Signal Processing 140 (2017) 177–189

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Robust coupled dictionary learning with ℓ_1-norm coefficients transition constraint for noisy image super-resolution

Bo Yue[a], Shuang Wang[a,*], Xuefeng Liang[b], Licheng Jiao[a]

[a] Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, 710071, China
[b] IST, Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan

ARTICLE INFO

Article history:
Received 9 December 2016
Revised 17 April 2017
Accepted 19 April 2017
Available online 20 April 2017

Keywords:
Image super-resolution
Coupled dictionary learning
ℓ_1-norm
Non-linear mapping
Non-local self-similarity

ABSTRACT

Conventional coupled dictionary learning approaches are designed for noiseless image super-resolution (SR), but quite sensitive to noisy images. We find that the cause is the commonly used ℓ_F-norm coefficients transition term. In this paper, we propose a robust ℓ_1-norm solution by introducing two subterms: *LR coefficient sparsity constraint term* and *HR coefficient conversion term*, which are able to prevent the noise transmission from noisy input to output. By incorporating our simple yet effective non-linear model inspired by auto-encoder, the proposed ℓ_1-norm dictionary learning achieves a more accurate coefficients conversion. Moreover, to make the coefficients conversion more reliable in the iterative process, we bring the non-local self-similarity constraint to regularize the HR sparse coefficients updates. The improved sparse representation further enhances SR inference on both synthesized noisy and noiseless images. Using standard metrics, we show that results are significantly clearer than state-of-the-arts on noisy images and sharper on denoised images. In addition, experiments on real-world data further demonstrate the superiority of our method in practice.

1. Introduction

The goal of single image super-resolution (SISR) is to reconstruct a high resolution (HR) image from a low resolution (LR) input. This problem is inherently ill-posed, thus very challenging in computer vision. To solve it, different forms of prior knowledge have been explored. In particular, learning-based (example-based) strategy [1–14], which captures the prior information by learning one or more mapping functions from generic datasets, has received a great attention in the past decade. It assumes that the lost details in LR images can be recovered by the prior learned from millions external LR-HR image patch pairs due to the richness of real-world images.

The dominant approach, namely coupled dictionary learning, utilizes the sparse representation to model the relation between the sparse coefficients of LR and HR patches over an over-complete dictionary pair ($\mathbf{D}_l$, $\mathbf{D}_h$). It first extracts overlapped patches from the LR image, which are then encoded as the higher dimensional sparse vectors with respect to a LR dictionary $\mathbf{D}_l$. Next, the LR sparse coefficients are projected to HR ones via a learned mapping function. Finally, the HR sparse coefficients are passed into a HR dictionary $\mathbf{D}_h$ for reconstructing HR patches. To learn the coefficients conversion, various assumptions have been imposed on the underlying mapping model from the initial sparse representation invariance [1–3,8], through to the linear mapping [4,5], and finally to the statistical dependence [6,9]. This task is carried out through the ℓ_F-norm coefficients transition term $\|\mathbf{W}_l^{-1}\alpha_l - \mathbf{W}_h^{-1}\alpha_h\|_F^2$ in the shared objective function (1) by optimizing the projection matrices ($\mathbf{W}_l^{-1}$, $\mathbf{W}_h^{-1}$) under varied assumptions. With the ℓ_F-norm minimizing the mapping error, these models achieve quite convincing performances for the noiseless image SR. Unfortunately, LR images often contain certain noise in the real-world applications. We find all above approaches are sensitive to noise, even the noise is weak. Fig. 1 demonstrates the noise sensitivities of one representative of sparse representation approach (Yang et al. [1]), two other recent approaches (Wang et al. [4] and A$^+$ method [8]) and the proposed method.

In this paper, we explore the mechanism of noise sensitivity of conventional approaches by studying the convex objective function (1) in Section 3.1. One can see that the coefficients conversion is carried out by the ℓ_F-norm due to the fact that ℓ_F-norm is easy to be optimized and its solution is unique. Despite its popularity, ℓ_F-norm is not a right choice for noisy image SR. Our analysis reveals that the ℓ_F-norm coefficients transition term causes the sparsity of

* Corresponding author.
E-mail addresses: yuebo313@126.com (B. Yue), shwang.xd@gmail.com (S. Wang), xliang@i.kyoto-u.ac.jp (X. Liang), LCHJiao@mail.xidian.edu.cn (L. Jiao).

http://dx.doi.org/10.1016/j.sigpro.2017.04.015
0165-1684/

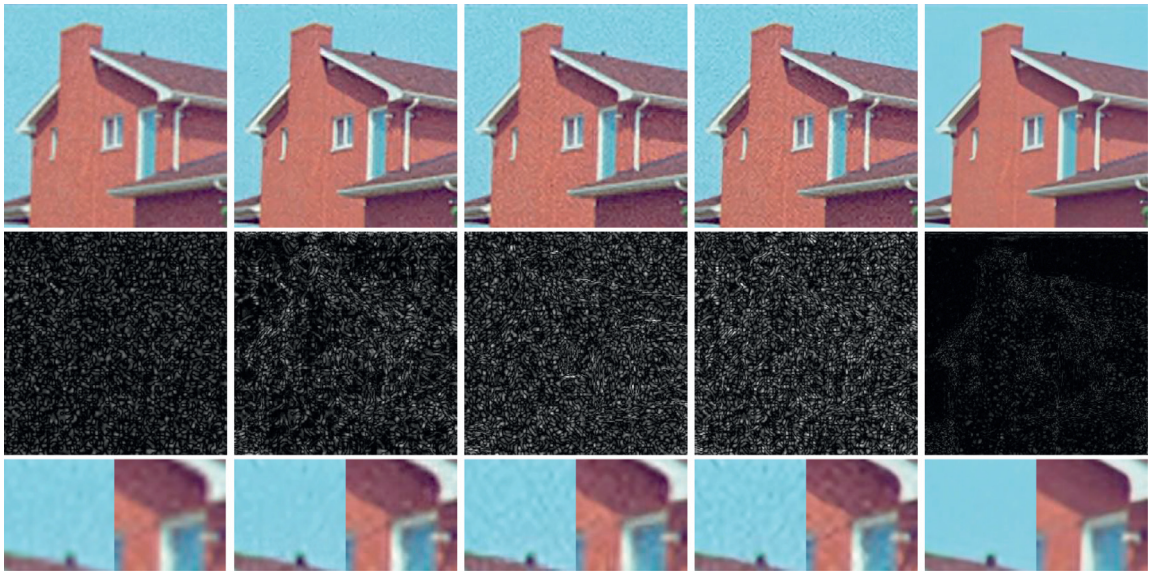

Fig. 1. Sensitivities to noise in SR, where Gaussian noise with standard deviation of 5 is added to the LR input. **Top:** SR outputs. **Middle:** noise residual maps visualizing the difference between the SR outputs when the input is a noisy LR image or the corresponding noiseless LR image. **Bottom:** the magnified local patches of SR output. (a) LR input interpolated by Bicubic operator; (b) Result by Yang et al. [1]; (c) Result by Wang et al. [4]; (d) Result by A$^+$ approach [9]; (e) Result by proposed method.

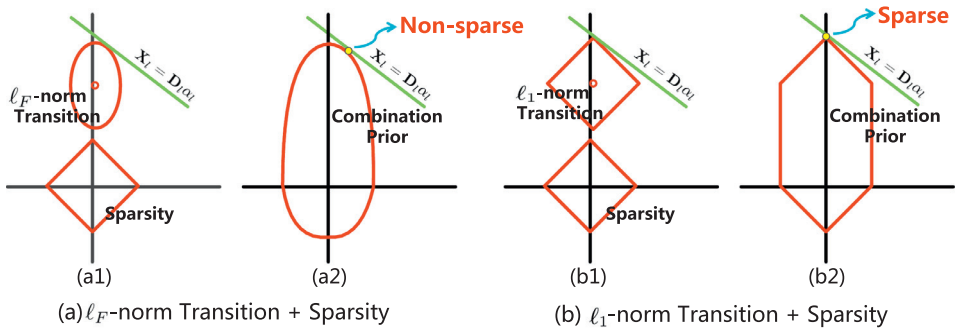

Fig. 2. The combination of transition term and sparsity-inducing term acting on the state in sparse-coding. (a) ℓ_F-norm regularization: (a1) The ellipse represents the ℓ_F-norm coefficients transition term, the diamond denotes the sparsity-inducing term. (a2) The joint solution takes place off the axis, which is not sparse. (b) ℓ_1-norm regularization: (b1) The upper diamond represents the ℓ_1-norm coefficients transition term. (b2) The joint solution takes place on the axis, which is much sparser.

LR sparse coefficients poorly maintained, and further encodes noise from the LR sparse coefficients into the HR sparse coefficients. A 2D geometric interpretation is given in Fig. 2.(a) by solely considering the solution to LR sparse coefficients. Owing to the convexity of ℓ_F-norm, the solution/intersection mostly takes place off the axis with all non-zero coordinates, which is not sparse.

With this in mind, we thus put emphasis on the upgrade of coefficients transition term. To well maintain the sparsity of LR coefficients, the intersection is desired to take place on the axis. We find that the ℓ_1-norm is a better choice, please refer to Fig. 2.(b). Nevertheless, simply replacing the ℓ_F-norm by a ℓ_1-norm to the entire transition term brings two problems. Firstly, it will downgrade the performance on the calculation of HR coefficients, because ℓ_1-norm restrains the non-zero sum other than the mean squared error in ℓ_F-norm. To address this problem, we introduce two auxiliary variables to split the transition term into two subterms: LR coefficient sparsity constraint term with ℓ_1-norm and HR coefficient conversion term with ℓ_F-norm in Section 3.2. Secondly, the presence of two ℓ_1-norm terms (the transition term and the sparsity-inducing term) cannot apply the standard optimization techniques to sparse-coding. To alleviate this problem, we propose a smooth proximal gradient method to approximate the transition term, and then are able to employ the efficient solvers, e.g. Fast Iterative Shrinkage Thresholding algorithm [15] in Section 3.4 (**Update α_l, α_h**).

Furthermore, considering after applying ℓ_1-norm on LR coefficient and ℓ_F-norm on HR coefficient, the coefficients conversion learning becomes more complicated. Either the sparse representation invariance [1–3,8] or the linear mapping [4,5] is incompetent to carry out an accurate relation modeling task, which is crucial to the dictionary learning SR approaches. Instead, we propose a non-linear method inspired by the auto-encoder that possesses an encoding process and a decoding process. It is simple yet effectively embedded into our optimization procedure in Section 3.2. Once established, we can iteratively update the LR patch representation and the estimated HR patch in the inference procedure. Unfortunately, it does not guarantee the HR patch free of noise at the first few iterations. Indeed, if an accurate initial estimation is given, the proposed method is able to steadily improve its output. Thus, we introduce a non-local constraint into the dictionary learning that correlates the similarity among non-local patches to regularize the HR sparse coefficients updates in Section 3.3. This constraint is able to distinguish the texture from noise effectively and results in a better noise suppression.

To the best of our knowledge, it is the first work in the literature to use coupled dictionary learning method on noisy image

super-resolution. Meanwhile, we show improved PSNR and SSIM over the competing sparse-coding approaches for a wide range of noise levels. In particular, given noise level from 0 to 20, our performance degrades only 2.53 dB in PSNR and 0.1140 in SSIM comparing with the least decrease (5.96 dB in PSNR and 0.3345 in SSIM) of others. In configuration of denoising + SR, our method also outperforms others on denoised inputs by improving PSNR 0.53 dB, SSIM 0.0122 at noise level 20. To show the generalization, our method is compared on the real noisy images and noiseless data against state-of-the-arts as well.

The rest of the paper is organized as follows. In the next section we briefly review the previous related works. The proposed model and its parameters training are presented in Section 3. The experimental results on several different image SR tasks are demonstrated in Section 4. Section 5 provides some conclusions and discussions.

2. Related work

Among the coupled dictionary learning based SR approaches, Yang's approaches [1,2] and Zeyde's method [3] are the most widely used, which assume that the sparse representation is invariant over the LR and HR dictionary pair. However, this simple assumption restricts its ability of the SR recovery. After that, there have been several attempts to go beyond the invariance assumption to improve the flexibility for accurate relationship learning. Wang et al. [4] proposed a semi-coupled dictionary training model by considering a linear mapping between the LR-HR coefficients. Similar to Wang's work, a beta process model [6] was proposed to learn the dictionaries whose sparse representations have the same sparsity but different values. It reported a more consistent and accurate mapping result. Alternatively, Huang et al. [5] presented a joint model which incorporated a common feature space learning into the coupled dictionary scheme to better describe the relationship. Recently, A^+ approach [8] exploits the same invariance assumption as Yang's works and combines the dictionaries with neighbor embedding methods to obtain improvements in both quality and speed. Dai et al. [16] jointly learned a collection of invariant mapping functions to alleviate the inability of a universal regressor for modeling the complex relationship. More recently, inspired by the deep learning achieving great success in many computer vision application, deep convolutional neural networks [17–21] are exploited to directly learn the mapping from LR input to HR output and achieved state-of-the-art performance.

Notwithstanding the demonstrated success in the noiseless image SR, none of the conventional coupled dictionary learning approaches is robust to noisy LR images. To our best knowledge, very few works address on this issue. A straightforward thought could do image denoising first, and super-resolve the denoised image afterward. Our test shows that many high frequency details are inevitably lost by denoising process, which is absolutely important to SR process. SRNI [22] tried to integrate the merits of image denoising and image SR. They first super-resolve the noisy LR image and the denoised LR image. Then, a convex combination of the denoised HR image and noisy HR image is exploited to obtain the final HR image. Recently, single dictionary learning approaches [23–25]are proposed for SR and denoising. It regularizes the degradation model $\mathbf{y} = \mathbf{AHx} + \mathbf{n}$ (where $\mathbf{y}$ is the noisy LR image, $\mathbf{A}$ is a downsampling, $\mathbf{H}$ is a blurring operator, $\mathbf{x}$ is the HR image to be recovered and $\mathbf{n}$ is the noise) with some sparsity priors (such as centralized sparsity [26], group sparsity [27,28], etc) by robust matrix factorization strategy [29–32]. Then, the final result is obtained through a maximum a posteriori (MAP).

Unlike all previous approaches, we provide a ℓ_1-norm solution to suppress the noise, meanwhile, incorporate a non-linear model inspired by auto-encoder and a non-local self-similarity constraint for accurate and reliable coefficients conversion. As was expected, our method out-competes previous approaches on noisy data by a big margin.

3. Robust coupled dictionary learning

This section describes our solution for noisy image SR. We start by analyzing the cause of noise sensitivity in conventional approaches that use ℓ_F-norm to regularize the relation between the sparse coefficients (α_l, α_h) in LR and HR images. The ℓ_F-norm is found to be the reason of encoding the noise from α_l to α_h. Next, we provide our ℓ_1-norm solution by introducing two auxiliary variables to make the objective function solvable. To model the relation between α_l and α_h more precise under the new regularization, we design a non-linear mapping approach. Afterward, we apply the non-local constraint into our dictionary learning framework to stabilize the estimation procedure and further improve the output performance. Finally, we detail the optimization procedure followed by the learning and the inference algorithms.

3.1. The cause of noise sensitivity

It has been known that noiseless images can be approximated as a linear combination of a few elementary atoms, which is named sparse representation/coding. This find lays the foundation of the coupled dictionary learning methods for SR. Their general objective function is formulated as follows:

$$\begin{aligned}\underset{\mathbf{D}_l,\mathbf{D}_h,\mathbf{W}_l,\mathbf{W}_h,\alpha_l,\alpha_h}{\operatorname{argmin}} & \frac{1}{2}(\|\mathbf{X}_l - \mathbf{D}_l\alpha_l\|_F^2 + \|\mathbf{X}_h - \mathbf{D}_h\alpha_h\|_F^2) \\ & + \lambda\|\mathbf{W}_l^{-1}\alpha_l - \mathbf{W}_h^{-1}\alpha_h\|_F^2 + \mu(\|\alpha_l\|_1 + \|\alpha_h\|_1),\end{aligned} \tag{1}$$

where α_l and α_h are the sparse coefficients over the LR-HR dictionary pair ($\mathbf{D}_l$, $\mathbf{D}_h$), respectively, $\|\alpha_l\|_1$ and $\|\alpha_h\|_1$ are the sparsity-inducing terms. Since $\mathbf{X}_l$ and $\mathbf{X}_h$ are in two different spaces, one of the tasks in function (1) is to establish the relation to link sparse coefficients in two spaces. The *coefficients transition term* $\|\mathbf{W}_l^{-1}\alpha_l - \mathbf{W}_h^{-1}\alpha_h\|_F^2$ is designed to learn the relation, where $\mathbf{W}_l^{-1}$ and $\mathbf{W}_h^{-1}$ are the projection matrices who project α_l and α_h into a common feature space. Here we use the inverse operation to regularize the projection matrices to prevent the model from learning the trivial solution (where the projection matrices are zeros). λ and μ are the regularization parameters to balance the terms. Several algorithms have been proposed to solve the objective function (1), such as the joint dictionary training algorithm [1,3,7], the coordinate descent algorithm [4,5] and the Bayesian algorithm [6].

In practice, most LR images contain certain noise. With a well-designed dictionary, the image signals can be easily separated from the noise via sparse representation. Thus, the sparsity-inducing term $\|\alpha_l\|_1$ in the function (1) is expected to suppress the noisy. Extensive experiments using the conventional approaches, however, do not show the desired efficacy.

By analyzing the convex solution of function (1), we consider that the ℓ_F-norm on the coefficients transition term makes the sparsity of α_l poorly maintained and causes the failed noise suppression. The 2D geometric interpretation is shown in Fig. 2. Let's only consider the LR space because the HR images are assumed to be free of noise. The ellipse (ℓ_p *ball*) in Fig. 2.(a1) represents $\|\mathbf{W}_l^{-1}\alpha_l - \mathbf{W}_h^{-1}\alpha_h\|_F^2$, the diamond denotes $\|\alpha_l\|_1$, and the green line is $\mathbf{X}_l = \mathbf{D}_l\alpha_l$. When we jointly solve them, the combination of ellipse and diamond becomes the convex shape in Fig. 2.(a2). One can see that the intersection (the problem solution α_l) of the red curve and the green line takes place off the axis which is not sparse, with all non-zero coordinates. The noise in the LR images, therefore, is transmitted/encoded into α_h through the coefficient transition term.

To suppress the noise, the intersection of the green line and the ℓ_p ball is expected to take place on the axis. This situation happens only if $0 < p \leq 1$. While $p < 1$, the solution is no longer convex and difficult to be solved. Thus, ℓ_1-norm becomes a better choice which is convex and the tendency to sparsity we are referring to. As we move the coefficient transition term from ℓ_F-norm towards ℓ_1-norm, Fig. 2.(b) shows that the intersection takes place on the axis, where most coordinates are zeros. Thus, it leads to a sparser solution. More specifically, we are able to prevent the noise from α_l to be encoded into α_h estimation using a ℓ_1-norm coefficient transition term.

3.2. ℓ_1-norm and coefficients conversion learning

However, simply applying a ℓ_1-norm to the coefficient transition term brings two problems. In this section, we focus on the first problem of an inaccurate α_h calculation, and provide our solution in below. For the second one of optimizing two ℓ_1-norm terms in the objective function (1), we propose a smooth proximal gradient algorithm to solve it in the Section 3.4 (**Update α_l, α_h**).

Applying a ℓ_1-norm to the entire coefficient transition term leads to an inaccurate HR coefficients calculation, $\alpha_h = \mathbf{W}_h\mathbf{W}_l^{-1}\alpha_l$, because $\|\mathbf{W}_l^{-1}\alpha_l - \mathbf{W}_h^{-1}\alpha_h\|_1$ restrains the non-zero sum rather than the mean square error in ℓ_F-norm. This is one of the reasons that ℓ_F-norm is widespread in various problems. To solve it, we introduce two auxiliary variables $\mathbf{P}_h$ and $\mathbf{P}_l$ to split the transition term into two sub-terms: *LR coefficient sparsity constraint term* with ℓ_1-norm for ensuring the sparsity, and *HR coefficient conversion term* with ℓ_F-norm for the accuracy guarantee. The new definition is given as follows:

$$\begin{aligned} &\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 + \|\alpha_h - \mathbf{T}_h^T\mathbf{P}_l\|_F^2 \\ &\quad s.t. \quad \mathbf{P}_h = \mathbf{T}_h\alpha_h, \mathbf{P}_l = \mathbf{T}_l\alpha_l, \end{aligned} \tag{2}$$

where $\mathbf{T}_h$ and $\mathbf{T}_l$ are the transition matrices of α_h and α_l respectively. With the new transition term ready, we note that the joint dictionary learning approaches in [1–3,8] are special cases when $\mathbf{W}_h = \mathbf{W}_l = I$, where I is a unit matrix. The approaches in [4,6] are similar but by having $\mathbf{W}_h = I$ only. They assume that the relations should be linear. Considering the complicated mapping between α_h and α_l after applying ℓ_1-norm on LR coefficient and ℓ_F-norm on HR coefficient, we prefer a non-linear model instead of a linear one. Our idea is to model the relations as a combination of an encoding process and a decoding process similar to auto-encoder. Specifically, we encode the input (α_l/α_h) non-linearly, and decode the latent representation to the output (α_h/α_l) linearly. Thus, the function (2) becomes:

$$\begin{aligned} &\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 + \|\alpha_h - \mathbf{T}_h^T\mathbf{P}_l\|_F^2 \\ &\quad s.t. \quad \mathbf{P}_h = soft(\mathbf{T}_h\alpha_h), \mathbf{P}_l = soft(\mathbf{T}_l\alpha_l), \end{aligned} \tag{3}$$

where the non-linear mapping is done by a soft-thresholding operator $soft(\cdot)$. In this situation, the auxiliary variables $\mathbf{P}_h$ and $\mathbf{P}_l$ are regarded as the latent representations in the auto-encoder network, which is obtained by projecting the input to the feature space. In fact, our coefficients conversion learning via auto-encoder is a generalization of the conventional linear mapping model.

3.3. Non-local constraint on HR sparse coefficient

When introducing the function (3), we assume the HR patch $\mathbf{X}_h$ has been free of noise and then update its estimation through minimizing the error by ℓ_F-norm. Nevertheless, the iterative optimization does not guarantee this, particularly, at the first few iterations. The encoded noise in α_l still affects α_h and the results can be arbitrarily bad. To make our solution steadily converge at a better result, the HR patch coefficient α_h is also desired to be regularized in both the dictionary learning and the image SR inference.

It has been found that the local image structures tend to repeat within a large region. These similar patches at different locations in the image are regarded as multiple observations of the target patch. Such non-local similarity provides additional information for estimating the target patch and inspired the non-local means (NLM) methods [33,34] that had been applied to image denoising. Thus, a non-local constraint, which correlates the similarities among patches, can be used to regularize the sparse decomposition to prevent the iterative estimates from exhibiting unpleasant noise amplification and artifacts.

In this work, we incorporate the non-local constraint (the correlation of patch similarities) into regularizing the HR sparse coefficients update during the optimization, which is formulated as:

$$\underset{\alpha_h}{\operatorname{argmin}} \sum_{i=1}^{n} \|\alpha_h^i - \sum_{j=1}^{n} \mathbf{N}_{ij}\alpha_h^j\|_F^2 = tr(\alpha_h\mathbf{L}\alpha_h^T), \tag{4}$$

where $\mathbf{N}$ is the weight matrix, $\mathbf{N}_{ij}$ denotes the similarity between patches $\mathbf{x}_i$ and $\mathbf{x}_j$, and $\mathbf{L} = (\mathbf{I}_{n\times n} - \mathbf{N})^T(\mathbf{I}_{n\times n} - \mathbf{N})$. We construct $\mathbf{N}$ through connecting every patch to its k most similar image patches and compute the weights of connected patches by Gaussian kernel function [33]. In addition, we do this on the HR output only because the HR image has more details, the weight matrix constructed from it is more accurate. Finally, the non-local constraint $tr(\alpha_h\mathbf{L}\alpha_h^T)$ is employed to regularize our dictionary learning and the image SR inference to further enhance the SR performance.

3.4. Optimization and inference

By incorporating all the above new constraints into our coupled dictionary learning, the proposed method needs to solve the following optimization problem:

$$\begin{aligned} \underset{\mathbf{D}_l,\mathbf{D}_h,\mathbf{T}_l,\mathbf{T}_h,\alpha_l,\alpha_h}{\operatorname{argmin}} &\frac{1}{2}(\|\mathbf{X}_l - \mathbf{D}_l\alpha_l\|_F^2 + \|\mathbf{X}_h - \mathbf{D}_h\alpha_h\|_F^2) \\ &+ \mu(\|\alpha_l\|_1 + \|\alpha_h\|_1) + \gamma tr(\alpha_h\mathbf{L}\alpha_h^T) \\ &+ \lambda(\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 + \|\alpha_h - \mathbf{T}_h^T\mathbf{P}_l\|_F^2) \\ s.t. \quad &\mathbf{P}_h = soft(\mathbf{T}_h\alpha_h), \mathbf{P}_l = soft(\mathbf{T}_l\alpha_l). \end{aligned} \tag{5}$$

While the objective function (5) is not jointly convex to $\mathbf{D}_l$, $\mathbf{D}_h$, $\mathbf{T}_l$, $\mathbf{T}_h$, α_l and α_h, it is convex with respect to each of them if the remaining variables are fixed. Given the training data $\mathbf{X}_l$, $\mathbf{X}_h$ and $\mathbf{L}$, we apply a coordinate descent algorithm (as shown in Algorithm 1)

Algorithm 1 Dictionaries and transition matrices learning

Input: Training data matrices $\mathbf{X}_l \in \mathbb{R}^{p_1\times q}$ and $\mathbf{X}_h \in \mathbb{R}^{p_2\times q}$, non-local matrix $\mathbf{L} \in \mathbb{R}^{q\times q}$.

1. Initialize $\mathbf{D}_l^0$, $\mathbf{D}_h^0$, α_l^0 and α_h^0 by [1], transition matrices $\mathbf{T}_l^0$ and $\mathbf{T}_h^0$ as $\mathbf{I}$.
2. Let $\mathbf{P}_l^0 \leftarrow soft(\mathbf{T}_l^0\alpha_l^0)$ and $\mathbf{P}_h^0 \leftarrow soft(\mathbf{T}_h^0\alpha_h^0)$, $k = 0$.

while not converged **do**

3. Update $\mathbf{D}_l^{k+1}$ and $\mathbf{D}_h^{k+1}$ by (6) with α_l^k and α_h^k derived from the previous iteration.
4[1]. Update α_l^{k+1} and α_h^{k+1} by (7) with $\mathbf{D}_l^{k+1}$, $\mathbf{D}_h^{k+1}$, $\mathbf{T}_l^k$ and $\mathbf{T}_h^k$.
5. Update $\mathbf{T}_l^{k+1}$ and $\mathbf{T}_h^{k+1}$ by (14) with $\mathbf{D}_l^{k+1}$, $\mathbf{D}_h^{k+1}$, α_l^{k+1} and α_h^{k+1}.
6. $\mathbf{P}_l^{k+1} \leftarrow soft(\mathbf{T}_l^{k+1}\alpha_l^{k+1})$ and $\mathbf{P}_h^{k+1} \leftarrow soft(\mathbf{T}_h^{k+1}\alpha_h^{k+1})$.
7. $k \leftarrow k + 1$.

end while

Output: $\mathbf{D}_l$, $\mathbf{D}_h$, $\mathbf{T}_l$ and $\mathbf{T}_h$.

to optimizing the dictionaries $\{\mathbf{D}_l, \mathbf{D}_h\}$, transition matrices $\{\mathbf{T}_l, \mathbf{T}_h\}$ and coefficients $\{\alpha_l, \alpha_h\}$, respectively. We now discuss how to update these variables in each iteration.

Update $\mathbf{D}_l$, $\mathbf{D}_h$: We first apply the algorithm of joint dictionary learning [1] to the initialization of $\mathbf{D}_l$ and $\mathbf{D}_h$ for the optimization process. When updating the two dictionaries during each iteration, we consider the sparse coefficients $\{\alpha_l, \alpha_h\}$ and transition matrices $\{\mathbf{T}_l, \mathbf{T}_h\}$ as constants. As a result, the problem of (5) can be simplified into the following forms:

$$\begin{aligned}&\underset{\mathbf{D}_l^{k+1}}{\operatorname{argmin}} \|\mathbf{X}_l - \mathbf{D}_l^{k+1}\alpha_l^k\|_F^2 + \eta\|\mathbf{D}_l^{k+1} - \mathbf{D}_l^k\|_F^2,\\ &\underset{\mathbf{D}_h^{k+1}}{\operatorname{argmin}} \|\mathbf{X}_h - \mathbf{D}_h^{k+1}\alpha_h^k\|_F^2 + \eta\|\mathbf{D}_h^{k+1} - \mathbf{D}_h^k\|_F^2.\end{aligned} \tag{6}$$

After updated by an iterative scheme (e.g. the Conjugate Gradient method), the dictionaries $\mathbf{D}_l$ and $\mathbf{D}_h$ are column normalized to avoid any trivial solutions.

Update α_l, α_h: Similar to dictionary updates, the transition matrices $\{\mathbf{T}_l, \mathbf{T}_h\}$ and dictionaries $\{\mathbf{D}_l, \mathbf{D}_h\}$ are fixed when we calculate the solutions of sparse coefficients $\{\alpha_l, \alpha_h\}$. Thus, the objective functions are written as follows:

$$\begin{aligned}&\underset{\alpha_l}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{X}_l - \mathbf{D}_l\alpha_l\|_F^2 + \lambda\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 + \mu\|\alpha_l\|_1,\\ &\underset{\alpha_h}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{X}_h - \mathbf{D}_h\alpha_h\|_F^2 + \lambda\|\alpha_h - \mathbf{T}_h^T\mathbf{P}_l\|_F^2\\ &\quad + \mu\|\alpha_h\|_1 + \gamma tr(\alpha_h\mathbf{L}\alpha_h^T).\end{aligned} \tag{7}$$

Besides the standard sparse-coding formulation, here is an additional coefficient sparsity constraint term $\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1$ with respect to α_l. So the commonly used minimization algorithms are no longer applicable. We propose a smooth proximal gradient method motivated by [35,36]. Specifically speaking, the non-smooth ℓ_1-norm coefficient sparsity constraint term is transformed into its dual domain, where it can be approximated by a smooth ℓ_∞ term using Nestrov's algorithm. Since the output is convex and differentiable with respect to α_l with a sparsity constraint, we are able to solve it efficiently using a proximal method, e.g. Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [15].

We first rewrite the sparsity constraint term $\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1$ using the dual form of ℓ_1-norm as

$$\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 = \underset{\|\delta\|_\infty \le 1}{\operatorname{argmax}} \delta^T(\alpha_l - \mathbf{T}_l^T\mathbf{P}_h), \tag{8}$$

where δ is the dual variable. Then, we approximate above solution using Nestrov's algorithm as

$$\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 \approx \underset{\|\delta\|_\infty \le 1}{\operatorname{argmax}} \delta^T(\alpha_l - \mathbf{T}_l^T\mathbf{P}_h) - \rho d(\delta), \tag{9}$$

where $d(\cdot) = \frac{1}{2}\|\delta\|_2^2$ is a smoothing operator and ρ is a smoothness parameter. We can see that the approximate function is convex and differentiable. Thus, the approximate gradient $\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1$ over α_l is:

$$\nabla_{\alpha_l}\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1 \approx S\left(\frac{\alpha_l - \mathbf{T}_l^T\mathbf{P}_h}{\rho}\right), \tag{10}$$

where $S(\mathbf{x})$ is a function projecting $\mathbf{x}$ onto an ℓ_∞-ball,

$$S(\mathbf{x}) = \begin{cases}\mathbf{x}, & -1 \le \mathbf{x} \le 1\\ 1, & \mathbf{x} > 1\\ -1, & \mathbf{x} < -1\end{cases}. \tag{11}$$

Finally, the gradient with respect to α_l is written as:

$$\nabla_{\alpha_l} h_1(\alpha_l) \approx \mathbf{D}_l^T(\mathbf{D}_l\alpha_l - \mathbf{X}_l) + \lambda S\left(\frac{\alpha_l - \mathbf{T}_l^T\mathbf{P}_h}{\rho}\right), \tag{12}$$

where $h_1(\alpha_l) = \frac{1}{2}\|\mathbf{X}_l - \mathbf{D}_l\alpha_l\|_F^2 + \lambda\|\alpha_l - \mathbf{T}_l^T\mathbf{P}_h\|_1$.

For the optimization of α_h, we also use the proximal methods like FISTA. Its gradient is given as follows:

$$\begin{aligned}\nabla_{\alpha_h} h_2(\alpha_h) \approx{}& \mathbf{D}_h^T(\mathbf{D}_h\alpha_h - \mathbf{X}_h) + 2\lambda(\alpha_h - \mathbf{T}_h^T\mathbf{P}_l)\\ &+ \gamma\alpha_h(\mathbf{L} + \mathbf{L}^T),\end{aligned} \tag{13}$$

where $h_2(\alpha_h) = \frac{1}{2}\|\mathbf{X}_h - \mathbf{D}_h\alpha_h\|_F^2 + \lambda\|\alpha_h - \mathbf{T}_h^T\mathbf{P}_l\|_F^2 + \gamma tr(\alpha_h\mathbf{L}\alpha_h^T)$.

Update $\mathbf{T}_l$, $\mathbf{T}_h$: When updating the transition matrices, only the terms associating with $\{\mathbf{T}_l, \mathbf{T}_h\}$ are considered in the optimization. With fixed $\{\mathbf{D}_l, \mathbf{D}_h\}$ and $\{\alpha_l, \alpha_h\}$, we solve the following problem for update,

$$\begin{aligned}&\underset{\mathbf{T}_l^{k+1}}{\operatorname{argmin}} \|\alpha_l - {\mathbf{T}_l^{k+1}}^T\mathbf{P}_h\|_1 + \eta\|\mathbf{T}_l^{k+1} - \mathbf{T}_l^k\|_F^2,\\ &\underset{\mathbf{T}_h^{k+1}}{\operatorname{argmin}} \|\alpha_h - {\mathbf{T}_h^{k+1}}^T\mathbf{P}_l\|_F^2 + \eta\|\mathbf{T}_h^{k+1} - \mathbf{T}_h^k\|_F^2.\end{aligned} \tag{14}$$

This can be done by the Conjugate Gradient method as well.

Inference: Once the optimization is complete, the derived model is employed for image SR reconstruction. We first partition the LR image into overlapped patches, and initialize the sparse coefficients α_l of $\mathbf{X}_l$ via solving:

$$\underset{\alpha_l^0}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{X}_l - \mathbf{D}_l\alpha_l^0\|_F^2 + \mu\|\alpha_l^0\|_1. \tag{15}$$

Once α_l is calculated, we initialize α_h in the derived feature space:

$$\alpha_h^0 = \mathbf{T}_h^T soft(\mathbf{T}_l\alpha_l^0). \tag{16}$$

After iteratively updating α_h through the objective function (5), we have $\mathbf{X}_h = \mathbf{D}_h\alpha_h$ as the final SR output. The procedure is given in Algorithm 2.

Algorithm 2 Image SR inference

Input: Dictionaries $\mathbf{D}_l$ and $\mathbf{D}_h$, transition matrices $\mathbf{T}_l$ and $\mathbf{T}_h$, a low-resolution image $\mathbf{Y}$.

1. Extract LR patches $\mathbf{X}_l \in \mathbb{R}^{p_1 \times M}$ from LR image $\mathbf{Y}$ with 1˜pixel overlap in each direction.
2. Initiate α_l^0 by (15) and α_h^0 by (16).
3. Let $\mathbf{P}_l^0 \leftarrow soft(\mathbf{T}_l^0\alpha_l^0)$ and $\mathbf{P}_h^0 \leftarrow soft(\mathbf{T}_h^0\alpha_h^0)$.
4. Initialize: $\mathbf{X}_h^0 \leftarrow \mathbf{D}_h\alpha_h^0$, $\mathbf{L}^0$ computed from $\mathbf{X}_h^0$, $k = 0$.

while not converged **do**

5[1]. Update α_l^{k+1} and α_h^{k+1} by (7) with $\mathbf{X}_h^k$, $\mathbf{L}^k$, $\mathbf{P}_l^k$ and $\mathbf{P}_h^k$.

6. Update $\mathbf{P}_l^{k+1} \leftarrow soft(\mathbf{T}_l^k\alpha_l^{k+1})$, $\mathbf{P}_h^{k+1} \leftarrow soft(\mathbf{T}_h^k\alpha_h^{k+1})$, $\mathbf{X}_h^{k+1} \leftarrow \mathbf{D}_h\alpha_h^{k+1}$ and $\mathbf{L}^{k+1}$ computed from $\mathbf{X}_h^{k+1}$.
7. $k \leftarrow k + 1$.

end while

8. Put the patches $\mathbf{X}_h$ into a high-resolution image $\mathbf{X}$.

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

4. Experiments and discussion

To validate the robustness of our method to noise and the effectiveness on detail recovery, the performance of our method is evaluated on various types of image SR tasks in Section 4.2, which include synthesized noisy images, real noisy images, denoised images and noiseless images. We also compare our approach with CSR [26], GSR [27] and SRNI [22], which are specifically designed for noisy SR application. We select the test images set including standard databases Set5, Set14, BD100 from [7], BD50 from [22] and newly created one as shown in Fig. 3. The quantitatively validations are calculated in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) from their HR outputs. To show the sparsity efficiency of the ℓ_1-norm coefficients transition term, we also provide some visualization results in Section 4.3. Furthermore, we conduct some discussions on the computational complexity in Section 4.4 and parameters determination in Section 4.5.

[1] The matrix $\mathbf{L}$ (non-local constraint) is applied to regularize α_h as the function (4).

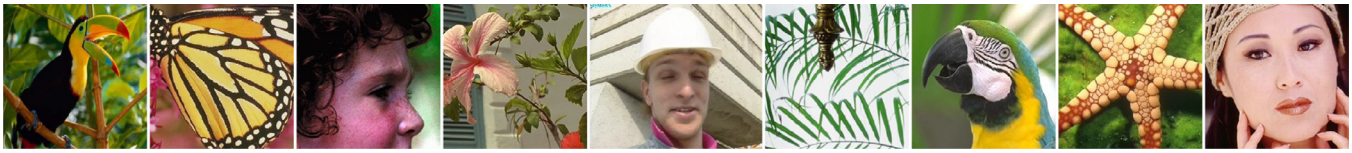

Fig. 3. The test images.

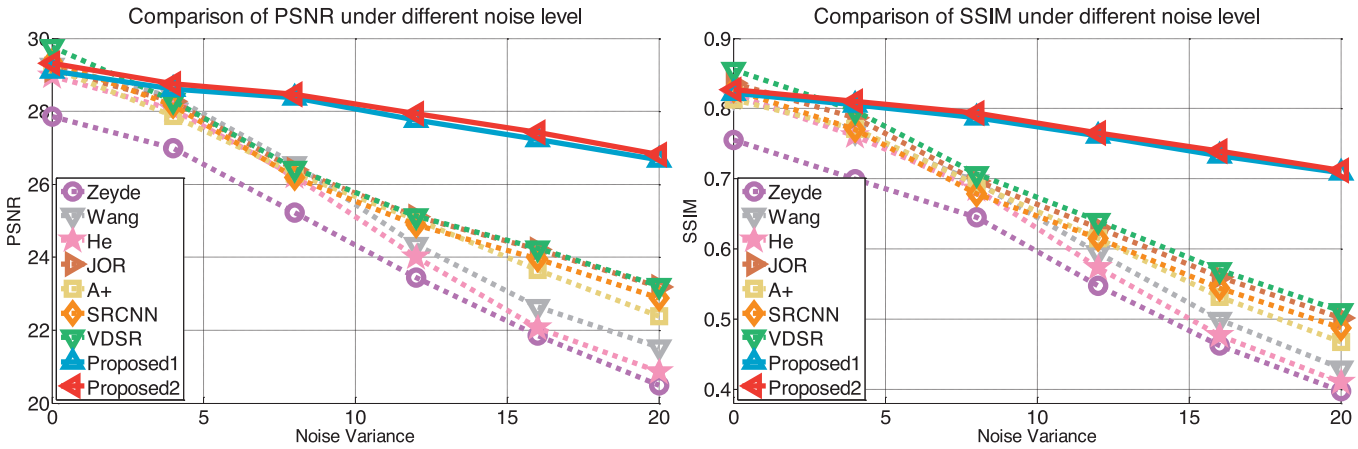

Fig. 4. PSNR and SSIM of seven competing methods and our methods when noise level varies from 0 up to 20.

4.1. Experimental settings

To have synthesized noisy/noiseless LR images, we blur and down-sample HR images using Matlab function "imresize()" with a scaling factor, and then add Gaussian white noise with varied standard deviations (noiseless image with a deviation 0). Considering that human eyes are more sensitive to the luminance channel, the proposed method is only performed on Y channel in YCbCr color space. The color is directly magnified to the desired size later by Bicubic interpolation. For the training dataset, 500,000 LR-HR image patch pairs are randomly extracted from database [1], where the LR patches are interpolated up to the same size as the HR training data. The HR patch size is fixed as 9×9 with one pixel overlap between adjacent patches. For the parameters in objective function (5), $\{\lambda, \mu, \gamma\}$ is set to $\{0.5, 0.1, 0.05\}$. The dictionary size is defined as 1024. Moreover, our approach finds $k = 10$ most similar patches to calculate the weight matrix $\mathbf{N}$ for the non-local constraint. The empirical studies of the parameters in the proposed algorithm will be discussed in the following subsection.

4.2. Experimental results

In this Section 4.2.1–4.2.4, several state-of-the-art methods are used as comparison baselines, which can be divided into two classes: (i) The conventional coupled dictionary learning approaches, such as Zeyde's approach [3] based on the sparse representation invariance, Wang's approach [4] based on a semi-coupled dictionary, He's approach [6] based on the Beta process, Dai's approach (JOR) [16] based on jointly optimized regressors and A^+ [8] based on an anchored neighborhood regression. (ii) Deep convolutional networks based methods, such as Super-resolution Convolutional Neural Network (SRCNN) [17,21] and Very Deep Convolutional Networks for Image Super-Resolution (VDSR) [20]. To have a better insight into our proposed method, we deliberately create two versions: **proposed1**, a partial version of only optimization on the ℓ_1-norm coefficients transition term to suppress noise; **proposed2**, a full version including both the ℓ_1-norm and the non-local constraint for improving SR quality. Moreover, effective noisy image SR methods: CSR [26] which uses the centralized sparse representation to regularize the ill-posed SR problem, GSR [37] which establishes a framework for image SR using group sparsity and proposes a robust dictionary learning method to suppress the noise, SRNI [22] which integrates the merits of image denoising and image SR by a convex combination are compared to demonstrate the advantage of our proposed method.

4.2.1. Results on synthetic noisy images

In this subsection, we use the test images shown in Fig. 3 to conduct experiments to explore how noise affects the performances of our method and the other state-of-the-art SR methods. The quantitative comparisons in both PSNR and SSIM are shown in Fig. 4 when varying the standard deviation of Gaussian noise from 0 up to 20 with step size 4. From Fig. 4 one can see other seven approaches degrade much quicker when noise level increases. In contrast, proposed1 and proposed2 perform well across all noise level, even when the input noise level is high. We further provide the visual comparisons on noisy image *parrots* with noise level 20 and image *butterfly* with noise level 4 shown in Figs. 5 and 6, respectively. In Fig. 5, one can see that Zeyde's method, A^+, SRCNN and VDSR generate plausible details and sharper edges along *parrots* beak, but are poor at suppressing the noise, particularly on the smooth regions. Wang's approach and He's approach generate over-smoothed results and meanwhile fail to suppress the noise effect. JOR produces relatively pleasing result by incorporating multiple mapping functions, and outperforms the similar approach, e.g. Wang's approach and He's approach. Our proposed1 (Fig. 5.i) produces visually comparable result that has little remaining noise. Moreover, the proposed2 with the non-local constraint (Fig. 5.j) further improves SR performance by finely synthesizing the high-frequency details and eliminating artifacts. Fig. 5 shows all other competing approaches produce unacceptable results, where the noise is magnified rather than suppressed. The reason is that the noise in LR input is first upscaled by interpolation, then encoded into HR result through the ℓ_F-norm coefficients transition term in the conventional coupled dictionary learning approaches and through the convolution operation in the deep convolutional networks based methods. Wang's approach generates a relatively finer result due to a non-local post-processing. However,

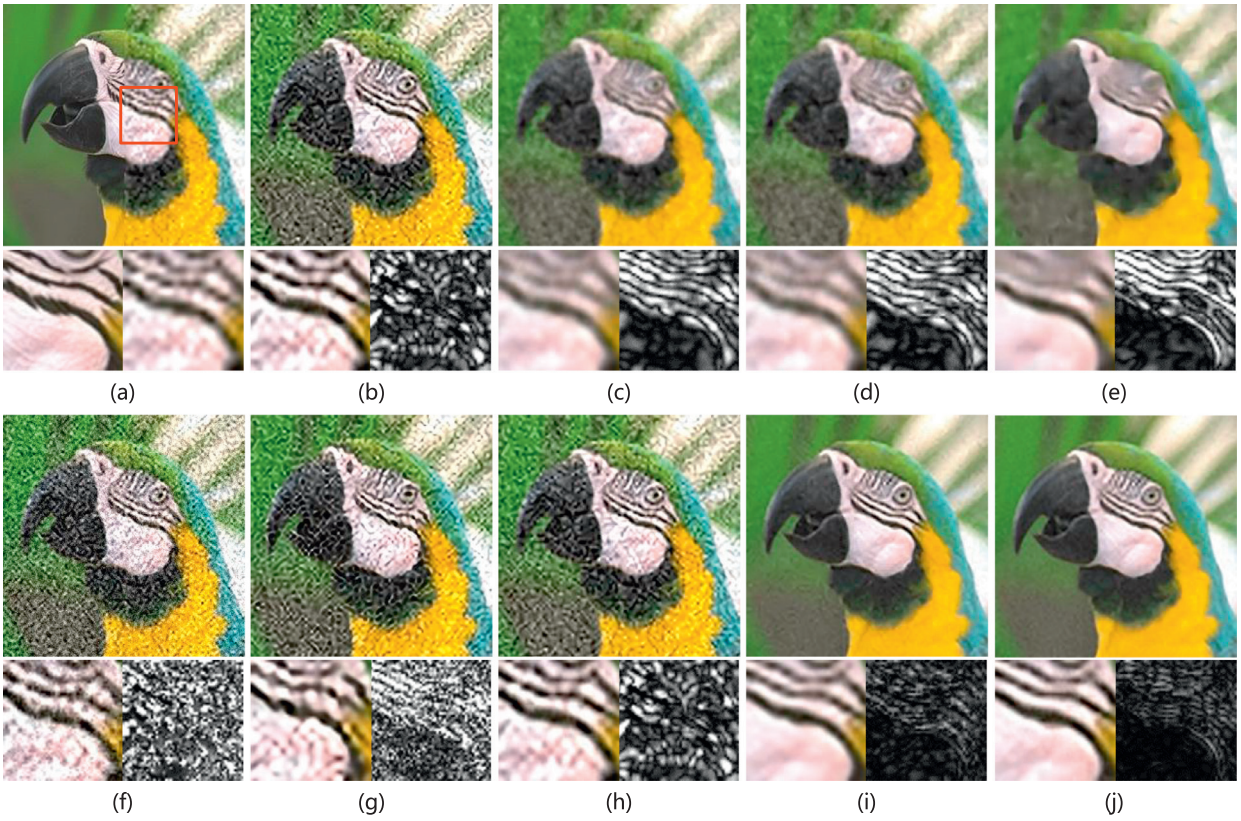

Fig. 5. Results of synthetic noisy image *parrots* (× 3) with a noise level of 20. **Top:** Visual SR results. **Bottom:** *left.* Magnified local SR patches. *right.* Noise residual maps visualizing the difference between SR patches of the noiseless LR inputs and their noisy ones, where the residual signal (white dots) is magnified by 10 times. (a) Ground-truth (left) and its interpolated LR version (right); (b) Zeyde's approach [3]; (c) Wang's approach [4]; (d) He's approach [6]; (e) JOR [16]; (f) A^+ [8]; (g) SRCNN [17,21]; (h) VDSR [20]; (i) Proposed1; (j) Proposed2.

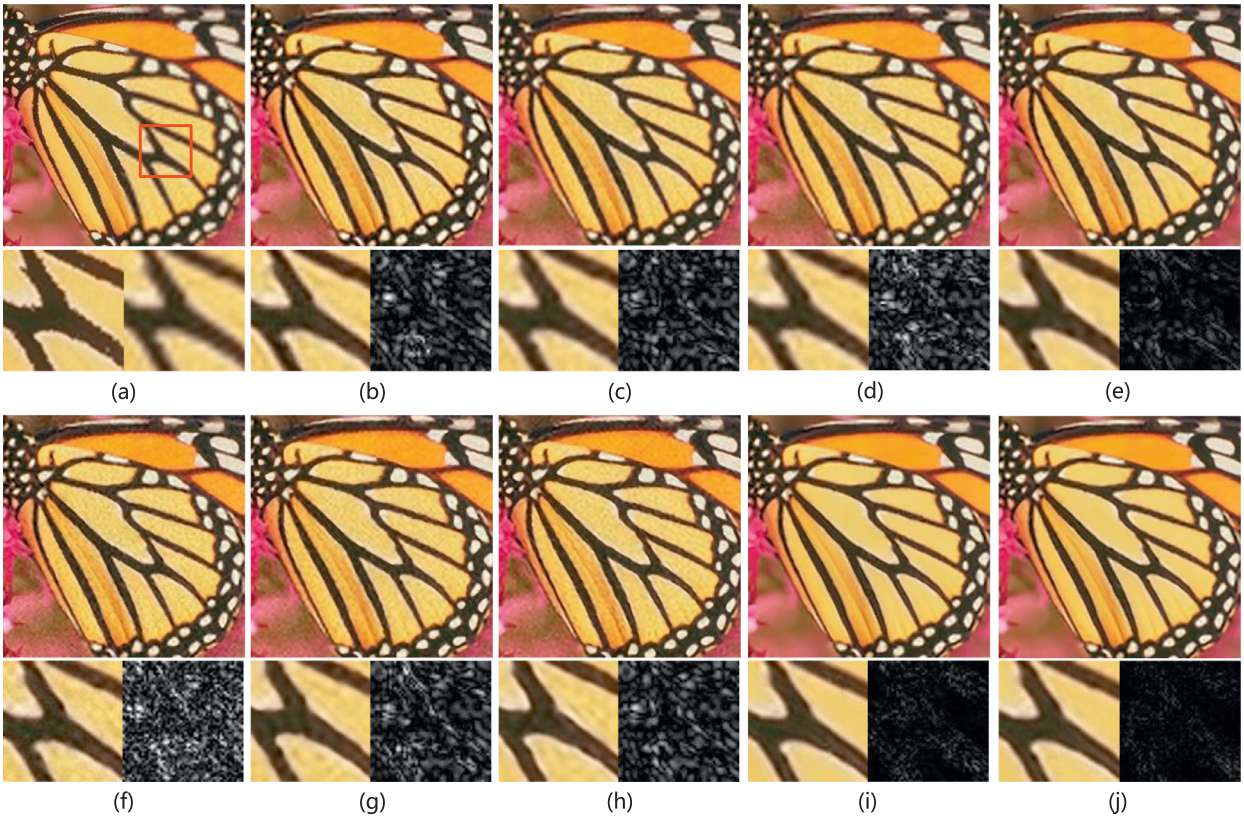

Fig. 6. Results of synthetic noisy image *butterfly* (× 3) with a noise level of 4. Please refer to Fig. 5 for the description of subfigures.

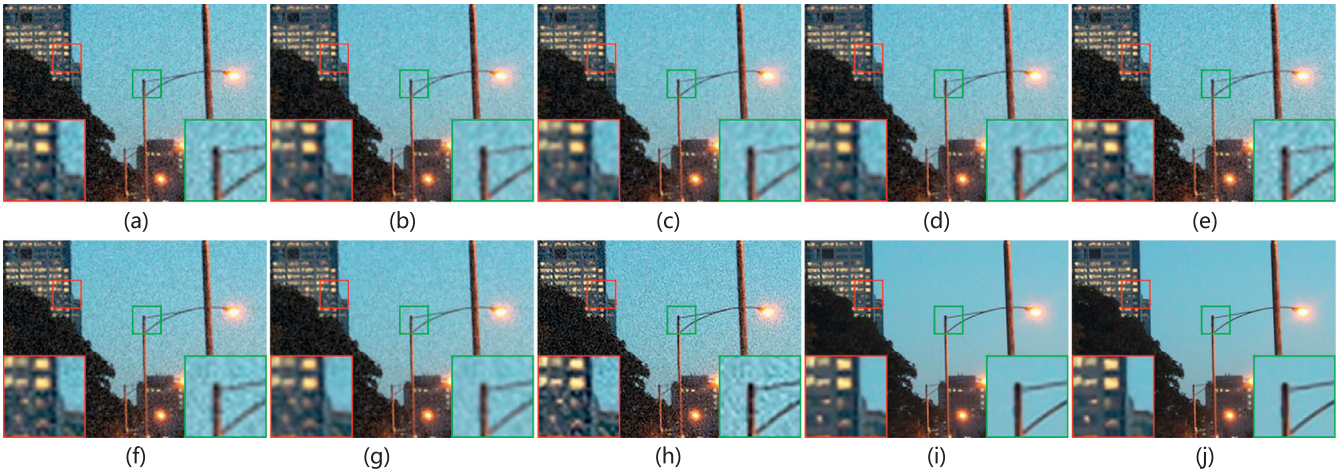

Fig. 7. Results of the real LR noise image (× 3). (a) Bicubic interpolation; (b) Zeyde's approach [3]; (c) Wang's approach [4]; (d) He's approach [6]; (e) JOR [16]; (f) A+ [8]; (g) SRCNN [17,21]; (h) VDSR [20]; (i) Proposed1; (j) Proposed2.

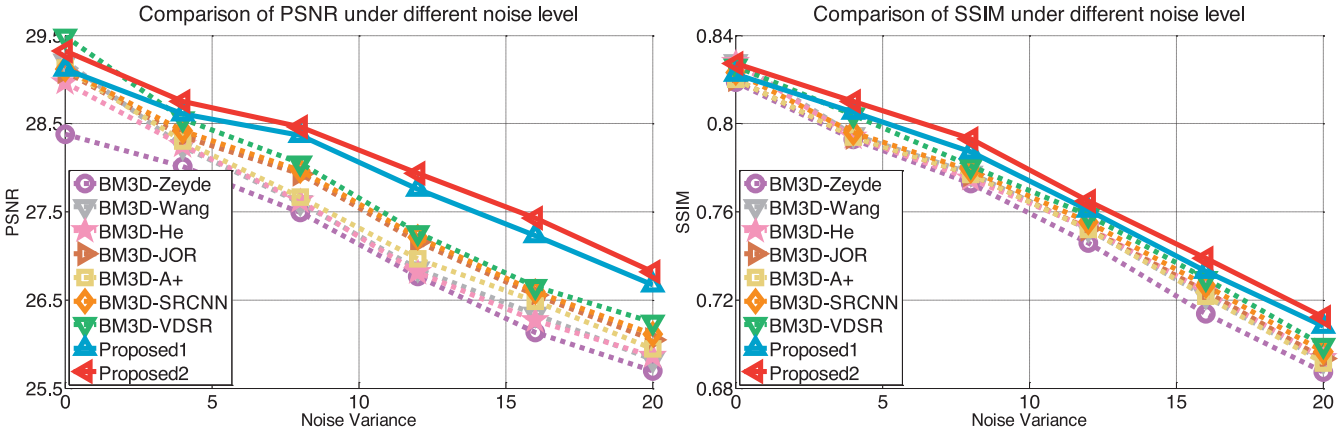

Fig. 8. PSNR and SSIM of seven competing methods and our methods, where the inputs of other seven have been denoised, and noise level varies from 0 up to 20.

it does the sparse-coding and non-local constraint regularization separately, which restricts the noisy SR performance. JOR produces an over-smoothed HR result because its class-label selection becomes ambiguous when the noise is severe. In contrast our methods (Fig. 5.i and j) successfully suppress the noise and produce visually pleasing SR results. The Fig. 6 further validates our summary of above results. More visual comparison can be found in supplementary document, Fig. C_1–C_8.

4.2.2. Results on real noisy examples

To validate the effectiveness of our proposed algorithm, we test its performance on real-world noisy images and the results are shown in Fig. 7. The results produced by the competing methods are sensitive to noise. For example, obvious noise can be observed in the sky area. Meanwhile, we can see that the competing methods (except VDSR) smooth much the edges of the buildings. Though VDSR generates sharp edges and fine details, it looks somewhat unnatural. Overall, our approach produces visually more satisfying results, involving sharp edges, noise suppression. More visual comparison can be found in supplementary document, Fig. D_1 and D_2.

4.2.3. Results on denoised images

In this subsection, we further perform a comparison on denoised synthetic noisy images (same as those used in the Section 4.2.1) and denoised real noisy data. The noisy images are first denoised by BM3D algorithm [38], and then input into seven competing methods. The averaged quantitative performance of denoised synthetic noisy images is summarized in Fig. 8. From this figure, we can see that our proposed method, without resorting to additional denoising, still achieves an obvious improvement than other state-of-the-art approaches. Apart from quantitative results, we also show visual comparisons of the denoised-SR results shown in Fig. 9 (denoised synthetic noisy images) and Fig. 10 (denoised real noisy images). In Fig. 9, the original noise level is 12. Thanks to denoising, the competing methods all produce rather clear result with little noise in Fig. 9 but tend to smooth high frequency details as shown along the brim of hat and destroy the image structure as shown cheek region within the face. This is in large part due to the missing or smoothing out edges in the denoising procedure, which restricts the subsequent SR step performance. By contrast, our method produces much sharper edges and more faithful details because of the inherent ℓ_1-norm who preserves sparsity of input signal. From Fig. 10, one can see that the visual comparison is analogous to that described in the results of the denoised synthetic noisy images, with an exception that VDSR generates over-sharp edges, which is unnatural to the observer. More visual comparison can be found in supplementary document, Fig. E_1–E_4.

4.2.4. Results on noiseless images

To verify the effectiveness of our proposed method to noiseless images, we also compare it against the aforementioned seven approaches, and summarize the average PSNR and SSIM scores on three test datasets with different magnification factors in

Fig. 9. Results of denoised synthetic noisy image *foreman* (× 3) with a noise level of 12. **Top:** Visual SR results. **Bottom:** *left.* Magnified local denoised SR patches. *right.* Magnified local noiseless SR patches. (a) Ground-truth (left) and BM3D-Bicubic (right); (b) BM3D-Zeyde's approach [3]; (c) BM3D-Wang's approach [4]; (d) BM3D-He's approach [6]; (e) BM3D-JOR [16]; (f) BM3D-A$^+$ [8]; (g) BM3D-SRCNN [17,21]; (h) BM3D-VDSR [20]; (i) Proposed1; (j) Proposed2.

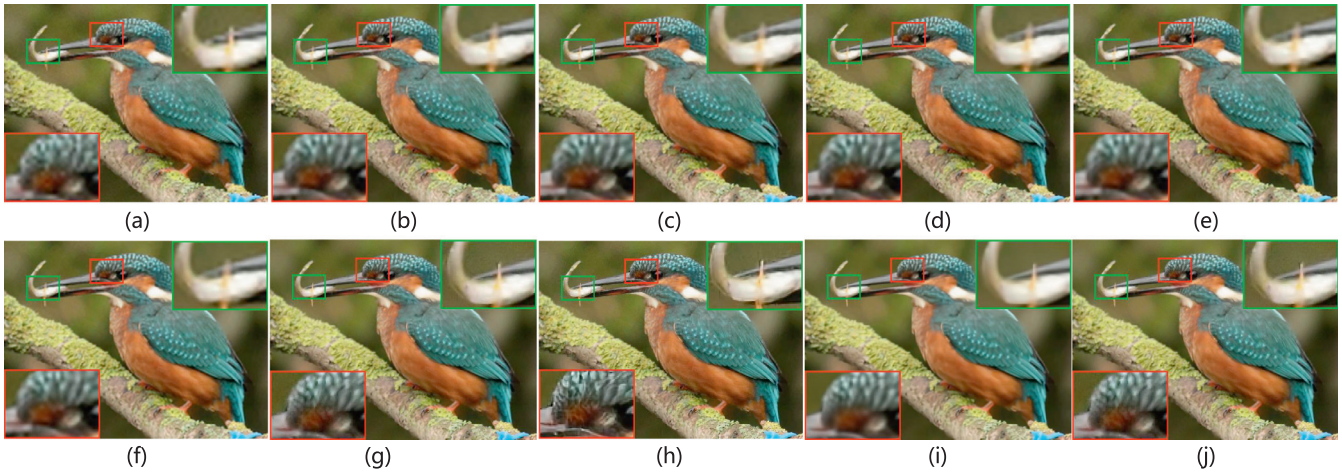

Fig. 10. Results of the denoised real LR noise image (× 3). (a) BM3D-Bicubic; b) BM3D-Zeyde's approach [3]; (c) BM3D-Wang's approach [4]; (d) BM3D-He's approach [6]; (e) BM3D-JOR [16]; (f) BM3D-A$^+$ [8]; (g) BM3D-SRCNN [17,21]; (h) BM3D-VDSR [20]; (i) Proposed1; (j) Proposed2.

Table 1
Average PSNR, SSIM and running time (× 3 magnification) on test data set Set5, Set14, BD100.

Benchmark		Zeyde [3]	Wang [4]	He [6]	JOR [16]	A$^+$ [8]	SRCNN [17,21]	VDSR [20]	Proposed1 (ours)	Proposed2 (ours)
Set5	PSNR	31.90	32.46	32.38	32.55	32.59	32.75	**33.66**	32.65	32.94
	SSIM	0.9088	0.9045	0.9037	0.9064	0.9088	0.9090	**0.9213**	0.9083	0.9105
	Time (s)	34.6	715.7	414.7	35.3	8.2	13.5	**6.6**	312.4	563.2
Set14	PSNR	28.67	29.26	28.97	29.09	29.13	29.30	**29.77**	29.15	29.44
	SSIM	0.8188	0.8204	0.8175	0.8194	0.8188	0.8215	**0.8314**	0.8178	0.8223
	Time (s)	71.3	1384.3	803.4	70.6	15.3	25.7	**12.4**	604.5	1124.4
BD100	PSNR	27.87	28.32	28.15	28.17	28.29	28.41	**28.82**	28.24	28.49
	SSIM	0.7808	0.7961	0.7953	0.7951	0.7835	0.7863	**0.7976**	0.7954	0.7975
	Time (s)	60.2	1044.3	626.4	56.3	12.4	17.7	**8.8**	454.2	782.4

Table 2
Average PSNR, SSIM and running time (× 4 magnification) on test data set Set5, Set14, BD100.

Benchmark		Zeyde [3]	Wang [4]	He [6]	JOR [16]	A^+ [8]	SRCNN [17,21]	VDSR [20]	Proposed1 (ours)	Proposed2 (ours)
Set5	PSNR	29.69	30.27	30.16	30.19	30.29	30.49	**31.35**	30.35	30.63
	SSIM	0.8603	0.8589	0.8612	0.8604	0.8603	0.8628	**0.8838**	0.8615	0.8645
	Time (s)	32.8	705.6	395.4	31.1	7.5	11.2	**6.1**	293.5	538.6
Set14	PSNR	26.88	27.26	27.13	27.26	27.33	27.50	**28.01**	27.41	27.73
	SSIM	0.7491	0.7485	0.7512	0.7506	0.7491	0.7513	**0.7674**	0.7504	0.7542
	Time (s)	64.5	1335.6	784.7	65.6	13.3	21.4	**11.3**	584.1	1108.8
BD100	PSNR	26.51	26.64	26.51	26.74	26.82	26.90	**27.29**	26.85	27.01
	SSIM	0.7085	0.7043	0.7035	0.7065	0.7087	0.7101	**0.7251**	0.7095	0.7143
	Time (s)	56.2	1024.9	602.4	48.3	10.5	12.4	**7.4**	432.6	758.5

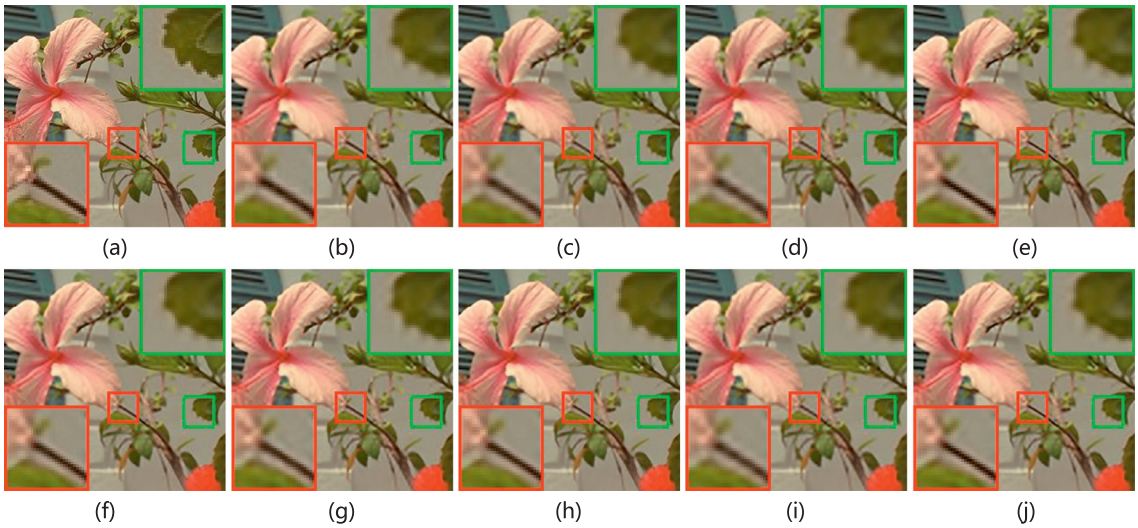

Fig. 11. Results of the noiseless LR image *flower* (× 3). (a) Ground-truth; (b) Zeyde's approach [3]; (c) Wang's approach [4]; (d) He's approach [6]; (e) JOR [16]; (f) A^+ [8]; (g) SRCNN [17,21]; (h) VDSR [20]; (i) Proposed1; (j) Proposed2.

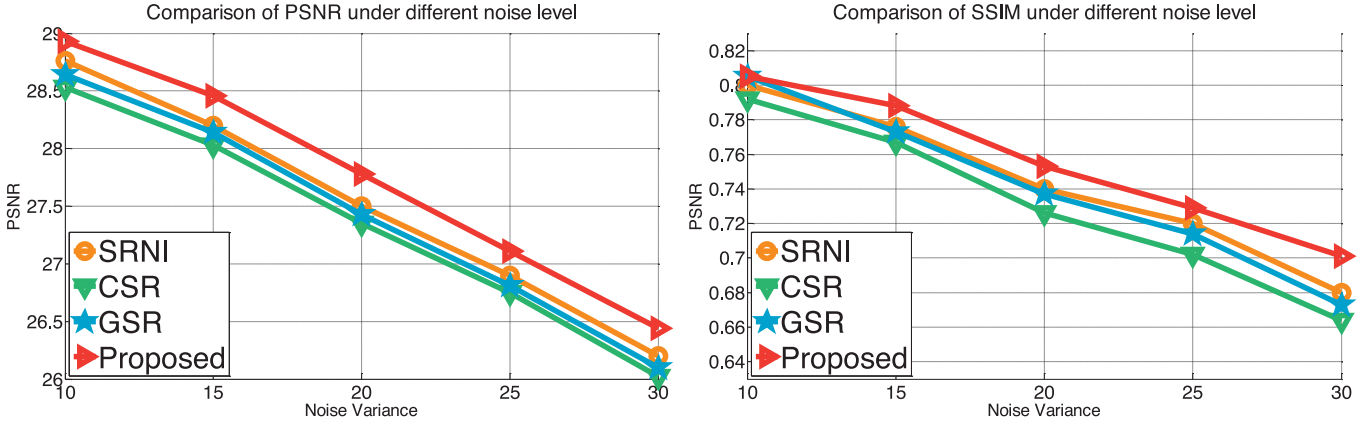

Fig. 12. PSNR and SSIM of SRNI [22], CSR [26], GSR [37] and our proposed method on BD50 under varied noise levels (from 10 up to 30).

Tables 1 and 2. One can see our methods are still able to outperform all other coupled dictionary based approaches across all datasets. Compared with deep learning based method SRCNN, the proposed1 is slightly inferior to the SRCNN. By applying the non-local similarity constraint, our best method (proposed2) achieves comparable performance to SRCNN (outperforms it over 0.2dB in PSNR, 0.0020 in SSIM for × 3 magnification and 0.16dB in PSNR, 0.0030 in SSIM for × 4 magnification), but is slightly inferior to VDSR by an average loss of 0.46dB in PSNR, 0.0067 in SSIM for × 3 magnification and 0.42dB in PSNR, 0.0144 in SSIM for × 4 magnification. Fig. 11 shows a visual comparison. One can see that Zeyde's approach [3] produces blurred textural details and zigzag edges because of their simple assumption of the invariant sparse representation. He's approach [6] is very competitive in terms of visual quality compared to Zeyde's by relaxing the representation relationship assumption to a beta process prior for the coupled dictionary learning. Wang's approach [4] is a large improvement by introducing a more complicated assumption that there is a linear mapping between $\boldsymbol{\alpha}_l$ and $\boldsymbol{\alpha}_h$. Although it produces a better HR image with many fine details, some unpleasant artifacts still can be found along major edges. JOR [16] improves the SR performance significantly compared to Zeyde's approach by involving multiple regressors. Using a similar framework (sparse representation invariance assumption) of Zeyde's, A^+ [8] combines the

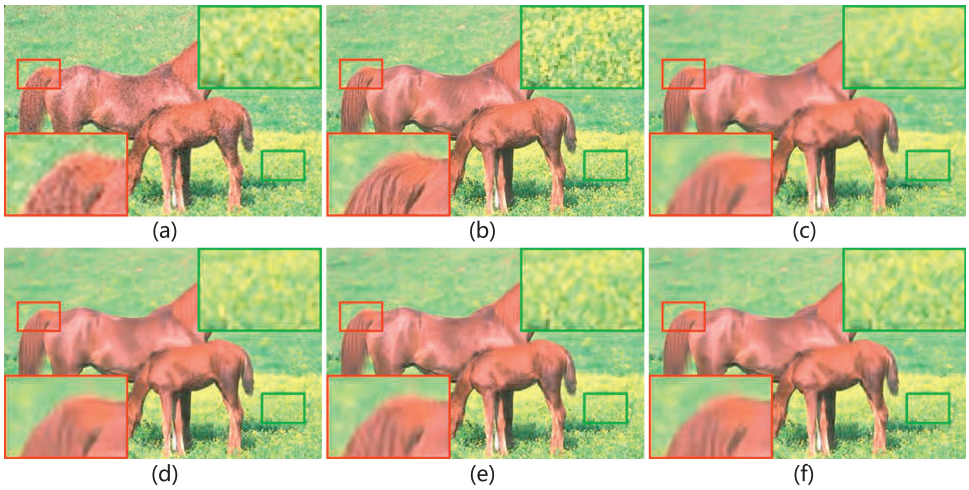

Fig. 13. Results of the noisy LR image *horse* (× 2) with a noise level of 20. (a) Bicubic interpolation. (b) Ground-truth. (c) SRNI [22]. (d) CSR [26]. (e) GSR [37]. (f) Proposed2.

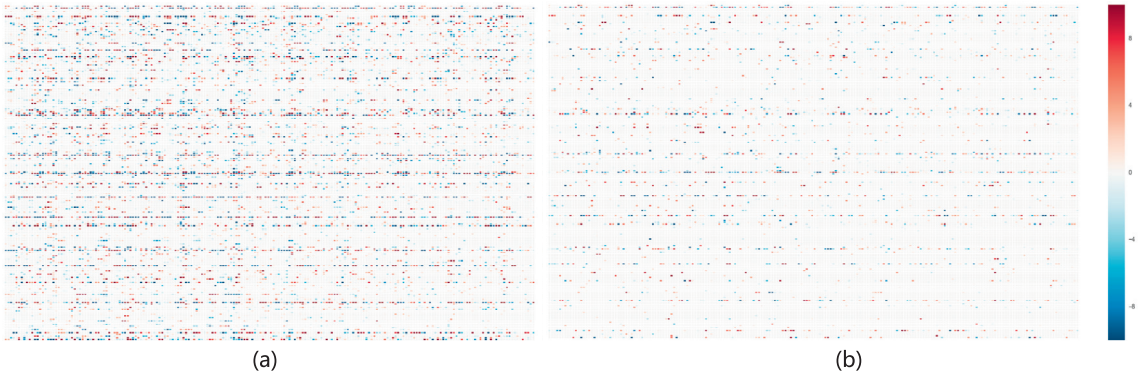

Fig. 14. Visualization of the sparse representation of LR image patches. For clarity, we only show 200× 200 matrix, which is the subset of size 1024 × M of the representation coefficients. Each column is the sparse coefficient of an image patch. Red indicates positive values, blue indicates negative values and white denotes exact zero. (a) with ℓ_F-norm coefficients transition term constraint; (b) with ℓ_1-norm coefficients transition term constraint. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

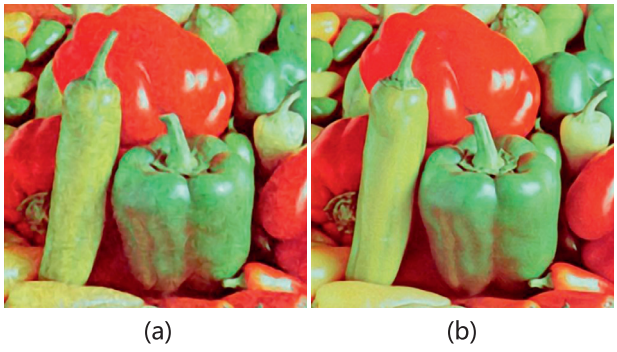

Fig. 15. Results of our proposed method with different transition term constraints (noise variance 12). (a) with ℓ_F-norm coefficients transition term constraint; (b) with ℓ_1-norm coefficients transition term constraint.

learned sparse dictionaries and neighbor embedding to improve the SR performance. However, noticeable zigzags are created along dominant edges, and some fine image structures are missing. The deep learning based methods SRCNN and VDSR generate much sharper edges and suppress noticeable artifacts in the SR result. This is mainly due to a more accurate mapping function learned by deep nets. Our proposed1 produces a slightly worse reconstruction quality with a few artifacts along the edges shown in Fig. 11.i. But, our proposed2 recovers a HR image with richer details shown in Fig. 11.j, superior to all other SR approaches including VDSR. It suggests that the non-local similarity constraint effectively improves the stability of sparse decomposition, which benefits the SR performance. More visual comparison can be found in supplementary document, Fig. F_1–F_3.

4.2.5. Comparison with specialized noisy SR methods

To further illustrate the advantage of our proposed method on the noisy images, we compare to the state-of-the-art noisy SR methods: CSR [26], GSR [37] and SRNI [22]. For a fair comparison, we quantitatively evaluate our performance on dataset BD50 in Fig. 12, which is used in the work of SRNI. From it, one can see that our method is consistently better than other methods at all the levels of noise. We also give the visual comparison in Fig. 13. As can be seen in Fig. 13, SRNI tends to remove textures besides noise and exhibits some high-frequency artifacts. CSR can suppress the noise, but some artifacts still can be found along edges. GSR achieves very similar performances as CSR. In contrast, our proposed method can successfully remove noise from textures and generate visually pleasing high-frequency details. More visual comparison can be found in supplementary document, Fig. G_1–G_4.

4.3. Visualization of the sparse representation

To order to explicitly make a comparison between ℓ_F-norm coefficients transition term and ℓ_1-norm one, we visualize the learned sparse representation of LR image patches with different

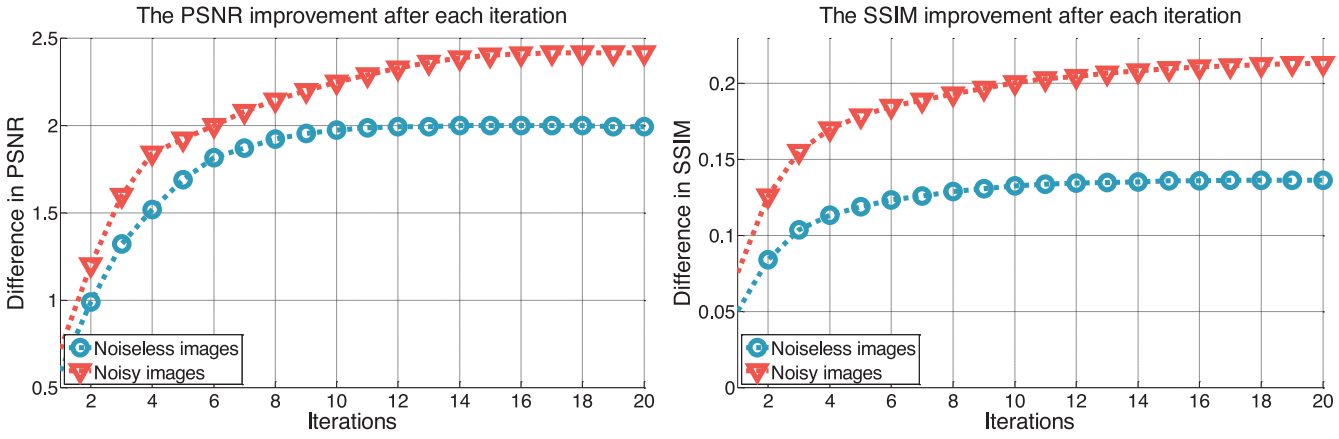

Fig. 16. The improvement in the SR results after each iteration.

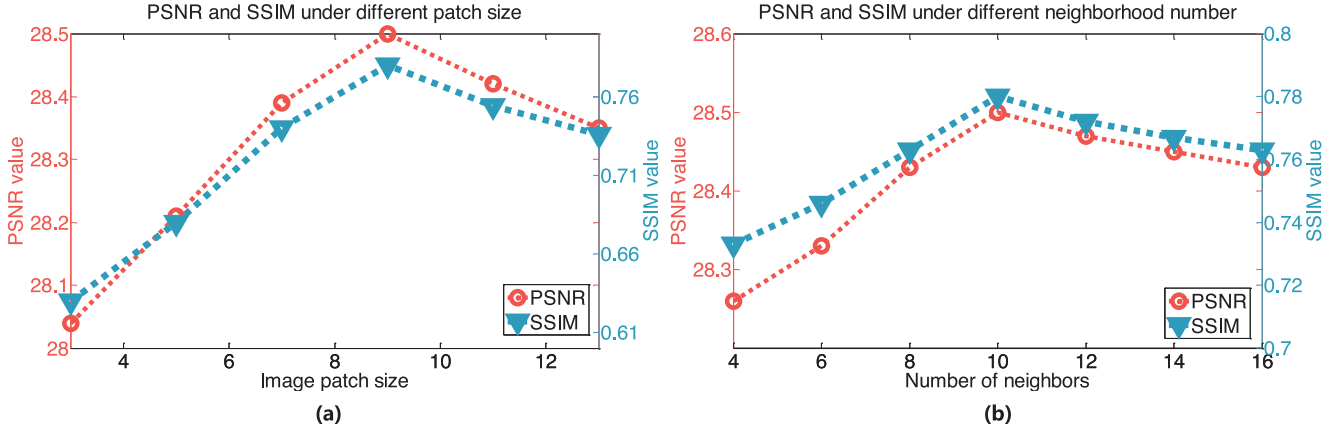

Fig. 17. Performance results with different parameters setting. (a) with different patch sizes; (b) with different neighbor numbers.

constraints in Fig. 14. In the heatmap, these LR coefficient matrices are visualized by the coefficients value, where the coefficient of an image patch is arranged as a column vector. The nonzero elements are denoted as colored points and white denote zero ones. It is obvious that our method with ℓ_1-norm yields sparser codes, which fully demonstrates the superiority of ℓ_1-norm over ℓ_F-norm, and validates the effectiveness of our proposed method. Also, Fig. 15 illustrates the SR output of our method with different transition term constraints.

4.4. Algorithm complexity and computational time

In this subsection we provide the complexity of our proposed method. The detailed inference algorithm is described in Algorithm 2. We can see that the main computations are in the similar patches searching and the gradient computing in every iteration. Assume that the number of similar patches is S_k, the dictionary size is D_s, the number of outer loop iterations is I_1 and the number of inter loop iterations is I_2. Thus, the total complexity of our method is $O(I_1(S_k P_1^2 logM + I_2 D_s M(D_s + M)))$. In practice, the inter loop is terminated after a fixed toleration (10^{-2} of the relative error change) is reached. In terms of outer loop iteration number determination, we plot the evolutions of PSNR and SSIM versus iteration numbers in Fig. 16. One can see that all the curves start to increase dramatically and get stable in about 12 iterations. Consequently, we use $I_1 = 12$ for the experiments in this paper. To super-resolve an 85 × 85 image with a magnification of 3, the proposed method requires about 10 minutes, on an Intel Core i5-3470 PC under Matlab R2012a environment. Also, we report the running time for the comparative methods under the same testing condition in the Table 1 and 2.

4.5. Parameter determination

To achieve a reasonable parameter determination, in this subsection, we provide an analysis of the sensitivity to the setting of parameters. We first investigate the influence of image patch size and Fig. 17 (a) gives the changing results of PSNR and SSIM from 3 to 13 with step 2. We see that these curves first ascend and then decline with the increase of patch size. Clearly, the bigger the patch size, the more expressive the input. Thus, increasing the patch size improves the performance. However, the quality cures have a slight drop when input size exceeds 9 × 9, due to the bigger input size (higher dimension) increasing the computational cost. Fig. 17 (b) shows the curves of the averaged PSNR and SSIM values with varying the number of similar neighbors in the nonlocal constraint, whose change trends are similar to Fig. 17 (a). As shown, both PSNR and SSIM values increase within the range from 4 to 10. Nonetheless, the curves then start to drop due to the need of massive amounts of similar patches leading to inaccurate patch matching. Furthermore, we analyze the influence of the other parameters (λ, μ, γ) by grid search strategy shown in Fig. 18 and find that there is a single peak for each parameter. Thus, the proper parameter values are selected with the best performances.

5. Conclusions

We propose a robust dictionary learning for noisy image SR, in which our ℓ_1-norm solution on coefficients transition term pre-

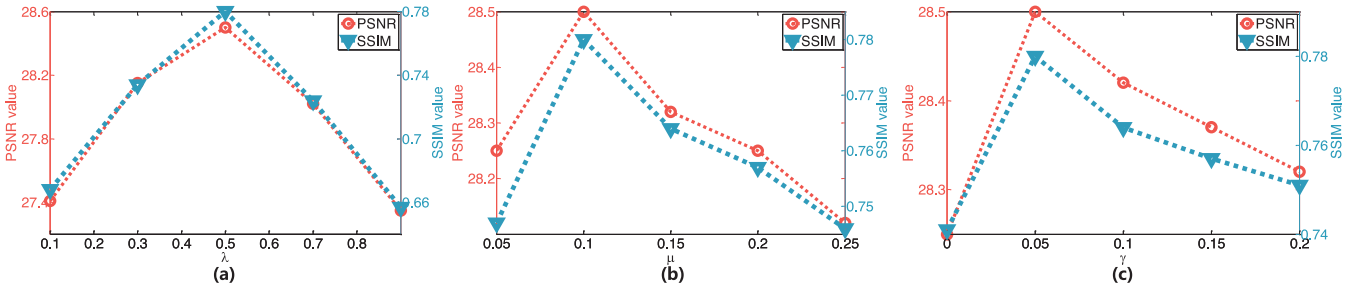

Fig. 18. Performance results with different parameters setting. (a) with different λ values; (b) with different μ values; (c) with different γ values.

vents the noise to be transmitted from noisy LR input to HR output. By incorporating the non-local constraint on HR sparse coefficient into our dictionary learning framework, the improved sparse representation further enhances SR inference. Results on noisy, denoised and noiseless data validate the superiority of the proposed method.

Although ℓ_1-norm effectively suppresses the noise, it requires sufficient iterations during SR inference. In our next work, we will try to use the dictionaries and transition matrices learned within much fewer iterations as the initialization of a deep neural network framework. It may reduce the training cost of the network and speed up the SR inference simultaneously.

Acknowledgments

This work is supported by the National Basic Research Program (973 Program) of China (No. 2013CB329402), the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) (No. B07048), the Program for Cheung Kong Scholars and Innovative Research Team in University(No. IRT 15R53), and JSPS Grants-in-Aid for Scientific Research C (No. 15K00236) for funding.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.sigpro.2017.04.015.

References

[1] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[2] J. Yang, Z. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution, IEEE Trans. Image Process. 21 (8) (2012) 3467–3478.
[3] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: Curves and Surfaces, Springer, 2012, pp. 711–730.
[4] S. Wang, L. Zhang, Y. Liang, Q. Pan, Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis, in: CVPR, IEEE, 2012, pp. 2216–2223.
[5] D.-A. Huang, Y.-C. F. Wang, Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition, in: ICCV, IEEE, 2013, pp. 2496–2503.
[6] L. He, H. Qi, R. Zaretzki, Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution, in: CVPR, IEEE, 2013, pp. 345–352.
[7] R. Timofte, V. De, L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in: ICCV, IEEE, 2013, pp. 1920–1927.
[8] R. Timofte, V. De Smet, L. Van Gool, A+: adjusted anchored neighborhood regression for fast super-resolution, in: ACCV, Springer, 2014, pp. 111–126.
[9] T. Peleg, M. Elad, A statistical prediction model based on sparse representations for single image super-resolution, IEEE Trans. Image Process. 23 (6) (2014) 2569–2582.
[10] S. Schulter, C. Leistner, H. Bischof, Fast and accurate image upscaling with super-resolution forests, in: CVPR, IEEE, 2015, pp. 3791–3799.
[11] C. Huang, Y. Liang, X. Ding, C. Fang, Generalized joint kernel regression and adaptive dictionary learning for single-image super-resolution, Els. Signal Process. 103 (2014) 142–154.
[12] S. Zhao, H. Liang, M. Sarem, A generalized detail-preserving super-resolution method, Els. Signal Process. 120 (2016) 156–173.
[13] K. Zhang, X. Gao, J. Li, H. Xia, Single image super-resolution using regularization of non-local steering kernel regression, Els. Signal Process. 123 (2016) 53–63.
[14] B. Yue, S. Wang, X. Liang, L. Jiao, Robust noisy image super-resolution using ℓ_1-norm regularization and non-local constraint, ACCV Workshops, Springer, 2016.
[15] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[16] D. Dai, R. Timofte, L. Van Gool, Jointly optimized regressors for image super-resolution, in: Computer Graphics Forum, Wiley Online Library, 2015, pp. 95–104.
[17] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: ECCV, Springer, 2014, pp. 184–199.
[18] Z. Cui, H. Chang, S. Shan, B. Zhong, X. Chen, Deep network cascade for image super-resolution, in: ECCV, Springer, 2014, pp. 49–64.
[19] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: ICCV, IEEE, 2015, pp. 370–378.
[20] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, CVPR, IEEE, 2016.
[21] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.
[22] A. Singh, F. Porikli, N. Ahuja, Super-resolving noisy images, in: CVPR, IEEE, 2014, pp. 2846–2853.
[23] M. Aharon, M. Elad, A. Bruckstein, K-svd: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[24] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Process. 20 (7) (2011) 1838–1857.
[25] Y. Wang, C. Xu, S. You, C. Xu, D. Tao, Dct regularized extreme visual recovery, IEEE Trans. Image Process. (2017).
[26] W. Dong, L. Zhang, G. Shi, Centralized sparse representation for image restoration, in: ICCV, IEEE, 2011, pp. 1259–1266.
[27] J. Zhang, D. Zhao, W. Gao, Group-based sparse representation for image restoration, IEEE Trans. Image Process. 23 (8) (2014) 3336–3351.
[28] X. Li, H. He, R. Wang, D. Tao, Single image super-resolution via directional group sparsity and directional features, IEEE Trans. Image Process. 24 (9) (2015) 2874–2888.
[29] Y. Hu, D. Zhang, J. Ye, X. Li, X. He, Fast and accurate matrix completion via truncated nuclear norm regularization, IEEE Trans. Pattern Anal. Mach. Intell. 35 (9) (2013) 2117–2130.
[30] T. Liu, D. Tao, On the performance of manhattan nonnegative matrix factorization, IEEE Trans. Neural Netw. Learn. Syst. 27 (9) (2016) 1851–1863.
[31] T. Liu, M. Gong, D. Tao, Large-cone nonnegative matrix factorization, IEEE Trans. Neural Netw. Learn. Syst. (2016).
[32] Y. Wang, C. Xu, C. Xu, D. Tao, Beyond rpca: flattening complex noise in the frequency domain, AAAI (2017) 500–505.
[33] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: CVPR, IEEE, 2005, pp. 60–65.
[34] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction, IEEE Trans. Image Process. 18 (1) (2009) 36–51.
[35] X. Chen, Q. Lin, S. Kim, J.G. Carbonell, E.P. Xing, et al., Smoothing proximal gradient method for general structured sparse regression, Ann. Appl. Stat. 6 (2) (2012) 719–752.
[36] R. Chalasani, J.C. Principe, Deep predictive coding networks, ArXiv:1301.3541 arXiv preprint (2013).
[37] K. Zhang, X. Gao, D. Tao, X. Li, Single image super-resolution with non-local means and steering kernel regression, IEEE Trans. Image Process. 21 (11) (2012) 4544–4556.
[38] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, IEEE Trans. Image Process. 16 (8) (2007) 2080–2095.