Self-Supervised Image Denoising Using Implicit Deep Denoiser Prior

Huangxing Lin[1], Yihong Zhuang[1], Xinghao Ding[1], Delu Zeng[2], Yue Huang[1], Xiaotong Tu[1*], John Paisley[3]

[1]School of Informatics, Xiamen University, China
[2]School of Mathematics, South China University of Technology, China
[3]Department of Electrical Engineering, Columbia University, USA
{hxlin, zhuangyihong}@stu.xmu.edu.cn, {dxh, yhuang2010, xttu}@xmu.edu.cn, dlzeng@scut.edu.cn, jwp2128@columbia.edu

Abstract

We devise a new regularization for denoising with self-supervised learning. The regularization uses a deep image prior learned by the network, rather than a traditional predefined prior. Specifically, we treat the output of the network as a "prior" that we again denoise after "re-noising." The network is updated to minimize the discrepancy between the twice-denoised image and its prior. We demonstrate that this regularization enables the network to learn to denoise even if it has not seen any clean images. The effectiveness of our method is based on the fact that CNNs naturally tend to capture low-level image statistics. Since our method utilizes the image prior implicitly captured by the deep denoising CNN to guide denoising, we refer to this training strategy as an Implicit Deep Denoiser Prior (IDDP). IDDP can be seen as a mixture of learning-based methods and traditional model-based denoising methods, in which regularization is adaptively formulated using the output of the network. We apply IDDP to various denoising tasks using only observed corrupted data and show that it achieves better denoising results than other self-supervised denoising methods.

Introduction

Images captured by various devices are prone to noise and corruption due to limitations of imaging environments such as low light and slow shutter speed. Denoising is the problem whereby we recover the underlying clean image from its noisy measurements. While sometimes the purpose may be for aesthetic reasons, downstream computer vision tasks such as detection and segmentation are also greatly facilitated by having minimally corrupted inputs.

A noisy image y is usually modeled as

$$y = x + n, \tag{1}$$

where n represents the noise and x is the clean image to be restored. The inverse problem of learning x is challenging because the statistics of n are usually unknown and complex.

In general, model-based image denoising algorithms address this problem by optimizing functions of the form

$$x^* = \arg\min_x \alpha \left\|x - y\right\|_2^2 + \beta R(x), \tag{2}$$

where $\|x - y\|_2^2$ is a data fidelity term that ensures the solution agrees with the observation, $R(x)$ is a regularizer, α and β are trade-off parameters. $R(x)$ is also referred to as a prior term without probabilistic connotations. The first challenge of Eq. (2) is choosing $R(x)$, which encodes the preknown image properties and directs the solution towards a more plausible image. Some common priors for constructing $R(x)$ include total variation (Selesnick 2017), non-local self-similarity (Dabov et al. 2007), and others (Dong et al. 2015; Meng and De La Torre 2013). However, these preselected priors often do not model the finer properties of natural images, and so can lead to degraded results.

Recently, supervised deep networks have achieved unprecedented success in image denoising by learning image priors and noise statistics from pairs of noisy/clean images (Zhang et al. 2019a). Most CNN denoisers such as DnCNN (Zhang et al. 2017a) and VDN (Yue et al. 2019) achieve superior performance over model-based denoisers. Unfortunately, in most cases collecting a large number of realistic paired data is difficult, which limits the application of supervised CNNs. This motivates learning to denoise without paired data – a common approach for non-deep models, but less addressed by neural networks. Some self-supervised methods (Laine et al. 2019; Pang et al. 2021) require only noisy data to train a denoising network under certain minor assumptions, such as zero mean. One of these methods that has attracted much attention is called a deep image prior (DIP) (Ulyanov, Vedaldi, and Lempitsky 2018). DIP found that the structure of a CNN is naturally an implicit regularizer for image denoising. Specifically, DIP uses a deep CNN to reparameterize a noisy image y,

$$\min_\theta \left\|F_\theta(z) - y\right\|_2^2, \tag{3}$$

where z is a fixed random vector, $F_\theta(\cdot)$ represents the deep CNN and θ is its parameters. The optimization trajectory of Eq. (3) tends to produce a good local optimum (i.e., a clean image) before overfitting to y. So denoising can be completed by stopping training early. Moreover, we found that similar results (see Figure 1) can be observed even if z in Eq. (3) is replaced by y,

$$\min_\theta \left\|F_\theta(y) - y\right\|_2^2. \tag{4}$$

DIP demonstrates that the CNN naturally tends to capture low-level image statistics, while showing strong

*Corresponding author.

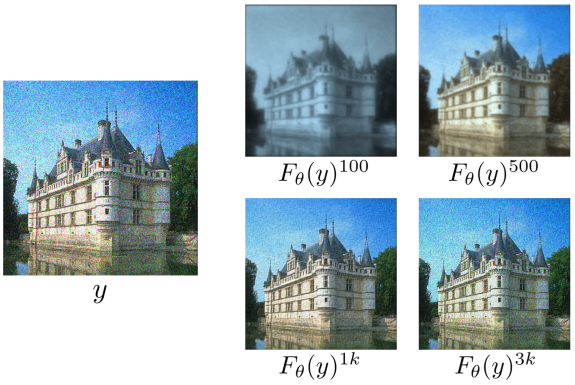


Figure 1: Visual results obtained by minimizing Eq. (4). The superscript represents the number of iterations in training.

"impedance" to unstructured noise (Ulyanov, Vedaldi, and Lempitsky 2018). It also shows the potential of neural networks to learn image priors without clean labels.

Motivation

While DIP is an effective unsupervised approach, a drawback for real-world applications is its time-consuming training that requires human intervention for stopping. Inspired by DIP, in this paper we explore a more flexible use of the neural network as image prior. Similar to DIP, we only focus on self-supervised cases, the difference being that our goal is to formulate an explicit regularization for denoising using the prior determined by the network.

We use a deep CNN to perform the denoising task in Eq. (2), so Eq. (2) can be rewritten as

$$\theta^* = \arg\min_\theta \alpha \left\| F_\theta(y) - y \right\|_2^2 + \beta R(F_\theta(y)). \quad (5)$$

Since DIP confirms the affinity of the deep CNNs for learning low-level image statistics, we expect that this network will also first capture low-level image statistics in early iterations before attempting to fit the noise. Since the output of the network $F_\theta(y)$ contains low-level statistics required to restore the image, we consider using $F_\theta(y)$ as the prior for denoising. Put another way, we ask: Can the output of the denoising network be used to construct a regularization for itself? In this paper, we give a feasible means for performing this task.

Our Proposed Contribution

Suppose there is new noise pattern n_2 that is sampled from a pre-known noise distribution and is specific to the denoised image $F_\theta(y)$. Based on n_2 and $F_\theta(y)$, we devise a new regularization term for image denoising by modifying Eq. (5) to

$$\theta^* = \arg\min_\theta \alpha \|F_\theta(y) - y\|_2^2 + \beta \left\| F_\theta\left(F_\theta\left(y\right) + n_2\right) - F_\theta\left(y\right) \right\|_2^2. \quad (6)$$

$\left\| F_\theta\left(F_\theta\left(y\right) + n_2\right) - F_\theta\left(y\right) \right\|_2^2$ is a regularization term, in which $F_\theta\left(y\right) + n_2$ can be regarded as a new, re-noised image. The purpose of this term is to restore $F_\theta(y)$ from $F_\theta\left(y\right) + n_2$. However, if $F_\theta(y)$ contains noise, the noise will be obscured by n_2. This makes restoring $F_\theta(y)$ from $F_\theta\left(y\right) + n_2$ is an impossible task for the network since it cannot recover random noise from a mixed signal. Therefore, minimizing $\left\| F_\theta\left(F_\theta\left(y\right) + n_2\right) - F_\theta\left(y\right) \right\|_2^2$ implicitly pushes $F_\theta(y)$ to be a noise-free image. With this regularization, it is not necessary to stop training early to prevent overfitting. However, the noise distribution is generally unknown in real-world applications, which hinders the acquisition of new noise n_2.

To circumvent this problem, we propose to use the output of the network to synthesize n_2 adaptively. This is done by an adaptive noise degradation module. In particular, we impose zero-mean and spatially independent constraints on the synthesized n_2, which help stabilize the training. In $\left\| F_\theta\left(F_\theta\left(y\right) + n_2\right) - F_\theta\left(y\right) \right\|_2^2$, $F_\theta(y)$ contains the deep image prior learned by the network. The minimization of $\left\| F_\theta\left(F_\theta\left(y\right) + n_2\right) - F_\theta\left(y\right) \right\|_2^2$ can be seen as using the image prior implicitly captured by the deep denoising network to guide the denoising. Hence, we call this method a *implicit deep denoiser prior* (IDDP). IDDP is self-supervised in that it relies only on the noisy images and does not require knowledge of the noise distribution. To estimate $F_\theta(y)$ and θ, IDDP employs a Siamese network as backbone. We demonstrate the denoising performance of IDDP in various experiments involving both synthetic and real noise. IDDP significantly outperforms state-of-the-art self-supervised methods.

Related Work

Existing image denoising methods can be roughly divided into model-based methods and learning-based methods.

Most model-based denoising methods focus on using handcrafted priors (Xu, Zhang, and Zhang 2018; Buades, Coll, and Morel 2005; Meng and De La Torre 2013), (Zhao et al. 2014) to construct regularization. Over the past decades, various ideas have been proposed, all aiming to identify sources of inner structure in visual data. For example, the non-local self-similarity (NSS) prior (Yao et al. 2019) is widely used in image denoising; some well-known NSS-based methods include BM3D (Dabov et al. 2007) and WNNM (Gu et al. 2014). Other prominent techniques, such as wavelet coring (Simoncelli and Adelson 1996), total variation (Selesnick 2017) and low-rank assumptions (Dong et al. 2015) are also standard approaches.

In recent years, supervised deep learning with CNNs has shown excellent denoising performance (Zhang et al. 2021). Many methods adopt sophisticated network structures to achieve better denoising results (Zhang, Zuo, and Zhang 2018; Yue et al. 2020; Liu et al. 2018). Their effectiveness is due to their ability to learn image priors from large amounts of paired data (Zhang et al. 2020). However, they require paired noisy/clean images. To mitigate this issue, Lehtinen *et al.* (Lehtinen et al. 2018) demonstrate that the denoising CNN can be trained with pairs of independent noisy measurements of the same scene. This training strategy is called Noise2Noise, and achieves denoising performance on par with general supervised learning methods.

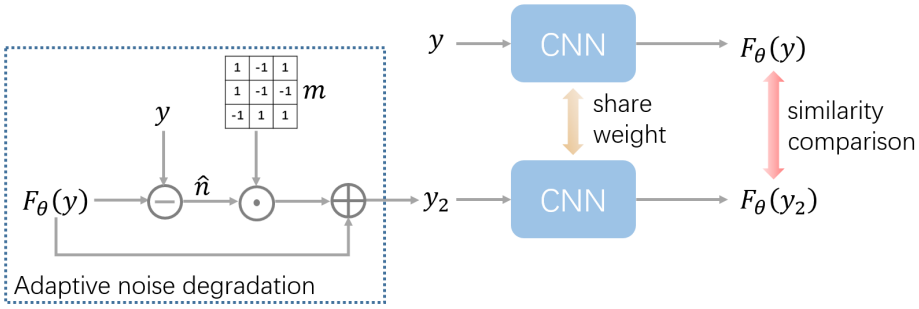


Figure 2: Illustration of IDDP training scheme. IDDP divides the training of the denoising network into two steps: prior estimation and parameter update. In the top branch, the network denoises the noisy image y to get $F_\theta(y)$, which is treated as the prior captured by the network. In the bottom branch, a re-noised image y_2 is produced by the adaptive noise degradation module. The CNN denoiser is optimized to maximize the similarity between $F_\theta(y_2)$ and its prior $F_\theta(y)$. The dimensions of m and $\hat{n}$ are the same. For better visualization, only nine elements in m are shown.

To relax the supervised requirement, networks trained without paired data have recently been considered (Laine et al. 2019; Soltanayev and Chun 2018). One strategy is to use unpaired noisy and clean images to learn the transformation between the noise domain and the clean domain (Du, Chen, and Yang 2020). Methods in this category often integrate noise modeling and removal into the same deep learning framework (*e.g.* DBSN (Wu et al. 2020), C2N (Jang et al. 2021) and Noise2Grad (Lin et al. 2021)). Although effective, unpaired data is not accessible in some real-world situations, such as ultrasound imaging (Mei et al. 2019). Therefore, interest is growing in self-supervised methods using only noisy data. For instance, Noise2Void (Krull, Buchholz, and Jug 2019) proposes a blind spot strategy, which trains a network to predict masked pixels using their neighbors so as to achieve denoising. The blind spot strategy has strong denoising ability, and is adopted by many self-supervised methods, such as Self2Self (Quan et al. 2020), Noise2Self (Batson and Royer 2019), AP-BSN (Lee, Son, and Lee 2022) and Blind2Unblind (Wang et al. 2022). These blind-spot networks cannot exploit the complete information in the image, which inevitably compromises their ability to preserve image details. Inspired by Noise2Noise, Neighbor2Neighbor (Huang et al. 2021) shows that two different subsamples of a noisy input can also be the training data for denoising. Since the backgrounds of different subsamples are not perfectly registered, the performance of Neighbor2Neighbor is not satisfactory. Recorrupted-to-Recorrupted (Pang et al. 2021) builds paired noisy/noisy images by adding additional noise to the noisy input, but for the case where the noise distribution is unknown, we cannot sample an additional noise for training. CVF-SID (Neshatavar et al. 2022) uses a cyclic network to disentangle the noisy image into three layers, but due to the lack of a strong constraint, the clean images produced by CVF-SID lose many image details.

Methodology

Given a noisy image dataset $\Omega^{noisy} = \{y^i\}_{i=1}^N$, we aim to learn to denoise without requiring clean images. The noise in the data are assumed to be zero-mean and spatially independent. We propose a new training scheme called Implicit Deep Denoiser Prior (IDDP), as shown in Figure 2. Our method is derived from Eq. (6), in which the regularization term plays a major role. Unlike DIP (Ulyanov, Vedaldi, and Lempitsky 2018), which trains a new network for each image, we use all noisy images to train a shared denoising network using stochastic gradient descent. For simplicity, we set the batch size to 1 in the objective function of IDDP, but it is straightforward to extend to the case where the batch size is greater than 1. After training, the denoising CNN in IDDP is directly applied to new noisy images without additional learning.

Overview of IDDP

Eq. (6) can be done with a Siamese network (Chen and He 2021). We first input y to a CNN to obtain a denoised image $F_\theta(y)$ and its removed noise $\hat{n} = y - F_\theta(y)$. In Eq. (6), a new noise n_2 is required to form the regularization term. However, the noise distribution is generally unknown, which hinders the acquisition of n_2. To solve this problem, we use $\hat{n}$ to synthesize a new n_2. Based on the assumption of zero mean and spatially independent noise, an adaptive noise degradation module is used to produce n_2,

$$\begin{aligned} n_2 &= m \odot \hat{n} \\ &= m \odot (y - F_\theta(y)), \end{aligned} \tag{7}$$

where $\odot$ represents element-wise multiplication and m is a random binary ± 1 mask that is independent of $\hat{n}$ but has the same dimension as $\hat{n}$. In each training iteration, a new m is sampled. Each element in m is 1 with a probability of $p = 0.5$, and -1 otherwise. By superimposing n_2 onto $F_\theta(y)$, a re-noised image y_2 can be expressed as

$$y_2 = F_\theta(y) + n_2. \tag{8}$$

We note that since n_2 is synthesized using Eq. (7), it possesses some desired properties:

1. Some image details may remain in $\hat{n}$, which can be removed by multiplying with a random mask m making n_2 a more naturally noisy (see Figure 3).

2. $\mathbb{E}(m) = 0$ leads to $\mathbb{E}(n_2) = \mathbb{E}(m)\mathbb{E}(\hat{n}) = 0$ (m and $\hat{n}$ are independent). Therefore, the noise of y_2 meets the zero-mean assumption. This zero-mean assumption makes training stable.
3. Since n_2 is composed of $\hat{n}$, the statistical characteristics of n_2 and $\hat{n}$ should be similar, as shown in Figure 4. In addition, m is random, so $n_2 \neq \hat{n}$. If $n_2 = \hat{n}$, the regularization in Eq. (6) is useless.

Substituting Eq. (8) into Eq. (6), we have

$$\theta^* = \arg\min_{\theta} \alpha \left\| F_\theta(y) - y \right\|_2^2 + \beta \left\| F_\theta(y_2) - F_\theta(y) \right\|_2^2 . \tag{9}$$

Since y_2 is synthesized using the output of the network, we realize that the identity map $F_\theta(y) = y$ is also a potential solution to Eq. (9). If the network is an identity, then $\hat{n}$ is a zero matrix and $y_2 = y$. To avoid this bad solution, we smooth y with an averaging filter to get a noise-free but blurry image y_{blur}. By replacing y in the fidelity term with y_{blur}, Eq. (9) is rewritten as

$$\theta^* = \arg\min_{\theta} \alpha \left\| F_\theta(y) - y_{blur} \right\|_2^2 + \beta \left\| F_\theta(y_2) - F_\theta(y) \right\|_2^2 . \tag{10}$$

The first term in Eq. (10) prevents the network from turning into an identity map, but it tends to smooth the denoised image. To reduce this negative impact, we set α to 0.001 and $\beta = 1$. However, Eq. (10) still cannot assure high-quality denoising. Therefore, inspired by the Siamese network based approach in Chen and He (2021), we further apply a "stop gradient" operation in the second term of Eq. (10) to facilitate training. Eq. (10) is reformulated as

$$\begin{aligned} \theta^* = \arg\min_{\theta} & \alpha \left\| F_\theta(y) - y_{blur} \right\|_2^2 + \\ & \beta \left\| F_\theta(y_2) - \text{Stopgrad}\left(F_\theta(y)\right) \right\|_2^2 . \end{aligned} \tag{11}$$

"Stopgrad" indicates that the gradient does not backpropagate to $F_\theta(y)$. The stop-gradient procedure is a common strategy for learning Siamese networks (Chen and He 2021), which can promote network convergence and prevent trivial solutions. In Eq. (11), stop-gradient is essential to obtain good denoising. Without stop-gradient, both $F_\theta(y_2)$ and $F_\theta(y)$ in $\left\| F_\theta(y_2) - F_\theta(y) \right\|_2^2$ will contribute to the update of θ, resulting in unstable training. We will show that the denoising network in IDDP converges to a poor result without using Stopgrad.

Interpretation of IDDP

IDDP trains the denoising network by solving Eq. (11). Since β is much larger than α, the regularization term plays a major role in denoising. Since low-level image statistics such as edges and textures are easier to capture by deep CNNs than noise, we use the image prior implicit in $F_\theta(y)$ to guide denoising. In $\left\| F_\theta(y_2) - \text{Stopgrad}\left(F_\theta(y)\right) \right\|_2^2$, y_2 and $F_\theta(y)$ can be regarded as "paired" data where y_2 is a noisy image and $F_\theta(y)$ is the corresponding label. By maximizing the similarity between $F_\theta(y_2)$ and $F_\theta(y)$, the denoising ability of the network is improved. Next, we further discuss why IDDP is a practical and effective denoising method and why "stop gradient" is necessary.

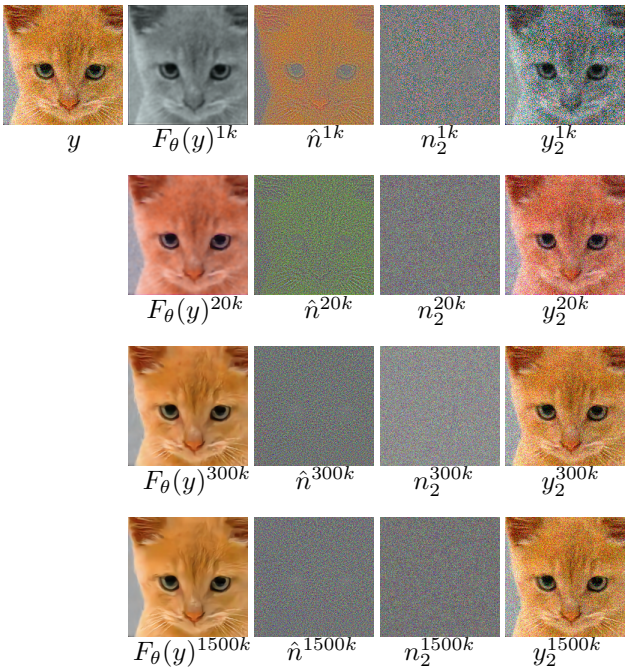


Figure 3: Visual examples produced by IDDP during training. The superscript represents the number of training steps. By multiplying with a random m, $\hat{n}$ is transformed into n_2.

Inspired by traditional methods based on the denoiser prior (Venkatakrishnan, Bouman, and Wohlberg 2013), Eq. (10) can be transformed into the following constrained optimization problem by introducing an auxiliary variable η_y,

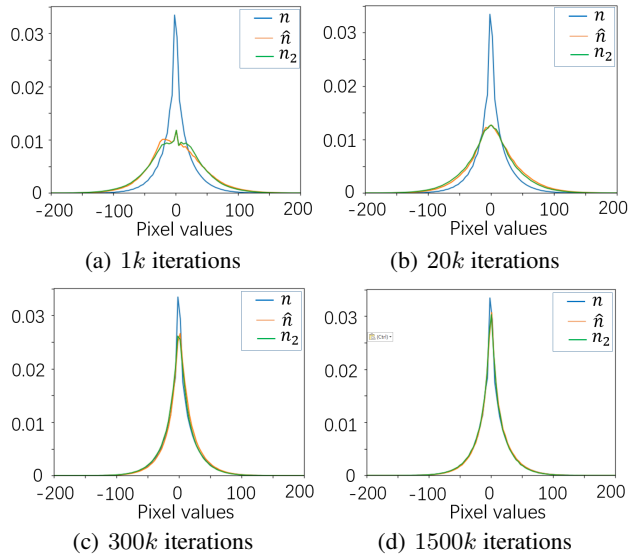
$$\begin{aligned} \theta^* = \arg\min_{\theta} & \alpha \left\| F_\theta(y) - y_{blur} \right\|_2^2 + \beta \left\| F_\theta(\eta_y + n_2) - \eta_y \right\|_2^2 , \\ \text{subject to} \quad & \eta_y = F_\theta(y) . \end{aligned} \tag{12}$$

Using the half quadratic splitting method (Zhang et al. 2017b), Eq. (12) is equivalent to

$$\begin{aligned} \arg\min_{\theta, \eta_y} \alpha \left\| F_\theta(y) - y_{blur} \right\|_2^2 + & \beta \left\| F_\theta(\eta_y + n_2) - \eta_y \right\|_2^2 + \\ & \mu \left\| \eta_y - F_\theta(y) \right\|_2^2 , \end{aligned} \tag{13}$$

where μ is a trade-off parameter. Eq. (13) involves two variables, θ and η_y, and solves the two underlying sub-problems of prior estimation and network parameter update. The problem in Eq. (13) can be solved by an alternating algorithm, fixing one variable and solving for the other. The presence of the stop-gradient in Eq. (11) is the consequence of introducing the extra variable. At each training step t, we randomly sample a noisy image y from the given noisy dataset Ω^{noisy}. Then, we alternately solve the two sub-problems:

$$\begin{aligned} \eta_y^{(t)} = \arg\min_{\eta_y} & \mu \left\| \eta_y - F_{\theta^{(t-1)}}(y) \right\|_2^2 + \\ & \beta \left\| F_{\theta^{(t-1)}}(\eta_y + n_2) - \eta_y \right\|_2^2 \end{aligned} \tag{14}$$



(a) $1k$ iterations (b) $20k$ iterations

(c) $300k$ iterations (d) $1500k$ iterations

Figure 4: Statistical histograms for $20,000$ noise examples with a size of 128×128. n is the real noise, $\hat{n}$ and n_2 are the noise produced by IDDP. As the training progresses, the statistical distributions of $\hat{n}$ and n_2 gradually move to the distribution of n.

$$\theta^{(t)} = \arg\min_{\theta} \alpha \left\| F_\theta\left(y\right) - y_{blur} \right\|_2^2 + \beta \left\| F_\theta\left(\eta_y^{(t)} + n_2\right) - \eta_y^{(t)} \right\|_2^2 + \mu \left\| \eta_y^{(t)} - F_\theta\left(y\right) \right\|_2^2 \tag{15}$$

Solving for η_y**.** Eq. (14) is equivalent to the problem of denoising $F_{\theta^{(t-1)}}(y)$. Specifically, $\|\eta_y - F_{\theta^{(t-1)}}(y)\|_2^2$ is the fidelity term and $\|F_{\theta^{(t-1)}}(\eta_y + n_2) - \eta_y\|_2^2$ is the regularization term. Previous literature on denoising priors suggests that the result of Eq. (14) can be replaced by the denoising result of an external denoiser for $F_{\theta^{(t-1)}}(y)$ (Zhang et al. 2017b; Romano, Elad, and Milanfar 2017). However, defining a good auxiliary denoiser is a tricky task. Fortunately, Eq. (15) constrains the output of the network to be noise-free. Therefore, $F_{\theta^{(t-1)}}(y)$ does not contain noise and does not need to be denoised again. Based on these analyses, we omit the optimization in Eq. (14) and simply redefine $\eta_y^{(t)}$ to be $F_{\theta^{(t-1)}}(y)$,

$$\eta_y^{(t)} = F_{\theta^{(t-1)}}\left(y\right). \tag{16}$$

Then, $\eta_y^{(t)}$ serves as a prior in Eq. (15).

Solving for θ**.** In Eq. (15), the network parameter θ is updated by gradient descent. Typically, $\theta^{(t-1)}$ is assigned to θ as the initial state for solving Eq. (15). Therefore, the third term $\|\eta_y^{(t)} - F_\theta(y)\|_2^2$ of Eq. (15) is equal to 0. Eq. (15) is simplified to

$$\theta^{(t)} = \arg\min_{\theta} \alpha \left\| F_\theta\left(y\right) - y_{blur} \right\|_2^2 + \beta \left\| F_\theta\left(\eta_y^{(t)} + n_2\right) - \eta_y^{(t)} \right\|_2^2. \tag{17}$$

Eq. (17) is consistent with Eq. (11). By solving Eq. (17), the network learns to denoise and learns the image prior implicit in $\eta_y^{(t)}$. $\eta_y^{(t)}$ is fixed in this subproblem, so the gradient does not backpropagate to $\eta_y^{(t)}$.

In short, the two equations (16) and (17) describe the core idea of IDDP. One output of the denoising network is obtained at training step t, which serves as a "prior" target for denoising. $\eta_y^{(t)}$ contains the low-level image statistics required to restore a clean image. Therefore, we use $\eta_y^{(t)}$ as the label and y_2 as the input to improve the denoising ability of the network. This leads the network to produce a better prior at time $t+1$. Also, the noise statistics of $n_2^{(t+1)}$ are closer to the real noise than $n_2^{(t)}$. By performing prior estimation and parameter updates alternately, the output of the network will approximate the true clean x, as shown in Figure 3.

Training Details

The denoising CNN in IDDP is a simple U-Net (Ronneberger, Fischer, and Brox 2015). We use PyTorch and Adam (Kingma and Ba 2014) with a batch size of 1 to train the network. The training images are randomly cropped into 128×128 patches before being input to the network. The learning rate is fixed to 0.0002 for the first 1,000,000 iterations and linearly decays to 0 for the next 1,000,000 iterations. After training, the CNN with fixed parameters θ^* is directly applied to new noisy images.

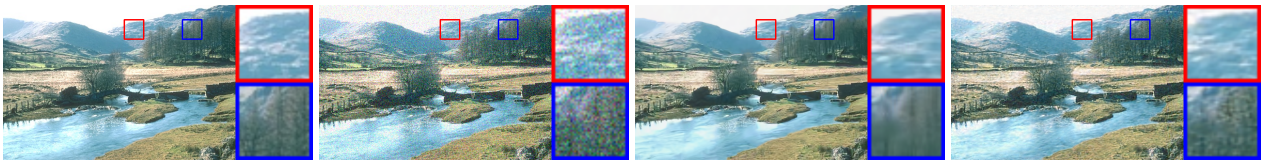
Experiments

We evaluate IDDP on various denoising tasks involving both synthetic and real-world noise.

Synthetic Noise

We collected 4744 images from the Waterloo Exploration Database (Ma et al. 2016) to synthesize noisy images for training. Several state-of-the-art denoising methods are adopted for performance comparison, including a model-based method BM3D (Dabov et al. 2007), self-learning methods DIP (Ulyanov, Vedaldi, and Lempitsky 2018), Noise2Void (N2V) (Krull, Buchholz, and Jug 2019), Self2Self (S2S) (Quan et al. 2020), Neighbor2Neighbor (Nb2Nb) (Huang et al. 2021), CVF-SID (Neshatavar et al. 2022) and Blind2Unblind (B2U) (Wang et al. 2022). Other deep learning methods include Noise2Noise (N2N) (Lehtinen et al. 2018) and a common fully-supervised U-Net. U-Net is trained with noisy/clean image pairs, while N2N uses paired noisy/noisy images. BSD300 (Martin et al. 2001) is the test set of the following experiments. We adopt peak signal to noise ratio (PSNR) and structural similarity index (SSIM) as evaluation metrics.

Gaussian noise. We first study the effectiveness of IDDP on additive Gaussian noise. Each training example is corrupted by zero-mean Gaussian noise with a random standard deviation $\sigma \in (0, 50]$. For testing, we synthesize two sets of noisy images, using a fixed noise level $\sigma = 25$ and a variable $\sigma \in (0, 50]$. We present the quantitative results in Tables 1-2. Visual comparisons can be found in Figure 5.



(a) Clean | PSNR, SSIM (b) Input | 23.89, 0.661 (c) BM3D | 28.31, 0.849 (d) DIP | 26.60, 0.793

(e) Nb2Nb | 29.01, 0.870 (f) B2U | 29.08, 0.868 (g) U-Net | **29.47**, **0.876** (h) Our | 29.31, 0.875

Figure 5: Example results for Gaussian denoising with $\sigma = 25$.

	Test noise level	BM3D	DIP	N2V	S2S	Nb2Nb	CVF-SID	B2U	N2N	U-Net	IDDP
Gaussian	$\sigma = 25$	30.90	29.99	30.56	30.63	31.52	27.14	31.60	31.79	**31.81**	31.70
	$\sigma \in (0, 50]$	31.69	29.66	31.88	31.76	33.08	27.92	33.21	33.40	**33.44**	33.32
Speckle	$v = 0.1$	26.64	26.73	28.40	28.60	30.62	26.12	30.61	31.33	**31.49**	30.96
	$v \in (0, 0.2]$	26.70	27.06	28.77	28.93	30.90	26.08	30.95	31.64	**31.86**	31.22
Poisson	$\lambda = 30$	27.70	28.77	29.58	29.83	30.81	27.03	30.89	31.07	**31.09**	30.91
	$\lambda \in [5, 50]$	27.23	27.98	28.76	29.01	30.19	26.69	30.28	30.42	**30.44**	30.30

Table 1: PSNR results (dB) on BSD300 data with Gaussian, Speckle and Poisson noise. IDDP outperforms all other self-supervised models and non-deep models.

	Test noise level	BM3D	DIP	N2V	S2S	Nb2Nb	CVF-SID	B2U	N2N	U-Net	IDDP
Gaussian	$\sigma = 25$	0.877	0.850	0.880	0.882	0.890	0.808	0.890	0.896	**0.897**	0.897
	$\sigma \in (0, 50]$	0.881	0.833	0.883	0.885	0.894	0.813	0.894	0.898	**0.902**	0.900
Speckle	$v = 0.1$	0.796	0.775	0.855	0.862	0.887	0.729	0.888	0.901	**0.905**	0.893
	$v \in (0, 0.2]$	0.783	0.780	0.858	0.863	0.891	0.741	0.890	0.902	**0.907**	0.895
Poisson	$\lambda = 30$	0.819	0.826	0.861	0.864	0.880	0.756	0.881	0.884	**0.887**	0.884
	$\lambda \in [5, 50]$	0.815	0.801	0.840	0.846	0.860	0.738	0.863	0.865	**0.866**	0.864

Table 2: SSIM results on BSD300 data with Gaussian, Speckle and Poisson noise.

In addition, Figure 4 shows the statistical histograms of the noise removed by IDDP during training.

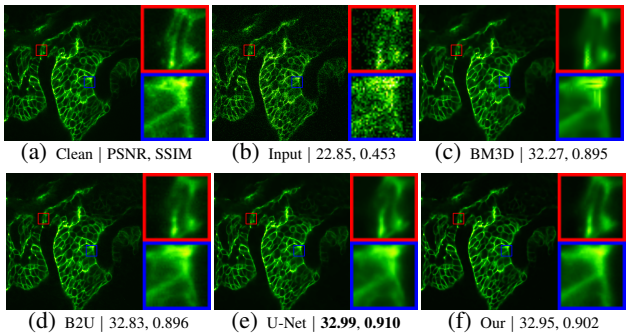
Speckle noise. Multiplicative speckle noise is signal dependent and is often observed in medical ultrasonic images and radar images. The speckle noise in the image is modeled as the random value multiplied by the pixel value of the latent signal x, which can be expressed as $y = x + x \cdot n$. In this model, n is uniform noise with a mean of 0 and variance of v. We vary the noise variance $v \in (0, 0.2]$ during training. Quantitative results are shown in Tables 1-2. Visual results can be found in the supplementary material.

Poisson noise. In our third experiment we consider Poisson noise, which can be used to model photon noise in imaging sensors. Poisson noise is also signal-dependent because its expected magnitude depends on the pixel brightness. We randomize the noise magnitude $\lambda \in [5, 50]$ separately for each training example. Quantitative comparisons are reported in Tables 1-2.

Discussion. As can be seen, the model-based method BM3D is inferior to deep learning methods. This significant performance gap suggests that the image priors implicitly captured by CNNs can better model the properties of natural images than traditional hand-crafted priors. Other self-supervised methods (DIP, N2V, S2S, Nb2Nb, CVF-SID, B2U) give decent denoising results. Their effectiveness stems from some noise statistics assumptions (e.g., additive noise) or special network architectures (e.g., blind spot networks). Although effective, they are still substantially inferior to supervised methods. Our IDDP outperforms the model-based and other self-supervised methods both qualitatively and quantitatively. The images produced by IDDP are of high quality and without visible noise. These results demonstrate that the output of the denoising network can be a prior for denoising, as this paper conjectures. By iteratively updating the prior and network parameters, our IDDP

Method	PSNR	SSIM	Method	PSNR	SSIM
BM3D	35.43	0.932	N2N	38.01	0.953
Nb2Nb	37.52	0.944	U-Net	38.15	0.955
B2U	37.60	0.946	IDDP	37.81	0.951

Table 3: Quantitative results on FMD data.



(a) Clean | PSNR, SSIM (b) Input | 22.85, 0.453 (c) BM3D | 32.27, 0.895

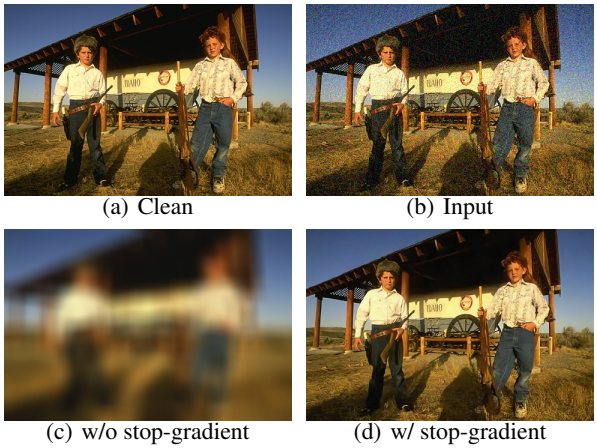
(d) B2U | 32.83, 0.896 (e) U-Net | **32.99**, **0.910** (f) Our | 32.95, 0.902

Figure 6: Example results for FM denoising.

achieves strong denoising performance. We also notice that IDDP is slightly inferior to supervised U-Net and N2N. This is reasonable considering that IDDP has no access to paired data for supervision. In short, IDDP consistently exhibits encouraging denoising performance for various types of noise. The denoised images are clean and sharp. More importantly, IDDP does not rely on paired data or pre-known noise statistics, which shows its potential value for many practical applications.

Real-world Noise

We further evaluate the performance of IDDP on a real Fluorescence Microscopy Denoising (FMD) dataset (Zhang et al. 2019b). The noisy microscopy images in this dataset were obtained by commercial confocal, two-photon and wide-field microscopes. These noisy images cover 5 different noise levels and the corresponding ground-truth images are synthesized by image averaging. However, the noise in widefield images is spatially correlated and the brightness of two-photon BPAE images is not well registered. Therefore, we cull them from the dataset and use the rest for training and testing. The training set and test set contain 30,000 and 120 pairs of images, respectively. IDDP is compared with BM3D, N2N, Nb2Nb, B2U and a fully supervised U-Net. Quantitative results are reported in Table **??**. IDDP achieves better denoising results than BM3D, Nb2Nb and B2U. Visual results are shown in Figure 6. BM3D tends to generate unrealistic artifacts. B2U leaves some minor noise in the denoised image, which degrades the visual quality. The visual results of IDDP and supervised U-Net are comparable, both are high-quality and noise-free images. For this real-world denoising task, IDDP still gives impressive results. These results further confirm the effectiveness of IDDP for various types of noise.



(a) Clean (b) Input

(c) w/o stop-gradient (d) w/ stop-gradient

Figure 7: IDDP with and without stop-gradient. The noise in (b) is Gaussian ($\sigma = 25$).

Stop-gradient

IDDP employs a Siamese network to learn denoising. A common trick for Siamese networks is to use "stop-gradients," which promote convergence and prevent collapsed solutions (Chen and He 2021). In IDDP, the stop-gradient operation is a natural consequence of introducing an extra variable. Here, our aim is to show that the stop-gradient is critical to preventing IDDP from converging to bad results. Figure 7 presents a comparison with and without using the stop-gradient. Without stop-gradients, the denoising network quickly reaches a compromise between the fidelity and regularization terms in the objective function. The denoising network can effectively suppress noise, but it produces blurry images without high-frequency image details. With stop-gradients, the training of IDDP is stable. The denoising network uses its output as a prior for learning to denoise. By alternately updating the prior and network parameters, the network becomes a good denoiser that can efficiently remove noise and restore high-quality images.

Conclusion

We have demonstrated that the output of a neural network can be a prior for image denoising. Based on this prior, we developed a new regularization where, even without any clean data, the network can learn to denoise in a self-supervised manner. We demonstrated the effectiveness and broad applicability of the proposed method on several denoising tasks involving different noise properties and data types. We believe that this study hints at the advantage of exploring the use of CNNs to generate deep image priors for various computer vision tasks.

Acknowledgements

The study was supported partly by the National Natural Science Foundation of China under Grants 82172033, U19B2031, 61971369, 52105126, 82272071.

References

Batson, J.; and Royer, L. 2019. Noise2self: Blind denoising by self-supervision. In *ICML*.

Buades, A.; Coll, B.; and Morel, J.-M. 2005. A non-local algorithm for image denoising. In *CVPR*.

Chen, X.; and He, K. 2021. Exploring simple siamese representation learning. In *CVPR*.

Dabov, K.; Foi, A.; Katkovnik, V.; and Egiazarian, K. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8): 2080–2095.

Dong, W.; Li, G.; Shi, G.; Li, X.; and Ma, Y. 2015. Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising. In *ICCV*.

Du, W.; Chen, H.; and Yang, H. 2020. Learning Invariant Representation for Unsupervised Image Restoration. In *CVPR*.

Gu, S.; Zhang, L.; Zuo, W.; and Feng, X. 2014. Weighted nuclear norm minimization with application to image denoising. In *CVPR*.

Huang, T.; Li, S.; Jia, X.; Lu, H.; and Liu, J. 2021. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *CVPR*.

Jang, G.; Lee, W.; Son, S.; and Lee, K. M. 2021. C2n: Practical generative noise modeling for real-world denoising. In *ICCV*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krull, A.; Buchholz, T.-O.; and Jug, F. 2019. Noise2void-learning denoising from single noisy images. In *CVPR*.

Laine, S.; Karras, T.; Lehtinen, J.; and Aila, T. 2019. High-quality self-supervised deep image denoising. In *NeurIPS*.

Lee, W.; Son, S.; and Lee, K. M. 2022. AP-BSN: Self-Supervised Denoising for Real-World Images via Asymmetric PD and Blind-Spot Network. In *CVPR*.

Lehtinen, J.; Munkberg, J.; Hasselgren, J.; Laine, S.; Karras, T.; Aittala, M.; and Aila, T. 2018. Noise2noise: Learning image restoration without clean data. In *ICML*.

Lin, H.; Zhuang, Y.; Huang, Y.; Ding, X.; Liu, X.; and Yu, Y. 2021. Noise2Grad: Extract Image Noise to Denoise. In *IJCAI*.

Liu, D.; Wen, B.; Fan, Y.; Loy, C. C.; and Huang, T. S. 2018. Non-local recurrent network for image restoration. In *NeurIPS*.

Ma, K.; Duanmu, Z.; Wu, Q.; Wang, Z.; Yong, H.; Li, H.; and Zhang, L. 2016. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2): 1004–1016.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*.

Mei, K.; Hu, B.; Fei, B.; and Qin, B. 2019. Phase asymmetry ultrasound despeckling with fractional anisotropic diffusion and total variation. *IEEE Transactions on Image Processing*, 29: 2845–2859.

Meng, D.; and De La Torre, F. 2013. Robust matrix factorization with unknown noise. In *ICCV*.

Neshatavar, R.; Yavartanoo, M.; Son, S.; and Lee, K. M. 2022. CVF-SID: Cyclic multi-Variate Function for Self-Supervised Image Denoising by Disentangling Noise from Image. In *CVPR*.

Pang, T.; Zheng, H.; Quan, Y.; and Ji, H. 2021. Recorrupted-to-Recorrupted: Unsupervised Deep Learning for Image Denoising. In *CVPR*.

Quan, Y.; Chen, M.; Pang, T.; and Ji, H. 2020. Self2self with dropout: Learning self-supervised denoising from single image. In *CVPR*.

Romano, Y.; Elad, M.; and Milanfar, P. 2017. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4): 1804–1844.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*.

Selesnick, I. 2017. Total variation denoising via the Moreau envelope. *IEEE Signal Processing Letters*, 24(2): 216–220.

Simoncelli, E. P.; and Adelson, E. H. 1996. Noise removal via Bayesian wavelet coring. In *ICIP*.

Soltanayev, S.; and Chun, S. Y. 2018. Training deep learning based denoisers without ground truth data. In *NeurIPS*.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2018. Deep image prior. In *CVPR*.

Venkatakrishnan, S. V.; Bouman, C. A.; and Wohlberg, B. 2013. Plug-and-play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, 945–948.

Wang, Z.; Liu, J.; Li, G.; and Han, H. 2022. Blind2Unblind: Self-Supervised Image Denoising with Visible Blind Spots. In *CVPR*.

Wu, X.; Liu, M.; Cao, Y.; Ren, D.; and Zuo, W. 2020. Unpaired Learning of Deep Image Denoising. In *ECCV*.

Xu, J.; Zhang, L.; and Zhang, D. 2018. A trilateral weighted sparse coding scheme for real-world image denoising. In *ECCV*.

Yao, J.; Meng, D.; Zhao, Q.; Cao, W.; and Xu, Z. 2019. Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing. *IEEE Transactions on Image Processing*, 28(6): 2991–3006.

Yue, Z.; Yong, H.; Zhao, Q.; Meng, D.; and Zhang, L. 2019. Variational denoising network: Toward blind noise modeling and removal. In *NeurIPS*.

Yue, Z.; Zhao, Q.; Zhang, L.; and Meng, D. 2020. Dual adversarial network: Toward real-world noise removal and noise generation. In *ECCV*.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017a. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7): 3142–3155.

Zhang, K.; Zuo, W.; Gu, S.; and Zhang, L. 2017b. Learning deep CNN denoiser prior for image restoration. In *CVPR*.

Zhang, K.; Zuo, W.; and Zhang, L. 2018. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9): 4608–4622.

Zhang, Y.; Li, K.; Li, K.; Sun, G.; Kong, Y.; and Fu, Y. 2021. Accurate and Fast Image Denoising via Attention Guided Scaling. *IEEE Transactions on Image Processing*, 30: 6255–6265.

Zhang, Y.; Li, K.; Li, K.; Zhong, B.; and Fu, Y. 2019a. Residual non-local attention networks for image restoration. In *ICLR*.

Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; and Fu, Y. 2020. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7): 2480–2495.

Zhang, Y.; Zhu, Y.; Nichols, E.; Wang, Q.; Zhang, S.; Smith, C.; and Howard, S. 2019b. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *CVPR*.

Zhao, Q.; Meng, D.; Xu, Z.; Zuo, W.; and Zhang, L. 2014. Robust principal component analysis with complex noise. In *ICML*.