Learning-Based Joint Super-Resolution and Deblocking for a Highly Compressed Image

Li-Wei Kang, *Member*, *IEEE*, Chih-Chung Hsu, *Student Member*, *IEEE*, Boqi Zhuang, Chia-Wen Lin, *Senior Member*, *IEEE*, and Chia-Hung Yeh, *Senior Member*, *IEEE*

***Abstract*—A highly compressed image is usually not only of low-resolution, but also suffers from compression artifacts (blocking artifact is treated as an example in this paper). Directly performing image super-resolution (SR) to a highly compressed image would also simultaneously magnify the blocking artifacts, resulting in unpleasing visual experience. In this paper, we propose a novel learning-based framework to achieve joint single-image SR and deblocking for a highly-compressed image. We argue that individually performing deblocking and SR (i.e., deblocking followed by SR, or SR followed by deblocking) on a highly compressed image usually cannot achieve a satisfactory visual quality. In our method, we propose to learn image sparse representations for modeling the relationship between low and high-resolution image patches in terms of the learned dictionaries for image patches with and without blocking artifacts, respectively. As a result, image SR and deblocking can be simultaneously achieved via sparse representation and MCA (morphological component analysis)-based image decomposition. Experimental results demonstrate the efficacy of the proposed algorithm.**

***Index Terms*—image super-resolution, sparse representation, dictionary learning, self-learning, image decomposition, morphological component analysis (MCA).**

I. INTRODUCTION

WITH the rapid development of multimedia and network technologies, delivering and sharing multimedia contents over the Internet and heterogeneous devices has become more and more popular. However, due to limited channel bandwidth and storage capability, most images distributed over the Internet exist in low-quality versions degraded from the sources. The most common image degradations come from downscaling and compression. Downscaling reduces the spatial resolution in an image, whereas compression further reduces the redundancy in the spatial and temporal domains. Although downscaling together with compression can greatly reduce the required bandwidth and storage for images, they would also lead to significant information loss and unpleasing visual artifacts, including blocking, ringing, or blurring [1].

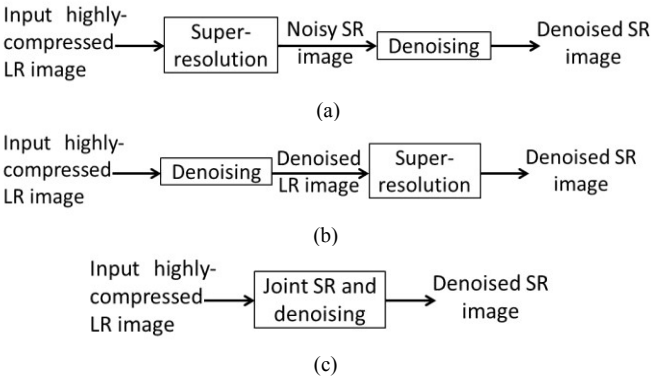

Fig. 1. SR methods for a highly compressed image: **(a)** cascading structure I: SR followed by denoising; **(b)** cascading structure II: denoising followed by SR; and **(c)** joint SR and denoising.

A. Image Super-resolution

There has been a great demand for improving the perceptual quality of images in terms of the spatial resolution enhancement of an image, also known as image super-resolution (SR). The goal of image SR is to recover a high-resolution (HR) image from one or multiple low-resolution (LR) input images, which is essentially an ill-posed inverse problem [2]. There are mainly two categories of approaches for image SR: (i) traditional approaches and (ii) exemplar/learning-based approaches. In the traditional approaches, one sub-category is reconstruction-based schemes, where a set of LR images of the same scene are aligned with sub-pixel accuracy to generate a HR image [3]. Such kind of approaches mainly rely on multi-frame alignment, which is usually time-consuming and inaccurate, and cannot be used for single-image SR since it requires multiple input LR images. The other sub-category is frame interpolation [4], which has been shown to generate overly smooth images with ringing and jaggy artifacts.

The exemplar/learning-based methods [5]–[14] hallucinate the high frequency details of a LR image based on the

Manuscript received March 25, 2014, revised November 2, 2014, revised February 17, 2015, accepted May 10, 2015. This work was supported in part by the Ministry of science and technology), Taiwan, under Grants MOST 101-2221-E-007-121-MY3, 103-2221-E-007-046-MY3, and MOST 103-2221-E-224-034-MY2.

Li-Wei Kang is with the Graduate School of Engineering Science and Technology-Doctoral Program, and the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan.

Chih-Chung Hsu is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan.

Boqi Zhuang is with the Altek Corporation, Hsinchu, Taiwan.

Chia-Wen Lin (corresponding author) is with the Department of Electrical Engineering and the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan, and with the Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan. (e-mail: cwlin@ee.nthu.edu.tw).

Chia-Hung Yeh is with the Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan.

(a) (b) (c) (d)

Fig. 2. An example illustrating the main problems caused by existing SR approaches for a highly compressed image: **(a)** the original HR image; **(b)** a LR and highly compressed version of (a); the SR version of (a) obtained by **(c)** directly applying the bicubic interpolation; and **(d)** applying image deblocking first, followed by SR (the Cascading-based approach used for comparison with the proposed method, described in Sec. IV). It can be observed from (c) that the blocking artifacts are significantly magnified by SR, while from (d) that some details are blurred by deblocking, resulting in poor SR result.

co-occurrence prior between LR and HR image patches in a training set, which has proven to provide much finer details compared to the traditional approaches. More specifically, for a LR input, exemplar-based methods [5]–[8] search for similar image patches from a pre-collected training LR image dataset or the same image itself based on self-examples, and use their corresponding HR versions to produce the final SR output. Nevertheless, the fine details reconstructed by such exemplar-based methods, though looking visually similar, cannot be guaranteed to provide the true missing HR details. Hence, the performance of this approach highly relies on the similarity between the training set and test set or the self-similarity in the image itself.

Moreover, learning-based SR approaches [9]–[14] focus on modeling the relationship between different resolutions of images. For example, Yang *et al.* [9] proposed to apply sparse coding techniques to learn a compact representation for HR/LR patch pairs for SR based on pre-collected HR/LR image pairs. Then, Yang *et al.* [10] advances [9] to propose a coupled dictionary training approach for SR based on patch-wise sparse recovery, where the learned couple dictionaries relate the HR/LR image patch spaces via sparse representations. It is guaranteed that the sparse representation of a LR image patch can well reconstruct its underlying HR image patch, which cannot be guaranteed in [9]. Moreover, a sparse representation based framework was proposed in [11] for image deblurring and SR based on adaptive sparse domain selection and adaptive regularization, where two adaptive regularization terms are introduced. In addition, Ren *et al.* [12] proposed to utilize context-aware sparsity prior to enhance the performance of sparsity-based restoration approach for image denoising and SR. This method mitigates the artifacts in the produced HR images based on incorporating the structural correlations of dictionary atoms into the employed sparse coding algorithm. On the other hand, self-learning frameworks based on self-similarity of an image were introduced for SR in [13], [14].

B. Motivation of SR for a Highly Compressed Image

The above-mentioned SR approaches, however, only consider that an input LR image is only degraded by down-scaling or at most an additional blurring operation (with a known or well-estimated blurring kernel). It is not always practical in a network environment, where image compression is usually adopted, which yields additional compression artifacts such as blocking and ringing. For image search engines, compression helps reduce the image size by up to 50% without obvious perceptual quality loss presented in the LR form of an image. Nevertheless, if SR is directly performed on the compressed LR image, compression artifacts will be simultaneously magnified and therefore the perceptual quality of resulting HR image would be poor [15]. Hence, a high-performance SR scheme for highly-compressed images is desirable for enhancing the resolution of image/video degraded by both down-scaling and compression. In [15], a unified framework was proposed to simultaneously improve the resolution and perceptual quality of low-quality web image/video degraded by down-scaling and compression. This approach combines adaptive regularization and learning-based SR, where the regularization strength is determined by the JPEG compression quality factor (QF) of an input image. QF is an integer ranging from 0 and 100, which controls the degree of compression for JPEG images. The larger the QF is, the lower the compression ratio and the higher the image quality are [24].

On the other hand, an exemplar-based SR algorithm of compressed videos in DCT (discrete cosine transform) domain was proposed in [16]. The inputs to the system include a compressed LR video together with a HR still image of similar content which is assumed to be available in advance. In addition, a unified framework achieving simultaneous denoising and SR for noisy video was proposed in [17]. This approach assesses the visual quality with respect to fidelity preservation, detail preservation, and spatio-temporal smoothness. More works for low-quality video SR can be found in [18], [19].

Existing SR approaches for low-quality images/videos [15]–[19], however, are all designed for some special purposes. For example, the SR framework proposed in [15] is mainly designed for JPEG compressed web images with known QFs. Moreover, most other related approaches are designed for video [16]–[19]. In addition, strictly speaking, most existing SR

methods for low-quality images usually rely on two cascading structures: 1) SR followed by a denoising operation [cascading structure I in Fig. 1(a)], and 2) denoising followed by SR [cascading structure II in Fig. 1(b)] . In the first cascading structure, the SR operation would significantly magnify the compression artifacts [e.g., the magnified blocking artifacts in Fig. 2(c)], making it difficult to eliminate the magnified artifacts using the next denoising operation. On the other hand, the second cascading structure (i.e., denoising followed by SR), would inevitably lose some image details caused by denoising (e.g., deblocking), thereby degrading the performance of SR, as exemplified in Fig. 2(d). Therefore, a joint SR and denoising structure [see Fig. 1(c)] is desirable as it not only avoids magnifying compression artifact but also reconstructs finer image details so as to achieve good-quality SR even the input image is highly compressed.

To achieve joint SR and deblocking, we proposed in the preliminary conference version [23] of this paper a self-learning-based SR framework to achieve joint SR and blocking artifact removal for a single LR image. The proposed method self-learns from the input image itself the sparse representations for modeling the relationship between LR and HR image patches based on the dictionaries learned from image patches with and without blocking artifacts, respectively. Based on the learned sparse representations, MCA (morphological component analysis)-based image decomposition [28], [29] is then applied to decompose and remove blocking artifacts so as to achieve image SR and deblocking simultaneously.

C. Contribution of Proposed Method

It should be noted that most research works on pure SR usually simulate the production of an input LR image as two degradation steps. For example, in [9], [10], it is assumed that an observed LR image is a blurred and down-sampled version of its HR version. In addition, the work proposed in [11] considers a more general image restoration problem, where an observed LR image is a blurred and down-sampled version of its HR version with additional additive noise. Moreover, the context-aware approach in [12] was proposed for LR images with Gaussian noise of known standard deviation. However, the image artifacts (e.g., additive Gaussian noise or blurring effect) considered in these works are significantly different from blocking artifact considered in this paper.

In our SR framework, we focus on a single input LR image/video frame being highly compressed by a block-based image/video compression algorithm (e.g., JPEG [24], H.264/AVC [25], or HEVC [26]), which has been widely used in current multimedia applications (e.g., images from WWW or video frames from YouTube [27]). To this end, we propose a learning-based sparse representation framework to achieve joint single-image SR and deblocking for a highly compressed image. Blocking artifact is one of the major visually unpleasing artifacts introduced by transform block-based compression and hence, deblocking has been extensively studied in the literature [1], [20]–[22]. The main contribution of this paper is three-fold: i) to the best of our knowledge, we are among the first to propose a unified framework for joint blocking artifact removal and single-image SR via sparse coding-based image decomposition; ii) the proposed framework can be adapted to any block transform-based compressed image/video without the need of any prior knowledge about the data source, coding bitrates, and compression algorithms; and iii) the proposed framework can be easily integrated into existing self-learning-based SR approaches [13]–[14], where no extra training samples are required, or non-self-learning-based ones [5]–[12], which require additional training samples collected offline.

TABLE I
NOTATION

Symbols	Meanings
I	Input LR image
I^d	Down-scaled version of I
I_{LF}, I_{HF}	Low-frequency and high-frequency parts of I
I^d_{LF}, I^d_{HF}	Low-frequency and high-frequency parts of I^d
x_i, y_i	Pair of the i-th corresponding HR and LR patches
$\mathcal{X}$, $\mathcal{Y}$	HR and LR patch sets
$\{\mathcal{X}^B, \mathcal{Y}^B\}$, $\{\mathcal{X}^N, \mathcal{Y}^N\}$	Blocking and non-blocking HR/LR patch pairs
D_B, D_N	Two sets of coupled dictionaries used for SR of blocking and non-blocking patches
D_N^{HR}, D_N^{LR}	Pair of dictionaries of corresponding HR/LR atoms for non-blocking patches, forming D_N
D_B^{HR}, D_B^{LR}	Pair of dictionaries of corresponding HR/LR atoms, forming D_B, where $D_B^{\mathrm{HR}} = [D_{B\_N}^{\mathrm{HR}} \mid D_{B\_B}^{\mathrm{HR}}]$ and $D_B^{\mathrm{LR}} = [D_{B\_N}^{\mathrm{LR}} \mid D_{B\_B}^{\mathrm{LR}}]$.
$D_{B\_N}^{\mathrm{HR}}$, $D_{B\_N}^{\mathrm{LR}}$	Pair of non-blocking dictionaries consisting of corresponding HR/LR atoms
$D_{B\_B}^{\mathrm{HR}}$, $D_{B\_B}^{\mathrm{LR}}$	Pair of blocking dictionaries consisting of corresponding HR/LR atoms
I_{HF}^N, I_{HF}^B	HR non-blocking and HR blocking components of I_{HF}
b_p^N (b_p^B), B_p^N (B_p^B)	The p-th non-blocking (blocking) patch in I_{HF} and its SR result
θ_p^N (θ_p^B), $\hat{\theta}_p^N$ ($\hat{\theta}_p^B$)	Sparse representation of b_p^N (b_p^B) and reconstructed sparse representation of b_p, used to estimate B_p^N (B_p^B)
$\hat{\theta}_{p\_N}^B$	Sparse representation of b_p^B with the coefficients corresponding to the atoms in $D_{B\_B}^{\mathrm{LR}}$, being set to zeros.

Compared with the preliminary conference version [23], this paper has been significantly extended in the following aspects: (i) In this paper, we further extend the proposed framework to non-self-learning-based SR approaches which was not discussed in [23]. (ii) We conduct subjective quality evaluations based on paired comparisons for different SR and deblocking schemes. (iii) This paper provides in-depth analyses and interpretations about the experimental results and computational complexity to offer good insights about the proposed method.

The rest of this paper is organized as follows. Sec. II reviews some related works. Sec. III presents the proposed SR framework for highly compressed images. In Sec. IV, experimental results are demonstrated. Finally, Sec. V concludes this paper.

II. PRELIMINARIES

In this section, we briefly review the concept of the general MCA-based framework for image decomposition and the utilized sparse representation and dictionary learning techniques. The major idea of MCA is to utilize the

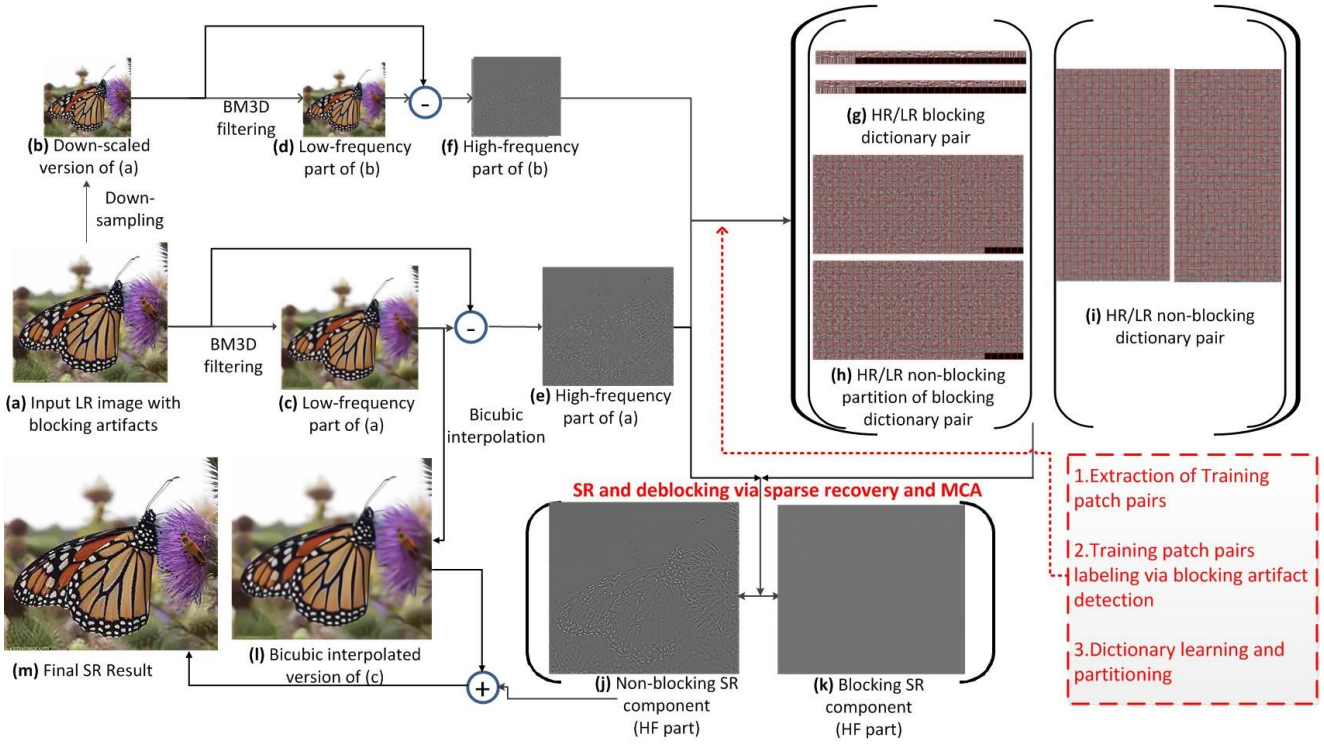

Fig. 3. Flowchart of the proposed self-learning-based super-resolution framework for a highly compressed image.

morphological diversity of different features contained in the data to be decomposed and to associate each morphological component to a dictionary of atoms. The MCA-based image decomposition technique described below will be integrated in the proposed SR framework mainly for image deblocking.

A. MCA-based Image Decomposition

Suppose an image I of N pixels is a superposition of S layers (called morphological components), denoted by $I = \sum_{s=1}^{S} I_s$, where I_s denotes the s-th component, such as the geometric or textural component of I. To decompose image I into $\{I_s\}_{s=1}^{S}$, the MCA algorithms [22], [28]–[31] iteratively minimize the following energy function:

$$E(\{I_s\}_{s=1}^{S}, \{\theta_s\}_{s=1}^{S}) = \frac{1}{2}\left\| I - \sum_{s=1}^{S} I_s \right\|_2^2 + \tau \sum_{s=1}^{S} E_s(I_s, \theta_s), \quad (1)$$

where θ_s denotes the sparse coefficients corresponding to I_s with respect to dictionary D_s, τ is a regularization parameter, and E_s is the energy function defined according to the type of D_s (global or local dictionary). For convenience, the symbols used in this paper are listed in TABLE I.

B. Sparse Representation and Dictionary Learning

Sparse coding [32], [33] is a technique of finding a sparse representation for a signal with a small number of nonzero or significant coefficients corresponding to the atoms in a dictionary. To construct a dictionary D_s to sparsely represent each patch b_s^k extracted from component I_s of image I, we may use a set of available training exemplars $y^k, k = 1,2,\ldots,P$, to learn dictionary D_s sparsifying y^k by solving the following optimization problem:

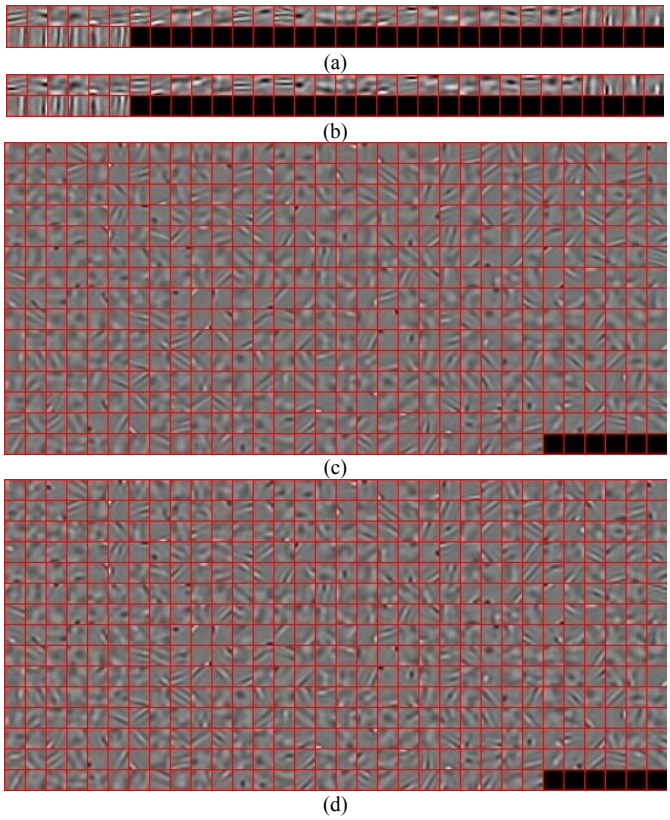

Fig. 4. Highlight of the two learned dictionary pairs based on the extracted HR/LR training patch pairs with blocking artifacts from Fig. 3(a): **(a)** the dictionary of HR atoms with blocking artifacts ($D_{B\_B}^{\mathrm{HR}}$); **(b)** the dictionary of LR atoms with blocking artifacts ($D_{B\_B}^{\mathrm{LR}}$); **(c)** the dictionary of HR atoms without blocking artifacts ($D_{B\_N}^{\mathrm{HR}}$), i.e., non-blocking partition of D_B^{HR}; and **(d)** the dictionary of LR atoms without blocking artifacts ($D_{B\_N}^{\mathrm{LR}}$), non-blocking partition of D_B^{LR}.

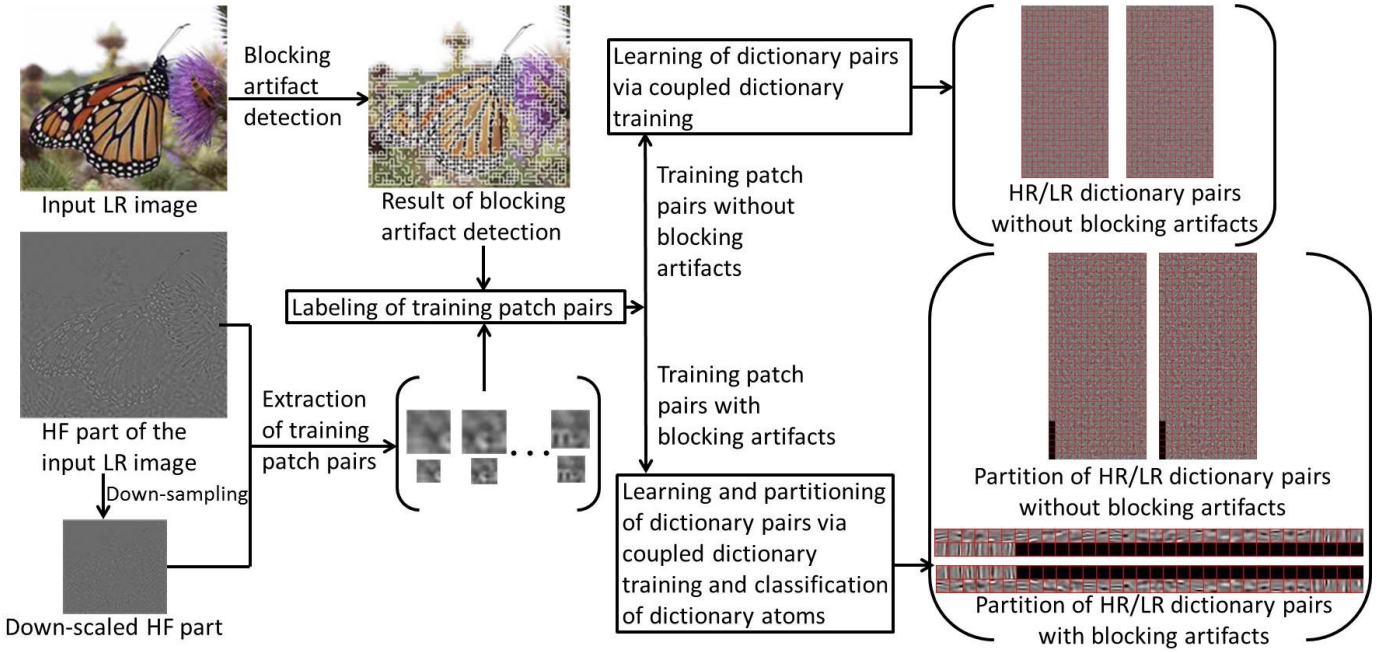

Fig. 5. Illustration of the proposed training patch extraction and dictionary learning processes in the proposed framework (highlight and illustration of the red dotted line region in Fig. 3).

$$\min_{D_s,\theta^k} \sum_{k=1}^{P} \left(\frac{1}{2} \|y^k - D_s\theta^k\|_2^2 + \lambda \|\theta^k\|_1 \right), \qquad (2)$$

where θ^k denotes the sparse coefficient vector of y^k with respect to D_s and λ is a regularization parameter.

III. PROPOSED FRAMEWORK FOR JOINT SR AND DEBLOCKING

In this section, we first explain the proposed framework for self-learning-based SR. Then, in Sec. III.E, we will show how to extend the proposed framework to non-self-learning-based SR schemes. Fig. 3 depicts the proposed framework for self-learning-based joint SR and deblocking for enhancing a downscaled and highly compressed image. Our method is to formulate the image enhancement problem as an MCA-based image decomposition problem via sparse representation.

As illustrated in Fig. 3, an input LR image I with blocking artifacts [Fig. 3(a)] and its downscaled version I^d [Fig. 3(b)] are first roughly decomposed into their corresponding low-frequency (LF) parts, I_{LF} and I_{LF}^d, and high-frequency (HF) parts, I_{HF} and I_{HF}^d, respectively, via a filter. As a result, the respective essential visual information will be retained in the LF parts [Figs. 3(c) and 3(d)], while the blocking artifacts and edge/texture details will be included in the HF parts [see Figs. 3(e) and 3(f)] of the images. Then, we classify all patches in I_{HF} together with their corresponding patches in I_{HF}^d into two clusters: the "blocking" and "non-blocking" HR/LR patch pairs, respectively. Based on the two training sets of patch pairs extracted from the input image itself, we learn two sets of coupled dictionaries, D_B and D_N, used for the SR of blocking and non-blocking patches, respectively, as illustrated in Figs. 3(g)–3(i).

To achieve the SR of I_{HF}, we perform patch-wise sparse reconstruction with the coupled dictionary set D_N [Fig. 3(i)] for each patch without blocking artifacts in I_{HF}. For each patch with blocking artifacts in I_{HF}, we perform SR reconstruction based on D_B (which consisits of two sub-dictionaries of HR and LR atoms, respectively), and MCA-based image decomposition to obtain the deblocked HR patch. As a result, the HF part I_{HF} can be simultaneously upscaled and decomposed into HR non-blocking and HR blocking components [Figs. 3(j) and 3(k)]. We then add the HR non-blocking component of I_{HF} to the bicubic-interpolated I_{LF} [Fig. 3(l)] to obtain the final SR result of I, as illustrated in Fig. 3(m). The detailed method will be elaborated below.

A. Preprocessing and Problem Formulation

Without the need of pre-collecting enormous additional training patches for SR (e.g., in [9], [10], collection of additional training patches is required), the proposed method intends to extract training patches from an input LR image itself. Moreover, to achieve joint SR and deblocking, we convert the problem into the high-frequency domain of the input image and conduct the following preprocessing tasks. To model the relationship between LR and HR image patches, for an input LR image I with blocking artifacts, we first down-scale I to obtain its down-scaled version I^d, and then apply the BM3D (block-matching and 3D filtering) algorithm [34] to decompose I into the LF part (I_{LF}) and HF part (I_{HF}), and decompose I^d into I_{LF}^d and I_{HF}^d, that is $I = I_{\mathrm{LF}} + I_{\mathrm{HF}}$, and $I^d = I_{\mathrm{LF}}^d + I_{\mathrm{HF}}^d$. Then, we identify two sets of HR/LR patch pairs as the training samples for learning the dictionaries used for SR and deblocking, where we extract each patch ($x_i \in \mathcal{X}$) in the higher scale (I_{HF}) and its corresponding patch ($y_i \in \mathcal{Y}$) in the lower scale (I_{HF}^d) with a certain magnification factor to form a coupled training patch pair. Then, we perform blocking artifact

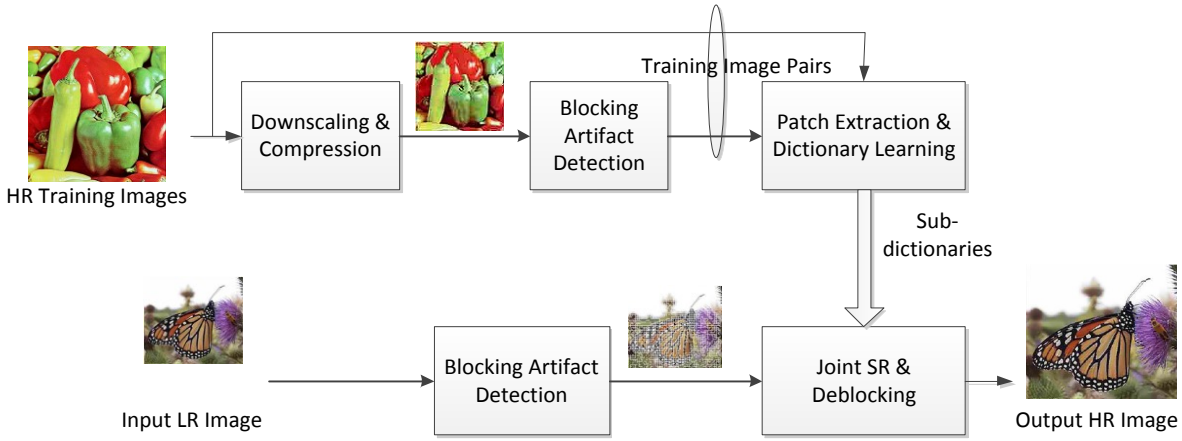

Fig. 6. Extension of the proposed framework to non-self-learning-based SR methods.

detection [20] on the HR part of each coupled patch pair (i.e., a patch from I_{HF}) and classify all coupled patch pairs into two sets of "blocking" and "non-blocking" patch pairs: $\{\mathcal{X}^B, \mathcal{Y}^B\}$ and $\{\mathcal{X}^N, \mathcal{Y}^N\}$.

Based on the two sets of training samples, we propose to learn two sets of dictionaries, respectively, for SR of non-blocking patches and joint SR and deblocking of blocking patches, as detailed in Sec. III.B and Sec. III.C. Then, we formulate the SR of each input LR non-blocking patch b_p^N (in I_{HF}) as the following sparse representation problem:

$$\hat{\theta}_p^N = \arg\min_{\theta_p^N} \left\| b_p^N - D_N^{\mathrm{LR}} \theta_p^N \right\|_2^2 + \lambda \left\| \theta_p^N \right\|_1, \tag{3}$$

where D_N^{LR} denotes the learned LR dictionary of non-blocking atoms, θ_p^N is the sparse presentation of b_p^N with respect to D_N^{LR}, $\hat{\theta}_p^N$ is the solution of θ_p^N for minimizing (3), and λ is a parameter controlling the sparsity penalty and representation fidelity. As a result, the SR result B_p^N of b_p^N can be obtained as follows:

$$B_p^N = D_N^{HR} \hat{\theta}_p^N, \tag{4}$$

where D_N^{HR} is the learned HR dictionary of non-blocking atoms, which leads to the same sparse representation $\hat{\theta}_p^N$ for an HR patch as that for its corresponding LR patch with respect to D_N^{LR}.

Moreover, we formulate the joint SR and deblocking for each input LR blocking patch b_p^B (in I_{HF}) as an MCA-based image decomposition problem via sparse representation:

$$\hat{\theta}_p^B = \arg\min_{\theta_p^B} \left\| b_p^B - D_B^{\mathrm{LR}} \theta_p^B \right\|_2^2 + \lambda \left\| \theta_p^B \right\|_1, \tag{5}$$

where the definitions of θ_p^B and $\hat{\theta}_p^B$ are similar to those in (3), and $D_B^{\mathrm{LR}} = \left[D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}} \right]$ is the learned LR dictionary based on the training samples with blocking artifacts, which can be further partitioned into $D_{B\_N}^{\mathrm{LR}}$ and $D_{B\_B}^{\mathrm{LR}}$, consisting of the non-blocking and blocking atoms, respectively. As a result, the joint SR and deblocking result B_p^B of b_p^B can be obtained by

$$B_p^B = D_{B\_N}^{HR} \hat{\theta}_{p\_N}^B, \tag{6}$$

where $D_{B\_N}^{\mathrm{HR}}$ is obtained from the partition of the learned HR dictionary D_B^{HR} based on the training samples with blocking artifacts, which can be further partitioned into $D_{B\_N}^{\mathrm{HR}}$ and $D_{B\_B}^{\mathrm{HR}}$, including the non-blocking and blocking atoms, respectively, and $\hat{\theta}_{p\_N}^B$ is the sparse representation of b_p^B obtained by solving (5) with the coefficients corresponding to the atoms in $D_{B\_B}^{\mathrm{LR}}$, being set to zero.

B. *Dictionary Learning for Single Image SR*

Based on the extracted HR/LR training patch pairs without blocking artifacts ($\{\mathcal{X}^N, \mathcal{Y}^N\}$) from I_{HF} itself, we intend to learn a couple of dictionaries (D_N^{HR} and D_N^{LR}) to model the relationships between HR and LR image patches. Similar to the coupled dictionary training method proposed in [10], we treat the set $\mathcal{Y}^N$ of LR training patches as the observation space, and the set $\mathcal{X}^N$ of HR training patches as the latent space, where the patches have sparse representations with respect to certain dictionaries. Patches (LR) in $\mathcal{Y}^N$ are observations, while patches (HR) in $\mathcal{X}^N$ are what we want to recover. The problem is to find a coupled dictionary pair D_N^{HR} and D_N^{LR} for spaces $\mathcal{X}^N$ and $\mathcal{Y}^N$, respectively, such that given an input LR patch $y_i \in \mathcal{Y}^N$, we can use its sparse representation with respect to D_N^{LR} to reconstruct the corresponding latent HR patch $x_i \in \mathcal{X}^N$ with respect to D_N^{HR}. Hence, for each coupled patch pair $\{x_i, y_i\}$, a desired pair of coupled dictionaries D_N^{HR} and D_N^{LR} should satisfy

$$\hat{\theta}_i = \arg\min_{\theta_i} \left\| y_i - D_N^{\mathrm{LR}} \theta_i \right\|_2^2 + \lambda \| \theta_i \|_1, \tag{7}$$

$$\hat{\theta}_i = \arg\min_{\theta_i} \left\| x_i - D_N^{\mathrm{HR}} \theta_i \right\|_2^2, \tag{8}$$

where $\hat{\theta}_i$ is the sparse representation of y_i with respect to D_N^{LR}, and also the sparse representation of x_i for recovering x_i with respect to D_N^{HR}.

Given input y_i, to learn a coupled dictionary pair D_N^{HR} and D_N^{LR} for reconstructing x_i with respect to D_N^{HR}, the problem can be formulated as [10]

$$\min_{D_N^{\mathrm{HR}}, D_N^{\mathrm{LR}}} \frac{1}{P} \sum_{i=1}^{P} L(D_N^{\mathrm{HR}}, D_N^{\mathrm{LR}}, x_i, y_i) \tag{9}$$

$$\text{s.t. } \hat{\theta}_i = \arg\min_{\theta_i} \left\| y_i - D_N^{\mathrm{LR}} \theta_i \right\|_2^2 + \lambda \| \theta_i \|_1, i = 1,2,\cdots,P,$$

where P is the number of training patch pairs, and $L(\cdot)$ represents a cost function for ensuring a good representation of y_i with respect to D_N^{LR}, and minimizing the reconstruction error of x_i with respect to D_N^{HR}, defined as follows:

TABLE II
PROPOSED JOINT SR AND DEBLOCKING ALGORITHM BASED ON SIX LEARNED DICTIONARIES

Input: a LR image I with blocking artifacts.
Output: the blocking artifacts-removed HR version of I.

Learning phase

1. Extraction of training HR/LR patch pairs from the HF part (I_{HF}) of I.
2. Classification of the training patches into training patch pairs ($\{\mathcal{X}^N, \mathcal{Y}^N\}$) without blocking artifacts and training patch pairs ($\{\mathcal{X}^B, \mathcal{Y}^B\}$) with blocking artifacts.
3. Learning of HR/LR dictionary pairs (D_N^{HR} and D_N^{LR}) without blocking artifacts via coupled dictionary training based on $\{\mathcal{X}^N, \mathcal{Y}^N\}$.
4. Learning of HR/LR dictionary pairs (D_B^{HR} and D_B^{LR}) with blocking artifacts via coupled dictionary training based on $\{\mathcal{X}^B, \mathcal{Y}^B\}$.
5. Partitioning D_B^{HR} and D_B^{LR} into $D_{B\_N}^{\mathrm{HR}}$ and $D_{B\_B}^{\mathrm{HR}}$, and $D_{B\_N}^{\mathrm{LR}}$ and $D_{B\_B}^{\mathrm{LR}}$, respectively, to obtain HR/LR dictionary pairs ($D_{B\_N}^{\mathrm{HR}}$ and $D_{B\_N}^{\mathrm{LR}}$) without blocking artifacts, and HR/LR dictionary pairs ($D_{B\_B}^{\mathrm{HR}}$ and $D_{B\_B}^{\mathrm{LR}}$) with blocking artifacts, where $D_{B\_B}^{\mathrm{HR}}$ can be discarded.

Reconstruction phase

1. For each LR patch without blocking artifacts of I_{HF}, perform sparse decomposition with D_N^{LR} via (3) and reconstruct its HR version with D_N^{HR} via (4).
2. For each LR patch with blocking artifacts of I_{HF}, perform sparse decomposition with $D_B^{\mathrm{LR}} = \left[D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}}\right]$ via (5) and keep only the sparse coefficients corresponding to $D_{B\_N}^{\mathrm{LR}}$. Then, reconstruct its blocking artifacts-removed HR version with $D_{B\_N}^{\mathrm{HR}}$ via (6).
3. Return the integration of the reconstructed blocking artifacts-removed HR version of I_{HF} and the bicubic-interpolated LF part (I_{LF}) of I.

$$L = \frac{1}{2}\left[\alpha\left\|D_N^{\mathrm{HR}}\hat{\theta}_i - x_i\right\|_2^2 + (1-\alpha)\left\|D_N^{\mathrm{LR}}\hat{\theta}_i - y_i\right\|_2^2\right], \quad (10)$$

where α $(0 < \alpha \leq 1)$ is used to balance the two reconstruction errors. The objective function in (9) can be minimized by alternatively optimizing over D_N^{HR} and D_N^{LR}, while keeping the other fixed. More details about the coupled dictionary learning can be found in [10]. The main difference between the proposed method and the approach proposed in [10] is that all of the training patches are extracted from an input LR image itself in our approach while enormous additional training patches are required in [10].

C. Dictionary Learning for Image Decomposition

For the extracted HR/LR training patch pairs with blocking artifacts, we need to not only learn a coupled dictionary pair D_B^{HR} and D_B^{LR} for SR purpose, but also identify the "blocking/non-blocking atoms" in the two dictionaries for achieving MCA-based deblocking. Similar to the employed coupled dictionary learning technique described in Sec. III.B, we first learn D_B^{HR} and D_B^{LR} by solving (9), where the parameters are replaced accordingly. Then, we analyze the atoms constituting D_B^{HR}, and find that these atoms can be roughly divided into two clusters (sub-dictionaries) for representing the non-blocking and blocking components of I_{HF}, respectively. That is, even if the training patches extracted from I_{HF} are classified into blocking patches based on [20], two different kinds of atoms in terms of blocking and non-blocking artifacts can still be analyzed after the dictionary learning process.

Based on our previous work in [22], we modify the HOG (histograms of oriented gradients) descriptor [35] to describe each atom in D_B^{HR} (the HOG features of the atoms in the dictionary learned from HR patches should be more significant than those from LR patches). Since blocking artifacts are visually noticeable changes in pixel values along block boundaries, it is only required to consider the horizontal and vertical gradients in each dictionary atom. Hence, we modify the original HOG to calculate only the histogram over the intervals of angles around 0°, 180°, 90°, and 270°. After extracting the HOG feature for each atom in D_B^{HR}, we first apply the K-means algorithm to classify all of the atoms in D_B^{HR} into two clusters, D_1 and D_2, based on their horizontal HOG feature descriptors. Then, we calculate the variance of gradient directions for each atom d_{ij} in cluster D_i, as VG_{ij}, $i = 1, 2$. We subsequently calculate the mean of VG_{ij} for each cluster D_i as MVG_i. Based on the fact that the edge directions of the samples of an atom with horizontal (or vertical) blocking artifacts should be consistent, *i.e.*, the variance of gradient directions for a "horizontal (or vertical) blocking" atom should be generally smaller than those of the other atoms with no remarkably dominating edge direction, we identify the cluster with the smaller MVG_i as horizontal blocking sub-dictionary. Then, similarly the other cluster can be further classified to obtain the "vertical blocking" sub-dictionary and non-blocking sub-dictionary $D_{B\_N}^{\mathrm{HR}}$. In addition, the identified horizontal and vertical blocking sub-dictionaries constitute the blocking sub-dictionary $D_{B\_B}^{\mathrm{HR}}$. That is, D_B^{HR}=$\left[D_{B\_N}^{\mathrm{HR}} | D_{B\_B}^{\mathrm{HR}}\right]$. Meanwhile, the atoms in D_B^{LR} can also be classified into two clusters according to their corresponding atoms in D_B^{HR} to form the blocking and non-blocking dictionaries, $D_{B\_N}^{\mathrm{LR}}$ and $D_{B\_B}^{\mathrm{LR}}$, of LR atoms, i.e., $D_B^{\mathrm{LR}} = \left[D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}}\right]$. As an example highlighting Figs. 3(g) and 3(h), Fig. 4 shows the two learned dictionary pairs based on the extracted HR/LR training patch pairs with blocking artifacts from Fig. 3(a). To summarize the training patch extraction and dictionary learning processes in our framework, the red dotted line region in Fig. 3 is highlighted and illustrated in Fig. 5.

D. Joint SR and Deblocking via Sparse Reconstruction

After learning the six dictionaries, D_N^{HR}, D_N^{LR}, $D_{B\_N}^{\mathrm{HR}}$, $D_{B\_B}^{\mathrm{HR}}$, $D_{B\_N}^{\mathrm{LR}}$, and $D_{B\_B}^{\mathrm{LR}}$, based on the training patches extracted from I_{HF}, joint SR and deblocking of I_{HF} can be efficiently achieved via patch-wise sparse recovery as follows.

1) SR for non-blocking patches: In this case, only SR is required to be performed for each patch. For each input LR non-blocking patch b_p^N in I_{HF}, we first calculate its sparse representation $\hat{\theta}_p^N$ with respect to D_N^{LR} via (3). Based on the employed coupled dictionary learning algorithm described in Sec. III-B, the spare representation can also be used to recover its HR version B_p^N with respect to D_N^{HR} via (4).

2) Joint SR and deblocking for blocking patches: In this case, both SR and deblocking are jointly performed for each

patch, For each input LR blocking patch b_p^B in I_{HF}, we first perform MCA-based image decomposition via sparse representation with respect to $D_B^{\mathrm{LR}} = [D_{B\_N}^{\mathrm{LR}} | D_{B\_B}^{\mathrm{LR}}]$ to find the atoms which actually contribute for representing the non-blocking part of b_p^B via (5). We then set the coefficients corresponding to the atoms in $D_{B\_B}^{\mathrm{LR}}$ in $\hat{\theta}_p^B$ (the solved sparse representation of b_p^B) to zero to obtain $\hat{\theta}_{p\_N}^B$ (the estimated sparse representation of the non-blocking component of b_p^B). As a result, the joint SR and deblocking of b_p^B can be achieved by using the sparse representation $\hat{\theta}_{p\_N}^B$ with respect to $D_{B\_N}^{\mathrm{HR}}$ via (6).

Then, the recovered and deblocked HR patches are tiled together to reconstruct the HR version I_{HF}^N of I_{HF}, where the average of multiple reconstructs is taken for each pixel in the overlapping region as its final recovery. Finally, we add the HR details I_{HF}^N to the bicubic-interpolated I_{LF} to obtain the final SR result of I. To briefly summarize the proposed framework and the relationship between it and the six learned dictionaries, the proposed joint SR and deblocking algorithm is summarized in TABLE II.

E. Extension to Non-Self-Learning-Based SR

Besides the self-learning-based scheme mentioned above, the proposed framework can also be easily integrated into other non-self-learning SR methods as illustrated in Fig. 6. Note that the main difference is that the dictionary used in self-learning SR is learned from the input LR image itself, whereas the non-self-learning SR algorithms (e.g., the sparse-coding SR in [9]) needs to learn a dictionary from a pre-collected training image set. In the extension shown in Fig. 6, each HR image in the training set is first downscaled and compressed to obtain its LR version with blocking artifacts, which are then localized by the following blocking artifact detection, so as to form a HR/LR training pair. Subsequently, the training patch extraction and dictionary learning processes described in Fig. 5 and Sec. III.C are performed on the blocking and non-blocking patch pairs extracted from the training HR/LR image pairs to learn and partition the blocking and non-blocking HR/LR sub-dictionaries. Consequently, based on the learned sub-dictionaries, joint SR and deblocking can be achieved by the proposed image decomposition method described in Fig. 3 and Sec. 3.D. Based on the method shown in Fig. 6, we integrate our framework with the sparse-coding SR (SCSR) in [9], which is a state-of-the-art non-self-learning SR scheme, and will report the results in the following section.

IV. EXPERIMENTAL RESULTS

A. Visual Quality Evaluation

Several LR JPEG images and YouTube videos with blocking artifacts were used to evaluate the performance of the proposed method. All the test dataset and results can be downloaded from our project web site [38]. In our experiments, all of the test LR images were compressed by JPEG with quality factor (QF) ranging from 15 to 25. Different from [15], where the JPEG compression QF is required to be known in advance, our approach does not require any prior knowledge (including QF) about an input LR image. The parameter settings of the proposed method are described as follows. For each test LR image of size ranging from 140×140 to 442×400, the upscaling factor, HR/LR patch sizes, the number of training iterations for dictionary learning, and the size (number of atoms) of each learned dictionary (including D_N^{LR}, D_N^{HR}, D_B^{LR}, and D_B^{HR}) are set to 2, 16×16/8×8, 100, and 1024, respectively. Our framework learns two pairs of dictionaries, respectively, for training patches with and without blocking artifacts, where each dictionary pair contains 512 pairs of HR/LR atoms, resulting in the total dictionary size of 1024. Besides, the regularization parameter λ used in the dictionary learning step in (9) is empirically set to 0.15, and the number of non-zero coefficients to be solved in each sparse reconstruction step in our method is set to 20.

To highlight the advantage of the proposed joint self-learning SR and deblocking approach for a single image, we compare our method with four upscaling methods, including the bicubic interpolation [4], coupled dictionary training-based SR with extra training samples (denoted by coupled-training-based) [10], self-learning SR (denoted by SelfSR, similar to [14]), and a cascading approach including the BM3D deblocking method [35], followed by the SelfSR method (denoted by Cascading-based), where the latter three methods are all based on dictionary learning and their respective dictionary sizes are all set to 1024 for fair comparisons with our method. Note that based on our experiments for all of the compared methods, reducing the dictionary size to some extent (smaller than 1024) will significantly degrade SR results, whereas increasing the dictionary size (larger than 1024) will increase the computational complexity without significant improvement on SR visual quality. Moreover, the number of non-zero coefficients to be solved in each sparse reconstruction step in the three compared methods is also set to 20. Here, the SelfSR method used for comparison is based on our framework without performing blocking artifact removal, which mainly relies on the dictionary learning based on HR/LR training patch pairs proposed in [9], [10] and the employed self-similarity characteristic of an input LR image proposed in [14].

Fig. 7 and Fig. 8 show some SR results obtained by the proposed method and the compared methods. The enlarged images in Fig. 7 and Fig. 8 illustrate that the visual qualities of our SR results are visually superior to the results obtained by the pure SR approaches, which show significantly magnified blocking artifacts. Moreover, our method also outperforms the cascading approach based on the fact that the application of deblocking prior to SR would inevitably lead to loss of some image details, which is usually hard to be recovered in the following SR stage.

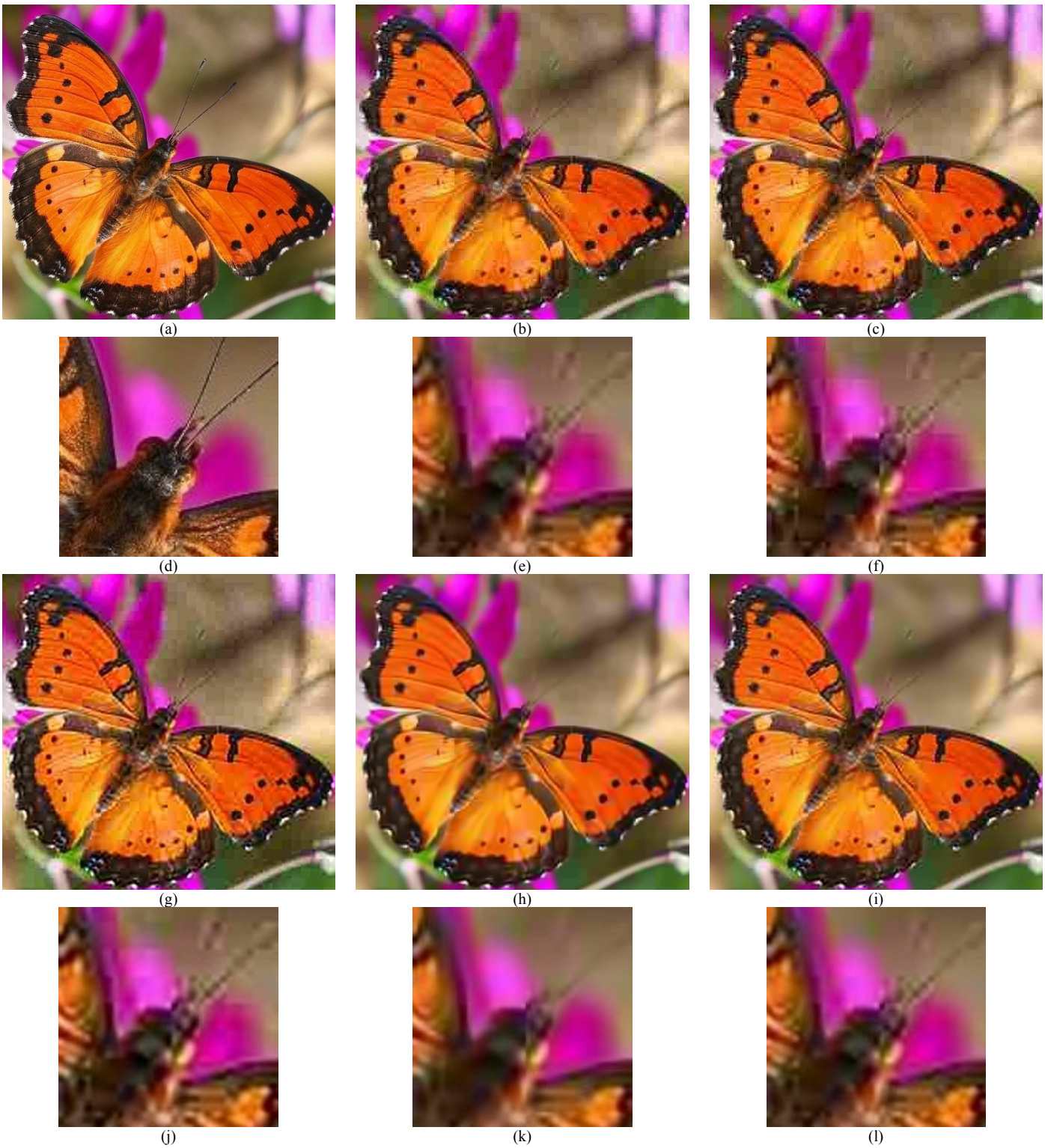

Fig. 7. SR results: **(a)** the original HR image; and the SR versions of the input LR image (LR and compressed version of (a), QF = 15) via the: **(b)** bicubic [4]; **(c)** Coupled-training-based [10]; **(d)** local highlight of (a); **(e)** local highlight of (b); **(f)** local highlight of (c); **(g)** SelfSR [14]; **(h)** Cascading-based; **(i)** proposed method; **(j)** local highlight of (g); **(k)** local highlight of (h); and **(l)** local highlight of (i).

We then further evaluate the performance of the proposed self-learning-based method on some real-world low-quality LR images/video frames with significant blocking artifacts, downloaded from the Internet (e.g., from YouTube [27]). These testing data might be downscaled by any down-sampling operations and highly compressed via any block-based compression algorithm. Fig. 9 shows some snapshots of the SR results obtained by the proposed method and the bicubic interpolation method [4] for a YouTube video. It can be similarly observed from Fig. 9 that the visual quality of our SR results is significantly better than the results obtained by the bicubic interpolation, especially for very low-quality

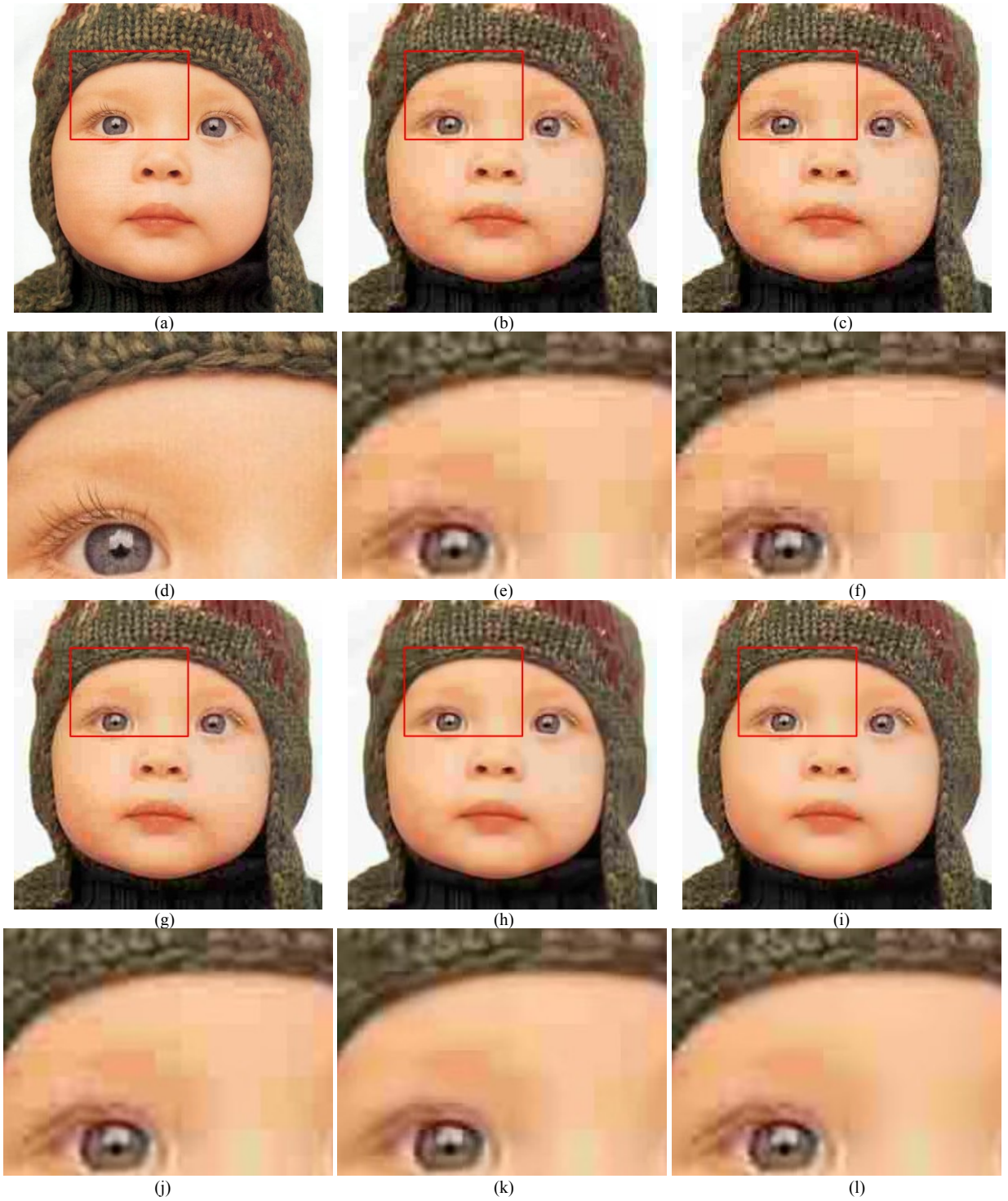

Fig. 8. SR results: **(a)** the original HR image; and the SR versions of the input LR image (LR and compressed version of (a), QF = 15) via **(b)** bicubic [4]; **(c)** Coupled-training-based [10]; **(d)** local highlight of (a); **(e)** local highlight of (b); **(f)** local highlight of (c); **(g)** SelfSR [14]; **(h)** Cascading-based; **(i)** proposed method; **(j)** local highlight of (g); **(k)** local highlight of (h); and **(l)** local highlight of (i).

image/video frames. More test results on some videos downloaded from YouTube can be found in our project web site [38], where all the source inks are provided. Due to the space limit, we do not show the results of the non-self-learning-based schemes (see [38] for the results).

We also conduct a subjective user study to evaluate the performances of various SR schemes (with/without deblocking) for 9 test images [38]. In the subjective paired comparisons test [39], 20 subjects are shown with two side-by-side upscaled images obtained using different SR methods (in a random order) at a time, and are asked to choose their preference from the two

(a) (b) (c)

Fig. 9. SR results: **(a)** the original low-quality video frame obtained from the Internet; and the SR versions of (a) via **(b)** bicubic [4]; and **(c)** the proposed method.

compared images. The subjects include 15 males and 5 females, whose ages range from 25 to 30. The test device is a full-HD 24-inch LCD display with color temperature 6500K. In the subjective test, we compare the proposed method with the original SR method [self-learning SR (SelfSR) or sparse-coding SR (SCSR)] without deblocking, Cascade type I (SR followed by deblocking), Cascade type II (deblocking followed by SR) for 9 test images. Each SR method is pairwisely compared with the others by totally $3 \times 9 \times 20 = 540$ times, meaning that 180 comparisons are made between two methods for the 9 test images. The ratio between the preference vote for one method and that for another method is used to compare the performances between the two methods. For example, if one method is preferred 80% of the time, the SR versions of all test images obtained by this method receive 80% of the preference vote, indicating that the method is visually preferred to the compared method in general.

TABLE III and TABLE IV show the winning frequency matrices [39] of the 20 subjects' preferences in subjective paired comparisons on the SR results obtained using the four

TABLE III
WINNING FREQUENCY MATRIX OF SUBJECTIVE PAIRED COMPARISONS BETWEEN SELF-LEARNING SR (SELFSR), BM3D+ SELFSR, SELFSR+BM3D AND THE PROPOSED JOINT SELFSR+DEBLOCKING

	Proposed SelfSR	Cascade I	Cascade II	SelfSR	Average
Proposed SelfSR	–	**146 (81%)**	**146 (81%)**	**168 (93%)**	**85.2%**
Cascade I (SelfSR+BM3D)	34 (19%)	–	88 (49%)	139 (77%)	48.3%
Cascade II (BM3D+SelfSR)	34 (19%)	92 (51%)	–	137 (76%)	48.7%
SelfSR	12 (7%)	41 (23%)	43 (24%)	–	17.8%

TABLE IV
WINNING FREQUENCY MATRIX OF SUBJECTIVE PAIRED COMPARISONS BETWEEN SPARSE CODING SR (SCSR), BM3D+ SCSR, SCSR+BM3D AND THE PROPOSED JOINT SCSR+DEBLOCKING

	Proposed SCSR	Cascade I	Cascade II	SCSR	Average
Proposed SCSR	–	**127 (71%)**	**148 (82%)**	**170 (94%)**	**82.4%**
Cascade I (SCSR+BM3D)	53 (29%)	–	91 (51%)	147 (82%)	53.9%
Cascade II (BM3D+SCSR)	32 (18%)	89 (49%)	–	143 (79%)	48.9%
SCSR	10 (6%)	33 (18%)	37 (21%)	–	14.8%

TABLE V
WINNING FREQUENCY MATRIX OF SUBJECTIVE PAIRED COMPARISONS BETWEEN SCSR, SELFSR, THE PROPOSED JOINT SCSR+DEBLOCKING AND JOINT SELFSR+DEBLOCKING

	Proposed SelfSR	Proposed SCSR	SelfSR	SCSR	Average
Proposed SelfSR	–	**117 (65%)**	**168 (93%)**	**171 (95%)**	**84.4%**
Proposed SCSR	63 (35%)	–	159 (88%)	170 (94%)	72.69%
SelfSR	12 (7%)	21 (12%)	–	100 (56%)	24.6%
SCSR	9 (5%)	10 (6%)	80 (44%)	–	18.3%

different schemes based on the SelfSR and SCSR, respectively, where the numbers in each row indicate the times of the method preferred than the compared in the column. Both tables indicate that, compared to the other schemes, our method achieves better subjective visual quality in 78%~88% and 80%~92% of the paired comparisons based on the two different SR frameworks, respectively. Moreover, TABLE V lists the subjective paired comparison results of the SelfSR, SCSR, Proposed SelfSR (joint SelfSR+Deblocking), and Proposed SCSR (joint SCSR+deblocking) methods. the results shows that, the proposed SelfSR is preferred to the proposed SCSR subjectively in 64% of paired comparisons, and significantly outperforms the original SelfSR and SCSR methods in terms of subjective visual quality.

Based on our experimental results for individual test images (as detailed in [38]), the performances of the proposed methods are dependent on the content characteristics (edges, textures, and smooth regions) of input LR image, as summarized below:

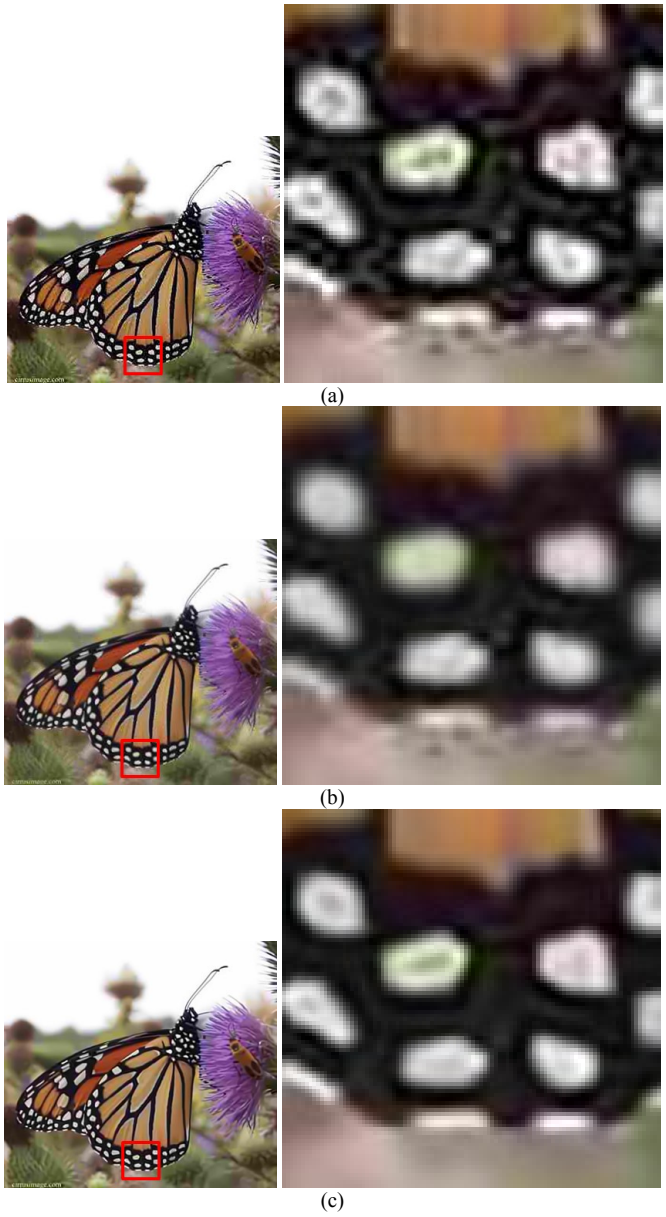

Fig. 10. SR results obtained by **(a)** the proposed method without de-ringing; **(b)** the proposed SR+Deblocking+Deringing method; and **(c)** the proposed SR+Deblocking method followed by BM3D filtering [34] to remove ringing artifacts (with some details blurred).

1. Compared to SR without deblocking (i.e., the original SCSR and SelfSR), our methods are significantly more effective for images with strong edges, textures, and smoothing regions.
2. Compared to SR with deblocking, our methods achieve significant visual quality improvement for image regions with strong edges and textures. As for smooth regions, since blocking artifacts in such regions can be effectively mitigated using the deblocking part of the cascading-based approaches as well, our methods achieve comparable visual quality with the cascading-based approaches.
3. In general, our methods achieve overall visual quality improvement over existing SR methods with or without deblocking.

B. Limitations

In our SR framework, we focus on performing SR for a highly compressed image while removing the visible blocking artifacts of the image. Nevertheless, the ringing artifact, not considered in this work, also usually occurs in a compressed image. Note, de-ringing can also be incorporated into the proposed MCA-based framework by using ringing artifact detection (e.g., [37]) to localize the ringing artifacts to learn the HOG characteristics of ringing artifacts and then use the learned characteristics to further partition an additional ringing sub-dictionary in the structure shown in Fig. 3(g). However, our experimental results show that, although the proposed framework can still mitigate ringing artifacts to some extent, the de-ringing performance is not satisfactory. Fig. 10 illustrates an example of ringing artifacts removal based on the proposed framework where Fig. 10(a) shows an image after joint SR and deblocking (without de-ringing) and Fig. 10(b) shows the version after joint SR, deblocking and de-ringing using the proposed framework. We can observe from Fig. 10(b) that the ringing artifacts, although being less severe compared to Fig. 10(a), are still clearly visible. One of the main reasons is that the structure of ringing artifact is usually irregular. Hence, it is not easy to well capture the characteristics of ringing artifact by the HoG features and automatically self-learn a pair of dictionaries of ringing atoms. Besides, ringing detection is not accurate enough, making it difficult to extract enough good exemplar patches for learning. To achieve better visual quality of a reconstructed image, performing a post-processing step to remove the ringing artifacts might be a choice, which would, however, also blur some image details, as exemplified in Fig. 10(c). For future work, integrating de-ringing into our SR framework to achieve satisfactory de-ringing performance still needs further study.

C. Computational Complexity

The proposed method was implemented in 32-bit MATLAB® on a personal computer equipped with Intel® Core™ i7-4470 processor and 32 GB memory. TABLE VI lists the run-time of individual steps, including the dictionary learning and sparse reconstruction, for the test image shown in Fig. 7 based on the SelfSR approach and the proposed SelfSR method, respectively. In the dictionary learning, our method learns an additional pair of dictionaries (either for training patches with or without blocking artifacts separately), whereas the SelfSR method learns only one pair of dictionaries. In our implementation, the size of each dictionary pair is set to 512 (total 1024 in size of two pairs of dictionary). The computational complexity for learning each dictionary pair of size, 512, is significantly lower than that for learning a dictionary pair of size, 1024, used in the SelfSR method. Based on TABLE VI, the complexity of sequentially learning two pairs of dictionaries in our method is still lower than that for learning a single dictionary pair of size 1024 used in the SelfSR method. Moreover, the complexity of our method can be further reduced by learning two pairs of dictionaries in parallel should a multi-core processor be used.

TABLE VI
RUN-TIME (IN SECONDS) OF THE DICTIONARY LEARNING AND SPARSE RECONSTRUCTION PHASES IN THE SELFSR METHOD AND THE PROPOSED SELFSR METHOD

Evaluated methods	Dictionary learning	Sparse reconstruction	Total time
SelfSR	229.14 s	31.32 s	260.46 s
Proposed SelfSR	123.64 s	30.10 s	153.74 s

TABLE VII
RUN-TIME (IN SECONDS) OF THE DICTIONARY LEARNING AND SPARSE RECONSTRUCTION PHASES IN THE SCSR METHOD AND THE PROPOSED SCSR METHOD

Evaluated methods	Dictionary learning (Offline)	Sparse reconstruction
SCSR	45541 s	123.7 s
Proposed SCSR	42331 s	121.9 s

On the other hand, the complexity of sparse reconstruction phase mainly relies on the dictionary size and the number of non-zero coefficients to be solved which is set to be the same for all evaluated dictionary learning-based methods. Similarly, the complexity of the sparse reconstruction required for the two pairs of learned dictionaries of size 512 for each in our method is lower than that of the SelfSR method with a dictionary size of 1024. The sparse reconstruction operations with respect to the two dictionary pairs in our method can be also performed in parallel to further reduce the complexity.

TABLE VII compares the run-time complexities of the SCSR and the proposed SCSR+deblocking methods. Both the two methods need to train dictionaries from an external training dataset. Therefore, their dictionary leaning is an offline process which consumes significantly higher complexity than that of SelfSR-based schemes, whereas the sparse reconstruction is performed on-line. TABLE VII shows that our method is still of lower time complexity, compared to the original SCSR.

The other overheads in the learning phase of our methods include the blocking artifact detection and dictionary partitioning, whose complexities are usually insignificant. By properly fusing the cores of the SR and deblocking operations into a unified MCA-based framework, the proposed framework can simultaneously achieve both SR and deblocking while reducing computational complexity, compared with a pure learning-based SR scheme. Moreover, the computational complexity of our method is obviously lower than that of cascading-based approach, if a similar SR scheme is used.

V. CONCLUSION

In this paper, we proposed a learning-based SR framework to achieve joint single-image SR and deblocking for image sparse representation for modeling the relationship between LR and HR image patches in terms of the learned dictionaries, respectively, for image patches with and without blocking artifacts. As a result, image SR and deblocking can be simultaneously achieved via sparse representation and MCA-based image decomposition. Our experimental results demonstrate the efficacy of the proposed algorithm on self-learning SR and sparse-coding SR. Furthermore, our method can be naturally extended to remove other types of noise or structured patterns while performing SR for images.

REFERENCES

[1] M. Y. Shen and C. C. J. Kuo, "Review of postprocessing techniques for compression artifacts removal," *J. Vis. Commun. Image Rep.*, vol. 9, no. 1, pp. 2–14, Mar. 1998.

[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21−36, May 2003.

[3] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327−1344, Oct. 2004.

[4] H. S. Hou and H. C. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Trans. Signal Process.*, vol. 26, no. 6, pp. 508–517, Dec. 1978.

[5] 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.

[6] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sept. 2009, pp. 349−356.

[7] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, Article 12, 2011.

[8] C.-C. Hsu, L.-W. Kang, and C.-W. Lin, "Temporally coherent super-resolution of textured video via dynamic texture synthesis," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 919−931, Mar. 2015.

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

[10] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. S. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467−3478, Aug. 2012.

[11] 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, July 2011.

[12] J. Ren, J. Liu, and Z. Guo, "Context-aware sparse decomposition for image denoising and super-resolution," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1456−1469, Apr. 2013.

[13] C.-Y. Yang, J.-B. Huang, and M.-H. Yang, "Exploiting self-similarities for single frame super-resolution," in *Proc. Asian Conf. Comput. Vis.*, Queenstown, New Zealand, Nov. 2010, pp. 497−510.

[14] M.-C. Yang and Y.-C. F. Wang, "A self-learning approach to single image super-resolution," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 498–508, April 2013.

[15] Z. Xiong, X. Sun, and F. Wu, "Robust web image/video super-resolution," *IEEE Trans. Image Process.*, vol. 19, no. 8, pp. 2017−2028, Aug. 2010.

[16] T. Q. Pham, L. J. van Vliet, and K. Schutte, "Resolution enhancement of low quality videos using a high-resolution frame," in *Proc. SPIE*, San Jose, CA, USA, Jan. 2006.

[17] F. Liu, J. Wang, S. Zhu, M. Gleicher, and Y. Gong, "Noisy video super-resolution," in *Proc. ACM Multimedia*, Vancouver, Canada, Oct. 2008, pp. 713−716.

[18] T. Yamaguchi, H. Fukuda, R. Furukawa, H. Kawasaki, and P. F. Sturm, "Video deblurring and super-resolution technique for multiple moving objects," in *Proc. Asian Conf. Comput. Vis.*, Queenstown, New Zealand, Nov. 2010, pp. 127−140.

[19] C. Liu and D. Sun, "A Bayesian approach to adaptive video super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, USA, June 2011, pp. 209−216.

[20] G. A. Triantafyllidis, D. Tzovaras, and M. G. Strintzis, "Blocking artifact detection and reduction in compressed data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 877−890, Oct. 2002.

[21] P. List, A. Joch, J. Lainema, G. Bjontegaard and M. Karczewicz, "Adaptive deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 614−619, July 2003.

[22] C.-H. Yeh, Y.-W. Chiou, L.-W. Kang, C.-W. Lin, and S.-J. Fan-Jiang, "Self-learning-based post-processing for image/video deblocking via sparse representation," *J. Visual Commun. Image Represent.*, vol. 25, no. 5, pp. 891−903, July 2014.

[23] L.-W. Kang, B.-C. Chuang, C.-C. Hsu, C.-W. Lin, and C.-H. Yeh, "Self-learning-based single image super-resolution of a highly compressed image," in *Proc. IEEE Workshop Multimedia Signal Process.*, Sardinia, Italy, Sept. 2013, pp. 224−229.
[24] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30−44, April 1991.
[25] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560−576, July 2003.
[26] G. J. Sullivan, J.-R. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649−1668, Dec. 2012.
[27] YouTube [online] http://www.youtube.com/
[28] J. L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.
[29] J. M. Fadili, J. L. Starck, J. Bobin, and Y. Moudden, "Image decomposition and separation using sparse representations: an overview," *Proc. IEEE*, vol. 98, no. 6, pp. 983–994, June 2010.
[30] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742−1755, Apr. 2012.
[31] D.-A. Huang, L.-W. Kang, Y.-C. F. Wang, and C.-W. Lin, "Self-learning based image decomposition with applications to single image denoising," *IEEE Trans. Multimedia* vol. 16, no. 1, pp. 83−93, Jan. 2014.
[32] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 13, pp. 607–609, June 1996.
[33] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.
[34] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080−2095, Aug. 2007.
[35] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, June 2005, pp. 886–893.
[36] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
[37] H. Lim and H. W. Park, "A ringing-artifact reduction method for block-DCT-based image resizing," *IEEE Trans. Circuits Syst. Video Techn.* vol. 21, no. 7, pp. 879–889, July 2011.
[38] NTHU Image/Video SR+Denoising Project. [Online]. Available: http://www.ee.nthu.edu.tw/cwlin/SR_Denoising/index.html.
[39] C.-C. Hsu, C.-W. Lin, Y. Fang, and W. Lin, "Objective quality assessment for image retargeting based on perceptual geometric distortion and information loss," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 377−389, June 2014.

Li-Wei Kang (S'05-M'06) received his B.S., M.S., and Ph.D. degrees in Computer Science from National Chung Cheng University, Chiayi, Taiwan, in 1997, 1999, and 2005, respectively.

Since February 2013, he has been with the Graduate School of Engineering Science and Technology-Doctoral Program, and the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan, as an Assistant Professor. Before that, he worked for the Institute of Information Science, Academia Sinica, Taipei, Taiwan, as an Assistant Research Scholar during 2010–2013, and as a postdoctoral research fellow during 2005–2010. His research interests include multimedia content analysis and multimedia communications.

Dr. Kang has served as the Editor-in-Chief of the Gate to Multimedia Processing (Science Gate Publishing). He served on the editorial board of the International Journal of Distributed Sensor Networks, an Editorial Advisory Board Member of a book entitled Visual Information Processing in Wireless Sensor Networks: Technology, Trends and Applications (IGI Global), a Guest Editor of the International Journal of Electrical Engineering (Taiwan), Special Session Co-Chair of APSIPA ASC 2012, and Registration Co-Chair of APSIPA ASC 2013. He serves as Demo/Exhibition Co-Chair of IEEE ICCE-TW 2015. He received a top 10% Paper Award from IEEE MMSP 2013.

Chih-Chung Hsu received his B.S. degree in Information management from Ling-Tung University of Science and Technology, Taiwan, in 2004, and M.S. and Ph.D. degrees in Electrical Engineering from National Yunlin University of Science and Technology and National Tsing Hua University (NTHU), Taiwan, in 2007 and 2014, respectively.

Dr. Hsu is a postdoctoral researcher with the the Institute of Communications Engineering, NTHU. His research interests mainly lie in computer vision, and image and video processing. He received a top 10% paper award from IEEE MMSP 2013.

Boqi Zhuang received his B.S. degree in Communication Engineering from National Central University, Taoyuan, Taiwan, in 2011, and M.S. degree in Electrical Engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2013.

Since August 2013, he has been with Altek Corporation, Hsinchu, Taiwan, as a software engineer. His research interests include image/video processing and compression. He received a top 10% paper award from the IEEE MMSP 2013.

Chia-Wen Lin (S'94-M'00-SM'04) received his Ph.D. degree in electrical engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2000.

He is currently a Professor with the Department of Electrical Engineering and the Institute of Communications Engineering, NTHU. He is also an Adjunct Professor with the Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan. He was with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, during 2000–2007. Prior to that, he worked for the Information and Communications Research Labs, Industrial Technology Research Institute, Hsinchu, Taiwan, during 1992–2000. His research interests include image/video processing and video networking.

Dr. Lin has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE MULTIMEDIA, and the *Journal of Visual Communication and Image Representation*. He also serves as a Steering Committee member of the IEEE TRANSACTIONS ON MULTIMEDIA. He is currently Chair of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He served as Technical Program Co-Chair of the IEEE ICME in 2010, and Special Session Co-Chair of IEEE ICME in 2009. He is Distinguished Lecturer of Asia-Pacific Signal and Information Processing Association (APSIPA). His paper won the Young Investigator Award presented by VCIP 2005. He received the Young Investigator Awards presented by National Science Council, Taiwan, in 2006.

Chia-Hung Yeh (M'03-SM'12) received his B.S. and Ph.D. degrees from National Chung Cheng University, Taiwan, in 1997 and 2002, respectively, both in electrical engineering.

He joined the Department of Electrical Engineering, National Sun Yat-sen University (NSYSU) in 2007, where he is a Professor. His research interests include multimedia communication and multimedia signal processing.

Dr. Yeh serves on the Editorial Board of the Journal of Visual Communication and Image Representation, and the EURASIP Journal on Advances in Signal Processing. In addition, he has organized several international/domestic conferences serving as keynote speaker, session chair, and technical program committee and program committee member. He has co-authored more than 150 technical international conferences and journal papers and holds 42 patents in the U.S., Taiwan, and China. He received the 2007 Young Researcher Award of NSYSU, the 2011 Distinguished Young Engineer Award from the Chinese Institute of Electrical Engineering, the 2013 Distinguished Young Researcher Award of NSYSU, and the 2013 IEEE MMSP Top 10% Paper Award.