IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 17, 2019, accepted July 12, 2019, date of publication July 18, 2019, date of current version September 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929773

Learning Shared and Cluster-Specific Dictionaries for Single Image Super-Resolution

TINGTING YAO[1], ZHIYONG WANG[2], YU LUO[1], YUE LIANG[1], QING HU[1], AND DAVID DAGAN FENG[2]

[1]College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
[2]School of Computer Science, The University of Sydney, Sydney, NSW 2006, Australia

Corresponding author: Tingting Yao (ytt1030@dlmu.edu.cn)

This work was partially supported by National Natural Science Foundation of China (No. 31700742) and Fundamental Research Funds for the Central Universities.

ABSTRACT Recently, many multi-dictionary-based sparse representation methods have been proposed for single-image super-resolution (SISR) by learning a sub-dictionary for each specific type of visual content (i.e., a cluster). Although promising reconstruction results have been achieved for certain scenarios, due to the similarity shared among different types of visual content, each sub-dictionary is not learned as discriminatively as expected, which could compromise the visual quality of reconstructed high-resolution (HR) images. In this paper, we propose a novel shared and cluster-specific dictionaries learning method for SISR, where cluster-specific dictionaries can be learned as separate as possible and a dictionary shared among clusters is learned explicitly to explore similarity among clusters. First, low-resolution (LR) and HR image patches extracted from training images are jointly grouped into visual clusters. Then, a shared sub-dictionary and a set of cluster-specific sub-dictionaries are learned simultaneously under the sparse representation framework. In addition, the group sparsity constraint, the locality constraint, and the incoherence penalty term are incorporated into a unified framework to preserve the relationship and structure among the training data in dictionary learning. Finally, an anchored neighborhood regression method is devised to pre-calculate a projection matrix of each learned dictionary atom from LR to its HR space so that an HR image can be more efficiently recovered in the reconstruction stage. Comprehensive experimental results on two widely used benchmark datasets demonstrate the effectiveness and robustness of our method against a number of state-of-the-art methods. The reconstruction results of real-world maritime surveillance images further indicate that the proposed method is well suitable for practical applications.

INDEX TERMS Dictionary learning, sparse representation, anchored neighborhood regression, single image super-resolution.

I. INTRODUCTION

High-resolution (HR) images are helpful for a wide range of applications such as security surveillance and image based medical diagnosis. However, image resolution is often practically constrained by the limitation of image acquisition and transmission systems. Therefore, single image super-resolution (SISR) has been investigated to recover a high resolution version of a low resolution (LR) input image [1], [2]. Due to the ill-posed nature, this task has been very challenging.

Inspired by intrinsic sparse property of natural images, compressive sensing-based methods have attracted lots of attentions in recent years [3]–[5]. The sparsity-based regularization terms have been introduced and HR image patches could be recovered by sparse coding coefficients calculated from the counterparts of an input LR image using the learned dictionary. As the estimation of an HR image depends on a learned dictionary, an appropriate dictionary is critical for deriving high quality recovery results.

Conventional compressive sensing-based SISR methods usually learn an over-complete dictionary from training data with different constraints [6], [7]. As the contents of training images vary significantly, the learned dictionary atoms are usually not directly relevant to image patches, which may compromise the recovery results. In order to improve the representation capacity of a learned dictionary, multiple sub-dictionaries based learning frameworks have been

The associate editor coordinating the review of this manuscript and approving it for publication was Qingchun Chen.

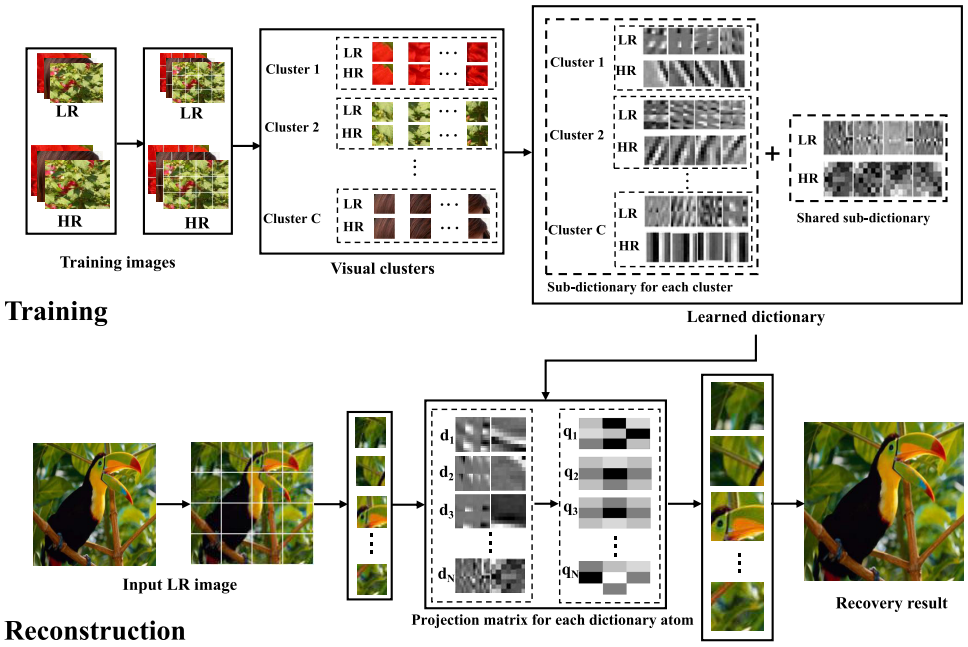


FIGURE 1. Illustrative diagram of our proposed shared and cluster-specific dictionary learning based SISR method.

proposed [8], [9]. The patches with similar structure are clustered into different subspaces and a sub-dictionary for each cluster is learned separately. Many efforts have been devoted to learning more representative sub-dictionaries by incorporating distinctive image characteristics, such as edge sharpness and directional orientations [10], [11]. However, different sub-dictionaries often share a large number of similar atoms, which could lead to unstable dictionary learning and deteriorate reconstruction results.

Recently, since super-resolution convolutional neural networks (SRCNN) have achieved superior performance over traditional sparse coding methods, neural networks have been widely utilized for super resolution. While deeper networks usually learn higher-level features for high quality image reconstruction, they are computationally expensive to train with a large number of parameters. Furthermore, for practical SISR applications in the real world, the reconstruction results of deep networks would be compromised due to limited training data.

To address the above-mentioned issue, in this paper, a novel shared and cluster-specific dictionary learning based SISR method is proposed. As shown in Fig.1, firstly, the LR and HR image patches extracted from training images are jointly clustered into groups. Secondly, a set of sub-dictionaries for individual clusters are learned together with a sub-dictionary shared across all the clusters. That is, instead of independently learning a sub-dictionary for each cluster, our proposed dictionary learning method exploits the inter-relationship among different clusters for deriving a more discriminant dictionary. As a result, the non-local self-similarity of natural images could be more effectively characterized and the unique representation of each sub-dictionary will be enhanced simultaneously. Three constraints including group sparsity constraint, locality constraint and incoherence penalty term are further incorporated in a unified framework to preserve the relationship and structure among training data in our proposed dictionary learning method. Finally, in order to reduce the complexity of HR image reconstruction, the corresponding projection matrix of each dictionary atom from LR to its HR space is pre-calculated with anchored neighborhood regression approach. Hence, instead of directly computing the sparse representation in reconstruction stage, an HR image could be estimated with ridge regression efficiently. Our proposed method could effectively improves SISR performance under different noisy conditions with only a few of training data, which is well suitable for practical applications in maritime domain.

The rest of this paper is organized as follows. Section II reviews the related works of learning-based SISR methods and specifically compressive sensing-based SISR methods. Section III explains the proposed dictionary learning framework in detail. Section IV presents experimental results and discussions in comparison with several state-of-the-art methods to evaluate the effectiveness and robustness of our proposed method. The reconstruction results of real-world images automatically extracted from a maritime surveillance system further indicate the application of our proposed method. Section V concludes this work with discussions on our future research.

II. RELATED WORK

Generally, the learning-based SISR methods could be divided into three major categories: compressive sensing based methods, example based methods, and neighbor embedding based methods [12].

Note that, due to the ground breaking success of deep learning techniques on various vision tasks, a lot of deep learning based SISR methods have also been proposed with favorable results. For example, Dong et al. proposed a convolutional neural network to estimate HR images through an end-to-end mapping [13]. Extensions such as introducing prior information and increasing the depth of networks have also been proposed to further improve SISR performances [14]–[16]. Although impressive improvements have been achieved, these methods are usually data and computation intensive which make them not well suitable for practical applications. In this section, we mainly review the non deep learning-based methods, specifically compressive sensing-based SISR which is related to our method.

A. COMPRESSIVE SENSING-BASED SISR

Motivated by the sparse property of natural images, many compressive sensing-based methods have been proposed. Some of them focus directly on obtaining a more accurate sparse representation to produce better reconstruction results. Li et al. proposed a hybrid parametric sparse model which learned the sparse prior of HR images from both training set and input LR images [1]. Zhao et al. proposed an adaptive sparse coding method which considered both sparsity and correlation in representation to produce more suitable sparse coefficients for image recovery [5]. Gu et al. utilized convolutional filters to decompose an input LR image into patches, and then reconstructed the HR image with predicted feature mapping through filters, which enforced the consistency between image patches [17]. As the sparse representation of an HR image is calculated based on dictionary atoms, the recovery results could be improved with more accurate dictionaries.

There have been many efforts to learn a more effective dictionary from training data. Ding et al. proposed a convex dictionary learning method with convex constraint, so that the dictionary atoms could be better formed by a linear combination of training data [4]. Yang et al. proposed a coupled dictionary learning method which optimized the LR and HR dictionaries jointly [7]. As a result, more accurate reconstruction results were achieved with a better representation of LR and HR dictionaries. Zhang et al. proposed an algorithm that dual dictionaries were utilized to reconstruct the main and residual high-frequency components of an HR image, respectively [18]. Although more accurate dictionaries can be learned, the adaptability of local structure has not been taken into account in the above mentioned dictionary learning frameworks.

In recent years, there have been increasing interests on learning multiple sub-dictionaries to better incorporate the non-local self-similarity of training data. Dong et al. adopted K-means algorithm to partition training data into different clusters and learned a sub-dictionary for each cluster. The adaptive sparse domain selection and two adaptive regularization terms were further introduced to improve quality of recovered image [19]. Yang et al. clustered image patches into several groups and learned a geometric dictionary for each group. A clustering aggregation and patch aggregation were further introduced to improve the results of HR image recovery [20]. Multiple characterizations and measurements were also employed to obtain more accurate clustering results of training images. Yeganli et al. employed scale-invariant gradient-based sharpness measure to classify the training image patches into different clusters [21]. Sub-dictionary for each cluster was learned and sparse coefficients over selected sub-dictionary atoms were obtained for reconstruction. Yang et al. utilized non-local similarity to cluster image patches, and orthogonal sub-dictionaries were learned for better reconstruction results [9]. Ahmed et al. proposed a method to classify training images in terms of edge sharpness, and a mapping function of each sub-dictionary was formulated to calculate the projection between LR and HR spaces [10]. Although more complicated representations have been calculated with sub-dictionaries, the similar patterns shared among different clusters have been overlooked, which could compromise representation capability of learned dictionary.

B. EXAMPLE BASED AND NEIGHBOR EMBEDDING BASED SISR

The concept of example-based method was proposed by Freeman et al which learned the prior relationship among HR and LR image patches with Markov Random Field. Then, the missing high frequency details in HR image patches were recovered with the searched matching patches from training set [22]. Since then, many methods have been proposed to learn correlation between HR and LR image patches with well chosen training images. Yue et al. extracted features from external datasets to better characterize the neighboring and non-local information with a hybrid neural network [23]. To further exploit the self-similarity of images, Huang et al. employed 3D scene geometry and patch search space expansion to improve the recovery results with internal datasets [24]. Recently, the mapping models using both internal and external datasets were also proposed. Wang et al. utilized an adaptive weight to combine two loss functions of external and internal example-based approaches, and more accurate results were obtained [25]. Although the relationships between patches could be well modeled in these methods, the quality of reconstructed images highly relied on the similarity between input LR images and training data.

To solve the bottleneck of computational complexity and memory requirements of SISR, neighbor embedding-based methods have been proposed. Such methods were based on the manifold learning theory, which suggested that the HR and LR patches could form manifolds with similar local geometry and neighborhood relationship in two distinct feature spaces. Timofte et al. introduced anchored neighborhood regression and its variant to precompute the corresponding projection matrix of the learned dictionary to improve the computational efficiency of SR reconstruction [26], [27]. Sun et al. divided the training patches into groups and learned the mapping matrix for each group. With the partially supervised

strategy, the class information was considered for mapping matrix matching [28]. Although, different mapping functions have been proposed to obtain a more appropriate transformation between HR and LR spaces, the representation capacity of dictionary atoms is usually overlooked in the process of calculating projection matrix.

III. PROPOSED METHOD

In this section, firstly, the basics of compressive sensing-based SISR is introduced. Then, the detailed process of our proposed shared and cluster-specific dictionary learning based SISR is presented in terms of dictionary modeling, dictionary learning, and image reconstruction via neighborhood regression.

A. BASICS OF COMPRESSIVE SENSING-BASED SISR

Let z_{hi} denote the representation of the i-th image patch extracted from an HR training image. According to compressive sensing theory, z_{hi} can be sparsely represented with the learned HR dictionary D_h as: $z_{hi} \approx D_h\alpha_{hi}$, where α_{hi} denotes the sparse coefficient with only a few non-zero elements. Similarly, let z_{li} be the corresponding patch extracted from the LR version of this image. Then, the LR counterpart z_{li} can be also sparsely represented with the learned LR dictionary D_l in the same way as: $z_{li} \approx D_l\alpha_{li}$.

In general, the relationship between an HR image and its corresponding LR counterpart can be modeled by blurring and down-sampling operator ψ as: $z_{li} \approx \Psi z_{hi}$. If the dictionary D_h and D_l are learned jointly, based on the scale invariance assumption between sparse coefficients of α_{hi} and α_{li}, dictionaries D_h and D_l can be also related by the same operator as: $D_l \approx \Psi D_h$ [29]. Then, the LR image patch z_{li} could be approximated as follows:

$$z_{li} \approx \Psi z_{hi} \approx \Psi D_h\alpha_{hi} \approx D_l\alpha_{hi}. \tag{1}$$

According to Eq. (1), it can be concluded that $\alpha_{hi} \approx \alpha_{li}$. Thus, given an observed LR image patch z_{li}, with learned dictionaries D_h, D_l, the recovered HR image patch $\hat{z}_{hi}$ could be approximated as follows:

$$\begin{aligned} \hat{z}_{hi} &\approx D_h\alpha_i, \\ \alpha_i &= \arg\min_{\alpha_i} \left\{ ||z_{li} - D_l\alpha_i||_2^2 + \lambda||\alpha_i||_1 \right\}, \end{aligned} \tag{2}$$

where α_i denotes the sparse coefficient and λ is a weighting factor which balances the effect of sparsity regularization term. $l1$ norm as $||\alpha_i||_1$ is the most commonly used constraint due to its effectiveness and simplicity.

For dictionary learning, the HR and LR image features extracted from a set of training data are usually jointly trained to preserve the intrinsic relationship between LR and corresponding HR counterpart as follows:

$$\arg\min_{D,\alpha_i} \sum_{i=1}^{M} \left\{ ||z_i - D\alpha_i||_2^2 + \lambda||\alpha_i||_1 \right\}, \tag{3}$$

where $D = \begin{bmatrix} D_l \\ D_h \end{bmatrix}$, $z_i = \begin{bmatrix} z_{li} \\ z_{hi} \end{bmatrix}$ and M is the total number of extracted image patches.

In order to improve the quality of recovered images, multiple sub-dictionaries based methods have been further proposed. The patches with similar structures are clustered into the same group or cluster, and a sub-dictionary D^c is learned for each cluster c as follows:

$$\arg\min_{D^c,\alpha_i} \sum_{c=1}^{C} \sum_{i\in I_c} \left\{ ||z_i - D^c\alpha_i||_2^2 + \lambda||\alpha_i||_1 \right\}, \tag{4}$$

where $D^c = \begin{bmatrix} D_l^c \\ D_h^c \end{bmatrix}$, I_c denotes a set of data in the c-th cluster and C denotes the total number of clusters. The sparse coding coefficient or sparse code α_i and sub-dictionary D^c in the objective function Eq. (4) could be iteratively calculated via many optimization methods [6].

Nevertheless, the similar representation shared among different sub-dictionaries will lead unstable sparse recovery, which may eventually deteriorate the reconstruction results.

B. SHARED AND CLUSTER-SPECIFIC DICTIONARY LEARNING BASED SISR

1) DICTIONARY MODELING

To address the above issue, a shared and cluster-specific dictionary has been introduced to improve the recovery quality. For a given training dataset, the images are divided into overlapped patches with size $s * s$. Similar to [29], four filters are applied to extract the first and second order gradient maps of patches as the patch-based representation of images. Then, K-means algorithm is carried out to cluster the training data into C groups. To enhance the modeling capability of learned dictionary, a shared sub-dictionary $D^s = \begin{bmatrix} D_l^s \\ D_h^s \end{bmatrix}$ is learned together with a set of sub-dictionaries $D^c = \begin{bmatrix} D_l^c \\ D_h^c \end{bmatrix}$ for individual clusters simultaneously. Hence, the dictionary learning model can be formulated as follows:

$$\arg\min_{D^s,D^c,\alpha_i} \sum_{c=1}^{C} \sum_{i\in I_c} \left\{ \begin{array}{l} ||z_i - D\alpha_i||_2^2 + ||D^{\notin c,s}\alpha_i||_2^2 \\ +||z_i - D^c\alpha_i^c - D^s\alpha_i^s||_2^2 + \lambda||\alpha_i||_1 \end{array} \right\}, \tag{5}$$

where $D = \left[D^1, D^2, \ldots, D^C, D^s\right]$, and α_i^c and α_i^s are the sparse codes based on D^c and D^s, respectively. $D^{\notin c,s}$ denotes the dictionary atoms excluding D^c and D^s from D. The second $||D^{\notin c,s}\alpha_i||_2^2$ and third $||z_i - D^c\alpha_i^c - D^s\alpha_i^s||_2^2$ terms in Eq. (5) are to ensure that the patches belonging to c-th cluster will be recovered only with the atoms in D^c and D^s effectively. In this way, the shared pattern among different clusters will be explicitly represented with D^s and the unique representation of each D^c is obtained.

As some common patterns may appear in several clusters, the incoherence penalty term $\sum_{i=1}^{C}\sum_{j=1,j\neq i}^{C} \left\|(D^i)^T D^j\right\|_F^2$ [30] is

introduced as follows:

$$\arg\min_{D^s,D^c,\alpha_i}\sum_{c=1}^{C}\sum_{i\in I_c}\left\{\begin{array}{l}||z_i-D\alpha_i||_2^2+||D^{\notin c,s}\alpha_i||_2^2\\+||z_i-D^c\alpha_i^c-D^s\alpha_i^s||_2^2+\lambda||\alpha_i||_1\end{array}\right\}$$
$$+\eta\sum_{i=1}^{C}\sum_{j=1,j\neq i}^{C}\left\|(D^i)^TD^j\right\|_F^2, \quad (6)$$

where F denotes Frobenius norm and η is weighting factor. This term penalizes the similarity of the learned dictionary atoms among different clusters, so that the uniqueness of each sub-dictionary D^c will be promoted. Meanwhile, the common patterns will be captured with shared sub-dictionary D^s at same time, which makes each sub-dictionary more representative.

The optimization problem defined in Eq.(6) could be solved by iteratively updating the dictionary atom D and sparse coefficient α. However, the intrinsic relationships among training data are ignored, as the coefficient α_i is calculated independently for each image patch. Since the image patches with similar representation have been clustered into same group for sub-dictionary learning, the $l_{1,2}$ mixed-norm $\sum_{c=0}^{C}\left\|\alpha_i^c\right\|_2$ [31] is introduced to enhance the group-level sparsity of our learned dictionary. With the help of this constraint, the sparse coefficient α_i for each image patch will tend to be represented with dictionary atoms selected from the same cluster.

To further incorporate the structure information in dictionary modeling, locality constraint $\|P_i\odot\alpha_i\|_2^2$ [32] has also been added to ensure that similar image patches will be reconstructed with similar dictionary atoms. Denote $P_i=[p_{i1},p_{i2},\ldots,p_{iN}]$ as a locality adapter which penalizes the distance between input z_i with each dictionary atom, N as the total number of dictionary atom, and symbol $\odot$ as the element-wise multiplication. In this paper, we utilize Euclidean distance as a measurement:

$$p_{ij}=\exp\left(\frac{||z_i-d_j||_2}{\delta}\right), \quad (7)$$

where d_j is the j-th atoms in D and δ controls the decay speed of locality adapter P_i.

Thus, the complete objective function of our dictionary modeling can be formulated as follows:

$$\arg\min_{D^s,D^c,\alpha_i}\sum_{c=1}^{C}\sum_{i\in I_c}\left\{\begin{array}{l}||z_i-D\alpha_i||_2^2+||D^{\notin c,s}\alpha_i||_2^2\\+||z_i-D^c\alpha_i^c-D^s\alpha_i^s||_2^2\\+\lambda_1\sum_{c=0}^{C}\left\|\alpha_i^c\right\|_2+\lambda_2\left\|P_i\odot\alpha_i\right\|_2^2\end{array}\right\}$$
$$+\eta\sum_{i=1}^{C}\sum_{j=1,j\neq i}^{C}\left\|(D^i)^TD^j\right\|_F^2, \quad (8)$$

where λ_1, λ_2, and η are constant term coefficients.

2) DICTIONARY LEARNING

With the clustered training data, each sub-dictionary is initialized with K-SVD approach due to its effectiveness and efficiency. Then, with the initialized value, the dictionary atoms will be optimized and updated iteratively. In this way, the optimization problem formulated by Eq. (8) can be rewritten as follows:

$$\arg\min_{\alpha_i}\sum_{c=1}^{C}\sum_{i\in I_c}\left\{\begin{array}{l}||z_i-D\alpha_i||_2^2+||D^{\notin c,s}\alpha_i||_2^2\\+||z_i-D^c\alpha_i^c-D^s\alpha_i^s||_2^2\\+\lambda_1\sum_{c=0}^{C}\left\|\alpha_i^c\right\|_2+\lambda_2\left\|P_i\odot\alpha_i\right\|_2^2\end{array}\right\}. \quad (9)$$

Such objective function Eq.(9) can be further transformed into the following form:

$$\arg\min_{\alpha_i}\sum_{c=1}^{C}\sum_{i\in I_c}\left\{\begin{array}{l}||\tilde{z}_i-\tilde{D}\alpha_i||_2^2+\frac{\lambda_1}{\sqrt{2}}\sum_{c=0}^{C}\left\|\alpha_i^c\right\|_2\\+\frac{\lambda_2}{\sqrt{2}}\left\|p_i\odot\alpha_i\right\|_2^2\end{array}\right\}, \quad (10)$$

where $\tilde{z}_i=\begin{bmatrix}z_i\\z_i\\0\end{bmatrix}$, $\tilde{D}=\begin{bmatrix}D\\D^c,D^s\\D^{\notin c}\end{bmatrix}$.

In this way, the optimization problem determined in Eq. (10) has been transformed as a classical group LASSO problem which could be effectively calculated. In this paper, we utilize the projected gradient method proposed in [33] to determine the sparse coefficient α_i for each z_i.

With the determined sparse coefficient α_i, each dictionary atom will be updated while others are fixed. For the sub-dictionary D^c of cluster c, the objective function of dictionary update can be simplified as follows by omitting the independent terms:

$$\arg\min_{D^c}\sum_{i\in I_c}\left\{\begin{array}{l}\left\|z_i-\sum_{j=1,j\notin c}^{C}D^j\alpha_i^j-D^s\alpha_i^s-D^c\alpha_i^c\right\|_2^2\\+\left\|z_i-D^c\alpha_i^c-D^s\alpha_i^s\right\|_2^2\end{array}\right\}$$
$$+2\eta\left\{\sum_{j=1,j\neq c}^{C}\left\|(D^c)^TD^j\right\|_F^2+\left\|(D^c)^TD^s\right\|_F^2\right\}. \quad (11)$$

The method proposed in [34] is employed to calculate the first derivative of Eq. (11) with respect to each dictionary atom in D^c. Then, the value making the derivative equal to zero is chosen as the updated dictionary atom.

Similarly, the atoms in shared sub-dictionary are updated with the objective function as follows:

$$\arg\min_{D^s}\sum_{c=1}^{C}\sum_{i\notin I_c}\left\{\begin{array}{l}\left\|z_i-\sum_{j=1}^{C}D^j\alpha_i^j-D^s\alpha_i^s\right\|_2^2\\+\left\|z_i-D^c\alpha_i^c-D^s\alpha_i^s\right\|_2^2\end{array}\right\}$$
$$+2\eta\sum_{j=1}^{C}\left\|\left(D^s\right)^TD^j\right\|_F^2. \quad (12)$$

3) IMAGE SUPER-RESOLUTION VIA ANCHORED NEIGHBORHOOD REGRESSION

While it is straightforward to obtain sparse representation of HR image patch via Eq.(2), it would be very time-consuming, especially when the given LR image contains a large number

TABLE 1. Comparisons of HR reconstruction performances in terms of PSNR(dB) and SSIM for upscaling factors ×2, ×3 and ×4 on Set5 and Set14.

dataset	scale	Bicubic	RFL	ScSR	SDS	A+	PSANR	Wu	SRCNN	Liang	Ours
Set5	2	33.66	36.54	35.78	–	36.55	36.63	36.52	36.34	36.41	**36.76**
		0.9299	0.9537	0.9485	–	0.9544	–	0.9538	–	–	**0.9550**
	3	30.39	32.43	31.46	31.98	32.59	32.67	32.60	32.39	32.48	**32.69**
		0.8677	0.9057	0.8903	0.8988	0.9088	–	0.9088	–	–	**0.9091**
	4	28.42	30.14	29.18	–	30.30	30.35	30.34	30.09	30.10	**30.39**
		0.8099	0.8548	0.8312	–	0.8603	–	0.8616	–	–	**0.8618**
Set14	2	30.23	32.26	31.85	–	32.27	32.36	32.25	32.18	32.22	**32.41**
		0.8691	0.9040	0.8972	–	0.9056	–	0.9037	–	–	**0.9062**
	3	27.54	29.05	28.31	28.75	29.12	**29.20**	29.14	29.00	29.05	**29.20**
		0.7741	0.8164	0.7974	0.8078	0.8188	–	0.8186	–	–	**0.8203**
	4	26.00	27.24	26.46	–	27.31	**27.42**	27.35	27.20	27.25	27.39
		0.7023	0.7451	0.7205	–	0.7491	–	0.7501	–	–	**0.7518**

TABLE 2. Comparison of HR reconstruction performances on natural images for upscaling factor ×3 in terms of PSNR(dB) and SSIM.

Images	Bicubic	ScSR	A+	REDSR	DPSR	Ours
Butterfly	24.1	26.1	27.1	27.5	27.0	27.4
	0.8198	0.8809	0.9069	0.9084	0.8955	0.9070
Flowers	27.3	28.7	29.1	29.1	28.8	29.3
	0.8027	0.8450	0.8537	0.8524	0.8477	0.8555
Hat	29.2	30.6	31.1	31.1	30.8	31.3
	0.8281	0.8616	0.8720	0.8701	0.8653	0.8709
Leaves	23.4	25.5	26.1	26.4	26.1	26.5
	0.8026	0.8794	0.9032	0.9054	0.8937	0.9037
Lenna	30.1	31.5	32.1	31.9	31.7	33.6
	0.8430	0.8748	0.8845	0.8810	0.8777	0.8867
Average	26.8	28.5	29.1	29.2	28.9	**29.6**
	0.8192	0.8683	0.8841	0.8835	0.8760	**0.8848**

of sampled patches. In order to reduce the computational cost at recovery stage, the anchored neighborhood regression approach has been utilized.

The Euclidean distance between training samples and learned dictionary atom is calculated, and K nearest neighbors from training data are collected to constitute the local neighborhoods $k_l^{d_n}, k_h^{d_n}$ for dictionary atom d_n. Then, the projection from LR space to HR space for each dictionary atom could be pre-calculated as follows:

$$q_{d_n} = k_h^{d_n}({k_l^{d_n}}^T k_l^{d_n} + \lambda I)^{-1} {k_l^{d_n}}^T, \tag{13}$$

where I denotes the identity matrix.

As a result, for an input LR image patch x_{li}, Eq. (2) can be reformulated as:

$$\begin{aligned} \hat{x}_{hi} &\approx D_h \alpha_i, \\ \alpha_i &= \arg\min_{\alpha_i} \left\{ ||x_{li} - k_l^{d_n}\alpha_i||_2^2 + \lambda \|\alpha_i\|_2 \right\}. \end{aligned} \tag{14}$$

The problem formulated in Eq. (14) can be converted into ridge regression [27]. That is, instead of calculating the sparse representation via Eq. (14), the recovered HR image patch $\hat{x}_{hi}$ is effectively projected into the HR space with the nearest matched dictionary atom and stored matrix as follows:

$$\hat{x}_{hi} = Q x_{li}, \tag{15}$$

where $Q = [q_{d_1}, q_{d_2}, \ldots, q_{d_N}]$. In this way, the sparse coding calculation of each image patch is avoidable which further improve the computational efficiency of our method.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we firstly compare our proposed algorithm with a number of state-of-the-art methods on two widely used benchmark datasets to evaluate the effectiveness and robustness of our proposed method. Then, we utilize the real-word images automatically extracted by surveillance system to further evaluate the applicability of our method in the maritime domain.

A. EXPERIMENTAL SETTINGS

We conduct our experiments on the benchmark datasets with the same settings in [29] where 91 images were used as training data for dictionary learning. Set5 [35], Set14 [36] and the images provided in NCSR [37] are used as testing data to evaluate the performances of HR reconstruction. These images cover a wide range of categories (e.g, animals, faces, and natural scenes) with rich details (e.g., abundant texture). Furthermore, we utilize the maritime surveillance images to demonstrate the applicability of our method. As maritime vehicles are the most interesting objects of surveillance system, 96 images of different categories of boats selected from MarDCT dataset [38] are utilized as training data for dictionary learning. The real images automatically extracted from an intelligent surveillance system is utilized as testing data.

In the training stage, the patch size is set to $s = 3$ for LR images with the overlap of 1 pixel. The number of clusters in K-means algorithm is set to $C = 7$, and the number of atoms in both D^s and D^c is set to 256. Thus, the total number of atoms in our dictionaries is $N = (7+1) \times 256 = 2,048$. The extracted patch-based image representations have been normalized and according to the settings in the literature [39] and experimental performances, the parameters in Eq. (8) are set as follows: $\lambda_1 = 0.1, \lambda_2 = 0.5, \eta = 0.15, \delta = 1,000$. We fix them for all of the experiments in this paper and good performances have been achieved without any parameters adjusting process under different conditions. According to the setting in [27], the neighboring size of training data is set as $K = 2,048$. In our experiments, color images are transformed into YCbCr color space. As human vision system is more sensitive to luminance than chrominance, our method is only applied to the luminance channel, and the Cb and Cr channels are interpolated with bicubic interpolation method directly.

To evaluate the effectiveness and robustness of our proposed method, 13 state-of-the-art methods are chosen for comparison, including Bicubic interpolation, ScSR [29], Zeyde [36], ANR [26], A+ [27], DPSR [40], SRCNN [41], RFL [42], REDSR (RED-based SR) [43], Liang (SRCNN-Multitask-Pr) [14], SDS [44], Wu [45] and PSANR [28]. Two metrics, PSNR (peak signal-to-noise ratio) and SSIM (structural similarity index measure), are chosen for quantitative evaluation. In general, the higher values of PSNR and SSIM, the better quality of reconstructed HR images. For fair comparisons, the results of those 13 methods are excerpted from existing publications or online websites of the authors.

B. QUANTITATIVE EVALUATION

Table 1 shows comparison results of average PSNR and SSIM on Set5 and Set14 for upscaling factors $\times 2$, $\times 3$ and $\times 4$. The best performance is highlighted in boldface, and the second best one is underlined. The comparison algorithms could be roughly classified into three categories: (1) compressive sensing based methods (such as RFL [42], ScSR [29] and SDS [44]), (2) neighbor embedding based methods (such as A+ [27], PSANR [28] and Wu [45]), and (3) deep learning based methods (such as SRCNN [41] and Liang [14]).

For upscaling factor $\times 2$, compared with the second best results, our method achieves 0.13dB and 0.05dB improvements in term of PSNR on Set5 and Set14, respectively. Compared with the best results of other compressive sensing based methods (RFL [42]), our method achieves 0.22dB and 0.15dB improvements in term of PSNR, respectively. These indicate that rich details can be better represented with our shared and cluster-specific dictionary learning method for better reconstructed quality.

For upscaling factor $\times 3$, compared with neighbor embedding based methods (such as A+ [27], PSANR [28] and Wu [45]), our method achieves 0.1dB, 0.02dB and 0.09dB improvements in term of SSIM on Set5, respectively. On Set14, our method achieves 0.15dB, 0.89dB and 0.45dB higher than other compressive sensing methods (such as RFL [42], ScSR [29] and SDS [44]) in term of PSNR, respectively. Our method achieves the best performances in terms of average PSNR and SSIM against other state-of-the-art methods on both Set5 and Set14. That is, considering the similarity and different characteristics of dictionary atoms explicitly would further improve the accuracy of image super-resolution.

For upscaling factors $\times 4$, compared with deep learning based methods (such as SRCNN [41]), our method achieves 0.3dB and 0.19dB improvements in term of PSNR on Set5 and Set14, respectively. Compared with Liang [14] which incorporates additional prior information based on SRCNN network, our method could also achieves 0.29dB and 0.14dB improvements, respectively. These indicate that our shared and cluster-specific dictionary learning based method has better performances with less training data and higher computational efficiency.

The detailed reconstruction results for upscaling factor $\times 3$ of the most commonly used 5 images chosen from benchmark datasets are shown in Table 2. In average, our method performs 0.4 dB better than the second best result of REDSR [43] method in term of PNSR and 0.0007 better than A+ [27] method in term of SSIM. For each image, our method performs either best or second best among compared methods in terms of both PSNR and SSIM.

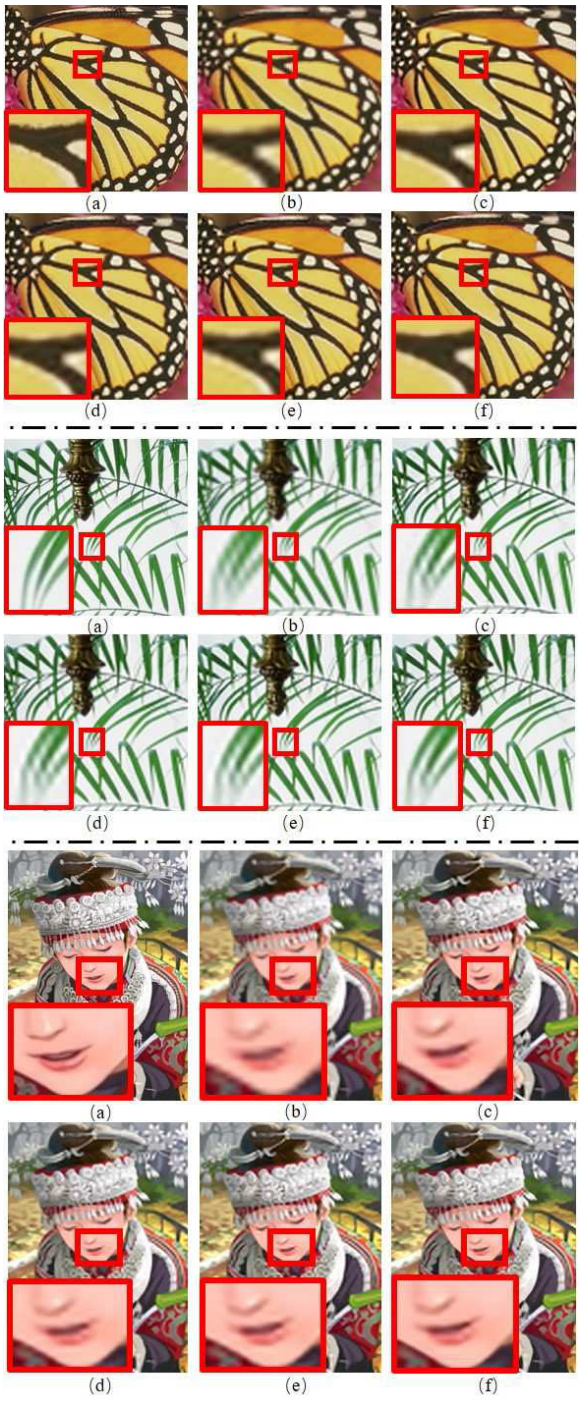


FIGURE 2. Reconstruction results for upscaling factor ×3 on three sample images ('butterfly', 'leaves' and 'comic' from Set5 [35], Set14 [36] and NCSR [37], respectively). (a) Ground truth. (b) Bicubic. (c) ScSR [29]. (d) Zeyde [36]. (e) ANR [26]. (f) Our method.

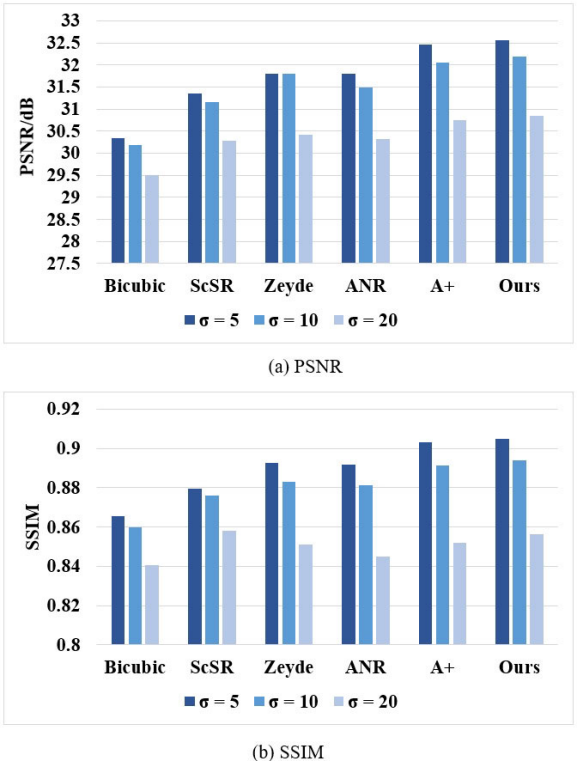


FIGURE 3. Average reconstruction results of different methods at different noise conditions for upscaling factor ×3 on Set5.

C. QUALITATIVE EVALUATION

In Fig.2, we show the reconstruction results for upscaling factor ×3 on sample images chosen from each benchmark testing dataset. It is clearly noticed that our proposed method preserves more detailed information (e.g., edges) and produces less artifacts than other methods. For example, the reconstruction results via the bicubic method generally contain a large amount of blur and tend to be over-smooth. Although existing compressive sensing-based methods (such as ScSR [29] and Zeyde [36]) are able to reconstruct more details, they are limited in recovering fine texture of natural images and usually introduce watercolor-like artifacts to reconstructed images.

D. ROBUSTNESS ANALYSIS

To further investigate the robustness of our proposed method, Gaussian noises with zero mean and different levels of variance ($\sigma = 5$, 10, and 20) are added to testing images. In Fig.3, we show the average reconstruction performances of different methods at different noise conditions for upscaling factor ×3 on Set5. It is noticed that with the increase of noise variance, the quality of reconstructed images will decrease for all the methods. As the HR image is recovered with projection matrix of effectively learned dictionary, our approach is not very sensitive to noise interference and performs best at different noise levels in terms of both PSNR and SSIM.

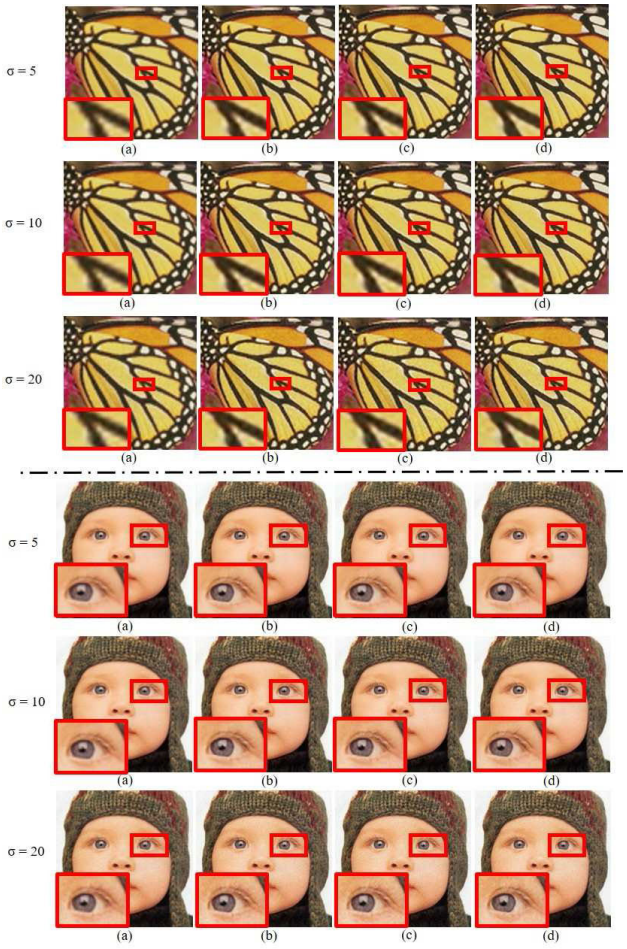


FIGURE 4. Reconstruction results of different methods at different noise conditions for upscaling factor ×3 on 'butterfly' and 'baby' from Set5. (a) Bicubic. (b) ScSR [29]. (c) ANR [26]. (d) Our method.

Visual results under different noise conditions are shown in Fig.4. The detailed images demonstrate that other methods (e.g., the compressive sensing-based method ScSR [29]) are more sensitive to noises for images containing finer structures. Even under the condition of noise variance $\sigma = 20$, our method could still preserve the edge well and recover smooth regions effectively.

E. SUPER-RESOLVING REAL-WORD MARITIME SURVEILLANCE IMAGES

Fig.5 shows the reconstruction results for upscaling factor ×2 on real-word images extracted from a maritime surveillance system. Note that, the blur kernel and ground truth HR images are not available in this case. In maritime environments, the captured surveillance images own strong characterizations consisting of various boats. As our method learns shared and cluster-specific dictionaries simultaneously, the representation of specific structures will be enhanced to reconstruct more favorable results for real-world images. Comparing with other methods, the hull mark and characters show that the recovery images produced by our

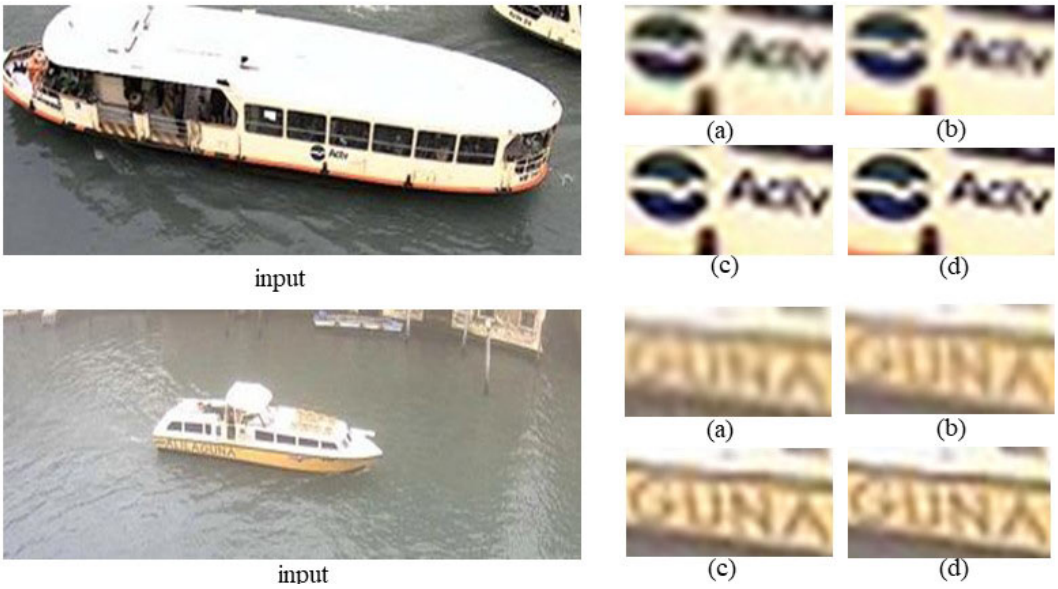


FIGURE 5. Reconstruction results for upscaling factor ×2 on real-world maritime surveillance images. (a) Bicubic. (b) ScSR [29]. (c) ANR [26]. (d) Our method.

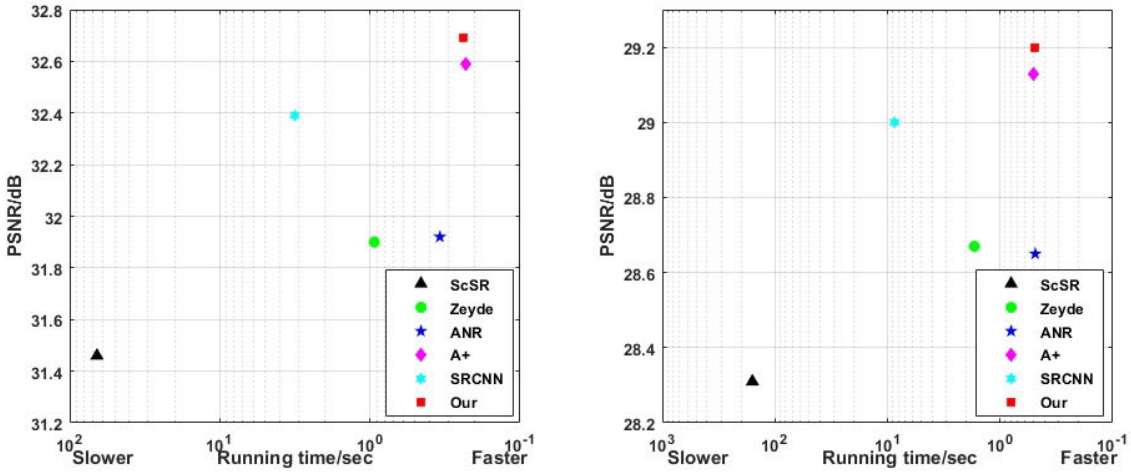


FIGURE 6. Speed and accuracy trade-off on Set5 and Set14 for upscaling factor ×3.

method contain less noise and block effect. The maritime image captured in foggy weather is also selected to evaluate the robustness of our method. It shows that our method could effectively reduce the fuzzy pattern between characters and boat edge in bad weather with the learned shared and cluster specific sub-dictionaries. Moreover, with the help of anchored neighborhood regression, the LR image could be more efficiently recovered in reconstruction stage.

F. COMPUTATIONAL TIME

Fig.6 shows the trade-offs between running time and accuracy in term of PSNR on Set5 and Set14 for upscaling factor ×3. We evaluate the speed of each state-of-the-art method on the same machine with 3.6 GHz Interl i7 CPU (16G Memory). As our method learns the common pattern with a shared dictionary explicitly in training stage, the cluster-specific dictionaries are more representative which leads to better results in reconstruction stage. With the help of anchored neighborhood regression, the computational efficiency of our method is further enhanced. Hence, the proposed method could achieve the best results with faster speed, which makes it more suitable for practical applications.

V. CONCLUSION

In this paper, we present a novel shared and cluster-specific dictionary learning method for single image super-resolution. Instead of learning a dictionary within each cluster individually, our proposed method takes the content similarity among clusters into account so that cluster-specific dictionaries are discriminant for specific content reconstruction. To preserve the relationship and structure information among image patches in dictionary modeling, group sparsity constraint, locality constraint and incoherence penalty term are introduced in a unified framework. In addition, at the

reconstruction stage, instead of calculating the sparse representation, a neighborhood regression method is devised so that an HR image can be efficiently reconstructed with precalculated projection matrix based on the learned dictionaries. Both quantitatively and qualitatively evaluations on two widely used benchmark datasets demonstrate the effectiveness and robustness of our proposed method over a large number of existing super-resolution methods. The reconstruction results of real-word maritime surveillance images further demonstrate the capacity of our method. Our future work will focus on devising more adaptive clustering algorithms under the proposed framework.

REFERENCES

[1] Y. Li, W. Dong, X. Xie, G. Shi, J. Wu, and X. Li, "Image super-resolution with parametric sparse model learning," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4638–4650, Sep. 2018.

[2] J. Zhang, L. Zhang, L. Xiang, Y. Shao, G. Wu, X. Zhou, D. Shen, and Q. Wang, "Brain atlas fusion from high-thickness diagnostic magnetic resonance images by learning-based super-resolution," *Pattern Recognit.*, vol. 63, pp. 531–541, Mar. 2017.

[3] R. Abiantun, F. Juefei-Xu, U. Prabhu, and M. Savvides, "SSR2: Sparse signal recovery for single-image super-resolution on faces with extreme low resolutions," *Pattern Recognit.*, vol. 90, pp. 308–324, Jun. 2019.

[4] P. L. K. Ding, B. Li, and K. Chang, "Convex dictionary learning for single image super-resolution," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 4058–4062.

[5] J. Zhao, H. Hu, and F. Cao, "Image super-resolution via adaptive sparse representation," *Knowl.-Based Syst.*, vol. 124, pp. 23–33, May 2017.

[6] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.

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

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

[9] W. Yang, J. Liu, S. Yang, and Z. Quo, "Image super-resolution via non-local similarity and group structured sparse representation," in *Proc. Vis. Commun. Image Process.*, 2015, pp. 1–4.

[10] J. Ahmed and R. Klette, "Coupled multiple dictionary learning based on edge sharpness for single-image super-resolution," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2016, pp. 3838–3843.

[11] J. Ahmed and M. A. Shah, "Single image super-resolution by directionally structured coupled dictionary learning," *EURASIP J. Image Video Process.*, vol. 36, no. 1, p. 36, 2016.

[12] K. Nasrollahi and T. B. Moeslund, "Super-resolution: A comprehensive survey," *Mach. Vis. Appl.*, vol. 25, no. 6, pp. 1423–1468, 2014.

[13] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[14] Y. Liang, J. Wang, S. Zhou, Y. Gong, and N. Zheng, "Incorporating image priors with deep convolutional neural networks for image super-resolution," *Neurocomputing*, vol. 194, pp. 340–347, Jun. 2016.

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

[16] L. Wang, Z. Huang, Y. Gong, and C. Pan, "Ensemble based deep networks for image super-resolution," *Pattern Recognit.*, vol. 68, pp. 191–198, Aug. 2017.

[17] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1823–1831.

[18] J. Zhang, C. Zhao, R. Xiong, S. Ma, and D. Zhao, "Image super-resolution via dual-dictionary learning and sparse representation," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2012, pp. 1688–1691.

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

[20] S. Yang, M. Wang, Y. Chen, and Y. Sun, "Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4016–4028, Sep. 2012.

[21] F. Yeganli, M. Nazzal, and H. Ozkaramanli, "Super-resolution using multiple structured dictionaries based on the gradient operator and bicubic interpolation," in *Proc. 24th Signal Process. Commun. Appl. Conf.*, May 2016, pp. 941–944.

[22] 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.

[23] B. Yue, S. Wang, X. Liang, and L. Jiao, "An external learning assisted self-examples learning for image super-resolution," *Neurocomputing*, vol. 312, pp. 107–119, Oct. 2018.

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

[25] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. S. Huang, "Learning super-resolution jointly from external and internal examples," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4359–4371, Nov. 2015.

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

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

[28] L. Sun, F. Han, C. Cai, and L. Su, "Partially supervised anchored neighborhood regression for image super-resolution through FoE features," *Neurocomputing*, vol. 275, pp. 2341–2354, Jan. 2018.

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

[30] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3501–3508.

[31] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.

[32] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3360–3367.

[33] M. Schmidt, K. Murphy, G. Fung, and R. Rosales, "Structure learning in random fields for heart motion abnormality detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[34] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 186–199.

[35] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. 23rd Brit. Mach. Vis. Conf. (BMVC)*, 2012, pp. 1–10.

[36] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surfaces*. Berlin, Germany: Springer, 2010, pp. 711–730.

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

[38] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, "ARGOS-Venice boat classification," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2015, pp. 1–6.

[39] T. Yao, Z. Wang, Z. Xie, J. Gao, and D. D. Feng, "Learning universal multiview dictionary for human action recognition," *Pattern Recognit.*, vol. 64, pp. 236–244, Apr. 2017.

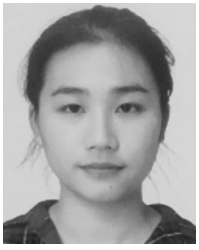
[40] Y. Zhu, Y. Zhang, and A. L. Yuille, "Single image super-resolution using deformable patches," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2917–2924.

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

[42] S. Schulter, C. Leistner, and H. Bischof, "Fast and accurate image upscaling with super-resolution forests," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3791–3799.
[43] T. Li, X. He, Q. Teng, and X. Wu, "Rotation expanded dictionary-based single image super-resolution," *Neurocomputing*, vol. 216, pp. 1–17, Dec. 2016.
[44] W. Lu, H. Sun, R. Wang, L. He, M. Jou, S. Syu, and J. Li, "Single image super resolution based on sparse domain selection," *Neurocomputing*, vol. 269, pp. 180–187, Dec. 2017.
[45] H. Wu, J. Zhang, and Z. Wei, "High resolution similarity directed adjusted anchored neighborhood regression for single image super-resolution," *IEEE Access*, vol. 6, pp. 25240–25247, 2018.

TINGTING YAO received the B.Eng. and Ph.D. degrees from the Hefei University of Technology, China. She is currently a Lecturer with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. Her research interests include image analysis and pattern recognition.

ZHIYONG WANG received the B.Eng. and M.Eng. degrees in electronic engineering from the South China University of Technology, Guangzhou, China, and the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong. He is currently an Associate Professor with the School of Computer Science, The University of Sydney, Australia. His research interest includes multimedia computing, including multimedia information processing, retrieval and management, Internet-based multimedia data mining, human-centered multimedia computing, and pattern recognition.



YU LUO is currently pursuing the B.Eng. degree with Dalian Maritime University. Her research interests include image processing and image super-resolution.

YUE LIANG received the B.Eng. degree from Shenyang Ligong University. He is currently pursuing the M.Eng. degree with Dalian Maritime University. His research interests include computer vision and image defogging.

QING HU received the Ph.D. degree in communication and information system from Dalian Maritime University, Dalian, Liaoning, China, where he is currently a Professor with the College of Information Science and Technology. His research interests include maritime communication and automatic identification systems.

DAVID DAGAN FENG received the M.E. degree in electrical engineering and computing science from Shanghai Jiao Tong University, in 1982, and the M.Sc. degree in biocybernetics and the Ph.D. degree in computer science from the University of California at Los Angeles, Los Angeles, in 1985 and 1988, respectively. He was an Assistant Professor with the University of California at Riverside, Riverside. He joined The University of Sydney as a Lecturer, in 1988, where he is currently a Professor with the School of Computer Science. He is also an Honorary Research Consultant with the Royal Prince Alfred Hospital, Sydney, the Chair Professor of information technology with The Hong Kong Polytechnic University, an Advisory Professor with Shanghai Jiao Tong University, and a Guest Professor with Northwestern Polytechnic University, Northeastern University, and Tsinghua University.