Denoising Prior Driven Deep Neural Network for Image Restoration

Weisheng Dong, Member, IEEE, Peiyao Wang, Wotao Yin, Member, IEEE, Guangming Shi, Senior member, IEEE, Fangfang Wu, and Xiaotong Lu

Abstract—Deep neural networks (DNNs) have shown very promising results for various image restoration (IR) tasks. However, the design of network architectures remains a major challenge for achieving further improvements. While most existing DNN-based methods solve the IR problems by directly mapping low quality images to desirable high-quality images, the observation models characterizing the image degradation processes have been largely ignored. In this paper, we first propose a denoising-based IR algorithm, whose iterative steps can be computed efficiently. Then, the iterative process is unfolded into a deep neural network, which is composed of multiple denoisers modules interleaved with back-projection (BP) modules that ensure the observation consistencies. A convolutional neural network (CNN) based denoiser that can exploit the multi-scale redundancies of natural images is proposed. As such, the proposed network not only exploits the powerful denoising ability of DNNs, but also leverages the prior of the observation model. Through end-to-end training, both the denoiser and the BP modules can be jointly optimized. Experimental results on several IR tasks, e.g., image denoising, super-resolution and deblurring show that the proposed method can lead to very competitive and often state-of-the-art results on several IR tasks, including image denoising, deblurring and super-resolution.

Index Terms—denoising-based image restoration, deep neural network, denoising prior, image restoration.

I. INTRODUCTION

Image restoration (IR) aiming to reconstruct a high quality image from its low quality observation has many important applications, such as low-level image processing, medical imaging, remote sensing, surveillance, etc. Mathematically, IR problem can be expressed as:

\[ y = Ax + n \]

where \( y \) denotes the degraded image and \( x \) the original image, \( A \) is the degradation matrix relating to the imaging/degradation system, and \( n \) the additive noise. Note that for different settings of \( A \), different IR problems can be expressed. For example, the IR problem is a denoising problem [1]–[5] when \( A \) is an identical matrix and becomes a deblurring problem [6]–[9] when \( A \) is a blurring matrix/operator, or a super-resolution problem [8], [10]–[12] when \( A \) is a subsampling matrix/operator. Essentially, restoring \( x \) from \( y \) is a challenging ill-posed inverse problem. In the past a few decades, the IR problems have been extensively studied. However, they still remain as an active research area.

Generally, existing IR methods can be classified into two main categories, i.e., model-based methods [1], [8], [9], [13]–[18] and learning-based methods [19]–[24]. The model-based methods attack this problem by solving an optimization problem, which is often constructed from a Bayesian perspective. In the Bayesian setting, the solution is obtained by maximizing the posterior \( P(x|y) \), which can be formulated as:

\[ x = \arg\max_x \log P(x|y) = \arg\max_x \log P(y|x) + \log P(x), \]

where \( \log P(y|x) \) and \( \log P(x) \) denote the data likelihood and the prior terms, respectively. For additive Gaussian noise, \( P(y|x) \) corresponds to the \( \ell_2 \)-norm data fidelity term, and the prior term \( P(x) \) characterizes the prior knowledge of \( x \) in a probability setting. Formally, Eq. (1) can be rewritten as:

\[ x = \arg\min_x ||y - Ax||^2 + \lambda J(x), \]

where \( J(x) \) denotes the regularizer associated with the prior term \( P(x) \). Then, the desirable solution is the one that minimizes both the \( \ell_2 \)-norm data fidelity term and the regularization term weighted by parameter \( \lambda \). Clearly, the regularization term plays a critical role in searching for high-quality solutions. Numerous regularizers have been developed, ranging from the well-known total variation (TV) regularizer [13], the sparsity-base regularizers with off-the-shelf transforms or learned dictionaries [1], [3], [14], [15], to the nonlocal self-similarity (NLSS) inspired regularizers [2], [8], [25]. The TV regularizer is good at characterizing the piecewise constant signals but unable to model more complex image edges and textures. The sparsity-based techniques are more effective in representing local image structures with a few elemental structures (called atoms) from an off-the-shelf transformation matrix (e.g., DCT and Wavelets) or a learned dictionary. Indeed, the IR community has witnessed a flurry of sparsity-based IR methods [1], [3], [11], [15] in the past decade. Motivated by the fact that natural images often contain rich repetitive structures, nonlocal regularization techniques [2], [4], [5], [8] combining the NLSS with the sparse representation and low-rank approximation, have shown significant improvements over their local counterparts. Using those carefully designed prior, significant progresses of IR have been achieved. In addition to these explicitly regularized IR methods, denoising-based IR methods have also been proposed [26]–[30]. In these methods, the original optimization problem is decoupled into two separated subproblems: one for dealing with the data fidelity term and the other for the regularization term, yielding.

W. Dong, P. Wang, G. Shi, Fangfang Wu, and X. Lu are with School of Electronic Engineering, Xidian University, Xi’an, 710071, China (e-mail: wsdong@mail.xidian.edu.cn).
W. Yin is with Department of Mathematics, University of California, Los Angeles, CA 90095.
simpler optimization problems. Specifically, the subproblem related to the regularization is a pure denoising problem, and thus other more complex denoising methods that cannot be expressed as regularization terms can also be adopted, e.g., BM3D [2], NCSR [8] and GMM [16] methods.

Different from the model-based methods that rely on a carefully designed prior, the learning-based IR methods learn mapping functions to infer the missing high-frequency details or desirable high-quality images from the observed image. In the past decade, many learning-based image super-resolution methods [19], [20], [22], [24] have been proposed, where mapping functions from the low-resolution (LR) patches to high-resolution (HR) patches are learned. Inspired by the great successes of the deep convolution neural network (DCNN) for image classification [31], [32], the DCNN models have also been successfully applied to image IR tasks, e.g., SRCNN [22], FSRCNN [33] and VDSR [34] for image super-resolution, and TNRD [35] and DnCNN [24] for image denoising. In these methods, a DCNN is used to learn the mapping function from the degraded images to the original images. Due to its powerful representation ability, the DCNN based methods have shown better IR performances than conventional optimization-based IR methods in various IR tasks [22], [24], [34], [35]. Though training of DCNN is very expensive, testing the DCNN is much more efficient than previous optimization-based IR methods. Though the DCNN models have shown promising results, the DCNN methods lack flexibilities in adapting to different image recovery tasks, as the data likelihood term has not been explicitly exploited. To address this issue, hybrid IR methods that combine the optimization-based methods and DCNN denoisers have been proposed. In [36], a set of DCNN models are pre-trained for image denoising task and are integrated into the optimization-based IR framework for different IR tasks. Compared with other optimization-based methods, the integration of the DCNN models has advantages in exploiting the large training dataset and thus leads to superior IR performance. Similar idea has also been exploited in the autoencoder-based IR method [37], where denoising autoencoders are pre-trained as a natural image prior and a regularizer based on the pre-trained autoencoder is proposed. The resulting optimization problem is then iteratively solved by gradient descent. Despite the effectiveness of the methods [36], [37], they have to iteratively solve optimization problems, and thus their computational complexities are high. Moreover, the CNN and autoencoder models adopted in [36], [37] are pre-trained and cannot be jointly optimized with other algorithm parameters.

In this paper, we propose a denoising prior driven deep network to take advantages of both the optimization- and discriminative learning-based IR methods. First, we propose a denoising-based IR method, whose iterative process can be efficiently carried out. Then, we unfold the iterative process into a feed-forward neural network, whose layers mimic the process flow of the proposed denoising-based IR algorithm. Moreover, an effective DCNN denoiser that can exploit the multi-scale redundancies is proposed and plugged into the deep network. Through end-to-end training, both the DCNN denoisers and other network parameters can be jointly optimized. Experimental results show that the proposed method can achieve very competitive and often state-of-the-art results on several IR tasks, including image denoising, deblurring and super-resolution.

II. RELATED WORK

We briefly review the IR methods, i.e., the denoising-based IR methods and the discriminative learning-based IR methods, which are related to the proposed method.

A. Denoising-based IR methods

Instead of using an explicitly expressed regularizer, denoising-based IR methods [26] allow the use of a more complex image prior by decoupling the optimization problem of Eq. (2) into two subproblems, one for the data likelihood term and the other for the prior term. By introducing an auxiliary variable $v$, Eq. (2) can be rewritten as

$$\min_{x,v} \frac{1}{2} \| y - Ax \|_2^2 + \lambda J(v), \ s.t. \ x = v. \quad (3)$$

In [26], [30], the ADMM technique is used to convert the above equally constrained optimization problem into two subproblems

$$x^{(t+1)} = \arg\min_x \frac{1}{2} \| y - Ax \|_2^2 + \frac{\mu}{2} \| x - v^{(t)} + u^{(t)} \|_2^2,$n$$

$$v^{(t+1)} = \arg\min_v \frac{\mu}{2} \| x^{(t+1)} - v + u^{(t)} \|_2^2 + \lambda J(v), \quad (4)$$

where $u$ denotes the augmented Lagrange multiplier updated as $u^{(t+1)} = u^{(t)} + \rho(x^{(t+1)} - v^{(t+1)})$. The $x$-subproblem is a simple quadratic optimization that admits a closed-form solution as

$$x^{(t+1)} = (A^T A + \lambda I)^{-1} (A^T y + \lambda (v^{(t)} - u^{(t)})). \quad (5)$$

The immediately reconstructed image $x^{(t+1)}$ depends on both the observation model and a fixed estimate of $v$. The $v$-subproblem is also called the proximity operator of $J(v)$ computed at point $x^{(t+1)} + u^{(t)}$, whose solution can be obtained by a denoising algorithm. By alternatively updating $x$ and $v$ until convergence, the original optimization problem of Eq. (2) is then solved. The advantage of this framework is that other state-of-the-art denoising algorithms, which cannot be explicitly expressed in $J(x)$, can also be used to update $v$, leading to better IR performance. For example, the well-known BM3D [2], Gaussian mixture model [16], NCSR [8] have been used for various IR applications [26–28]. In [36], the state-of-the-art CNN denoiser has also been plugged as an image prior for general IR. Due to the excellent denoising ability, state-of-the-art IR results for different IR tasks have been obtained. Similar to [37], an autoencoder denoiser is plugged into the objective function of Eq. (2). However, different from the variable splitting method described above, the objective function of [37] is minimized by gradient descent. Though the denoising-based IR methods are very flexible and effective in exploiting state-of-the-art image prior, they require a lot of iterations for convergence and the whole components cannot be jointly optimized.
B. Deep network based IR methods

Inspired by the great success of DCNNs for image classification [31], [32], object detection [38], semantical segmentation [39], etc., DCNNs have also been applied for low-level image processing tasks [22], [24], [34], [35]. Similar to the coupled sparse coding [11], DCNNs have been used to learn nonlinear mapping from the LR patch space to the HR patch space [22]. By designing very deep CNNs, state-of-the-art image super-resolution results have been achieved [34]. Similar network structures have also been applied for image denoising [24] and also achieved state-of-the-art image denoising performance. For non-blind image deblurring, multi-player perceptron network [40] has been developed to remove the deconvolution artifacts. In [41], Xu et al. propose to use DCNN for non-blind image deblurring. Though excellent IR performances have been obtained, these DCNN methods generally treat the IR problems as denoising problems, i.e., removing the noise or artifacts of the initially recovered images, and ignore the observation models.

There has been some attempts to leverage the domain knowledge and the observation model for IR. In [23], based on the learned iterative shrinkage/thresholding algorithm (LISTA) [42], Wang et al. developed a deep network whose layers correspond to the steps of the sparse coding based image SR. In [35], the classic iterative nonlinear reaction diffusion method is also implemented as a deep network, whose parameters are jointly trained. The DNN inspired from the ADMM-based sparse coding algorithm has also been developed for compressive sensing based MRI reconstruction [43]. In [44], DNNs constructed from truncated iterative hard thresholding algorithm has also been developed for solving $l_0$-norm sparse recovery problem. These model-based DNNs have shown significant improvements in terms of both efficiency and effectiveness over original iterative algorithms. However, the strict implementations of the conventional sparse coding based methods result in a limited receive field of the convolutional filters and thus cannot exploit the spatial correlations of the feature maps effectively, leading to limited IR performance.

III. PROPOSED DENOISING-BASED IMAGE RESTORATION ALGORITHM

In this section, we develop an efficient iterative algorithm for solving the denoising-based IR methods, based on which a feed-forward DNN will be proposed in the next section. Considering the denoising-based IR problem of Eq. (3), we adopt the half-quadratic splitting method, by which the equally constrained optimization problem can be converted into a non-constrained optimization problem, as

$$\begin{align*}
(x, v) &= \arg \min_{x,v} \frac{1}{2} \| y - Ax \|^2 + \eta \| x - v^{(t)} \|^2 + \lambda J(v).
\end{align*}$$

(6)

The above optimization problem can be solved by alternatively solving two sub-problems,

$$\begin{align*}
x^{(t+1)} &= \arg \min_{x} \| y - Ax \|^2 + \eta \| x - v^{(t)} \|^2, \\
v^{(t+1)} &= \arg \min_{v} \eta \| x^{(t+1)} - v \|^2 + \lambda J(v).
\end{align*}$$

(7)

The $x$-subproblem is a quadratic optimization problem that can be solved in closed-form, as $x^{(t+1)} = W^{-1}b$, where $W$ is a matrix related to the degradation matrix $A$. Generally, $W$ is very large, so it is impossible to compute its inverse matrix. Instead, the iterative classic conjugate gradient (CG) algorithm can be used to compute $x^{(t+1)}$, which requires many iterations for computing $x^{(t+1)}$. In this paper, instead of solving for an exact solution of the $x$-subproblem, we propose to compute $x^{(t+1)}$ with a single step of gradient descent for an inexact solution, as

$$\begin{align*}
x^{(t+1)} = x^t - \Delta A^\top (A x^t - y) + \eta (x^t - v^t) \\
&= A x^t + \Delta A^\top y + \eta v^t,
\end{align*}$$

(8)

where $\Delta = [(1 - \eta \Delta) I - \Delta A^\top A]$ and $\eta$ is the parameter controlling the step size. By pre-computing $A$, the update of $x^t$ can be computed very efficiently. As will be shown later, we do not have to solve the $x$-subproblem exactly. Updating $x^{(t+1)}$ once is sufficient for $x^t$ to converge to a local optimal solution. The $v$-subproblem is a proximity operator of $J(v)$ computed at point $x^{(t+1)}$, whose solution can be obtained by a denoiser, i.e., $v^{(t+1)} = f(x^{(t+1)})$, where $f(\cdot)$ denotes a denoiser. Various denoising algorithms can be used, including that cannot be explicitly expressed by the MAP estimator with $J(x)$. In this paper, inspired by the success of DCNN for image denoising, we choose a DCNN-based denoiser to exploit the large training dataset. However, different from existing DCNN models for IR, we consider the network that can exploit the multi-scale redundancies of natural images, as will be described in the next section. In summary, the proposed iterative algorithm for solving the denoising-based IR problems is summarized in Algorithm 1. We now discuss the convergence property of Algorithm 1.

Algorithm 1 Denoising-based IR Algorithm

- **Initialization:**
  1. Set observation matrix $A, \Delta, \eta > 0, \eta > 0, t = 0$;
  2. Initialize $x$ as $x^{(0)} = A^\top y, v^{(0)} = 0$;

- **While** not converge **do**
  1. Compute $x^{(t+1)} = A x^{(t)} + \Delta A^\top y + \eta v^{(t)}$
  2. Compute $v^{(t+1)} = f(x^{(t+1)})$

**End while**

Output: $x^{(t)}$

**Theorem 1.** Consider the energy function

$$\xi(x, v) := \frac{1}{2} \| y - Ax \|^2 \| v \|_2^2 + \frac{\eta}{2} \| x - v \|^2_2 + \lambda J(v).$$

Assume that $\xi$ is lower bounded and coercive$^1$. For Algorithm 1, $(x^{(t)}, v^{(t)})$ has a subsequence that converges to a stationary point of the the energy function provided that the denoiser $f(\cdot)$ satisfies the sufficient descent condition:

$$\begin{align*}
\eta \| x - v \|^2_2 + \lambda J(v) - \frac{\eta}{2} \| x - f(x) \|^2_2 - \lambda J(f(x)) \\
&\geq c_2 \| \nabla_v \xi(x, v) \|^2_2,
\end{align*}$$

(9)

$^1$ $\xi(x, v) \to \infty$ whenever $\| (x, v) \| \to \infty$. 

where \( c_2 > 0 \) and \( \nabla_v \xi(x, \cdot) \) is a continuous limiting subgradient of \( \xi \).

**Proof** See the Appendix.

Let us discuss the condition (9). We list some combinations of the function \( J \) and mapping \( f \) that satisfy (9):

1) \( J \) is \( L \)-Lipschitz differentiable, and \( f : (x, v) \mapsto v - \alpha \nabla_v \xi(x, v) \) is a gradient descent map, where \( \alpha \in (0, \frac{1}{L^2}) \) if \( \xi(x, v) \) is convex in \( v \) or \( \alpha \in (0, \frac{1}{L}) \) otherwise. Then, (9) follows from standard gradient analysis.

2) \( J \) is proper and lower semi-continuous, the function \( \xi'(u; x, v) := \frac{\mu}{2} ||x - u||^2_2 + \lambda J(u) + \frac{\beta}{2} ||v - u||^2_2 \) is at least \( \beta \)-strongly convex in \( u \), and \( f : (x, v) \mapsto v^+ := \arg\min_u \xi'(u; x, v) \). This \( f \) is known as the proximal mapping of \( \frac{\mu}{2} ||x - \cdot||^2_2 + J(\cdot) \). The properties of \( J \) ensures \( v^+ \) to be well defined. Then, by convexity and optimality condition of the "argmin" subproblem,

\[
\frac{\mu}{2} ||x - v^+||^2_2 + \lambda J(v^+) - \frac{\mu}{2} ||x - v||^2_2 - \lambda J(v) \\
\geq \beta ||v - v^+||^2_2 = \frac{1}{\beta} \mu ||x - v^+||^2_2 + \lambda \nabla_v J(v^+) ||^2_2 \\
= \frac{1}{\beta} ||\nabla_v \xi(x, v^+)||^2_2.
\]

This is different from (9) since the right-hand side uses \( v^+ \) rather than \( v \). However, applying the right-hand side term \( ||v - v^+||^2_2 \) in the proof yields \( \lim ||v(t) - v^+(t+1)||^2_2 = 0 \) and thus (9) is satisfied asymptotically and the proof results still apply.

3) Let \( \mathcal{M} \) denote a manifold of (noisless) images and \( J(v) := \text{dist}(v, \mathcal{M})^2 \) to be a function that measures a certain kind of squared distance between \( v \) and \( \mathcal{M} \). In particular, consider the squared Euclidean distance \( J(v) = \frac{1}{2} ||v - \Pi_{\mathcal{M}}(v)||^2_2 \), where \( \Pi_{\mathcal{M}}(v) \) denotes orthogonal projection of \( v \) to \( \mathcal{M} \). Then, for \( f(x) := \arg\min_u \frac{\mu}{2} ||x - u||^2_2 + \frac{1}{2} ||v - \Pi_{\mathcal{M}}(u)||^2_2 + \frac{\beta}{2} ||v - u||^2_2 \), we have \( f(x) = \frac{1}{\sqrt{\lambda + \beta \mu}} (\mu x + \beta v + \lambda \Pi_{\mathcal{M}}(\mu x + \beta v)) \).

Similar to the last point, we have (10) and thus (9) asymptotically.

4) For the same \( J \) in the last part, define \( J(x) = \delta_{\mathcal{M}}(x) \), which returns 0 if \( x \in \mathcal{M} \) and \( \infty \) if \( x \notin \mathcal{M} \). If the manifold \( \mathcal{M} \) is bounded and differentiable, then \( J(x) \) is known as restricted prox-regular. For \( f(x) := \arg\min_u \left( \frac{\mu}{2} ||x - u||^2_2 + \delta_{\mathcal{M}}(x) + \frac{\beta}{2} ||v - u||^2_2 \right) \), it is discussed in [45] that (10) holds and thus (9) holds in the asymptotic sense.

In parts 2–4 above, we can remove the proximity term \( \frac{\beta}{2} ||v - u||^2_2 \), which is used in defining the mapping \( f \), and still ensure the same result, i.e., subsequence convergence to a stationary point. However, the proof must be adapted to each \( J(v) \) separately. We leave this to our future work.

It has been shown in [46] that if \( \xi \) has the Kurdyka-Lojasiewicz property, the subsequence convergence can be upgraded to the convergence of full sequence, which has been a standard argument in recent convergence analysis. As shown in [47], functions satisfying the KL property include, but not limited to, real analytic functions, semi-algebraic functions, and locally strongly convex functions. Therefore, \((x^{(t)}, v^{(t)})\) converges to a stationary point. It is possible that the stationary point \((x^*, v^*)\) is a saddle point rather than a local minimizer. However, it is known that first-order methods almost always avoid saddle points assuming the initial solution is randomly selected [48]. Therefore, converging to a saddle point is extremely unlikely.

It has been shown in [49] that the denoiser autoencoder can be regarded as an approximately orthogonal projection of the noisy input \( y \) to the manifold of noisless images. Therefore, as shown in the above parts 2 and 3, Algorithm 1 with the mapping function \( f(\cdot) \) defined by the DCNN denoiser in a loose sense converges to a local minimizer, based on the above analysis.

**IV. Denoising Prior Driven Deep Neural Network**

In general, Algorithm 1 requires many iterations to converge and is computationally expensive. Moreover, the parameters and the denoiser cannot be jointly optimized in an end-to-end training manner. To address these issue, here we propose to unfold the Algorithm 1 into a deep network of the architecture shown in Fig. 1 (a). The network exactly executes \( K \) iterations of Algorithm 1. The input degraded image \( y \in \mathbb{R}^{nu} \) first goes through a linear layer parameterized by the degradation matrix \( A \in \mathbb{R}^{nu \times ms} \) for an initial estimate \( x^{(0)} \). \( x^{(0)} \) is then fed into the linear layer parameterized by matrix \( A \in \mathbb{R}^{ms \times ms} \), whose output is added with \( x^{(0)} \) weighted by \( \delta_0 \) via a shortcut connection. The updated \( x^{(1)} \) is fed into the denoising module, whose structure is shown in Fig. 1 (b). The denoised signal \( v^{(1)} \) is fed into the linear layer parameterized by \( A \), whose output is further added with \( x^{(0)} \) and \( v^{(1)} \) via two shortcut connections for the updated \( x^{(2)} \). Such a process is repeated \( K \) times. In our implementation, \( K = 6 \) was always used. Instead of using fixed weights, all the weights \((\delta_0, \delta_{k,1}, \delta_{k,2}, k = 1, 2, \ldots, K) \) involved in the \( K \) recurrent stages can be discriminatively learned through end-to-end training. Regarding the denoising module, as we are using a DCNN-based denoiser that contains a large number of parameters, we enforce all the denoising modules to share the same parameters to avoid over-fitting.

The linear layers \( A^T \) and \( A \) are also trainable for a typical degradation matrix \( A \). For image denoising, \( A = A^T = I \), and \( A \) also reduces to a weighted identity matrix \( \hat{A} = \lambda I \), where \( \lambda = 1 - \delta(1 + \eta) \). For image deblurring, the layer \( A^T \) can be simply implemented with a convolutional layer. The layer \( \hat{A} = -\delta A^T \) can also be computed efficiently by convolutional operations. The weight \( A \) and filters correspond to \( A^T \) and \( A \) can also be discriminatively learned. For image super-resolution, two types of degradation operators are considered: the Gaussian downsampling and the bicubic downsampling. For Gaussian downsampling, \( A = DH \), where \( H \) and \( D \) denote the Gaussian blur matrix and the downsampling matrix, respectively. In this case, the layer \( A^T = H^T D^T \) corresponds to first upsample the input LR image by zero-padding and then convolute the upsampled image with a filter. Layer \( A \) can also be efficiently computed with convolution, downsampling and upsampling operations. All convolutional filters involved in these operations can be discriminatively learned. For bicubic
downsampling, we simply use the bicubic interpolator function with scaling factor \( s \) and \( 1/s \) (\( s = 2, 3, 4 \)) to implement the matrix-vector multiplications \( A^\top y \) and \( Ax \), respectively.

A. The DCNN denoiser

Inspired by the recent advances on semantical segmentation [39] and object segmentation [50], the architecture of the denoising network is illustrated in Fig. 1(b). Similar to the U-net [51] and the sharpMask net [50], the proposed network contains two parts: the feature encoding and the decoding parts. In the feature encoding part, there are a series of convolutional layers followed by pooling layers to reduce the spatial resolution of the feature maps. The pooling layer helps increase the receipt field of the neurons. In the feature encoding stage, all the convolutional layers are grouped into \( L \) feature extraction blocks (\( L = 4 \) in our implementation), as shown by the blue blocks in Fig. 1(b). Each block contains four convolutional layers with ReLU nonlinearity and \( 3 \times 3 \) kernels. The first three layers generate 64-channel feature maps, while the last layer doubles the number of channels followed by a pooling layer to reduce the spatial resolution of the feature maps with scaling factor 0.5. In the pooling layers, the feature maps are first convoluted with \( 2 \times 2 \) kernels and then subsampled by a scaling factor of 2 along both axes.

The feature decoding part also contains a series of convolutional layers, which are also grouped into four blocks followed by an upsampling layer to increase the spatial resolution of the feature maps. As the finally extracted feature maps lose a lot of spatial information, directly reconstructing images from the finally extracted features cannot recover fine image details. To address this issue, the feature maps of the same spatial resolution generated in the encoding stage are fused with the upsampled feature maps generated in the decoding stage, for obtaining newly upsampled feature maps. Each reconstruction block also consists of four convolutional layers with ReLU nonlinearity and \( 3 \times 3 \) kernels. In each reconstruction block, the first three layers produce 128-channels feature maps and the fourth layer generate 512-channels feature maps, whose spatial resolutions are upsampled with a scaling factor of 2 by a deconvolution layer. The upsampled feature maps are then fused with the feature maps of the same spatial resolution from the encoding part. Specifically, the fusion is conducted by concatenating the feature maps. The last feature decoding block reconstructed the output image. A skip connection from the input image to the reconstructed image is added to enforce the denoising network to predict the residuals, which has been verified to be more robust [24].

B. Overall network training

Note that the DCNN denoisers do not have to be pre-trained. Instead, the overall deep network shown in Fig. 1 (a) is trained by end-to-end training. To reduce the number of parameters and thus avoid over-fitting, we enforce each DCNN denoiser to share the same parameters. Mean square error (MSE) based loss function is adopted to train the proposed deep network, which can be expressed as

\[
\Theta = \arg\min_{\Theta} \sum_{i=1}^{N} ||F(y_i; \Theta) - x_i||_2^2, \tag{11}
\]

where \( y_i \) and \( x_i \) denote the \( i \)-th pair of degraded and original image patches, respectively, and \( F(y_i; \Theta) \) denotes the reconstructed image patch by the network with parameter set \( \Theta \). It is also possible to train the network with other perceptual based loss functions, which may lead to better visual quality. We remain this as future work. The ADAM optimizer [52] is used to train the network with setting \( \beta_1 = 0.9, \beta_2 = 0.999 \) and \( \epsilon = 10^{-8} \). The learning rate is initialize as \( 10^{-4} \) and halved at every \( 2 \times 10^5 \) minibatch updates. The proposed network is implemented with framework and trained using 4 Nvidia Titan X GPUs, taking about one day to converge.

V. EXPERIMENTAL RESULTS

In this section, we perform several IR tasks to verify the performance of the proposed network, including image denoising, deblurring, and super-resolution. We trained each model for different IR tasks. We empirically found that implementing \( K = 5 \) iterations of Algorithm 1 in the network generally lead to satisfied IR results for image denoising, deblurring and super-resolution tasks. Thus, we fixed \( K = 5 \) for all IR tasks. To train the networks, we constructed a large training image set, consisting of 1000 images of size \( 256 \times 256 \) used in [6].

A. Image denoising

For image denoising, \( A = I \) and Algorithm 1 reduce to the iterative denoising process, i.e., the weighted noise image is added back to the denoised image for the next denoising...
process. Such iterative denoising has shown improvements over conventional denoising methods that only denoise once [3]. Here, we also found that implementing multiple denoising iterations in the proposed network improves the denoising results. To train the network, we extracted image patches of size 40 × 40 from the training images and added additive Gaussian noise to the extracted patches to generate the noisy patches. Totally $N = 450,000$ patches were extracted for training. Note that none of the test images was included into the training image set. The training patches were also augmented by flip and rotations. We compared the proposed network with several leading denoising methods, including three model-based denoising methods, i.e., BM3D method [2], the EPLL method [16], and the low-rank based method WNNM method [5], and two deep learning based methods, i.e., the TNRD method [35] and the DnCNN-S method [24].

Table I shows the PSNR results of the competing methods on a set of commonly used test images shown in Fig. 2. It can be seen that both the DnCNN-S and the proposed network outperform other methods. For most of the test images and noise levels, the proposed network outperforms the DnCNN-S method, which is the current state-of-the-art denoising method. On average, the PSNR gain over DnCNN-S can be up to 0.32 dB. To further verify the effectiveness of the proposed method, we also employ the Berkeley segmentation dataset (BSD68) that contains 68 natural images for comparison study. Table II shows the average PSNR and SSIM results of the test methods on BSD68. One can see that the PSNR gains over the other test methods become even larger for higher noise levels. The proposed method outperforms the DnCNN-S method by up to 0.78 dB on average on the BSD68, demonstrating the effectiveness of the proposed method. Parts of the denoised images by the test methods are shown in Figs. 3-4. One can see that the image edges and textures recovered by model-based methods, i.e., BM3D, WNNM and EPLL are over-smoothed. The deep learning based methods, TNRD, DnCNN-S and the proposed method produce much more visually pleasant image structures. Moreover, the proposed method generates even better results in recovering more details than TNRD and DnCNN-S.

### B. Image deblurring

To train the proposed network for image deblurring, we first convoluted the training images with a blur kernel to generate the blurred images and then extracted the training image patches of size $120 \times 120$ from the blurred images. The additive Gaussian noise of standard deviation $\sigma_n$ was also added to the blurred images. Patch augmentation with flips and rotations were adopted, generating total 450,000 patches for training. Two types of blur kernels were considered, i.e., the $25 \times 25$ Gaussian blur kernel of standard deviation 1.6 and two motion blur kernels adopted in [53] of sizes $19 \times 19$ and $17 \times 17$. We trained each model for different blur settings. We compared the proposed method with several leading deblurring methods, i.e., three leading model-based deblurring methods (EPLL [16], IDDBM3D [7] and NCSR [8]) and the current state-of-the-art denoising-based deblurring methods with CNN denoisers [36] (denoted as DD-CNN). The test images involved in this comparison study are shown in Fig. 5. In this experiment, we only conduct deconvolution for grayscale images. However, the proposed method can be easily extended for color image deblurring.

The PSNR results of the test deblurring methods are reported in Table III. For fair comparisons, all the PSNRs of the other methods are generated by the codes released by the authors or directly written according to their papers. From Table III, we can see that the DD-CNN method performs much better than conventional model-based EPLL, IDDBM3D and NCSR methods. For Gaussian blur, the proposed method outperforms DD-CNN by 0.27 dB on average. For other motion blur kernels with higher noise levels, the proposed method is slightly worse than DD-CNN method. Parts of the deblurred images by the competing methods are shown in Figs. 6-8. From Figs. 6-8, one can see that the proposed method not only produces more sharper edges but also recovers more details than the other methods.

### C. Image super-resolution

For image super-resolution, we consider two image subsampling operators, i.e., the bicubic downsampling and the Gaussian downsampling. For the former case, the HR images are downsampled by applying the bicubic interpolation function with scaling factor $1/s$ ($s = 2, 3, 4$) to simulate the LR images. For the latter case, the LR images are generated by applying the Gaussian blur kernel to the original images followed by subsampling. The $7 \times 7$ Gaussian blur kernel of standard deviation of 1.6 is used in this case. The LR/HR patch pairs are extracted from the LR/HR training image pairs and augmented by flip and rotations, generating 450,000 patch pairs. The LR patch size is 32 × 32, while the HR patch size is $32 \times s \times 32 \times s$. We train each network for the two downsampling cases. The image data sets commonly used in the image super-resolution (SR) literature are adopted for performance verification, including the set5, set14, the Berkeley segmentation dataset containing 100 images (denoted as BSD100), the Urban 100 dataset [34] containing 100 high-quality images. We compared the proposed method with several leading image SR methods, including two DCNN based SR methods (SRCNN [22] and VDSR [34]) and two denoising methods (TNRD [35] and DnCNN [24]), which produce the HR images by first upsampling the LR images with the bicubic interpolator and then denoising the upsampled images to recover the high-frequency details. For fair comparisons, the results of the others are directly borrowed from their papers or generated by the codes released by the authors.

The PSNR results of the test methods for bicubic downsampling are reported in Tables IV-V, from which we can see that the proposed method outperforms other competing methods. We observed that the PSNR gains over other methods becomes larger for large scaling factors, verifying the importance of observation consistencies for IR. The PSNR results of the test methods for Gaussian downsampling with scaling factor 3 are reported in Table VI. For this case, we compare the proposed
method with DD-CNN [36], which has much better results than their earlier DnCNN [24]. Since VDSR and SRCNN are trained for bicubic downsampling, it is unfair to directly apply these methods to the LR images generated by Gaussian downsampling and thus we didn’t include their results into this table. Parts of the reconstructed HR images by the test methods are shown in Fig. 9-11, from which we can see that the proposed method can produce sharper edges than other methods.

VI. CONCLUSION

In this paper, we have proposed a novel deep neural network for general image restoration (IR) tasks. Different from current deep network based IR methods, where the observation models are generally ignored, we construct the deep network based on a denoising-based IR framework. To this end, we first developed an efficient algorithm for solving the denoising-based IR method and then unfolded the algorithm into a deep network, which is composed of multiple denoising modules interleaved with back-projection modules for data consistencies. A DCNN-based denoiser exploiting multi-scale redundancies of natural images was developed. Therefore, the proposed deep network can exploit not only the effective

DCNN denoising prior but also the prior of the observation model. Experimental results show that the proposed method can achieve very competitive and often state-of-the-art results on several IR tasks, including image denoising, deblurring and super-resolution.

APPENDIX

CONVERGENCE

Theorem 2. Consider the energy function

\[
\xi(x, v) := \frac{1}{2} \| y - Ax \|^2_2 + \frac{\eta}{2} \| x - v \|^2_2 + \lambda J(v).
\]

Assume that \( \xi \) is lower bounded and coercive$^2$. For Algorithm I, \((x^{(t)}, v^{(t)})\) has a subsequence that converges to a stationary point of the energy function provided that the denoiser \( f(\cdot) \) satisfies the sufficient descent condition:

\[
\frac{\eta}{2} \| x - v \|^2_2 + \lambda J(v) - \frac{\eta}{2} \| x - f(x) \|^2_2 - \lambda J(f(x)) \geq c_2 \| \nabla_v \xi(x, v) \|^2_2,
\]

where \( c_2 > 0 \) and \( \nabla_v \xi(x, \cdot) \) is a continuous limiting subgradient of \( \xi \).

$^2$\( \xi(x, v) \to \infty \) whenever \( \| (x, v) \| \to \infty \).
ξ is lower bounded, adding (9) and (10) yields the contradiction since ξ is Lipschitz continuous with constant ∥A²A∥, it is well known that the gradient step on x with step size δ ∈ (0, 2∥A²A∥/η) satisfies the descent property

\[
ξ(x(t), v(t)) - ξ(x(t+1), v(t+1)) ≥ c_1 ∥x(t) - x(t+1)∥_2^2,
\]

where c₁ := 1/2 - ∥A²A∥/2 > 0. By assumption, the v-step satisfies

\[
ξ(x(t+1), v(t)) - ξ(x(t+1), v(t+1)) ≥ c_2 ∥∇_v ξ(x(t+1), v(t))∥_2^2.
\]

Since ξ(x, v) is coercive and, by (13) and (14), ξ(x(t), v(t)) is monotonically nonincreasing, the sequence (x(t), v(t))_{t=0,1,2,...} is bounded (otherwise, it would cause the contradiction ξ(x(t), v(t)) → ∞), so it has a convergent subsequence (x(t_k), v(t_k))_{k=0,1,...} → (x*, v*). Since ξ(x, v) is lower bounded, adding (9) and (10) yields

\[
ξ(x(t), v(t)) - ξ(x(t+1), v(t+1)) ≥ c_1 ∥x(t) - x(t+1)∥_2^2 + c_2 ∥∇_v ξ(x(t+1), v(t))∥_2^2.
\]

and, by telescopic sum over t = 0, 1, ..., and by monotonicity and boundedness of ξ(x(t), v(t)), we have the summability properties \(\sum_t ∥x(t) - x(t+1)∥_2^2 < ∞\) and \(\sum_t ∥∇_v ξ(x(t+1), v(t))∥_2^2 < ∞\), from which we conclude

\[
\lim_{t→∞} ∥x(t) - x(t+1)∥_2 = 0,
\]

\[
\lim_{t→∞} ∥∇_v ξ(x(t+1), v(t))∥_2 = 0.
\]

Based on \(x(t+1) - x(t) = δ ∇_v ξ(x(t), v(t))\), we get

\[
∇_v ξ(x*, v*) = \lim_{k→∞} ∇_v ξ(x(t_k), v(t_k)) = 0,
\]

where we have used the continuity of \(∇_v ξ(x, v)\) in x. Also, \(\lim_{k→∞} ∇_v ξ(x(t_k), v(t_k)) = lim_{k→∞} ∇_v ξ(x(t_k+1), v(t_k)) = 0\), where the first “=” follows from the continuity of \(∇_v ξ(x, v) = 2 μ (v - x) + λ ∇ J(v)\) in x and (16). Therefore, \((x*, v*)\) is a stationary point of ξ.

**REFERENCES**

[1] M. Elad and M. Aharon, “Image denoising via sparse and redundant representation over learned dictionaries,” *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[2] 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.

[3] W. Dong, X. Li, L. Zhang, and G. Shi, “Sparsity-based image denoising via dictionary learning and structural clustering,” in *Proc. of the IEEE CVPR*, 2011, pp. 457–464.

[4] W. Dong, G. Shi, and X. Li, “Nonlocal image restoration with bilateral variance estimation: a low-rank approach,” *IEEE Transactions on image processing*, vol. 22, no. 2, pp. 700–711, 2013.
Fig. 4: Denoising results for Lena image with noise level 50. The PSNR results: (b) BM3D [2] (29.05 dB); (c) WNNM [5] (29.25 dB); (d) TNRD [35] (28.93 dB); (e) DnCNN-S [24] (29.37 dB); (f) Ours (29.85 dB).

Fig. 5: The test images used for image deblurring.

[5] W. Dong, X. Li, L. Zhang, and G. Shi, “Weighted nuclear norm minimization with application to image denoising,” in Proc. of the IEEE CVPR, 2014, pp. 2862–2869.
[6] W. Dong, L. Zhang, G. Shi, and X. Wu, “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization,” IEEE Transactions on Image Processing, vol. 20, no. 7, pp. 1838–1857, 2011.
[7] A. Danielyan, V. Katkovnik, and K. Egiazarian, “Bm3d frames and variational image deblurring,” IEEE Transactions on Image Processing, vol. 21, no. 4, pp. 1715–1728, 2012.
[8] W. Dong, L. Zhang, G. Shi, and X. Li, “Nonlocally centralized sparse representation for image restoration,” IEEE Transactions on Image Processing, vol. 22, no. 4, pp. 1620–1630, 2013.
[9] W. Dong, G. Shi, Y. Ma, and X. Li, “Image restoration via simultaneous sparse coding: Where structured sparsity meets gaussian scale mixture,” International Journal of Computer Vision, pp. 1–16, 2015.
[10] A. Marquina and S. J. Osher, “Image super-resolution by tv-regularization and bregman iteration,” Journal of Scientific Computing, vol. 37, no. 3, pp. 367–382, 2008.
[11] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in Proc. of the IEEE CVPR, 2008, pp. 1–8.
[12] X. Gao, K. Zhang, D. Tao, and X. Li, “Image super-resolution with sparse neighbor embedding,” Image Processing, IEEE Transactions on, vol. 21, no. 7, pp. 3194–3205, 2012.
[13] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, “An iterative regularization method for total variation-based image restoration,” Multiscale Modeling and Simulation, vol. 4, no. 2, pp. 460–489, 2005.
[14] J. M. Bioucas-Dias and M. A. Figueiredo, “A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration,” IEEE Transactions on Image Processing, vol. 16, no. 12, pp. 2992–3004, 2007.
[15] J. Mairal, M. Elad, and G. Sapiro, “Sparse representation for color image restoration,” IEEE Transactions on Image Processing, vol. 17, no. 1, pp. 53–69, 2008.
[16] D. Zoran and Y. Weiss, “From learning models of natural image patches to whole image restoration,” in Proc. of the IEEE ICCV, 2011, pp. 479–486.
[17] S. Roth and M. J. Black, “Fields of experts,” International Journal of Computer Vision, vol. 82, no. 2, pp. 205–229, 2009.
[18] G. Yu, G. Sapiro, and S. Mallat, “Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity,” IEEE Transactions on Image Processing, vol. 21, no. 5, pp. 2481–2499, 2012.
[19] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution,” Computer Graphics and Applications, vol. 22, no. 2, pp. 56–65, 2002.
[20] R. Timofte, V. De Smet, and L. Van Gool, “A+- Adjusted anchored neighborhood regression for fast super-resolution,” in Asian Conference...
TABLE III: The PSNR results of the deblurred images by the test methods.

| Methods | $\sigma$ | Butterfly | Peppers | Parrot | starfish | Barbara | Boats | C.Man | House | Leaves | Lena | Average |
|---------|---------|-----------|---------|--------|----------|---------|-------|-------|-------|--------|------|---------|
| IDD-BM3D |       | 29.79 | 29.64 | 31.90 | 30.57 | 25.99 | 31.17 | 27.68 | 33.56 | 30.13 | 30.91 | 30.13 |
| EPLL    | 2      | 29.78 | 26.73 | 31.32 | 28.52 | 24.22 | 28.84 | 26.57 | 31.76 | 25.29 | 29.46 | 27.85 |
| NCSR    | 2.55   | 29.72 | 30.04 | 32.07 | 30.83 | 26.54 | 31.22 | 27.99 | 33.38 | 30.13 | 30.99 | 30.29 |
| DD-CNN  |        | 30.44 | **30.69** | 31.83 | 30.78 | 26.13 | 31.41 | 28.05 | 33.80 | **30.44** | 31.05 | 30.48 |
| Ours    | **30.67** | 30.18 | **32.40** | **32.00** | 26.47 | **31.54** | **28.24** | **34.25** | 30.23 | **31.48** | **30.75** |       |

| Methods | $\sigma$ | Butterfly | Peppers | Parrot | starfish | Barbara | Boats | C.Man | House | Leaves | Lena | Average |
|---------|---------|-----------|---------|--------|----------|---------|-------|-------|-------|--------|------|---------|
| EPLL    | 2.55   | 26.23 | 27.40 | 33.78 | 29.79 | 29.78 | 30.15 | 30.24 | 31.73 | 25.84 | 31.37 | 29.63 |
| EPLL    | 7.65   | 32.33 | 32.00 | 34.48 | 32.26 | 32.38 | 33.05 | 31.50 | 34.89 | 33.29 | 33.54 | 32.96 |
| Ours    | **32.58** | 32.05 | **34.98** | **32.71** | 32.59 | **33.39** | **31.70** | **35.34** | **32.99** | **33.01** | **33.19** |       |
| EPLL    | 2.55   | 24.27 | 26.15 | 30.01 | 26.81 | 26.95 | 27.72 | 27.57 | 29.89 | 23.81 | 31.37 | 29.63 |
| EPLL    | 7.65   | 28.51 | **28.88** | 31.07 | 27.86 | **28.18** | **29.13** | **32.03** | **28.42** | **29.52** | **29.17** |       |
| Ours    | **28.24** | 28.42 | 31.03 | **28.00** | 28.01 | **29.19** | **27.77** | **32.06** | 27.98 | 29.42 | 29.01 |       |

| Methods | $\sigma$ | Butterfly | Peppers | Parrot | starfish | Barbara | Boats | C.Man | House | Leaves | Lena | Average |
|---------|---------|-----------|---------|--------|----------|---------|-------|-------|-------|--------|------|---------|
| EPLL    | 2      | 26.48 | 27.37 | 33.88 | 29.56 | 28.29 | 29.61 | 29.66 | 32.97 | 25.69 | 30.67 | 29.42 |
| EPLL    | 7.65   | 31.97 | **31.89** | 34.46 | 32.18 | **32.00** | **33.06** | **31.29** | 34.82 | **32.96** | **33.35** | **32.80** |
| Ours    | **31.86** | 31.38 | 34.72 | **32.28** | 31.36 | 32.86 | 31.21 | **35.09** | 32.29 | **33.35** | **32.64** |       |
| EPLL    | 2.55   | 24.27 | 26.15 | 30.01 | 26.81 | 26.95 | 27.72 | 27.57 | 29.89 | 23.81 | 31.37 | 29.63 |
| EPLL    | 7.65   | 28.21 | **28.71** | **30.68** | 27.67 | **28.95** | **27.70** | **31.95** | **27.92** | **29.27** | **28.84** |       |
| Ours    | **27.47** | 28.02 | **30.46** | **27.82** | 26.86 | 28.84 | 27.48 | **31.91** | **27.28** | **29.25** | **28.54** |       |

(a) Original (b) IDDBM3D (c) EPLL (d) NCSR (e) DD-CNN (f) Ours

Fig. 6: Deblurring results for Cameraman image with $25 \times 25$ Gaussian blur kernel and $\sigma_n = 2$. (a) Original image; (b) IDD-DM3D [7] (27.68 dB); (C) EPLL denoiser [16] (26.57 dB); (d) NCSR [8] (27.99 dB); (e) DD-CNN [36] (28.05 dB); (f) Ours (28.24 dB).

[21] U. Schmidt and S. Roth, “Shrinkage fields for effective image restoration,” in Proc. of the IEEE CVPR, 2014, pp. 2774–2781.
[22] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in European Conference on Computer Vision. Springer, 2014, pp. 184–199.
[23] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, “Deep networks for image super-resolution with sparse prior,” in Proc. of the IEEE ICCV, 2015, pp. 370–378.
[24] 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.
[25] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in Proc. of the IEEE CVPR, 2005, pp. 60–65.
[26] S. Venkatakrishnan, C. Bouman, E. Chu, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in Proc. of IEEE Global Conference on Signal and Information Processing, 2013, pp. 945–948.
[27] A. Brifman, Y. Romano, and M. Elad, “Turning a denoiser into a super-resolver using plug and play priors,” in Proc. of the IEEE ICIP, 2016, pp. 1404–1408.
[28] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, “Image restoration and reconstruction using variable splitting and class-adapted
Fig. 7: Deblurring results for *house* image with $19 \times 19$ motion blur kernel 1 and $\sigma_n = 2.55$. (a) Original image; (b) EPLL [16] (31.73 dB); (c) DD-CNN [36] (34.89 dB); (d) Ours (35.34 dB).

Fig. 8: Deblurring results for *lena* image with $19 \times 19$ motion blur kernel 1 and $\sigma_n = 2.55$. (a) Original image; (b) EPLL [16] (31.37 dB); (c) DD-CNN [36] (33.54 dB); (d) Ours (33.80 dB).

TABLE IV: PSNR results of the reconstructed HR images by the test SR methods on set5 for bicubic downsampling.

| Images | Scaling factor | TNRD | SRCNN | VDSR | DnCNN | Ours |
|--------|----------------|------|-------|------|-------|------|
| Baby   | 2              | 38.53| 38.54 | 38.75| 38.62 | **38.88** |
| Bird   |                | 41.31| 40.91 | 42.42| 42.20 | **42.99** |
| Butterfly |            | 33.17| 32.75 | 34.49| 34.51 | **34.72** |
| Head   |                | 35.75| 35.72 | 35.93| 35.84 | **36.00** |
| Woman  |                | 35.30| 35.31 | 36.03| 35.99 | **36.24** |
| Average|                | 36.85| 36.66 | 37.33| 37.00 | **37.75** |
| Baby   | 3              | 35.28| 35.25 | 35.38| 35.42 | **35.57** |
| Bird   |                | 36.09| 35.48 | 36.66| 36.61 | **37.51** |
| Butterfly |            | 28.92| 27.95 | **29.96** | 29.95 | 29.81 |
| Head   |                | 33.75| 33.71 | 33.96| 33.92 | **34.03** |
| Woman  |                | 31.79| 31.77 | 32.35| 32.31 | **32.71** |
| Average|                | 33.17| 32.73 | 33.06| 33.64 | **33.95** |
| Baby   | 4              | 31.30| 31.13 | 33.41| 33.23 | **33.64** |
| Bird   |                | 32.99| 32.52 | 33.54| 33.06 | **34.09** |
| Butterfly |            | 26.22| 25.46 | 27.28| 26.94 | **27.68** |
| Head   |                | 32.51| 32.44 | 32.70| 32.36 | **32.88** |
| Woman  |                | 29.20| 28.89 | 29.81| 29.46 | **30.34** |
| Average|                | 30.85| 30.48 | 31.35| 31.01 | **31.72** |

image priors,” in *Proc. of IEEE ICIP*, 2016, pp. 3518–3522.

[29] M. E. Y. Romano 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.

[30] S. H. Chan, X. Wang, and O. A. Elgendy, “Plug-and-play admm for image restoration: Fixed-point convergence and applications,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. of the IEEE CVPR*, 2016, pp. 770–778.

[33] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *European Conference on Computer Vision*. Springer, 2016, pp. 391–407.

[34] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[35] 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.

[36] K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep cnn denoiser prior for image restoration,” in *Proc. of the IEEE CVPR*, 2017, pp. 2808–2817.

[37] S. A. Bigdeli and M. Zwicker, “Image restoration using autoencoding priors,” in *arXiv:1703.09964v1*, 2017.

[38] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: towards real-time...
TABLE V: The PSNR and SSIM results of reconstructed HR images by the test methods for the bicubic downsampling.

| Dataset | Scaling factor | TNRD PSNR | TNRD SSIM | SRCNN PSNR | SRCNN SSIM | VDSR PSNR | VDSR SSIM | DnCNN PSNR | DnCNN SSIM | Ours PSNR | Ours SSIM |
|---------|----------------|-----------|-----------|------------|------------|-----------|-----------|------------|------------|-----------|-----------|
| Set14   | 2              | 32.54     | 0.907     | 32.42      | 0.906      | 33.03     | 0.912     | 33.03      | 0.911      | 33.30     | 0.915     |
|         | 3              | 29.46     | 0.823     | 29.28      | 0.821      | 29.77     | 0.831     | 29.82      | 0.830      | 30.02     | 0.836     |
|         | 4              | 27.68     | 0.750     | 27.49      | 0.750      | 28.01     | 0.767     | 27.85      | 0.755      | 28.28     | 0.773     |
| BSD100  | 2              | 31.40     | 0.888     | 31.36      | 0.888      | 31.90     | 0.896     | 31.84      | 0.894      | 32.04     | 0.898     |
|         | 3              | 28.50     | 0.788     | 28.41      | 0.786      | 28.82     | 0.798     | 28.80      | 0.795      | 28.91     | 0.801     |
|         | 4              | 27.00     | 0.714     | 26.90      | 0.710      | 27.29     | 0.725     | 27.08      | 0.709      | 27.39     | 0.729     |
| Urban100 | 2            | 29.70     | 0.899     | 29.50      | 0.895      | 30.76     | 0.914     | 30.63      | 0.911      | 31.50     | 0.922     |
|         | 3              | 26.44     | 0.801     | 26.24      | 0.799      | 27.14     | 0.828     | 27.08      | 0.824      | 27.61     | 0.842     |
|         | 4              | 24.62     | 0.729     | 24.52      | 0.722      | 25.18     | 0.752     | 24.94      | 0.735      | 25.53     | 0.768     |

TABLE VI: The PSNR results of the reconstructed HR images by the test methods for the Gaussian downsampling with scaling factor 3.

| Dataset | NCSR PSNR | SRCNN PSNR | VDSR PSNR | DD-CNN PSNR | Ours PSNR |
|---------|-----------|------------|-----------|-------------|-----------|
| Set5    | 32.07     | -          | -         | 33.88       | 34.22     |
| Set14   | 29.30     | -          | -         | 29.63       | 29.88     |

Object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[39] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. of IEEE CVPR*, 2015, pp. 3431–3440.

[40] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: can plain neural networks compete with bm3d?” in *Proc. of IEEE CVPR*, 2012, pp. 2392–2399.

[41] L. Xu, J. S. Ren, C. Liu, and J. Jia, “Deep convolutional neural network for image deconvolution,” in *Advances in Neural Information Processing Systems*, 2014.

[42] K. Gregor and Y. LeCun, “Learning fast approximations of sparse coding,” in *Proc. of the IEEE ICML*, 2010.

[43] Y. Yang, J. Sun, H. Li, and Z. Xu, “Deep admn-net for compressive sensing mri,” in *Advances in Neural Information Processing Systems*, 2016.

[44] B. Xin, Y. Wang, W. Gao, and D. Wipf, “Maximal sparsity with deep networks?” in *Advances in Neural Information Processing Systems*, 2016.

[45] Y. Wang, W. Yin, and J. Zeng, “Global convergence of admn in nonconvex nonsmooth optimization,” *Journal of Scientific Computing*, 2018.

[46] J. Bolte, A. Daniilidis, and A. Lewis, “The lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems,” *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 1205–1223, 2007.

[47] Y. Xu and W. Yin, “A block coordinate descent method for regularized multicovex optimization with applications to nonnegative tensor factorization and completion,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[48] J. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. Jordan, and B. Recht, “First-order methods almost always avoid saddle points,” arXiv:1710.07406.

[49] G. Alain and Y. Bengio, “What regularized auto-encoders learn from the data-generating distribution,” *Journal of Machine Learning Research*, vol. 15, pp. 3743–3773, 2014.

[50] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollar, “Learning to refine object segments,” in *Proc. of ECCV*, 2016.

[51] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on*
Fig. 10: SR results for 6th image of Set14 for bicubic downsampling and scaling factor 3. The PSNR results: (b) TNRD [35] (32.51 dB); (c) SRCNN [22] (32.38 dB); (d) VDSR [34] (32.70 dB); (e) DnCNN [24] (32.36 dB); (f) Ours (32.88 dB).

Fig. 11: SR results for 11th image of Set14 for bicubic downsampling and scaling factor 3. The PSNR results: (b) TNRD [35] (30.77 dB); (c) SRCNN [22] (30.22 dB); (d) VDSR [34] (31.59 dB); (e) DnCNN [24] (31.30 dB); (f) Ours (32.22 dB).

Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.

[52] D. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in Proc. of ICLR, 2014.

[53] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding and evaluating blind deconvolution algorithms,” in Proc. of CVPR, 2009, pp. 1964–1971.