Enhanced Non-Local Total Variation Model and Multi-Directional Feature Prediction Prior for Single Image Super Resolution

Chao Ren, *Member, IEEE*, Xiaohai He, *Member, IEEE*, Yifei Pu, and Truong Q. Nguyen, *Fellow, IEEE*

***Abstract*— It is widely acknowledged that single image super-resolution (SISR) methods play a critical role in recovering the missing high-frequencies in an input low-resolution image. As SISR is severely ill-conditioned, image priors are necessary to regularize the solution spaces and generate the corresponding high-resolution image. In this paper, we propose an effective SISR framework based on the enhanced non-local similarity modeling and learning-based multi-directional feature prediction (ENLTV-MDFP). Since both the modeled and learned priors are exploited, the proposed ENLTV-MDFP method benefits from the complementary properties of the reconstruction-based and learning-based SISR approaches. Specifically, for the non-local similarity-based modeled prior [enhanced non-local total variation, (ENLTV)], it is characterized via the decaying kernel and stable group similarity reliability schemes. For the learned prior [multi-directional feature prediction prior, (MDFP)], it is learned via the deep convolutional neural network. The modeled prior performs well in enhancing edges and suppressing visual artifacts, while the learned prior is effective in hallucinating details from external images. Combining these two complementary priors in the MAP framework, a combined SR cost function is proposed. Finally, the combined SR problem is solved via the split Bregman iteration algorithm. Based on the extensive experiments, the proposed ENLTV-MDFP method outperforms many state-of-the-art algorithms visually and quantitatively.**

***Index Terms*— Super resolution, decaying kernel, stable group similarity reliability, enhanced non-local total variation, multi-directional feature prediction.**

I. Introduction

SINGLE image super-resolution (SISR) methods generate a high-resolution (HR) image using a single low-resolution (LR) observation. Since low-performance imaging equipments and poor imaging conditions have inherent limitations, acquiring an HR image at desired resolution level is not always an easy task in practical applications, such as remote sensing imaging, medical imaging, and video surveillance. By using the SISR technique, the limitations of both imaging devices and the environment can be reduced, and a desired HR image can be produced. In the highly ill-posed SISR problem, effective prior knowledge should be exploited to regularize the super-resolved images. Generally, the existing numerous SISR methods can be roughly categorized into three groups [1]–[4]: interpolation-based super-resolution (SR) algorithms [5]–[7], learning-based SR algorithms [3], [4], [8]–[32], and reconstruction-based SR algorithms [33]–[42]. We briefly review these major SR categories in the next section below, and then discuss our motivations and contributions.

A. A Review of Different Categories of SISR Methods

The interpolation-based methods [5]–[7] predict the unknown HR image by using interpolation kernels. Although they are simple and fast, their applications are limited as only the down-sampling degradation is considered.

To benefit from a large collection of training samples for high-quality image SR, the learning-based methods focus on learning the LR-HR mapping relationship or some training samples-driven priors. Since their introduction by Freeman *et al.* [43] around 2002, the learning-based SR methods have achieved great success. In general, these methods can be divided into two groups: unsupervised or supervised methods [44]. For the unsupervised methods, the training samples are collected from the input image itself. For example, by using similar patches across different scales and within the same scale simultaneously, an effective unsupervised learning-based method is presented in [45]. Huang *et al.* [16] further present a self-similarity based unsupervised SR algorithm that uses transformed self-exemplars. In contrast, the supervised methods collect the training samples from the external images. Most of the learning-based SR methods belong to this group. Specifically, they include the following categories based on different learning models: Markov random fields [43], sparse representation [3], [10]–[13], local linear embedding [14], [46], regression model [17]–[21], and deep convolutional neural network (CNN)-based methods [22]–[30].

Manuscript received May 8, 2018; revised November 23, 2018 and January 14, 2019; accepted February 23, 2019. Date of publication March 4, 2019; date of current version June 13, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61801316, in part by the National Postdoctoral Program for Innovative Talents of China under Grant BX201700163, and in part by the Post-Doctoral Research and Development Foundation of Sichuan University under Grant 2017SCU12003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Raja Bala. *(Corresponding author: Chao Ren.)*

C. Ren and X. He are with the College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China (e-mail: chaoren@scu.edu.cn; hxh@scu.edu.cn).

Y. Pu is with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: puyifei@scu.edu.cn).

T. Q. Nguyen is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: tqn001@eng.ucsd.edu).

Digital Object Identifier 10.1109/TIP.2019.2902794

For example, a coupled LR-HR dictionaries based method is proposed in [3], with the assumption that the LR-HR patch pairs share the same sparse representation coefficients. In [20], first, the local geometry prior is exploited to regularize the patch representation, and then the HR output is improved via the non-local means filter. Although the conventional learning-based methods are capable of reproducing fine details, they are also prone to introduce displeasing artifacts into the super-resolved results [2].

In contrast, the deep learning-based SISR methods can overcome these limitations to some extent. Nowadays, CNN is becoming more popular in SR applications [22]–[30], [44]. For example, Johnson *et al.* [27] propose the use of perceptual loss functions for training feed-forward networks for image transformation tasks, including style transfer and SISR. Although Johnson's method can achieve good visual performance, both PSNR and SSIM results are relatively low. An effective method that combines sparse prior with the deep networks is proposed for SISR in [47]. Simonyan and Zisserman [48] propose an SR method by using a very deep convolutional network inspired by VGG-net used for ImageNet classification. Indeed, deep convolutional neural networks (CNNs) have been attracting considerable attention recently. However, as the depth grows, the long term dependency problem is rarely realized for these very deep models. Motivated by the fact that human thoughts have persistency, Tai *et al.* [29] propose a very deep persistent memory network that introduces a memory block to explicitly mine persistent memory through an adaptive learning process. The above mentioned deep learning-based SISR methods all belongs to the supervised group. Shocher *et al.* [44] propose an unsupervised CNN-based SR method for the first time, which can handle non-ideal imaging conditions, and a wide variety of images and data types. Generally, since the network is becoming deeper and deeper, the complexity of the deep CNN-based method significantly increases. Designing efficient SISR method that significantly improves SR performance is a challenging problem.

Different from the learning-based methods which highly depend on training samples, the reconstruction-based approaches focus on designing some prior constraints for SR without using any training samples. Specifically, based on the maximum a posteriori (MAP) theory, these methods enforce consistency between the estimated HR image and the observed LR image by incorporating prior constraints in the cost function to regularize the solution spaces. Various image priors have been introduced into the SR problems. For instance, a smoothness prior is proposed in [49], which blurs out high-frequency details. To reproduce sharp edges, in [42], a robust edge-preserving smoothing prior is proposed, which preserves edges well and also reduces noise. In [36], an edge-directed prior is adopted to preserve edges. However, it is sensitive to noise and prone to unnatural results. To maintain edges and suppress artifacts, non-local means (NLM) [50] based regression prior is proposed in [33]. In addition, based on the non-local similarity, many non-local total variation (NLTV) priors [37], [51]–[53] are proposed. In [38], Ren *et al.* propose an adaptive high-dimensional non-local total variation (AHNLTV) to further improve the traditional NLTV. Moreover, the non-local means prior is integrated into the sparse coding based method in [2]. In summary, the reconstruction-based methods are typically only propitious to suppress artifacts and preserve edges, while fine image details may be smoothed out [2]. Consequently, it is difficult to make a trade-off between artifacts suppression and details recovery. Moreover, as many state-of-the-art image priors have been proposed, to simply modify the prior to achieve significant SR improvement is also challenging.

B. Motivations and Contributions

As reported in many SISR literatures [37]–[39], [54], [55], assembling multiple complementary priors can obtain superior SR performance comparing to individual single prior. In addition, the work in [2] claims that the combination of the reconstruction-based and learning-based methods can further improve the SR performance. Inspired by these works, we propose a combined SISR method, which bridges these two SISR methods and provides complementary regularization constraints. We propose an enhanced non-local total variation (ENLTV) model to suppress images noises and artifacts, as well as a deep CNN-based local multi-directional feature prediction (MDFP) prior to recover fine image details. By incorporating the two priors into a MAP-based framework, the HR estimate can be obtained via split Bregman iteration (SBI) algorithm. The proposed method has the following benefits: 1) Since ENLTV and MDFP constrain the non-local and local features of images respectively, the unknown HR image features can be well constrained; 2) ENLTV and MDFP can fully exploit the internal image (input image itself) and external images (training dataset) respectively for better HR image reconstruction; 3) ENLTV is a conventional modeled prior which performs well in eliminating artifacts and noises, and MDFP is a deep CNN-based learned prior which performs well in reconstructing fine details. Consequently, the proposed method takes advantages of the complementary ENLTV and MDFP priors.

Specifically, since the original AHNLTV [38] treat all the shifted target patches with different shift-distances equally, it ignores the effect of shift-distance. Intuitively, a large shift-distance should lead to a small weight, and vice versa, which means the weight function should have a spatially-adaptive decaying kernel with respect to shift-distance. Therefore, we propose a Decaying Kernel (DK) scheme for multi-shifted target patches. In addition, in AHNLTV, a group of similar pixels will be searched for each pixel. However, we find that the original AHNLTV cannot accurately calculate the reliability of the similar pixel group. To improve the effectiveness of AHNLTV, a more effective strategy that can adaptively tune the constraint strength for each similar pixel group is needed. Consequently, we propose the stable group similarity reliability (SGSR) scheme. By using both the DK scheme for shifted target patches and measuring the reliability of the searched similar pixel group with the SGSR scheme, an improved AHNLTV model, i.e., ENLTV, is proposed, and the artifacts can be

well suppressed. To reconstruct fine details, we construct and train a simple but effective feature prediction CNN for multi-directional features prediction. Next, the unknown HR multi-directional features are predicted via the proposed deep CNN. With these predicted features, we can formulate a deep learning-based learned prior MDFP. The main contributions are as follows:

1) We characterize the non-local similarity by incorporating the proposed DK and SGSR schemes into the AHNLTV model. These two schemes take both the shift of each target patch and the reliability of the searched similar pixel group into consideration, and lead to a powerful modeled prior ENLTV. This modeled prior can constrain the non-local features of images by using the internal image itself.

2) We propose to utilize the multi-directional feature prediction CNN to estimate the unknown HR local features by using external samples. The estimated local features are formulated as an effective learned prior MDFP.

3) By incorporating the modeled prior ENLTV and the learned prior MDFP, an effective combined SISR framework is developed. This framework combines the reconstruction-based and learning-based methods in a simple manner. The ENLTV-MDFP-driven minimization problem is solved via the SBI algorithm.

4) ENLTV is essentially a non-local and internal image-based modeled prior, which performs well in artifacts and noises suppression. MDFP is essentially a local and external images-based learned prior, which performs well in fine details recovery. Benefiting from these complementary properties of ENLTV and MDFP, HR images with better objective and subjective quality can be reconstructed. In addition, with the initial HR image estimation, the proposed method is also much faster than many existing state-of-the-art reconstruction-based methods.

C. Organization

The rest of the paper is structured as follows. First, Section II briefly reviews the related works. The proposed combined SR framework is presented in Section III. Experimental results of all competing methods are illustrated in Section IV. Finally, conclusions are drawn in Section V.

II. Related Works

In this section, we briefly introduce several related works, including AHNLTV and deep CNN.

A. Adaptive High-Dimensional Non-Local Total Variation

Recently, several NLTV models have been developed [37], [51]–[53] based on the non-local similarity throughout natural images. However, for the L nearest-neighbors search of each pixel X_i, those models only exploit a fixed non-shifted $p \times p$ target patch $\mathbf{P}_i \in \mathbb{R}^{p^2 \times 1}$ (search area size $r \times r$). The search accuracy decreases in regions where non-shifted similar neighbors are rare [38]. To fully model the non-local similarity, a multi-shifted similar-patch search (MSPS) based AHNLTV prior is proposed in our previous work [38]. Specifically, for X_i, p^2 shifted target patches are simultaneously used. An example of $p = 7$ is given in Fig. 1, where the non-shifted target patch $\mathbf{P}_i$, and the 1st, 2nd, 8th, and 49th shifted target patches ($\mathbf{P}_{i1}$, $\mathbf{P}_{i2}$, $\mathbf{P}_{i8}$, $\mathbf{P}_{i49}$) are given (the number of the shifted target patch is from left to right and up to down). For each shift, we search the L nearest neighbors, and thus for X_i, there are totally Lp^2 similar pixels (X_j-s). Denote the coordinates of these similar pixels as $\mathbb{N}_i \in \mathbb{R}^{Lp^2 \times 1}$, where the super-high dimension of $\mathbb{N}_i$ will significantly increase the computation complexity. Fortunately, we found that there are many repetitions in each $\mathbb{N}_i$, and by removing the repetition in $\mathbb{N}_i$, we can construct a new index set $\mathbb{N}_i^R \in \mathbb{R}^{\rho_i Lp^2 \times 1}$ with much lower dimension (ρ_i is the reduction ratio). The similar weight w_{ij} can be calculated by

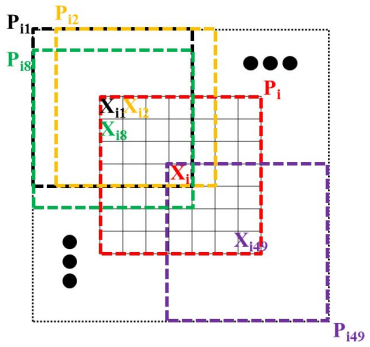


Fig. 1. Graphical illustration of multi-shifted target patches.

$$w_{ij} = w_c(i, j)w_d(i, j), j \in \mathbb{N}_i^R \tag{1}$$

where $w_c(i, j)$ is the probability weight and $w_d(i, j)$ is the pixel distance weight. $w_c(i, j)$ is defined as follows:

$$w_c(i, j) = c_{ij}/(\alpha + c_{ij}), j \in \mathbb{N}_i^R \tag{2}$$

where c_{ij} is the repetition number of j in $\mathbb{N}_i$. α is a constant, and is set to p^2. A reasonable assumption is that if two pixels are more similar than others, then the non-local similar pixels of these two pixels should also be similar. With this assumption, the weighted average of a given pixel and its non-local pixels can be used to redefine this given pixel for the stability of the pixel distance weight $w_d(i, j)$. Calculating the weighted average for all the pixels in the image is equivalent to construct the following weighted reference image $\tilde{\mathbf{X}}$:

$$\tilde{X}_i = \beta X_i + (1 - \beta) \sum_{l \in \mathbb{N}_i} (X_l/(Lp^2)) \tag{3}$$

where β is set to $1/(L + 1)$. By using $\tilde{\mathbf{X}}$, $w_d(i, j)$ can be defined by

$$w_d(i, j) = \exp(-(\tilde{X}_i - \tilde{X}_j)^2/2h^2), j \in \mathbb{N}_i^R \tag{4}$$

where $\exp(\cdot)$ is the exponential function, $\tilde{X}_j$ is the j-th pixel of $\tilde{\mathbf{X}}$, and h is a constant. With these notations, the AHNLTV model can be formulated as

$$\mathfrak{M}_A(\mathbf{X}) = \sum\nolimits_{i \in \Omega} \sqrt{\sum\nolimits_{j \in \mathbb{N}_i^R} w_{ij}(X_i - X_j)^2} \tag{5}$$

where Ω represents the index set for pixels of $\mathbf{X}$.

Algorithm 1 Decaying Kernel (DK) Scheme

Input: Estimated HR image $\mathbf{X}$, number of non-local similar patches L, and threshold τ_{dis}.

Output: $w^{DK}(i,j)$, $\widehat{\mathbb{N}_i^R}$, and ENLTV0 model $\mathfrak{M}_0(\mathbf{X})$.

1. Obtain $\mathbf{S}_{i,k}$-s, and stack their indices to form $\mathbb{N}_i$;
2. Calculate the index set $\mathbb{I}_j$ of the index j in $\mathbb{N}_i$;
3. Calculate $\mathbf{c}_i$ according to Eqs. (10) and (11);
4. Calculate $\widehat{\mathbb{N}_i^R}$ and $\hat{\mathbf{c}}_i$ via discarding the elements with probability smaller than the threshold τ_{dis} in $\mathbf{c}_i$ according to Eq. (12);
5. Calculate $w^{DK}(i,j)$-s according to Eq. (13);
6. Construct the ENLTV0 model $\mathfrak{M}_0(\mathbf{X})$ via Eq. (15).

We should note that, the computational complexity for the similar patches search is as high as $\mathcal{O}(MNp^2r^2)$ (the image size is $M \times N$). In AHNLTV, the integral image technique [38], [56], [57] is exploited to reduce the complexity. According to [38], the complexity is reduced to $\mathcal{O}(6MNr^2)$. We refer interested readers to [38] for more details.

B. Deep Convolutional Neural Network

Nowadays, deep CNN techniques are gaining significant popularity in image restoration applications [58]–[60]. On the whole, the deep CNN is an artificial neural network formed by a stack of distinct layers that transform the input volume into an output volume through a differentiable function. It can automatically learn the filters from external images, which is independent of prior knowledge, and performs extremely well in nonlinear fitting. Specifically, it consists of an input layer, an output layer, and multiple hidden layers. Typically, the hidden layers include convolution layers, normalization layers, pooling layers, etc. For CNN, many achievements have been gained, and some representative achievements related to our work will be introduced, i.e., residual learning [61] and Rectified Linear Unit (ReLU) [62]. The vanishing gradient problem is challenging in the deep CNN. Residual learning is proposed to address the performance degradation caused by increasing the network depth. ReLU is a commonly used activation function in deep learning, and it performs a threshold operation to each element of the input, where any value less than zero is set to zero. This operation is equivalent to

$$\mathcal{R}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{6}$$

ReLU has low complexity and can also alleviate the gradient vanishing problem. Consequently, a deeper CNN can be well trained with residual learning and ReLU.

III. Proposed Method

In this section, we propose a combined SISR method to significantly improve SR performance. On the one hand, by using the DK scheme and the SGSR scheme, the original AHNLTV prior is further improved for better non-local similarity modeling based on the internal image, which is essentially a non-local prior and performs well in artifacts and noises suppression. On the other hand, a local multi-directional feature prediction prior is learned from external images for better fine structures recovery, which is essentially a local prior. By combining these two complementary priors, higher quality images can be obtained.

A. Combined Single Image Super Resolution

The basic formula of our proposed combined method for SISR problem is

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \underbrace{\|\mathbf{Y} - \mathbf{DHX}\|_2^2}_{\text{fidelity term}} + \lambda \underbrace{\mathfrak{M}(\mathbf{X})}_{\text{modeled prior}} + \eta \underbrace{\mathfrak{L}(\mathbf{X})}_{\text{learned prior}} \tag{7}$$

where $\mathbf{H}$ is the blur matrix, $\mathbf{D}$ denotes the down-sampling matrix, λ and η represent the regularization parameters. The fidelity term enforces the estimated $\mathbf{X}$ to be consistent with the degraded LR input $\mathbf{Y}$ via back-projection; the modeled prior and the learned prior are used to regularize the solution space of $\mathbf{X}$ by using both the modeled features (e.g., non-local similarity) and the learned features (e.g., local multi-directional features).

B. Modeled Prior: Enhanced Non-Local Total Variation

1) Decaying Kernel (DK) for Multishifted Target Patches: In this subsection, we will analyze the influence of each shifted target patch in AHNLTV. For convenience, only the shift-distance is taken into consideration. The original AHNLTV applies a uniform weight to all the shifted target patches with different shift-distances. However, intuitively, a large shift-distance should lead to a small weight, and vice versa, which means the weight function should have a spatially-adaptive decaying kernel with respect to shift-distance. From Eqs. (1) and (2), we can conclude that w_{ij} would increase with the growth of c_{ij}. Consequently, we can operate on c_{ij} to implement the DK scheme. For pixel X_i, let the searched L similar pixels corresponding to the k-th shift be $\mathbf{S}_{i,k} = \{X_j | j \in \mathbb{N}_{i,k}^L\}$. Then, in our implementation, c_{ij} can be changed to the decaying weight summation of all the $\{\mathbf{S}_{i,k} | k = 1, 2, ..., p^2\}$ for X_i. To meet the previous requirement for DK, a $p \times p$ Gaussian kernel with standard-deviation σ can be utilized, and it is defined as

$$\mathbf{G}_p^\sigma = \{\exp(-\delta_\imath^2/\sigma^2)/\mathcal{Z}_\sigma | \imath = 1, 2, ..., p^2\} \in \mathbb{R}^{p^2 \times 1} \tag{8}$$

where $\delta_\imath$ is the shift-distance, and $\mathcal{Z}_\sigma$ is the normalization factor. A graphical illustration of DK for target patches is shown in Fig. 2, where the red patch is the non-shifted target patch $\mathbf{P}_i$, the blue patch is the 3rd shifted target patch $\mathbf{P}_{i3}$, and the yellow patch is the 33th shifted target patch $\mathbf{P}_{i33}$. As the shift-distance of $\mathbf{P}_{i3}$ ($\delta_3 = \sqrt{10}$) is larger than the shift-distance of $\mathbf{P}_{i33}$ ($\delta_3 = \sqrt{2}$), $\mathbf{P}_{i3}$ should have a smaller weight (i.e., $\mathbf{G}_p^\sigma(3)$) than $\mathbf{P}_{i33}$ (i.e., $\mathbf{G}_p^\sigma(33)$). Define $\widehat{\mathbf{G}}_p^\sigma$ as

$$\widehat{\mathbf{G}}_p^\sigma = [\underbrace{\mathbf{G}_p^\sigma(1)...\mathbf{G}_p^\sigma(1)}_{L}, ..., \underbrace{\mathbf{G}_p^\sigma(p^2)...\mathbf{G}_p^\sigma(p^2)}_{L}]^T \in \mathbb{R}^{Lp^2 \times 1} \tag{9}$$

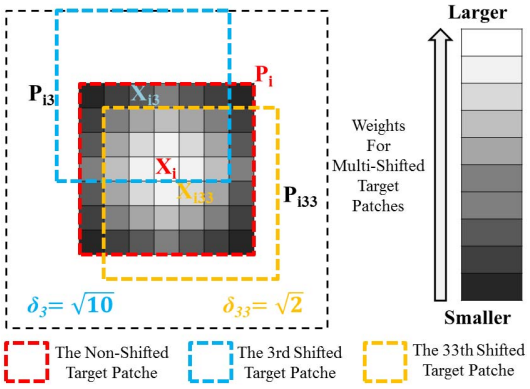


Fig. 2. Graphical illustration of decaying kernel.

As there are many repetitions for each index j in $\mathbb{N}_i$, we define $\mathbb{I}_j$ as the index set of j in $\mathbb{N}_i$. Then, the c_{ij} in original AHNLTV is changed to c_{ij}^{DK} as

$$c_{ij}^{DK} = \sum_{m \in \mathbb{I}_j} \widehat{\mathbf{G}}_p^{\sigma}(m) \tag{10}$$

Next, we define a probability vector as

$$\mathbf{c}_i = \{c_{ij}^{DK} | j \in \mathbb{N}_i^R\} \tag{11}$$

Rather than choosing all similar pixels, we discard similar pixels with small probability for both stability and dimension reduction. Thus, the final $\mathbf{c}_i$ is given as

$$\hat{\mathbf{c}}_i = \mathrm{Dis}(\mathbf{c}_i, \tau_{dis}) = \{c_{ij}^{DK} | j \in \widehat{\mathbb{N}_i^R}\} \tag{12}$$

where $\mathrm{Dis}(\mathbf{c}_i, \tau_{dis})$ is a function discarding the elements with probability smaller than a threshold τ_{dis} in $\mathbf{c}_i$, and $\widehat{\mathbb{N}_i^R} \in \mathbb{R}^{\hat{\rho}_i L p^2 \times 1}$ denotes the index set removing the indices of pixels whose c_{ij}^{DK} is smaller than τ_{dis} in $\mathbb{N}_i^R$.

Finally, the DK-based weight can be expressed as

$$w^{DK}(i,j) = w_c^{DK}(i,j) w_d(i,j), \; j \in \widehat{\mathbb{N}_i^R}. \tag{13}$$

where

$$w_c^{DK}(i,j) = c_{ij}^{DK} / (\alpha + c_{ij}^{DK}) \tag{14}$$

Thus, a new modeled prior (called ENLTV0 in this work) can be constructed. It is formulated as

$$\mathfrak{M}_0(\mathbf{X}) = \sum\nolimits_{i \in \Omega} \sqrt{\sum\nolimits_{j \in \widehat{\mathbb{N}_i^R}} w^{DK}(i,j)(X_i - X_j)^2} \tag{15}$$

2) Stable Group Similarity Reliability (SGSR) for $\mathfrak{M}_0(\mathbf{X})$: The non-local gradient function $\mathcal{G}_{w^{DK}}^i : \mathbb{R}^{MN \times 1} \to \mathbb{R}^{\hat{\rho}_i L p^2 \times 1}$ can be defined by

$$\mathcal{G}_{w^{DK}}^i(\mathbf{X}) \overset{\text{def}}{=} [(X_j - X_i)\sqrt{w^{DK}(i,j)} | j \in \widehat{\mathbb{N}_i^R}] \tag{16}$$

Then, the non-local gradient magnitude function $\mathcal{F}_{w^{DK}} : \mathbb{R}^{MN \times 1} \to \mathbb{R}^{MN \times 1}$ is defined as follows:

$$\mathcal{F}_{w^{DK}}(\mathbf{X}) \overset{\text{def}}{=} [\|\mathcal{G}_{w^{DK}}^1(\mathbf{X})\|_2, ..., \|\mathcal{G}_{w^{DK}}^{MN}(\mathbf{X})\|_2]^T \tag{17}$$

With these notations, Eq. (15) can be rewritten as

$$\mathfrak{M}_0(\mathbf{X}) = \|\mathcal{F}_{w^{DK}}(\mathbf{X})\|_1 \tag{18}$$

Define the summation S_i of the non-local weights for X_i, the normalized non-local weight $w_n^{DK}(i,j)$, and the confidence image vector $\mathbf{W}_{CI}^{Sum} \in \mathbb{R}^{MN \times 1}$ as

$$\begin{cases} S_i \overset{\text{def}}{=} \sum_{j \in \widehat{\mathbb{N}_i^R}} w^{DK}(i,j) \\ w_n^{DK}(i,j) \overset{\text{def}}{=} w^{DK}(i,j)/S_i \\ \mathbf{W}_{CI}^{Sum} \overset{\text{def}}{=} [\sqrt{S_1}, ..., \sqrt{S_{MN}}]^T \end{cases} \tag{19}$$

It is easy to prove that

$$\|\mathcal{G}_{w^{DK}}^i(\mathbf{X})\|_2 = \sqrt{S_i}\|\mathcal{G}_{w_n^{DK}}^i(\mathbf{X})\|_2 \tag{20}$$

where $\mathcal{G}_{w_n^{DK}}^i(\cdot)$ is the normalized non-local gradient function for X_i. Then,

$$\begin{aligned} \mathfrak{M}_0(\mathbf{X}) = \|\mathcal{F}_{w^{DK}}(\mathbf{X})\|_1 &= \left\| \begin{bmatrix} \|\mathcal{G}_{w^{DK}}^1(\mathbf{X})\|_2 \\ \vdots \\ \|\mathcal{G}_{w^{DK}}^{MN}(\mathbf{X})\|_2 \end{bmatrix} \right\|_1 \\ &= \left\| \begin{bmatrix} \sqrt{S_1} \\ \vdots \\ \sqrt{S_{MN}} \end{bmatrix} \odot \begin{bmatrix} \|\mathcal{G}_{w_n^{DK}}^1(\mathbf{X})\|_2 \\ \vdots \\ \|\mathcal{G}_{w_n^{DK}}^{MN}(\mathbf{X})\|_2 \end{bmatrix} \right\|_1 \\ &= \|\mathbf{W}_{CI}^{Sum} \odot \mathcal{F}_{w_n^{DK}}(\mathbf{X})\|_1 \end{aligned} \tag{21}$$

where "$\odot$" denotes the element-wise Hadamard product of two vectors, and $\mathcal{F}_{w_n^{DK}}(\cdot)$ is the normalized non-local gradient magnitude function. Eq. (21) reveals that the normalized non-local gradient magnitude vector $\mathcal{F}_{w_n^{DK}}(\mathbf{X})$ is weighted via $\mathbf{W}_{CI}^{Sum}$ in the ENLTV0 prior. However, we argue that $\mathbf{W}_{CI}^{Sum}$ does not accurately capture the reliability of the similar pixel set according to the results in Section IV. To improve the effectiveness of the ENLTV0 model, a more effective pixel-dependent vector $\mathbf{W}_{CI}^{Disp} \in \mathbb{R}^{MN \times 1}$ that can adaptively tune the constraint strength for each pixel is needed. In our method, we use the stable group similarity reliability (SGSR) to design $\mathbf{W}_{CI}^{Disp}$.

Before the introduction of $\mathbf{W}_{CI}^{Disp}$, we give the definition of the dispersion ζ_i for the group $\mathbf{S}_i$. As the dimension of $\mathbf{S}_i$ is greatly reduced in DK step, we denote the reduced version as $\hat{\mathbf{S}}_i \overset{\text{def}}{=} \{\tilde{X}_j | j \in \widehat{\mathbb{N}_i^R}\} \in \mathbb{R}^{\hat{\rho}_i L p^2 \times 1}$, $\mathbf{w}_i \overset{\text{def}}{=} \{w^{DK}(i,j) | j \in \widehat{\mathbb{N}_i^R}\} \in \mathbb{R}^{\hat{\rho}_i L p^2 \times 1}$, $w^{DK}(i,i) \overset{\text{def}}{=} \max(\mathbf{w}_i)$, $\tilde{\mathbf{w}}_i \overset{\text{def}}{=} \{w_n^{DK}(i,j) | j \in \widehat{\mathbb{N}_i^R}\} \in \mathbb{R}^{\hat{\rho}_i L p^2 \times 1}$, and $\mathbb{1}_i$ be an $\hat{\rho}_i L p^2$-dimensional column vector of all ones. The weighted mean vector can be calculated as follows (note that the center pixel $\tilde{X}_i$ is included):

$$\bar{\mathbf{X}}_i = \mathbb{1}_i \cdot \left(\frac{[\mathbf{w}_i^T, \quad w^{DK}(i,i)]}{\|[\mathbf{w}_i^T, \quad w^{DK}(i,i)]\|_1} \cdot \begin{bmatrix} \hat{\mathbf{S}}_i \\ \tilde{X}_i \end{bmatrix} \right) \tag{22}$$

After that, the dispersion ζ_i can be computed as

$$\zeta_i = \sqrt{\tilde{\mathbf{w}}_i^T (\hat{\mathbf{S}}_i - \bar{\mathbf{X}}_i)^{\odot 2}} \tag{23}$$

where "$\odot^2$" stand for the Hadamard power. With ζ_i-s, $\mathbf{W}_{CI}^{Disp}$ can generally be given as follows:

$$\mathbf{W}_{CI}^{Disp} \overset{\text{def}}{=} [\mathcal{J}(\zeta_1), ..., \mathcal{J}(\zeta_{MN})]^T \in \mathbb{R}^{MN \times 1} \tag{24}$$

where $\mathcal{J}(\cdot)$ is a positive weighting function associated with the i-th normalized non-local gradient magnitude $\|\mathcal{G}_{w_n^{DK}}^i(\mathbf{X})\|_2^1$.

If the corresponding $\hat{\mathbf{S}}_i$ is very reliable, i.e., the corresponding SGSR is high, $\mathcal{J}(\cdot)$ should be large, and vice versa. This leads to the following dispersion-based strategy: if the dispersion of $\hat{\mathbf{S}}_i$ is small, a large constraint should be imposed on X_i, and vice versa. It can be given by

$$\mathcal{J}(\zeta_i) \stackrel{\text{def}}{=} (1 + \mathcal{A}\zeta_i^{\mathcal{B}})^{-1} \tag{25}$$

where the two non-negative constants $\mathcal{A}$ and $\mathcal{B}$ are set to 2 and 0.75, respectively. As $\mathbf{W}_{CI}^{Disp}$ is dependent on the image structures, it should have strong local consistency. Therefore, we introduce the concept of confidence image filtering into this work. In addition, the confidence image filtering can also lead to better noise suppression. First, we reshape the weighting vector $\mathbf{W}_{CI}^{Disp}$ to form a 2D image, and we call it confidence image $\mathbf{W}_{CI}^{Disp}|_{2D}$. Since the steering kernel [63] is essential for capturing the structure information of images and is very robust to perturbations of local image data, we filter $\mathbf{W}_{CI}^{Disp}|_{2D}$ via local steering kernel to improve the local consistency and suppress noise. Next, in the local structure analysis window, we calculate the $p' \times p'$ steering kernel of the j-th neighbor of X_i according to the following equation:

$$w_{ij}^K = \frac{\sqrt{\det(\mathbf{C}_i)}}{2\pi\hbar^2} \times \exp\left(-\left(\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} - \begin{bmatrix} j_1 \\ j_2 \end{bmatrix}\right)^T \mathbf{C}_i \left(\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} - \begin{bmatrix} j_1 \\ j_2 \end{bmatrix}\right) \Big/ 2\hbar^2\right) \tag{26}$$

where w_{ij}^K is the j-th element of the steering kernel $\mathbf{w}_i^K$, $\mathbf{C}_i$ is the gradient covariance matrix, $\hbar$ represents the smoothing parameter, and $[i_1,\ i_2]^T$ and $[j_1,\ j_2]^T$ are the 2D forms of the coordinates i and j. (See [63] for more details). Furthermore, the filtered $\widetilde{\mathbf{W}_{CI}^{Disp}}|_{2D}$ can be calculated by

$$\widetilde{\mathbf{W}_{CI}^{Disp}}|_{2D} = \sum\nolimits_{i\in\Omega} R_i^T \left(R_i(\mathbf{W}_{CI}^{Disp}|_{2D}) * \mathbf{w}_i^K\right) \tag{27}$$

where $R_i(\cdot)$ is a function extracting the $p' \times p'$ patch centered at i. "$*$" denotes the convolution operation. For mathematical convenience, we use the matrix form to denote this operation. For the coordinate i, by defining the index set of the group of neighbors as $\mathbb{O}_i$, Eq. (27) can be reformulated as

$$\widetilde{\mathbf{W}_{CI}^{Disp}} = \mathbf{K}\mathbf{W}_{CI}^{Disp} \tag{28}$$

where

$$\mathbf{K}(i,j) = \begin{cases} w_{ij}^K, & j \in \mathbb{O}_i \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

Finally, the DK- and SGSR-based ENLTV can be given as

$$\mathfrak{M}(\mathbf{X}) = \|\widetilde{\mathbf{W}_{CI}^{Disp}} \odot \mathcal{F}_{w_n^{DK}}(\mathbf{X})\|_1 \tag{30}$$

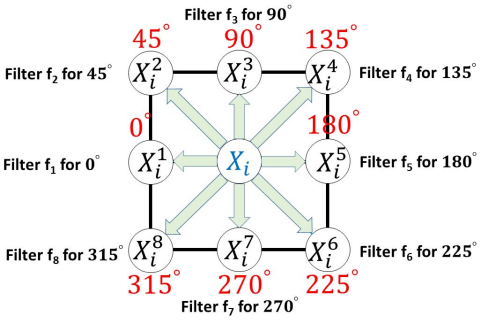


Fig. 3. Graphical illustration of multi-directional filters.

Algorithm 2 Stable Group Similarity Reliability (SGSR) Scheme

Input: Estimated HR image $\mathbf{X}$, non-local weight $w^{DK}(i,j)$-s, and steering kernel w_{ij}^K-s.

Output: $\widetilde{\mathbf{W}_{CI}^{Disp}}$ and ENLTV model $\mathfrak{M}(\mathbf{X})$.

1. Calculate $w_n^{DK}(i,j)$-s according to Eq. (19);
2. Construct $\hat{\mathbf{X}}_i$ and $\widetilde{\mathbf{w}}_i$, and then calculate the dispersion ζ_i of $\hat{\mathbf{S}}_i$ according to Eq. (23);
3. Calculate $\widetilde{\mathbf{W}_{CI}^{Disp}}$ according to Eqs. (24) and (25);
4. Calculate $\widetilde{\mathbf{W}_{CI}^{Disp}}$ according to Eq. (28);
5. Construct the ENLTV model $\mathfrak{M}(\mathbf{X})$ via Eq. (30).

C. Learned Prior: Multi-Directional Feature Prediction

1) Learned MDFP Prior: Natural images contain various underlying features. By constraining image structures via some predicted features, the images details can be well recovered. This constraint will lead to a learned prior. Let $\mathbf{Y}$ be the input LR image, $\mathbf{X}$ be the HR image, $E(\mathbf{X})$ be the extracted features from $\mathbf{X}$, and $\psi_E(\mathbf{Y})$ be the predicted HR features, then the learned prior can be written as

$$\mathfrak{L}(\mathbf{X}) = \|E(\mathbf{X}) - \psi_E(\mathbf{Y})\|_2^2 \tag{31}$$

Specifically, for each pixel X_i, eight filters $\{\mathbf{f}_1,\ \mathbf{f}_2, ...,\ \mathbf{f}_8\}$ can be used to extract its corresponding eight directional features, as shown in Fig. 3. The filters are defined as follows:

$$\mathbf{f}_1 = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{f}_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$\mathbf{f}_3 = \text{rot}(\mathbf{f}_1, 90°), \quad \mathbf{f}_4 = \text{rot}(\mathbf{f}_2, 90°), \quad \mathbf{f}_5 = \text{rot}(\mathbf{f}_1, 180°),$$
$$\mathbf{f}_6 = \text{rot}(\mathbf{f}_2, 180°), \ \mathbf{f}_7 = \text{rot}(\mathbf{f}_1, 270°), \ \mathbf{f}_8 = \text{rot}(\mathbf{f}_2, 270°) \tag{32}$$

where $\text{rot}(\mathbf{f}, \theta°)$ denotes rotating $\mathbf{f}$ clockwise by θ degrees. Define the k-th feature for X_i be the convolution of $\mathbf{X}$ and $\mathbf{f}_k$ at position i, and denote it as $E_k(\mathbf{X})_i$. Due to the symmetry, only four directional features (i.e., 0°, 45°, 90°, 135° which correspond to $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$, $\mathbf{f}_4$) are collected in our implementation. The impact of the number of features will be discussed at the end of subsection III-C. By using four features, $E(\mathbf{X})_i = \{E_k(\mathbf{X})_i | k = 1, 2, 3, 4\} \in \mathbb{R}^{4\times 1}$, and

$$E(\mathbf{X}) = [E(\mathbf{X})_1^T, E(\mathbf{X})_2^T, ..., E(\mathbf{X})_{MN}^T]^T \in \mathbb{R}^{4MN\times 1} \tag{33}$$

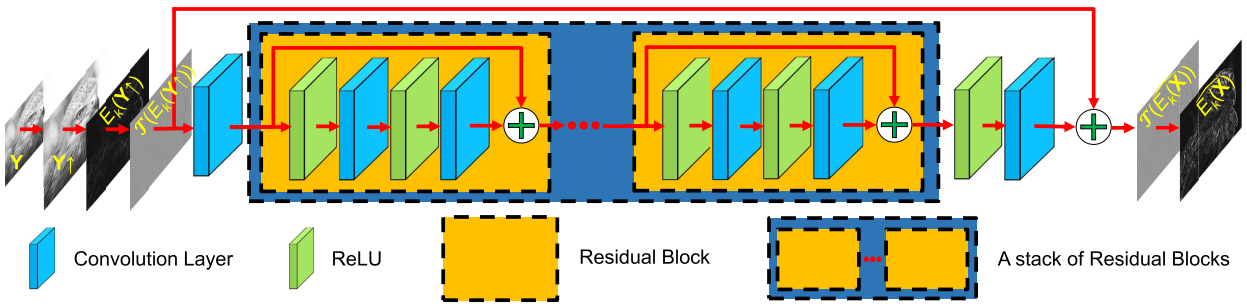


Fig. 4. The architecture of the proposed multi-directional feature prediction network with pre-activation residual block.

2) Deep CNN for MDFP: For the convenience of constructing MDFP CNN, we reshape $E(\mathbf{X})$ via the reshape function $S_h : \mathbb{R}^{4MN\times 1} \rightarrow \mathbb{R}^{4\times MN}$ as

$$S_h(E(\mathbf{X})) = \begin{bmatrix} E_1(\mathbf{X})_1 & E_1(\mathbf{X})_2 & \dots & E_1(\mathbf{X})_{MN} \\ E_2(\mathbf{X})_1 & E_2(\mathbf{X})_2 & \dots & E_2(\mathbf{X})_{MN} \\ E_3(\mathbf{X})_1 & E_3(\mathbf{X})_2 & \dots & E_3(\mathbf{X})_{MN} \\ E_4(\mathbf{X})_1 & E_4(\mathbf{X})_2 & \dots & E_4(\mathbf{X})_{MN} \end{bmatrix} = [E_1(\mathbf{X}),\ E_2(\mathbf{X}),\ E_3(\mathbf{X}),\ E_4(\mathbf{X})]^T \quad (34)$$

This reveals that we can individually train a CNN for each feature image $E_k(\mathbf{X})$. Let "$\uparrow$" be the bicubic up-sampling operator. Then, the CNN is focused on designing a mapping function $\psi_{E_k}(\mathbf{Y}) : E_k(\mathbf{Y}\uparrow) \longrightarrow E_k(\mathbf{X})$. After all $\psi_{E_k}(\mathbf{Y})$-s are obtained, $\psi_E(\mathbf{Y})$ can be calculated by

$$\psi_E(\mathbf{Y}) = S_h^{-1}\Big([\psi_{E_1}(\mathbf{Y}), \psi_{E_2}(\mathbf{Y}), \psi_{E_3}(\mathbf{Y}), \psi_{E_4}(\mathbf{Y})]^T\Big) \quad (35)$$

where S_h^{-1} denotes the inversion operator of S_h, which reshapes the $4 \times MN$ matrix to a $4MN \times 1$ vector.

To construct an effective feature prediction prior by using external images for details restoration, an MDFP CNN is proposed. First, the degraded LR input $\mathbf{Y}$ is upscaled as $\mathbf{Y}\uparrow$ via Bicubic. Second, the multi-directional features of $\mathbf{Y}\uparrow$ are calculated by feature extraction operators $E_k(\cdot)$-s. Third, by using transformation function $\mathcal{T}(x) = x/510+0.5$, the range of each feature is normalized from $[-255\ 255]$ to $[0\ 1]$. Fourth, each transformed LR feature $\mathcal{T}(E_k(\mathbf{Y}\uparrow))$ is mapped into the desired HR transformed feature $\mathcal{T}(\widehat{E_k(\mathbf{X})})$ via MDFP CNN. Finally, the predicted HR feature images $\{\widehat{E_k(\mathbf{X})}|k = 1,\ 2,\ 3,\ 4\}$ can be obtained via inverse transformation function $\mathcal{T}^{-1}(x) = 510x - 255$. In the following, the architecture of the proposed MDFP CNN will be introduced in detail.

For the MDFP CNN, since its input $\mathcal{T}(E_k(\mathbf{Y}\uparrow))$ (the LR transformed feature) is highly similar to its output $\mathcal{T}(\widehat{E_k(\mathbf{X})})$ (the HR transformed feature), the global residual learning [61] strategy is preferred. To ease the optimization of the MDFP CNN, and gain accuracy from increased depth, the local residual learning strategy is also utilized. Specifically, the first convolutional layer (64 filters of size $3\times3\times1$) is used to extract the features of the input $\mathcal{T}(E_k(\mathbf{Y}\uparrow))$. The output of this layer is served as the input of the following pre-activation residual block. The pre-activation residual block in MDFP CNN consists of two convolutional layers (64 filters of size $3\times3\times64$, where ReLU is placed before each convolutional layer). These layers predict the local residual and then added by the local input of the current residual block to obtain the local output. The analysis of our pre-activation residual block will be given at subsection III-C.3. After a stack of multiple pre-activation residual blocks (the residual block number is empirically set to 9), the output is rectified by the ReLU function $\mathcal{R}(\cdot)$. The last convolutional layer (1 filter of size $3\times3\times64$, where ReLU is placed before the convolutional layer) is applied on the residual features to produce the final residual image of our MDFP network.

Let the trainable parameter set for the k-th deep network be Θ_k^{FP}, and the residual image be $\Lambda_k = \mathcal{T}(E_k(\mathbf{X})) - \mathcal{T}(E_k(\mathbf{Y}\uparrow))$, we define the global residual mapping function as $\mathcal{H}_{FP}(\mathcal{T}(E_k(\mathbf{Y}\uparrow)); \Theta_k^{FP}) : \mathcal{T}(E_k(\mathbf{Y}\uparrow)) \rightarrow \Lambda_k$, which predicts the residual image Λ_k from the LR input $\mathcal{T}(E_k(\mathbf{Y}\uparrow))$. At last, the desired transformed feature $\mathcal{T}(\widehat{E_k(\mathbf{X})})$ can be calculated via the summation of the LR input $\mathcal{T}(E_k(\mathbf{Y}\uparrow))$ and the HR residual estimate Λ_k. Denote the $\mathcal{N}$ external LR-HR training pairs as $\{\mathcal{T}(E_k(\mathbf{Y}^i\uparrow)), \mathcal{T}(E_k(\mathbf{X}^i))\}_{i=1}^{\mathcal{N}}$. As the residual learning strategy is adopted, the loss function is given by

$$\mathcal{L}_k(\Theta_k^{FP}) = \frac{1}{\mathcal{N}}\sum_{i=1}^{\mathcal{N}} \|\mathcal{H}_{FP}(\mathcal{T}(E_k(\mathbf{Y}^i\uparrow)); \Theta_k^{FP}) - \Lambda_k^i\|_2^2 \quad (36)$$

We optimize the loss function via ADAM [64]. After the k-th network is trained, for the LR input $\mathbf{Y}$, the predicted feature $E_k(\mathbf{X})$ can be expressed as

$$\psi_{E_k}(\mathbf{Y}) = \mathcal{T}^{-1}(\mathcal{H}_{FP}(\mathcal{T}(E_k(\mathbf{Y}\uparrow)); \Theta_k^{FP}) + \mathcal{T}(E_k(\mathbf{Y}\uparrow))) \quad (37)$$

Finally, after $\psi_{E_k}(\mathbf{Y})$-s are obtained, $\psi_E(\mathbf{Y})$ can be calculated via Eq. (35).

3) Analysis of the Proposed Network:

(a) Pre-Activation Residual Block for MDFP CNN

As shown in Fig. 4, the number of convolution layers in the pre-activation residual block is 2. Mathematically, the pre-activation residual block can be formulated as:

$$\mathcal{U}_{l+1} = \mathcal{U}_l + \mathcal{H}(\mathcal{U}_l, \mathbf{F}_l, \mathbf{B}_l) \quad (38)$$

where $\mathcal{U}_l$ is the input feature to the l-th residual block. $\mathbf{F}_l = \{\mathbf{F}_{l,m}|m = 1,\ 2\}$ and $\mathbf{B}_l = \{\mathbf{B}_{l,m}|m = 1,\ 2\}$ are the sets of weights and biases corresponding to the l-th residual block, respectively. $\mathcal{H}(\cdot)$ represents the local residual function. In our residual block, $\mathcal{H}(\mathcal{U}_l, \mathbf{F}_l, \mathbf{B}_l)$ can be formulated as

$$\mathcal{H}(\mathcal{U}_l, \mathbf{F}_l, \mathbf{B}_l) = \mathcal{R}(\mathcal{R}(\mathcal{U}_l) * \mathbf{F}_{l,1} + \mathbf{B}_{l,1}) * \mathbf{F}_{l,2} + \mathbf{B}_{l,2} \quad (39)$$

where $\mathcal{R}(\cdot)$ is the ReLU function defined in Eq. (6). It is not difficult to prove that, for any deeper block ν and any

Fig. 5. Test images for multi-directional feature prediction via deep CNN.

TABLE I
AVERAGE PSNR (DB) AND SSIM RESULTS ON TEST IMAGES

Number of features	1	2	3	4
PSNR	28.76	29.05	29.17	29.20
SSIM	0.8459	0.8533	0.8552	0.8559
Number of features	5	6	7	8
PSNR	29.20	29.20	29.21	29.22
SSIM	0.8560	0.8561	0.8561	0.8563

shallower block l, the feature $\mathcal{U}_v$ can be represented as the feature $\mathcal{U}_l$ plus the summation of the outputs of all the residual functions between blocks v and l. Mathematically, the relationship between $\mathcal{U}_v$ and $\mathcal{U}_l$ can be given by

$$\mathcal{U}_v = \mathcal{U}_l + \sum_{i=l}^{v-1} \mathcal{H}(\mathcal{U}_i, \mathbf{F}_i, \mathbf{B}_i) \tag{40}$$

Next, we will prove that the cascade of multiple residual blocks (Eq.(40)) can lead to nice backward propagation characteristics. For the loss function $\mathcal{L}_k$, we have

$$\frac{\partial \mathcal{L}_k}{\partial \mathcal{U}_l} = \underbrace{\frac{\partial \mathcal{L}_k}{\partial \mathcal{U}_v}}_{\text{first term}} + \underbrace{\frac{\partial \mathcal{L}_k}{\partial \mathcal{U}_v}\frac{\partial}{\partial \mathcal{U}_l}\sum_{i=l}^{v-1} \mathcal{H}(\mathcal{U}_i, \mathbf{F}_i, \mathbf{B}_i)}_{\text{second term}} \tag{41}$$

where the first term $\frac{\partial \mathcal{L}_k}{\partial \mathcal{U}_v}$ propagates information back to any shallower unit l directly without concerning any convolution layers. While for the second term, it guarantees that the gradient $\frac{\partial \mathcal{L}_k}{\partial \mathcal{U}_l}$ will not be canceled out. The reason is that the second term cannot be always $-\frac{\partial \mathcal{L}_k}{\partial \mathcal{U}_v}$ for all samples. That is to say, the gradient of a convolution layer will not vanish even if the weights are arbitrarily small. Because of these nice properties, the proposed MDFP CNN can be well trained, and thus it can perform outstanding feature predictions.

(b) Number of Features for MDFP Prior

In order to analyze the impact of the number of features, we test the SR performance of different MDFP priors (with different number of features). The test images are shown in Fig. 5. The PSNR and SSIM [65] results are given in Table I, and the PSNR and SSIM gains vs. number of features are also shown in Fig. 6. We can conclude from the results that, the improvement in PSNR and SSIM is directly proportional to the number of feature. When the number is not larger than 4, the improvements are significant (0.46 dB/0.0100). While for larger number ($>$ 4), further increasing the number of features does not significantly improve the PSNR and SSIM performance (0.02 dB/0.0004). This is because of the symmetry of features. Symmetrically, the 180°, 225°, 270°, 315° directional features are similar to the 0°, 45°, 90°, 135° directional features, respectively. As a result, further increasing the number of features does not provide additional complementary constraints. Consequently, for computation efficiency, only four features (corresponding to 0°, 45°, 90°, 135°) are used in MDFP.

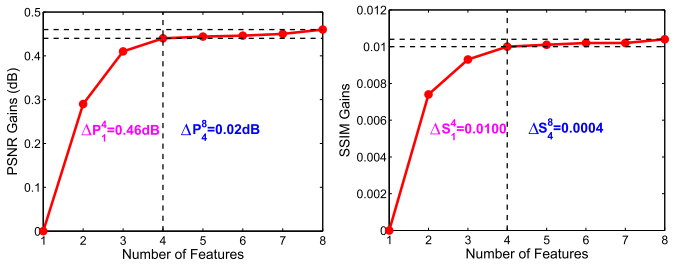


Fig. 6. The PSNR gain distributions of SR (left) and SSIM gain distributions of SR (right) experiments over different number of features.

Algorithm 3 Multi-Directional Feature Prediction

Input: Training samples $\{\mathbf{X}^i\}_{i=1}^{\mathcal{N}}$ and LR image $\mathbf{Y}$.
Output: Predicted feature $\psi_E(\mathbf{Y})$ and learned MDFP prior $\mathfrak{L}(\mathbf{X})$.
1. Construct the $\mathcal{N}$ external training sample pairs $\{\mathcal{T}(E_k(\mathbf{Y}^i \uparrow)), \mathcal{T}(E_k(\mathbf{X}^i))\}_{i=1}^{\mathcal{N}}$;
2. Train the MDFP network via Eq. (36);
3. Calculate $\psi_E(\mathbf{Y})$ according to Eqs. (35) and (37);
4. Construct the learned MDFP prior $\mathfrak{L}(\mathbf{X})$ via Eq. (31).

D. HR Image Estimation via Combined SISR

Since ENLTV is essentially a non-local prior and MDFP is essentially a local prior, and also to fully exploit the advantages of both reconstruction- and learning-based SISR methods, we combine the modeled prior and the learned prior to propose a combined SISR method. Detailed discussion is given in the following subsections:

1) Proposed Objective Cost Function: Inserting Eq. (30) and Eq. (31) into Eq. (7), the objective cost function for our combined SISR can be formulated as follows:

$$\begin{aligned}\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} &\|\mathbf{Y} - \mathbf{DHX}\|_2^2 + \lambda \|\widetilde{\mathbf{W}_{CI}^{Disp}} \odot \mathcal{F}_{w_n^{DK}}(\mathbf{X})\|_1 \\ &+ \eta \|E(\mathbf{X}) - \psi_E(\mathbf{Y})\|_2^2\end{aligned} \tag{42}$$

To help in developing the solution for the combined SISR problem, the matrix forms of $\widetilde{\mathbf{W}_{CI}^{Disp}}$ and $E(\mathbf{X})$ are needed. Let $\widetilde{\mathcal{Q} \in \mathbb{R}^{MN \times MN}}$ and $\mathbf{M}_E \in \mathbb{R}^{4MN \times MN}$ be the matrix forms of $\widetilde{\mathbf{W}_{CI}^{Disp}}$ and $E(\mathbf{X})$, respectively, and they can be given by

$$\mathcal{Q}(i,j) = \begin{cases} \widetilde{\mathbf{W}_{CI}^{Disp}}(i), & \text{if } \ j = i \\ 0, & \text{otherwise} \end{cases} \tag{43}$$

$$\mathbf{M}_E(i,j) = \begin{cases} 1, & \text{if } \ j = \xi_i \\ -1, & \text{if } \ j \neq \xi_i \ \text{ and } \ (i,j) \in \beth_i \\ 0, & \text{otherwise} \end{cases} \tag{44}$$

where $\xi_i = \lceil \frac{i-0.5}{4} \rceil$ ($\lceil x \rceil$ is the ceiling function, which maps x to the least integer greater than or equal to x). $\beth_i = \{(4\xi_i - 3, \xi_i + M), (4\xi_i - 2, \xi_i - 1), (4\xi_i - 1, \xi_i - M), (4\xi_i, \xi_i + 1)\}$. Using $\mathcal{Q}$ and $\mathbf{M}_E$, the cost function can be reformulated as

$$\begin{aligned}\hat{\mathbf{X}} &= \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{DHX}\|_2^2 \\ &\quad + \lambda \|\mathcal{Q}\mathcal{F}_{w_n^{DK}}(\mathbf{X})\|_1 + \eta \|\mathbf{M}_E\mathbf{X} - \psi_E(\mathbf{Y})\|_2^2 \\ &= \arg\min_{\mathbf{X}} \left\| \begin{bmatrix} \mathbf{Y} \\ \psi_E(\mathbf{Y}) \end{bmatrix} - \begin{bmatrix} \mathbf{DH} \\ \sqrt{\eta}\mathbf{M}_E \end{bmatrix} \mathbf{X} \right\|_2^2 + \lambda \|\mathcal{Q}\mathcal{F}_{w_n^{DK}}(\mathbf{X})\|_1 \\ &= \arg\min_{\mathbf{X}} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{K}}\mathbf{X}\|_2^2 + \lambda \|\mathcal{Q}\mathcal{F}_{w_n^{DK}}(\mathbf{X})\|_1 \end{aligned} \tag{45}$$

2) Initial HR Image Estimation for Modeled Prior: In Eq. (45), $\mathcal{Q}$ and $w_n^{DK}(i, j)$-s in the modeled prior depend on the unknown $\mathbf{X}$. Therefore, the cost function is non-convex, which makes Eq. (45) difficult to solve. To overcome this problem, the traditional SR methods [33], [34], [38], [66] often start from bicubic interpolation result. Because the estimate from bicubic interpolation result is not accurate, the data-adaptive $\mathcal{Q}$ and $w_n^{DK}(i, j)$-s need to be recomputed many times by using the previous HR image estimates. However, the multiple recomputing process will be time-consuming. In our algorithm, as ℓ_2-norm is used in $\mathfrak{L}(\mathbf{X})$, the optimization of single $\mathfrak{L}(\mathbf{X})$-based SISR problem can be implemented efficiently. In addition, as the MDFP CNN can well predict the HR feature structures, a good rough estimate $\mathbf{X}^{(0)}$ of the underlying unknown image $\mathbf{X}$ can be obtained. Consequently, we propose an initial HR image estimation via learned prior. By using the good estimate $\mathbf{X}^{(0)}$ for calculating $\mathcal{Q}$ and $w_n^{DK}(i, j)$-s, the need for regular update is eliminated and our method is significantly speeded up. Specifically, we formulate the $\mathfrak{L}(\mathbf{X})$-based SISR problem as

$$\mathbf{X}^{(0)} = \arg\min_{\mathbf{X}} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{K}}\mathbf{X}\|_2^2 \tag{46}$$

which yields the following closed-form solution

$$\mathbf{X}^{(0)} = (\tilde{\mathbf{K}}^T\tilde{\mathbf{K}})^{-1}\tilde{\mathbf{K}}^T\tilde{\mathbf{Y}} \tag{47}$$

However, directly calculating $(\tilde{\mathbf{K}}^T\tilde{\mathbf{K}})^{-1}$ requires very large computational complexity, making the SR problem cumbersome. Since the minimization problem in Eq. (46) is a convex quadratic function, we use the Templates for First-Order Conic Solvers (TFOCS) technique [67] to solve it. After $\mathbf{X}^{(0)}$ is obtained, $\mathcal{Q}$ and $w_n^{DK}(i, j)$-s can be calculated, and they are fixed during the optimization process.

3) Optimization of the Proposed Cost Function: The SBI algorithm [68] is extended to solve the problem in Eq. (45). First, we define a series of new matrices $\mathbf{W}_i^n \in \mathbb{R}^{\hat{\rho}_i Lp^2 \times MN}$-s as

$$\mathbf{W}_i^n(a, b) \stackrel{\text{def}}{=} \begin{cases} \sqrt{w_n^{DK}(i, j_a)}, & b = j_a \\ -\sqrt{w_n^{DK}(i, j_a)}, & b = i \\ 0, & \text{otherwise} \end{cases} \tag{48}$$

where j_a is the a-th entry of $\widehat{\mathbb{N}_i^R}$. Then, the cost function can be rewritten as

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{Y} - \tilde{\mathbf{Q}}\mathbf{X}\|_2^2 + \lambda \sum\nolimits_{i\in\Omega} \|\mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X}\|_2 \tag{49}$$

Replacing $\mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X}$ by $\mathbf{v}_i \in \mathbb{R}^{\hat{\rho}_i Lp^2 \times 1}$, defining $\mathbf{v} \stackrel{\text{def}}{=} [\mathbf{v}_i | i \in \Omega] \in \mathbb{R}^{(\sum_{i\in\Omega}\hat{\rho}_i Lp^2)\times 1}$, and utilizing the Bregman iteration process, Eq. (49) can be rewritten as

$$\begin{aligned}(\hat{\mathbf{X}}, \hat{\mathbf{v}}) = \arg\min_{\mathbf{X},\mathbf{v}} &\|\tilde{\mathbf{Y}} - \tilde{\mathbf{K}}X\|_2^2 + \lambda \sum_{i\in\Omega} \|\mathbf{v}_i\|_2 \\ &+ \mu \sum\nolimits_{i\in\Omega} \|\mathbf{v}_i - \mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X} - \mathbf{b}_i\|_2^2 \end{aligned} \tag{50}$$

where $\mathbf{b}_i \in \mathbb{R}^{\hat{\rho}_i Lp^2 \times 1}$-s are auxiliary variables, and μ is a penalty parameter. By letting

$$\begin{cases} \mathbf{b} \stackrel{\text{def}}{=} [\mathbf{b}_i | i \in \Omega] \in \mathbb{R}^{(\sum_{i\in\Omega}\hat{\rho}_i Lp^2)\times 1} \\ \mathbf{W}^n \stackrel{\text{def}}{=} \left[\mathbf{W}_i^n | i \in \Omega\right] \in \mathbb{R}^{(\sum_{i\in\Omega}\hat{\rho}_i Lp^2)\times MN} \\ F(\mathbf{X}) \stackrel{\text{def}}{=} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{K}}X\|_2^2 \end{cases} \tag{51}$$

and defining an extended proximal operator

$$\text{prox}_{\gamma\varphi}^{\mathbf{U}}(\mathcal{Y}) \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{\chi}} \frac{1}{2\gamma}\|\mathbf{U}\boldsymbol{\chi} - \mathcal{Y}\|_2^2 + \varphi(\boldsymbol{\chi}) \tag{52}$$

the optimization problem in Eq. (50) can be converted into the following $\mathbf{X}$ and $\mathbf{v}$ sub-problems:

(a) $\mathbf{X}$ Sub-Problem

Given $\mathbf{v}$, the $\mathbf{X}$ sub-problem becomes

$$\mathbf{X}^{(k+1)} = \text{prox}_{1/(2\mu)F}^{\mathcal{Q}\mathbf{W}^n}(\mathbf{v}^{(k)} - \mathbf{b}^{(k)}) \tag{53}$$

This extended proximal operator admits the following closed-form solution

$$\begin{aligned}\mathbf{X}^{(k+1)} = &(\tilde{\mathbf{K}}^T\tilde{\mathbf{K}} + \mu(\mathcal{Q}\mathbf{W}^n)^T\mathcal{Q}\mathbf{W}^n)^{-1} \\ &\times(\tilde{\mathbf{K}}^T\tilde{\mathbf{Y}} + \mu(\mathcal{Q}\mathbf{W}^n)^T(\mathbf{v}^{(k)} - \mathbf{b}^{(k)})) \end{aligned} \tag{54}$$

Similar to $(\tilde{\mathbf{K}}^T\tilde{\mathbf{K}})^{-1}$ in Eq. (47), directly calculating $(\tilde{\mathbf{K}}^T\tilde{\mathbf{K}} + \mu(\mathcal{Q}\mathbf{W}^n)^T\mathcal{Q}\mathbf{W}^n)^{-1}$ is cumbersome. Since the minimization problem in Eq. (53) is strictly convex, it can be solved by the TFOCS technique.

(b) $\mathbf{v}$ Sub-Problem

Given $\mathbf{X}$, the $\mathbf{v}$ sub-problem becomes

$$\mathbf{v}^{(k+1)} = \sum_{i\in\Omega} \text{prox}_{\lambda/(2\mu)\|\cdot\|_2}(\mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X}^{(k+1)} + \mathbf{b}_i^{(k)}) \tag{55}$$

Since $\mathbf{v}$ can be separated into $\mathbf{v}_i$-s, each $\mathbf{v}_i$ can be independently obtained via

$$\mathbf{v}_i^{(k+1)} = \text{prox}_{\lambda/(2\mu)\|\cdot\|_2}(\mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X}^{(k+1)} + \mathbf{b}_i^{(k)}) \tag{56}$$

According to [49], the solution of the proximal operator can be obtained via shrinkage operator

$$\mathbf{v}_i^{(k+1)} = \text{shrink}(\mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X}^{(k+1)} + \mathbf{b}_i^{(k)}, 2\mu/\lambda) \tag{57}$$

where $\text{shrink}(\mathbf{x}, \vartheta) = \max(\|\mathbf{x}\|_2 - 1/\vartheta, 0)\mathbf{x}/\|\mathbf{x}\|_2^1$.

Finally, $\mathbf{b}$ can be updated by

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} + \mathcal{Q}\mathbf{W}^n\mathbf{X}^{(k+1)} - \mathbf{v}^{(k+1)} \tag{58}$$

The proposed ENLTV-MDFP method is summarized in **Algorithm 4**, and the flowchart is shown in Fig. 7.

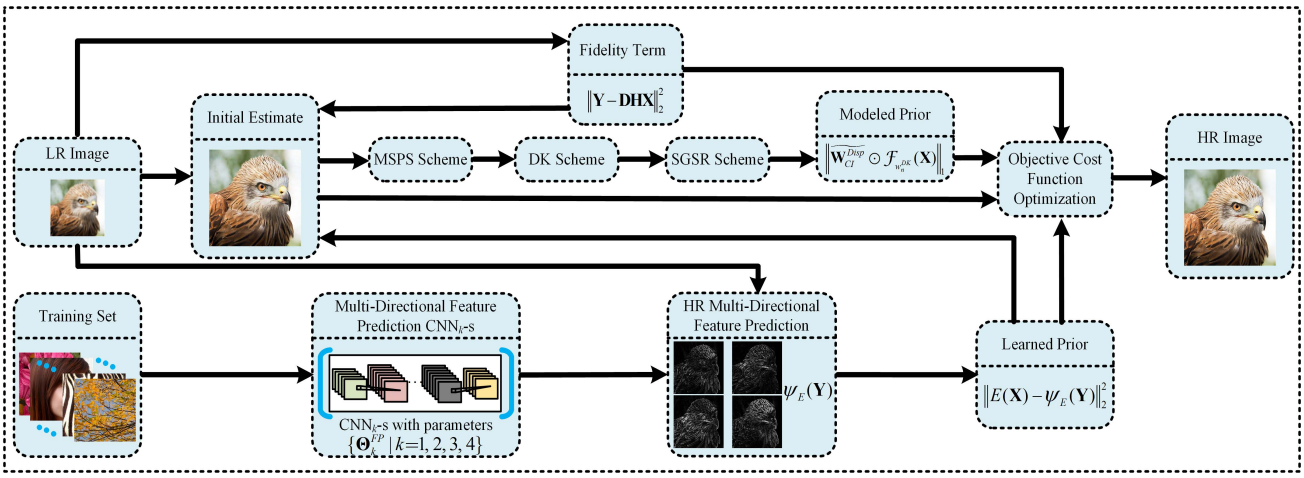


Fig. 7. Flowchat of the proposed method in SISR.

Algorithm 4 Optimization of the Proposed Cost Function in Eq. (45)

Input: LR image $\mathbf{Y}$, Blurring matrix $\mathbf{H}$, and Down-sampling matrix $\mathbf{D}$.

Output: HR image $\hat{\mathbf{X}}$.

1. Use Eq. (46) to obtain the initial estimate $\mathbf{X}^{(0)}$;
2. Update the data-adaptive $\mathcal{Q}$, $\mathbf{W}^n$, and calculate the modeled prior Eq. (30);
3. Predict the multi-directional features $\psi_E(\mathbf{Y})$ according to Eq. (35), and calculate the learned prior Eq. (31);
4. Set $k=0$, $\mathbf{v}^{(0)}=\mathbf{0}$, $\mathbf{b}^{(0)}=\mathbf{0}$, choose $\mu>0$.

for $k=0$ to T_I-1 **do**

a. Update the HR estimate $\mathbf{X}^{(k+1)}$ using Eq. (53):

$$\mathbf{X}^{(k+1)}=\text{prox}_{1/(2\mu)F}^{\mathcal{Q}\mathbf{W}^n}(\mathbf{v}^{(k)}-\mathbf{b}^{(k)});$$

b. Update $\mathbf{v}^{(k+1)}$ using Eq. (55):

$$\mathbf{v}^{(k+1)}=\sum_{i\in\Omega}\text{prox}_{\lambda/(2\mu)\|\cdot\|_2}(\mathcal{J}(\zeta_i)\mathbf{W}_i^n\mathbf{X}^{(k+1)}+\mathbf{b}_i^{(k)});$$

c. Update $\mathbf{b}^{(k+1)}$ using Eq. (58):

$$\mathbf{b}^{(k+1)}=\mathbf{b}^{(k)}+\mathcal{Q}\mathbf{W}^n\mathbf{X}^{(k+1)}-\mathbf{v}^{(k+1)}.$$

end for

return $\hat{\mathbf{X}}$.

IV. Numerical Experiments

In this section, the performance of the proposed ENLTV-MDFP method are demonstrated in several experiments. First, the advantage of ENLTV-MDFP over other state-of-the-art SISR methods are demonstrated in both noiseless and noisy cases. To thoroughly verify the robustness of the proposed method to a variety of natural images, we conduct statistical experiments on a large image dataset. In addition, to further demonstrate the robustness of our method to inaccurate blur kernels, corresponding experiments are conducted. Furthermore, the effectiveness of each step in the proposed method is also demonstrated. Finally, the running times of the proposed method and other baselines are also reported.

A. Experimental Settings

1) Degradation Models: In our SR experiments, the HR image is first blurred by a 7×7 Gaussian kernel with standard deviation 1.5. Then, the blurred image is decimated by a factor 3. At last, the additive Gaussian noise is added. In the noiseless case (configuration 1), the noise level is 0, while in the noisy case (configuration 2), the noise level is 5.

According to [38], the main parameters of ENLTV-MDFP are empirically set to $T_I=15$, $p=7$, $r=13$, $h=24$, $L=10$. For configuration 1, the regularization parameters λ, η and μ are set to 4.5×10^{-5}, 6×10^{-2} and 1.2×10^{-2}, respectively. For configuration 2, λ, η and μ are set to 4.1×10^{-3}, 1.776 and 1.18×10^{-1}, respectively.

2) Comparison Baselines: The comparison baselines include Bicubic, one anchored neighborhood regression-based methods A+ [17], two deep convolutional network-based methods SRCNN [69] and VDSR [25], two non-local variational methods NLTV [52] and AHNLTV-AGD (note that, we solve the NLTV-driven problem via SBI for fair comparison), a sparse representation based method (NCSR) [66], a joint prior based method (SKR-NLM) [34], a CNN denoiser prior based method IRCNN [70], and a multi-scale method (MSEPLL) [71]. For the three learning-based methods A+ [17], SRCNN [69], and VDSR [25], their models are retrained properly according to the degradation models in our tests. The resultant images of different methods are evaluated perceptually and quantitatively. For the quantitative comparison, PSNR and SSIM results are reported.

3) Test Images and Datasets: Fig. 8 shows the 10 test images (Set10) used in our experiments, which are widely used in SR literature. For color images, different SR methods are only applied on the luminance component. In order to evaluate the robustness of ENLTV-MDFP to various images, we perform statistical experiments on a combined dataset. This combined dataset is formed by four commonly used datasets in SR, and contains 250 images of various contents (5 images from Set5 [69], 14 images from Set14 [69], 100 images from B100,[1] and the rest from Flickr [33]). Referring to [33]

[1] Available:http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench.

TABLE II
PSNR (dB) AND SSIM RESULTS ON NOISELESS IMAGES WITH A SCALE FACTOR OF 3

Methods	*Bird*	*Butterfly*	*Chip*	*Flower*	*House*	*Leaves*	*Parrot*	*Plants*	*Woman*	*Yacht*	Average
Bicubic	30.74	22.64	27.86	26.34	29.62	21.97	26.63	29.85	27.16	26.67	26.95
	0.8958	0.7797	0.8963	0.7371	0.8319	0.7425	0.8510	0.8333	0.8600	0.7919	0.8219
A+ [17]	35.37	27.15	32.44	28.96	32.99	26.11	29.60	33.54	31.19	29.75	30.71
	0.9532	0.9075	0.9606	0.8435	0.8837	0.8984	0.9097	0.9145	0.9266	0.8858	0.9083
SRCNN [23]	35.15	27.87	32.78	28.99	32.88	26.48	29.81	33.50	31.20	29.92	30.86
	0.9491	0.9077	0.9590	0.8415	0.8793	0.8994	0.9093	0.9097	0.9244	0.8841	0.9063
NLTV [52]	35.50	28.63	33.05	29.30	33.01	27.37	30.31	34.06	31.86	29.82	31.29
	0.9535	0.9282	0.9636	0.8534	0.8812	0.9231	0.9113	0.9183	0.9289	0.8849	0.9146
NCSR [66]	36.10	28.24	33.94	29.37	33.45	27.67	30.19	33.98	31.91	30.33	31.52
	0.9588	0.9207	0.9670	0.8547	0.8872	0.9248	0.9131	0.9193	0.9321	0.8967	0.9174
SKR-NLM [34]	34.83	26.54	32.54	28.75	32.45	26.04	29.78	33.08	30.78	29.44	30.42
	0.9497	0.8925	0.9581	0.8336	0.8748	0.8906	0.9059	0.9030	0.9191	0.8784	0.9006
AHNLTV-AGD [38]	36.21	29.73	33.82	29.74	33.59	28.21	30.60	34.74	32.40	30.41	31.94
	0.9589	0.9384	0.9692	0.8636	0.8875	0.9350	0.9165	0.9256	0.9351	0.8972	0.9227
VDSR [25]	36.54	30.09	33.24	29.77	33.45	28.72	30.76	34.89	32.39	30.11	32.11
	0.9608	0.9440	0.9715	0.8686	0.8873	0.9453	0.9203	0.9293	0.9372	0.8966	0.9261
IRCNN [70]	36.47	29.40	33.74	29.86	**34.33**	28.12	30.54	34.22	31.91	30.40	31.90
	0.9630	0.9385	0.9721	0.8692	**0.8932**	0.9360	0.9200	0.9263	0.9355	0.9020	0.9256
MSEPLL [71]	35.45	28.27	33.23	29.08	32.89	26.42	29.89	33.44	31.00	29.71	30.94
	0.9517	0.9202	0.9597	0.8457	0.8801	0.9034	0.9083	0.9078	0.9229	0.8825	0.9082
ENLTV-MDFP	**36.92**	**30.39**	**34.61**	**29.89**	34.20	**29.51**	**30.92**	**35.22**	**32.70**	**30.77**	**32.52**
	0.9620	**0.9444**	**0.9738**	**0.8693**	0.8923	**0.9501**	**0.9205**	**0.9313**	**0.9389**	**0.9030**	**0.9286**

Fig. 8. Test images (Set10). From left to right and top to bottom: *Bird* (288 × 288), *Butterfly* (256 × 256), *Chip* (244 × 200), *Flower* (256 × 256), *House* (256×256), *Leaves* (256×256), *Parrot* (256×256), *Plants* (256×256), *Woman* (228 × 344), *Yacht* (512 × 480).

and [38], all the test images are cropped to acquire 256×256 sub-images.

B. Experimental Results on 10 Test Images (Set10)

In this subsection, we present SR results of all competing methods on Set10 (configuration 1). The quantitative evaluation results of PSNR and SSIM are shown in Table II. To compare the perceptual quality of the SR results of ENLTV-MDFP and other baselines, the results of *Plants* are used as an example in Fig. 9. It can be observed that the bicubic interpolation produces the worst visual quality with blurring and aliasing artifacts. Although SRCNN, SKR-NLM, A+, and MSEPLL are more competitive than bicubic in preserving image edges, their performance is lower than NLTV in terms of both perceptual and quantitative evaluations. NCSR, IRCNN and AHNLTV-AGD can better infer the missing high-frequency details and achieve higher objective assessment performance than other comparison baselines. However, blurred or distorted artifacts still exist (e.g., the petal and stem of the plant in the super-resolved images). VDSR achieves better results than other baseline methods, but its PSNR/SSIM values are lower than those of the proposed method. Overall, the proposed ENLTV-MDFP method achieves the best PSNR/SSIM values

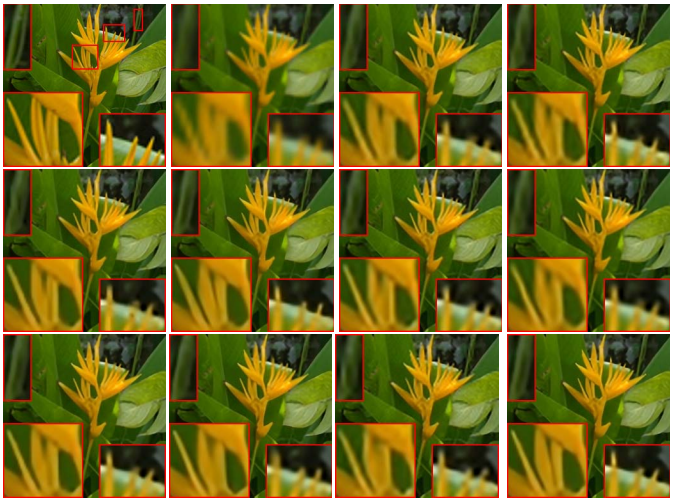


Fig. 9. SR results (×3, noise level 0) of *Plants* by different methods. From left to right and top to bottom: Original image, Bicubic (29.85, 0.8333), A+ (33.54, 0.9145), SRCNN (33.50, 0.9097), NLTV (34.06, 0.9183), VDSR (34.89, 0.9293), NCSR (33.98, 0.9193), SKR-NLM (33.08, 0.9030), AHNLTV-AGD (34.74, 0.9256), IRCNN (34.22, 0.9263), MSEPLL (33.44, 0.9078), and ENLTV-MDFP (35.22, 0.9313).

and the best visual quality among all these methods. The average PSNR/SSIM gains over the NCSR, IRCNN, VDSR, and AHNLTV-AGD are 1.00 dB/0.0112, 0.62 dB/0.0030, 0.41 dB/0.0025, and 0.58 dB/0.0059, respectively. These experimental results demonstrate the effectiveness of the proposed ENLTV-MDFP method in SR application.

C. Robustness to Noise

Since image noise will make the SR problem more challenging, we present SR results of all comparison methods on noisy images (configuration 2) to test the robustness of these methods against noise in this subsection.

Fig. 10. SR results (×3, noise level 5) of *Butterfly* by different methods. From left to right and top to bottom: Original image, LR image, Bicubic (22.58, 0.7652), NCSR (27.07, 0.8957), VDSR (28.12, 0.9068), AHNLTV-AGD (28.12, 0.9062), IRCNN (27.23, 0.8968), and ENLTV-MDFP (28.53, 0.9173).

TABLE III
AVERAGE PSNR (DB) AND SSIM RESULTS ON NOISY IMAGES WITH A SCALE FACTOR OF 3 (NOISE LEVEL: 5)

Bicubic	NCSR	AHNLTV -AGD	VDSR	IRCNN	ENLTV -MDFP
26.75	30.04	30.18	30.18	29.99	**30.67**
0.7993	0.8781	0.8757	0.8774	0.8766	**0.8917**

According to the experiment in subsection IV-B, VDSR, AHNLTV-AGD, IRCNN, and NCSR perform the 2nd, 3rd, 4th, and 5th best, respectively. We compare the proposed ENLTV-MDFP method with these four methods, and the corresponding average PSNR/SSIM scores are reported in Table III. For the visual quality comparison, the SR results of *Butterfly* are shown in Fig. 10. Specifically, in terms of the quantitative evaluation, IRCNN and NCSR perform on par with each other, while the edges are smoothed out to some extent. AHNLTV-AGD and VDSR can achieve relatively good performance with visually pleasant SR results. Overall, the proposed ENLTV-MDFP method has the best objective performance. Moreover, the recovered edges and high-frequency details by our method are much more accurate. For example, in our super-resolved *Butterfly* image, the edges and details of the patterns look much clearer than the results of the baseline methods. Consequently, the robustness of the proposed method to noise is well verified.

D. Experimental Results on Image Dataset

To comprehensively test the effectiveness and robustness of the proposed ENLTV-MDFP approach, we perform extensive experiments on the combined image dataset that contains 250 images in this subsection. The 2nd, 3rd, 4th, and 5th best comparison algorithms in Section IV-B (i.e., VDSR, AHNLTV-AGD, IRCNN, and NCSR) are selected as baselines.

We test these methods in both configuration 1 and configuration 2. The average objective results for these methods are tabulated in Table IV. We observe that, in both noiseless and noisy cases, the proposed ENLTV-MDFP method outperforms the competing methods. For configuration 1, the average PSNR/SSIM gains of the ENLTV-MDFP over NCSR, IRCNN, VDSR, and AHNLTV-AGD are 0.85 dB/0.0164, 0.47 dB/0.0058, 0.11 dB/0.0016, and 0.50 dB/0.0106, respectively. In configuration 2, the average PSNR/SSIM gains of the ENLTV-MDFP over NCSR, IRCNN, VDSR, and AHNLTV-AGD are 0.58 dB/0.0171, 0.46 dB/0.0139, 0.39 dB/0.0222, and 0.40 dB/0.0163, respectively. The SR results on dataset verify the robustness and superiority of the proposed ENLTV-MDFP method.

TABLE IV
AVERAGE PSNR (DB) AND SSIM RESULTS FOR ×3 MAGNIFICATION ON DATASET

Methods	NCSR	AHNLTV -AGD	VDSR	IRCNN	ENLTV -MDFP
Noiseless	27.95	28.30	28.59	28.33	**28.80**
	0.8173	0.8231	0.8321	0.8279	**0.8337**
Noise	27.08	27.26	27.27	27.20	**27.66**
	0.7714	0.7722	0.7663	0.7746	**0.7885**

TABLE V
AVERAGE PSNR (DB) AND SSIM RESULTS OF DIFFERENT EXTENDED VERSIONS ON SET10

Bicubic	MDFP	AHNLTV	ENLTV0	ENLTV	ENLTV -MDFP
26.95	32.27	31.59	31.69	31.98	**32.52**
0.8219	0.9269	0.9206	0.9210	0.9241	**0.9286**

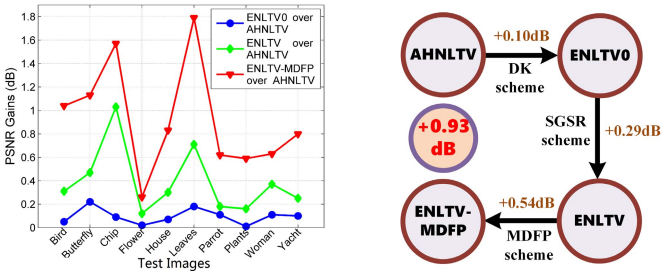


Fig. 11. The PSNR gains of different extended versions for each image (left) and the graphical illustration of the PSNR gain of each step (right).

E. Effectiveness of Each Step in ENLTV-MDFP

In this subsection, four extended methods are used to validate the effectiveness of each step in ENLTV-MDFP, including AHNLTV, ENLTV0 (improved AHNLTV with DK scheme), ENLTV (improved ENLTV0 with SGSR scheme), and ENLTV-MDFP (combining ENLTV with MDFP). We conduct experiments on Set10 in configuration 1, and report the average PSNR and SSIM scores of different extended versions in Table V. In addition, graphical illustrations of PSNR gains are given in Fig. 11. We can see from Fig. 11 that the average PSNR gain of ENLTV-MDFP over AHNLTV is as high as 0.93 dB. Moreover, for visual comparison, the results of *Leaves* are presented in Fig. 12.

1) Effectiveness of DK Scheme: We compare ENLTV0 with the original AHNLTV prior to verify the effectiveness of the DK scheme. As observed in Table V and Figs. 11 and 12, ENLTV0 outperforms AHNLTV in terms of PSNR and SSIM values (the average PSNR/SSIM gains are 0.10 dB/0.0004), which verifies the effectiveness of the DK scheme.

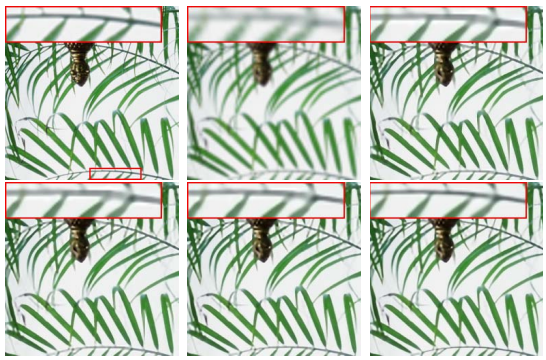


Fig. 12. SR results (×3, noise level 0) of *Leaves* by different extended versions of the proposed ENLTV-MDFP method. From left to right and top to bottom: Original image, Bicubic (21.97, 0.7425), AHNLTV (27.75, 0.9312), ENLTV0 (27.93, 0.9331), ENLTV (28.46, 0.9408), and ENLTV-MDFP (29.51, 0.9501).

2) Effectiveness of SGSR Shceme: First, we compare ENLTV0 with its normalized version. However, according to our tests, the PSNR/SSIM gains of ENLTV0 over its normalized version is minimal (only about 0.02 dB/0.0001). It reveals that $\mathbf{W}_{CI}^{Sum}$ is not powerful in measuring the group similarity reliability. In contrast, with the SGSR scheme, ENLTV can further produce sharper edges, providing significant improvements (0.29 dB/0.0031) over ENLTV0. Consequently, the effectiveness of the SGSR scheme is verified.

3) Effectiveness of Combining ENLTV With MDFP: Combining ENLTV with MDFP can achieve 0.54 dB/0.0045 improvement over using only the separated ENLTV prior. In terms of the visual quality, the combined method can produce sharp edges and fine structures, and suppresses the artifacts well. It demonstrates that the combination of ENLTV and MDFP can fully exploit their complementary advantages, and is helpful to enhance the quality of super-resolved images.

F. Discussion on Different Initialization Strategies

To show the effectiveness of the proposed initial HR estimation strategy in terms of both SR accuracy and computational time, we conduct the following experiments. Two different initialization strategies with different outer iteration number are tested on Set10. This implies that the values of $\mathcal{Q}$ and $w_n^{DK}(i, j)$-s are computed iteratively. Fig. 13(a) plots the progression curves of the PSNR results achieved by solving the objective function in Eq. (45) with initial HR estimation strategy and conventional bicubic strategy (the corresponding two average PSNR curves (labeled in pink) of Set10 are also plotted). The magnified views of the average PSNR in the rectangle regions of Fig. 13(a) are shown in Figs. 13(b)-(d). For the bicubic strategy, the PSNR curves converge when the outer iteration number reaches 3, and there is little PSNR improvement when the outer iteration number is larger than 3. For the initial HR estimation strategy, the PSNR curves almost converge even when the outer iteration number is 1. Additional iteration can achieve about 0.03dB on average. We conclude that the initial HR estimation strategy is more efficient and effective than the bicubic strategy in solving the objective

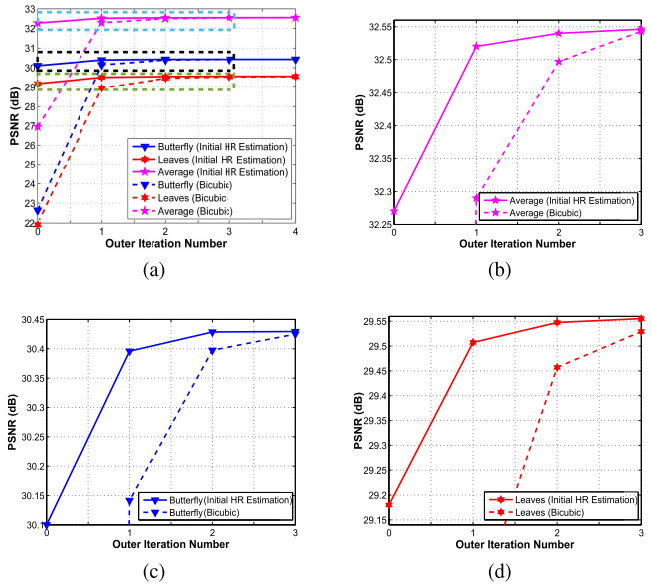


Fig. 13. The effect of different initialization strategies on SR accuracy. (a) Comparison between initial HR estimation strategy and bicubic strategy in terms of the progression of PSNR. (b) Magnified view of average PSNR in the rectangle region. (c) Magnified view of PSNR for *Butterfly* image in the rectangle region. (d) Magnified view of PSNR for *Leaves* image in the rectangle region.

function in Eq. (45). Obviously, a larger iteration number requires a longer running time. For example, when the outer iteration number is 3, the running time is 3 times longer than that when the outer iteration number is 1. Therefore, to best balance the SR quality and running time, we adopt initial HR estimation with iteration number 1 for the sake of fast implementation.

G. Robustness to Inaccurate Blur Kernels

Since all the competing methods in previous experiments belong to non blind SR, it is assumed that the blur kernel is known. In order to test the robustness of these methods when the blur kernel is inaccurate, the following experiments are carried out. To simplify the analysis, we only consider the estimation error of standard deviation. Specifically, the Gaussian kernel in subsection IV-A is used as an inaccurate estimation kernel in these tests. According to [37], the ratio p_{te} is used to metric the estimation error, and it is defined as

$$p_{te} = (\sigma_e - \sigma_t)/\sigma_t \times 100\% \tag{59}$$

where σ_t is the true standard deviation, and $\sigma_e = 1.5$ is the estimated standard deviation. Five competing methods, including ENLTV-MDFP, NCSR, IRCNN, VDSR, and AHNLTV-AGD, are tested for five different p_{te}-s (i.e., -20%, -10%, 0%, 10%, 20%). The average PSNR/SSIM results (Configuration 1) vs. p_{te} are given in Fig. 14. We observe that the performance of all SR methods is decreased as the estimation error increases. Among them, the CNN denoiser prior based method IRCNN is largely affected by p_{te}, and its SR ability will significantly decrease with a larger $|p_{te}|$. However, the results suggest that the ENLTV-MDFP method consistently outperforms NCSR, IRCNN, VDSR, and AHNLTV-AGD in

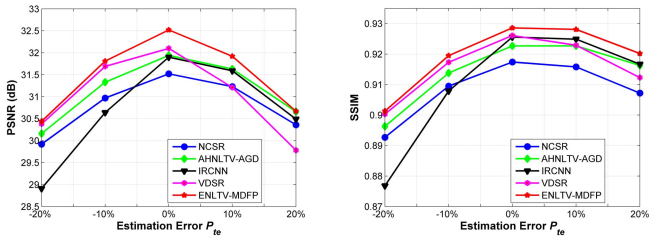


Fig. 14. Average PSNR and SSIM results of reconstructed HR images on Set10 with inaccurate blur kernels.

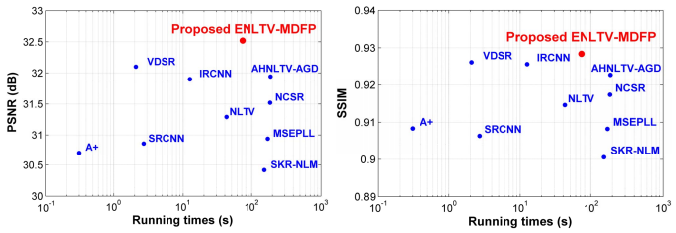


Fig. 15. Average PSNR/SSIM scores (Configuration 1) vs. running time (s).

all scenarios. Overall, the results of these experiments demonstrate the robustness of the proposed method to inaccurate blur kernels.

H. Discussion on Computational Time

To comprehensively evaluate the computational time of the proposed method, the results of PSNR/SSIM vs. running times of each method are reported in Fig. 15. As depicted in Fig. 15, for Set10, on an Intel Core i7 7700K CPU in Linux platform, ENLTV-MDFP achieves the best SR performance with an average of 72.1s. MSEPLL (171.4s), SKR-NLM (152.8s), NCSR (185.4s) and AHNLTV-AGD (187.6s) takes much more running times than the proposed method, and their PSNR and SSIM are also much lower. Although, A+, SRCNN, IRCNN (CPU mode), VDSR (CPU mode), and NLTV are faster than ENLTV-MDFP, their performance is relatively lower. In terms of the running times, we can conclude that the ENLTV-MDFP method is more efficient comparing to many state-of-the-art reconstruction-based SR methods, although it is relatively slower than those of learning-based methods. Overall, the proposed method produces the best SR results with reasonable computational time.

V. Conclusion

In this paper, we present a novel SR method (ENLTV-MDFP) by integrating the non-local variational prior and the learned local multi-directional feature prior into the reconstruction framework. The non-local similarity is modeled by the DK and SGSR-based variational method using the HR image itself (ENLTV), and the feature prior is learned by the deep CNN using external images (MDFP). Overall, the proposed combined SR framework fully exploits the advantages of these two complementary regularization terms, and achieves state-of-the-art SR performance. The quantitative and visual evaluation demonstrates that the proposed ENLTV-MDFP method achieves significant improvement over the competing methods. In our future work, we plan to investigate other methods to generate direction features such as wavelet features. We will also explore the extensions of the proposed framework on other image restoration applications, such as deblurring, denoising, deblocking, inpainting, and SR of compressed images.

References

[1] R. Timofte, R. Rothe, and L. Van Gool, "Seven ways to improve example-based single image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1865–1873.

[2] J. Yu, X. Gao, D. Tao, X. Li, and K. Zhang, "A unified learning framework for single image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 780–792, Apr. 2014.

[3] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[4] K. Zhang, D. Tao, X. Gao, X. Li, and Z. Xiong, "Learning multiple linear mappings for efficient single image super-resolution," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 846–861, Mar. 2015.

[5] R. G. Keys, "Cubic convolution interpolation for digital image processing. IEEE Trans acoust speech signal process," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Jan. 1982.

[6] Y. Romano, M. Protter, and M. Elad, "Single image interpolation via adaptive nonlocal sparsity-based modeling," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3085–3098, Jul. 2014.

[7] Z. Wei and K.-K. Ma, "Contrast-guided image interpolation," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4271–4285, Nov. 2013.

[8] R. Timofte, V. De, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[9] Y. Zhu, Y. Zhang, and A. L. Yuille, "Single image super-resolution using deformable patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2917–2924.

[10] Q. Ning, K. Chen, L. Yi, C. Fan, Y. Lu, and J. Wen, "Image super-resolution via analysis sparse prior," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 399–402, Apr. 2013.

[11] C. He, L. Liu, L. Xu, M. Liu, and M. Liao, "Learning based compressed sensing for SAR image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1272–1281, Aug. 2012.

[12] Z. Pan *et al.*, "Super-resolution based on compressive sensing and structural self-similarity for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4864–4876, Sep. 2013.

[13] Y. Sun, G. Gu, X. Sui, Y. Liu, and C. Yang, "Single image super-resolution using compressive sensing with a redundant dictionary," *IEEE Photon. J.*, vol. 7, no. 2, pp. 1–11, Apr. 2015.

[14] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1674–1684, Sep. 2016.

[15] Z. Xiong, D. Xu, X. Sun, and F. Wu, "Example-based super-resolution with soft information and decision," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1458–1465, Oct. 2013.

[16] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

[17] R. Timofte, V. D. Smet, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. 12th Asian Conf. Comput. Vis.* Springer, 2014, pp. 111–126.

[18] Y. Tian, F. Zhou, W. Yang, X. Shang, and Q. Liao, "Anchored neighborhood regression based single image super-resolution from self-examples," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2827–2831.

[19] D. Dai, R. Timofte, and L. V. Gool, "Jointly optimized regressors for image super-resolution," in *Proc. Comput. Graph. Forum*, vol. 34, no. 2, 2015, pp. 95–104.

[20] J. Jiang, X. Ma, C. Chen, T. Lu, Z. Wang, and J. Ma, "Single image super-resolution via locally regularized anchored neighborhood regression and nonlocal means," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 15–26, Jan. 2017.

[21] T. Lu, L. Pan, J. Jiangs, Y. Zhang, and Z. Xiong, "DLML: Deep linear mappings learning for face super-resolution with nonlocal-patch," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 1362–1367.

[22] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen, "Deep network cascade for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 49–64.

[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[24] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. Mccann, and M. Unser. (Sep. 6, 2017). "CNN-based projected gradient descent for consistent image reconstruction," [Online]. Available: https://arxiv.org/abs/1709.01809

[25] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[26] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jul. 2017, pp. 1132–1140.

[27] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[28] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans Image Process.*, vol. 25, no. 7, pp. 3194–3207, Jul. 2016.

[29] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.

[30] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5835–5843.

[31] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, and R. Hu, "SRLSP: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 27–40, Jan. 2017.

[32] J. Liu, W. Yang, X. Zhang, and Z. Guo, "Retrieval compensated group structured sparsity for image super-resolution," *IEEE Trans. Multimedia*, vol. 19, no. 2, pp. 302–316, Feb. 2017.

[33] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[34] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4544–4556, Nov. 2012.

[35] X. Li, H. He, R. Wang, and D. Tao, "Single image superresolution via directional group sparsity and directional features," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2874–2888, Sep. 2015.

[36] Q. Yan, Y. Xu, X. Yang, and T. Q. Nguyen, "Single image superresolution based on gradient profile sharpness," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3187–3202, Oct. 2015.

[37] C. Ren, X. He, Q. Teng, Y. Wu, and T. Q. Nguyen, "Single image super-resolution using local geometric duality and non-local similarity," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2168–2183, May 2016.

[38] C. Ren, X. He, and T. Q. Nguyen, "Single image super-resolution via adaptive high-dimensional non-local total variation and adaptive geometric feature," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 90–106, Jan. 2017.

[39] K. Chang, P. L. K. Ding, and B. Li, "Single image super resolution using joint regularization," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 596–600, Apr. 2018.

[40] T. Li, X. He, L. Qing, Q. Teng, and H. Chen, "An iterative framework of cascaded deblocking and superresolution for compressed images," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1305–1320, Jun. 2017.

[41] C. Ren, X. He, and T. Q. Nguyen, "Adjusted non-local regression and directional smoothness for image restoration," *IEEE Trans. Multimedia*, vol. 21, no. 3, pp. 731–745, Mar. 2019.

[42] S. Huang, J. Sun, Y. Yang, Y. Fang, P. Lin, and Y. Que, "Robust single-image super-resolution based on adaptive edge-preserving smoothing regularization," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2650–2663, Jun. 2018.

[43] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.

[44] A. Shocher, N. Cohen, and M. Irani, "'Zero-shot' super-resolution using deep internal learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3118–3126.

[45] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 349–356.

[46] K. Zhang, X. Gao, X. Li, and D. Tao, "Partially supervised neighbor embedding for example-based image super-resolution," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 230–239, Apr. 2011.

[47] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 370–378.

[48] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556

[49] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 492–526, 2010.

[50] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65.

[51] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, Nov. 2008.

[52] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM J. Imag. Sci.*, vol. 3, no. 3, pp. 253–276, 2010.

[53] S. Tang, W. Gong, W. Li, and W. Wang, "Non-blind image deblurring method by local and nonlocal total variation models," *Signal Process.*, vol. 94, pp. 339–349, Jan. 2014.

[54] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, p. 205, 2009.

[55] B. Wen, Y. Li, and Y. Bresler. (2018). "The power of complementary regularizers: Image recovery via transform learning and low-rank modeling." [Online]. Available: https://arxiv.org/abs/1808.01316

[56] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, pp. 511–518.

[57] J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G. J. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," in *Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro.*, May 2008, pp. 1331–1334.

[58] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[59] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[60] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777.

[61] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, p. 2012, 2013.

[63] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.

[64] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015.

[65] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[66] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[67] S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comput.*, vol. 3, no. 3, pp. 165–218, 2011.

[68] T. Goldstein and S. Osher, "The split bregman method for L1-regularized problems," *SIAM J. Imag. Sci.*, vol. 2, no. 2, pp. 323–343, 2009.

[69] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2014, pp. 184–199.

[70] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jul. 2017, pp. 2808–2817.

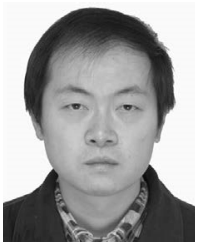
[71] V. Papyan and M. Elad, "Multi-scale patch-based image restoration," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 249–261, Jan. 2016.

Chao Ren (M'17) received the B.S. degree in electronics and information engineering and the Ph.D. degree in communication and information system from Sichuan University, Chengdu, China, in 2012 and 2017, respectively. From 2015 to 2016, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California at San Diego, CA, USA.

He is currently an Associate Research Professor with the College of Electronics and Information Engineering, Sichuan University. He has received support from the National Postdoctoral Program for Innovative Talents of China and National Natural Science Foundation of China. His research interests include inverse problems in image and video processing.

Xiaohai He (M'15) received the B.S. and M.S. degrees in electronics and electrical engineering and the Ph.D. degree in biomedical engineering from Sichuan University, Chengdu, China, in 1985, 1991, and 2002, respectively.

He is currently a Professor with the College of Electronics and Information Engineering, Sichuan University. His research interests include image processing, pattern recognition, computer vision, image compression, and software engineering. He is a Senior Member of the Chinese Institute of Electronics. He was an Editor of the *Journal of Data Acquisition & Processing* and is an Editor of the *Journal of Information and Electronic Engineering*.



Yifei Pu received the Ph.D. degree from the College of Electronics and Information Engineering, Sichuan University, in 2006.

He is currently a Professor with the College of Computer Science, Sichuan University. He is elected into the Thousand Talents Program of Sichuan Province as the Academic and Technical Leader of Sichuan Province. He focuses on the application of fractional calculus and fractional partial differential equation to signal analysis, image processing, circuits and systems, and machine intelligence. He has authored, with the first author's identity, about 20 papers indexed by SCI, in journals such as the *International Journal of Neural Systems*, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE ACCESS. He held positions in several research projects, such as the National Nature Science Foundation of China.

Truong Q. Nguyen (F'05) is currently a Professor with the Electrical and Computer Engineering Department, University of California at San Diego, San Diego, CA, USA. He has authored more than 400 publications and several MATLAB-based toolboxes on image compression, electrocardiogram compression, and filterbank design. He has co-authored (with Prof. G. Strang) a textbook *Wavelets and Filter Banks* (Wellesley-Cambridge, 1997). His research interests include 3D video processing and communications and their efficient implementation.

Prof. Nguyen received the IEEE TRANSACTIONS ON SIGNAL PROCESSING Paper Award (Image and Multidimensional Processing Area) for a paper that he co-authored with Prof. P. P. Vaidyanathan on linear-phase perfect reconstruction filter banks in 1992 and the NSF Career Award in 1995. He is also the Series Editor of *Digital Signal Processing* (Academic Press). He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1996, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1996 to 1997 and from 2001 to 2004, the IEEE SIGNAL PROCESSING LETTERS from 2001 to 2003, and the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2004 to 2005.