbf{X} \quad (18)$$

Plugging Eq. (14) into the left side of Eq. (18), we have:

$$Jf(\mathbf{X}) = [Jf(\mathbf{X})]^T \quad (19)$$

Thus, Eq. (17) is valid only when  $Jf(\mathbf{X})$  satisfies symmetry, which is proved in IV-C. Consequently, we obtain Eq. (17), which means our goal Eq. (5) is established.

<sup>1</sup>Please refer to [30] for proof.

TABLE II  
FROM TOP TO BOTTOM: COMPARISON ON: IMAGE INPAINTING, SUPER-RESOLUTION AND DEBLURRING. THE BEST PSNR IS MARKED IN RED.

Image Inpainting						
Method	GN:15+line	GN:25+grid	GN:50+mask	PN:15+line	PN:25+grid	PN:50+text
DIP [1]-(2018-CVPR)	26.83	27.12	23.78	22.57	21.34	17.17
DD [2]-(2018-ICLR)	26.64	27.60	23.76	22.63	21.63	17.41
NAS-DIP [3]-(2020-ECCV)	27.10	28.10	24.05	22.94	21.67	17.11
DeepRED [4]-(2019-ICCVW)	29.85	31.55	27.06	25.41	24.02	19.26
PIP [5]-(2022-Arxiv)	25.83	27.64	24.13	22.64	21.84	17.15
SHI [6]-(2022-IJCV)	27.30	28.71	25.11	23.68	21.91	18.06
DRP-DIP [7]-(2023-CVPR)	22.45	23.38	19.91	17.97	18.95	15.54
OURs	30.01	31.75	27.40	25.73	24.19	19.43
Image Super-Resolution						
Method	GN:15+S:2	GN:25+S:4	GN:50+S:8	PN:15+S:2	PN:25+S:4	PN:50+S:8
DIP [1]-(2018-CVPR)	25.67	21.78	17.84	21.88	18.82	15.28
DD [2]-(2018-ICLR)	25.54	21.77	17.90	22.12	18.69	15.06
NAS-DIP [3]-(2020-ECCV)	27.87	23.70	19.70	24.00	20.69	17.30
DeepRED [4]-(2019-ICCVW)	28.63	24.27	20.41	24.52	21.18	17.73
PIP [5]-(2022-Arxiv)	26.15	22.20	18.20	22.36	19.12	15.60
SHI [6]-(2022-IJCV)	26.24	22.15	18.23	22.51	19.10	15.62
DRP-DIP [7]-(2023-CVPR)	26.55	22.09	16.59	22.36	18.64	14.24
OURs	29.51	24.54	20.92	24.76	21.36	18.25
Image Deblurring						
Method	GN:15+gauss	GN:25+gauss	GN:50+gauss	PN:15+gauss	PN:25+gauss	PN:50+gauss
DIP [1]-(2018-CVPR)	24.20	23.81	22.59	22.89	21.80	19.27
DD [2]-(2018-ICLR)	24.25	23.71	22.66	22.51	21.59	18.90
NAS-DIP [3]-(2020-ECCV)	26.69	25.29	23.24	23.90	22.50	19.67
DeepRED [4]-(2019-ICCVW)	27.14	25.76	23.59	24.21	22.68	19.75
PIP [5]-(2022-Arxiv)	22.53	22.12	20.70	20.83	19.64	17.04
SHI [6]-(2022-IJCV)	22.62	22.15	20.78	20.95	19.82	17.40
DRP-DIP [7]-(2023-CVPR)	22.28	21.20	19.13	20.00	18.64	16.32
OURs	27.68	26.53	24.25	24.51	22.97	19.95

<sup>1</sup> The symbols in Table II are explained as follows: Top: image inpainting, GN and PN represent Gaussian and Poisson Noise respectively, and the following numbers represent the specific noise level. Line, grid and text represent three different covered masks. Middle: image super-resolution, S: up-sampling factor, and the following numbers represent the specific up-sampling scales. Bottom: image deblurring, gauss: image is deblurred by Gaussian kernel with size of  $25 \times 25$  and standard deviation 1.6.

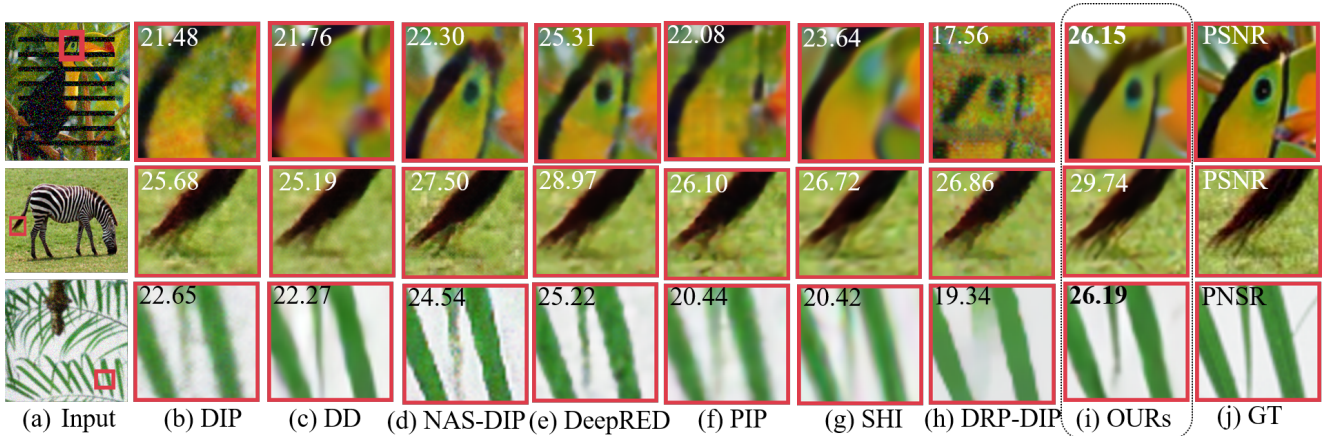


Fig. 4. Visual comparisons of image restoration results. **The first row:** Image inpainting, image mixed corrupted by line mask and poisson noise (PN:15). **The second row:** Image super-resolution, image mixed corrupted by gaussian noise (GN:15) and down-sample scale factor 2. **The third row:** Image deblurring, image mixed corrupted by gaussian noise (GN:15) and gauss blur. Our methods recovers much better cleaner images with fine detailed structures.

## V. EXPERIMENTS

### A. Implementation Details

**Compared methods and Metric:** We compare our method to another seven unsupervised learning image restoration approaches: DIP [1], DD [2], NAS-DIP [3], DeepRED [4], PIP [5], SHI [6] and DRP-DIP [7]. All the results are obtained from the codes provided by authors. Since all methods are iteratively, we select the highest PSNR during iterations in each method as the result for fair comparison. The source code of our method is available at: <https://github.com/weimin581/>

### demo\_DRED\_DIP.

**Image restoration task and dataset:** We focus on three representative tasks. Specifically, in image inpainting, testing image from Set5 [33] is occluded by a binary mask (line, grid or text) and simultaneously corrupted by noise (Gaussian or Poisson noise). For image super-resolution, image from Set14 [34] is down-scaled and simultaneously corrupted by noise (Gaussian or Poisson noise). For image deblurring, image from Set3 is degraded by Gaussian blur and concurrently contaminated by noise (Gaussian or Poisson noise). For specific experimental settings, see the footnote of Table II.

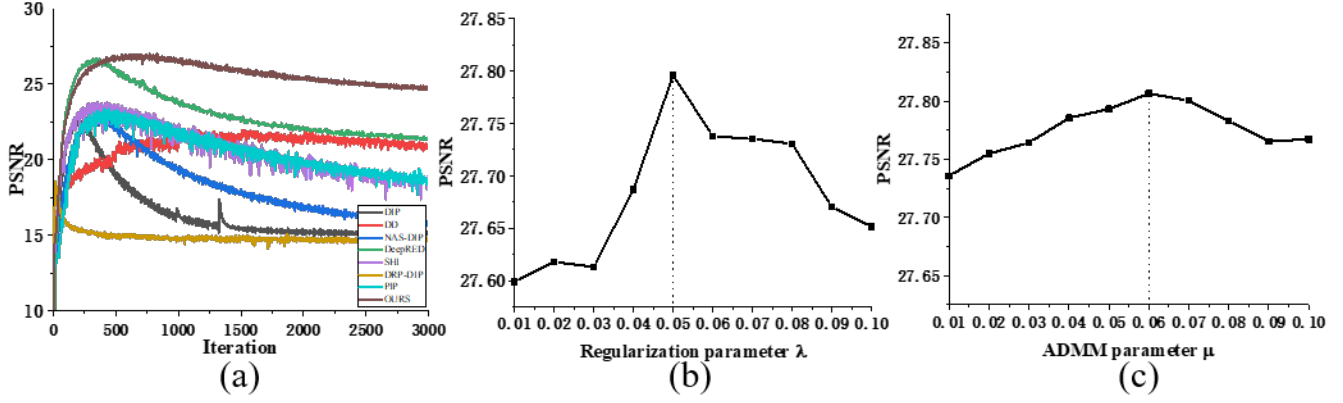


Fig. 5. Convergence (a) and parameter analysis (b,c) on Set5. Mixed degradation condition: text mask covered and Gaussian noise (level  $\sigma_n = 50$ ) corrupted.

**Parameters setting:** Regularization parameter  $\lambda$  and ADMM parameter  $\mu$  in Eq. (8) are set to 0.05, 0.06, respectively. Network  $T_{\Theta}(Z)$  in our method adopts the U-Net [26] with skip-connections architecture as DIP. We use the Adam optimizer [35] in our method. The learning rate is 0.001 and noise vector  $Z$  is randomly initialized.

### B. Quantitative and Qualitative Analysis

**Quantitative Analysis:** It can be observed from Table II, under six different settings in each task, our method performs **consistently** the best. As a lightweight version of DIP, DD, DRP-DIP and SHI effectively reduce the complexity of the network, but they do not significantly improve the recovery effect, or even worsen, which shows that changing the network architecture from the inside is not enough to deal with more complex mixed degraded images recovery. DeepRED shows the best results apart from our method. Its PNP over-fitting presentation uses BM3D or NLM as denoiser prior. However, as shown in Table I, the Jacobian matrix of BM3D or NLM do not satisfy symmetry to some extent. PIP uses a pixel-level MLP with much fewer parameters than U-Net equipped in DIP. Although PIP can achieve similar results to DIP, both are difficult to achieve satisfactory restoration results. **Qualitative Analysis:** As shown in Fig. 4. In the first row, DIP, DD, SHI, DRP-DIP and PIP are difficult to recover the original undamaged edges and textures of image. Although NAS-DIP and DeepRED are better at restoring the contour of the image, but filled with noise or visual artifacts. Our method can effectively recover the lost information and remove noise simultaneously. The results in the second and third rows demonstrate that our method is better at effectively recovering down-sampled information for super-resolution and noisy blurred information for deblurring, respectively.

TABLE III  
ABLATION STUDY ON CBSD68 DATASET.

Method	Inpainting GN:25+grid	Super-Resolution GN:25+S:4	Deblurring GN:25+gauss
DIP	25.61	21.11	22.37
DIP+FFDNet (PNP)	27.64	22.85	24.06
DIP+FFDNet (OURs)	<b>29.55</b>	<b>24.37</b>	<b>25.10</b>

### C. Ablation, Convergence, Complexity and Parameter Analysis

**Ablation Study:** We conduct ablation study to validate the effect of RED framework, powered by FFDNet. Table III lists the average PSNR for three tasks on CBSD68. Two conclusions can be drawn from Table III. 1. The incorporation of the RED component driven by FFDNet significantly enhances our method’s performance in restoring mixed degraded images, affirming its efficacy in handling such harsh scenarios. 2. In comparison to the PNP framework, introducing FFDNet as an explicit regularization term through the RED framework yields more effective results.

**Convergence Performance:** As for convergence performance shown in Fig. 5(a), after DIP reaches the optimal point, the over-fitting noise causes a rapid decline. The same phenomenon occurs in NAS-DIP, DeepRED, SHI and PIP. Although DRP-DIP improves the speed at which DIP reaches its peak value, it still cannot avoid over-fitting and cannot achieve satisfactory results. In contrast, our method exhibits two advantages, the over-fitting phenomenon is effectively avoided, the optimization process becomes stable. In addition, our method achieves the best recovery performance.

TABLE IV  
TIME CONSUMPTION PER ITERATION.

Method	DIP	DD	NAS-DIP	DeepRED	PIP	SHI	DRP-DIP	OURs
Time (s)	0.093	0.017	0.047	1.053	0.057	0.556	0.045	0.201

**Complexity Performance:** We compare the running time on image of size  $256 \times 256$  on a machine with an NVIDIA RTX 3090 GPU. Because different methods achieve optimal results at different numbers of iterations, for a fair comparison, we compare the time it takes per iteration. As shown in Table IV, by introducing the deep denoising engine on the top of DIP architecture, our algorithm inevitably consumes more time than the former, but faster than DeepRED, SHI.

**Parameter Analysis:** We analyze the parameter selection of the proposed method. There are two key parameters in Eq. (8), including regularization parameter  $\lambda$ , and ADMM parameter  $\mu$ . To illustrate the influence of each parameter, we plot the PSNR change curves with different parameters shown in Figs. 5(b, c). As the value of parameter  $\lambda$  rises, the reconstruction



effect starts to improve as shown in Figs. 5(b). When  $\lambda$  reaches 0.05 (marked by the vertical line), the optimal reconstructed performance is achieved and the curves begin to decrease. Therefore, we recommend that the optimal value of  $\lambda$  is 0.05. In Figs. 5(c), when  $\mu$  varies from 0.05 to 0.07, the curves change steadily. We recommend that the value of  $\mu$  is 0.06. All parameters are fixed in our method.

## VI. CONCLUSION

This paper focuses on enhancing the performance of DIP in restoring mixed degraded images. We achieve this by incorporating the regularization term of RED along with the pre-trained deep denoising engine FFDNet. This integration forms a new objective function capable of addressing mixed degradation efficiently, solved through the ADMM. We provide the detailed analysis of the feasibility of the proposed method. Experimental results on three representative image restoration tasks illustrate the superiority of our approach. Moreover, the potential applications of other low-vision tasks, such as de-hazing [36], [37], stereo image [38] restoration, should be explored in the future.

## REFERENCES

- [1] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.
- [2] R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," in *International Conference on Learning Representations*, 2019.
- [3] Y.-C. Chen, C. Gao, E. Robb, and J.-B. Huang, "Nas-dip: Learning deep image prior with neural architecture search," in *European Conference on Computer Vision*, 2020.
- [4] G. Mataev, P. Milanfar, and M. Elad, "Deepred: Deep image prior powered by red," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Oct 2019.
- [5] N. Shabtay, E. Schwartz, and R. Giryes, "PIP: Positional-encoding Image Prior," *arXiv e-prints*, p. arXiv:2211.14298, Nov. 2022.
- [6] Z. Shi, M. Pascal, and S. Subhransu, "On measuring and controlling the spectral bias of the deep image prior," *International Journal of Computer Vision*, vol. 130, no. 4, pp. 885–908, Apr. 2022.
- [7] T. Li, H. Wang, Z. Zhuang, and J. Sun, "Deep random projector: Accelerated deep image prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2023, pp. 18 176–18 185.
- [8] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 60–65 vol. 2.
- [9] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [10] D. Liu, J. Wang, Q. Shan, D. Smyl, J. Deng, and J. Du, "Deepdit: Deep image prior enabled electrical impedance tomography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9627–9638, 2023.
- [11] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Deep image prior," *International Journal of Computer Vision*, vol. 128, pp. 1867 – 1888, 2020.
- [12] K. Gong, C. Catana, J. Qi, and Q. Li, "Pet image reconstruction using deep image prior," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1655–1665, 2019.
- [13] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, "Time-dependent deep image prior for dynamic mri," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3337–3348, 2021.
- [14] D. O. Bager, J. Leuschner, and M. Schmidt, "Computed tomography reconstruction using deep image prior and learned reconstruction methods," *Inverse Problems*, vol. 36, no. 9, p. 094004, 2020.
- [15] Y. Liu, J.-P. Li, Y. Pang, D. Nie, and P.-T. Yap, "The devil is in the upsampling: Architectural decisions made simpler for denoising with deep image prior," *ArXiv*, vol. abs/2304.11409, 2023.
- [16] M. E. Arican, O. Kara, G. Bredell, and E. Konukoglu, "Isnas-dip: Image-specific neural architecture search for deep image prior," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1950–1958, 2021.
- [17] Y. Liu, J. Li, Y. Pang, D. Nie, and P.-T. Yap, "The devil is in the upsampling: Architectural decisions made simpler for denoising with deep image prior," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 12 408–12 417.
- [18] Z. Sun, F. Latorre, T. Sanchez, and V. Cevher, "A plug-and-play deep image prior," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 8103–8107.
- [19] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.
- [20] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.
- [21] J. Liu, Y. Sun, X. Xu, and U. S. Kamilov, "Image restoration using total variation regularized deep image prior," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 7715–7719.
- [22] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.
- [23] R. Fermanian, M. Le Pendu, and C. Guillemot, "Regularizing the deep image prior with a learned denoiser for linear inverse problems," in *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSp)*. IEEE, 2021, pp. 1–6.
- [24] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE global conference on signal and information processing*. IEEE, 2013, pp. 945–948.
- [25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, p. 1–122, 2011.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, p. 84–90, may 2017.
- [28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning - Volume 37*, 2015, p. 448–456.
- [29] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [30] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Transactions on Computational Imaging*, vol. 5, pp. 52–67, 2018.
- [31] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.
- [32] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [33] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
- [34] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*. Springer, 2012, pp. 711–730.
- [35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015.
- [36] W. Yuan, C. Meng, and X. Bai, "Weighted side-window based gradient guided image filtering," *Pattern Recognition*, vol. 146, p. 110006, 2024.
- [37] W. Yuan, Y. Wang, C. Meng, and X. Bai, "Guided image filtering: A survey and evaluation study," in *Proceedings of the 5th ACM International Conference on Multimedia in Asia Workshops*, 2023, pp. 1–5.
- [38] W. Yuan, C. Meng, X. Tong, and Z. Li, "Efficient local stereo matching algorithm based on fast gradient domain guided image filtering," *Signal Processing: Image Communication*, vol. 95, p. 116280, 2021.