Single-image Super-resolution based on Non-local Means and Double-sparsity Dictionaries

Xiuxiu Liao, Kejia Bai*, Qian Zhang, Xiping Jia, Jia Ouyang, Yunzhi Jiang

***Abstract*—Sparse representation models have been widely used in single-image super-resolution reconstruction. The construction of dictionaries is especially important in these models. Basically, approaches of dictionary construction in sparse representation can be divided into two categories: analytical and learning-based approaches. Analytical approaches are effective and fast, but they are unable to fit different types of data; learning-based approaches are adaptive, but their implementation takes a significant amount of time. In this study, an image super-resolution reconstruction algorithm based on double-sparsity dictionaries is proposed. The algorithm combines the efficiency of analytical approaches and adaptability of learning-based approaches. In addition to the sparsity prior, the non-local self-similarity prior is also considered in the algorithm. Non-local means filtering is used to be the constraints on regularized super-resolution reconstruction procedures, which improves the quality of super-resolution reconstruction results further, while the runtime of the algorithm is still acceptable. Experimental results demonstrate the advantage of the proposed algorithm.**

***Index Terms*—Double-sparsity Dictionaries, Non-local Means, Sparse Representation, Superresolution**

This work was supported by Young Creative Talents Project of Department of Education of Guangdong Province(Natural Science) (2016KQNCX092, 2017KQNCX117, 2016KQNCX089), National Natural Science Foundation of China (61772144, 61702119), Foreign Science and Technology Cooperation Plan Project of Guangzhou Science Technology and Innovation Commission (201807010059), Natural Science Foundation of Guangdong (2016A030313472, 2018A030313994, 2018A0303130187), Science and Technology Program of Guangzhou(201607010152).

Xiuxiu Liao is a lecturer of the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China(e-mail: xxliao2013@aliyun.com).

Kejia Bai is an associate professor of the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China(corresponding author to provide phone: 8620-38256730; e-mail: 191474470@qq.com).

Qian Zhang is a lecturer of the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China(e-mail: 234482377@qq.com).

Xiping Jia is an associate professor of the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China(e-mail: 260628072@qq.com).

Jia Ouyang is a lecturer of the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China(e-mail: oyj@imozo.cn).

Yunzhi Jiang is an associate professor of the School of Mathematics and Systems Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China(e-mail: jyzscut@gmail.com).

I. INTRODUCTION

In recent years, single-image super-resolution (SR) has gained increasingly attention from researchers. In image processing, obtaining larger and clearer images from the original images is a priority. SR reconstruction not only can improve the visual effect of an image but also has very important meaning for further applications, such as feature extraction, automatic recognition, and image analysis.

Single-image SR refers to the problem of using signal-processing techniques to estimate a high-resolution (HR) image **X** with better quality from an observed low-resolution (LR) image **Y**. The image-observation model is usually described as $\mathbf{Y} = \mathbf{SHX} + \mathbf{V}$, where **H** is a blurring filter, **S** a down-sampling operator, and **V** additive noise. The above image-degradation model shows that image SR reconstruction is an inverse problem and ill-posed. Many HR images **X** may satisfy the model obtained from a given LR image **Y**.

During the reconstruction processing, introducing an effective prior or constraint (regularization) is very efficacious in helping people obtain a well-defined solution. The Maximum a posterior (MAP) estimation was first introduced into the image SR reconstruction algorithm by Schultz *et al.* [1], and the Huber-Markov field (MRF) was used as the prior knowledge to improve the image resolution. Based on the MAP framework, Belekos *et al.* [2] utilized the state-of-the-art single-channel image prior and observation models, and defined a new multichannel image prior model. Mofidi *et al.* [3] introduced a set of weighting coefficients that control the contribution between the regularization and data-error term in each of the estimated HR pixels. The employed coefficients were defined according to the information of neighbors of the estimated pixel. Li *et al.* [4] obtained SR images of real beam scanning radar using an accelerated MAP method. Based on the first and second orders of difference information, a prediction vector was constructed before each iteration operation, and the convergence speed was improved without performance loss.

Another widely used image prior is the non-local self-similarity prior [5]. The self-similarity of non-local patches is an essential feature of natural images. Based on this observation, small image patches are redundantly reproduced on the same scale and between different scales. Buades *et al.* [6] established a mathematical framework for non-local means (NLM) filtering. The idea of the framework was rather simple:

image patches with similar patterns may be far apart in space, so it was possible to search for image patches with similar patterns throughout the entire image. It compared the geometrical configuration in an entire neighborhood and obtained a more robust comparison than neighborhood filters. NLM was used to be the constraints on regularized SR reconstruction problems, which further improved the quality of reconstructed images, while the time consumption was still acceptable. Dong *et al.* [7] utilized both the non-local and local priors to optimize the objective function. Zhang *et al.* [8] incorporated the global reconstruction constraint, non-local similarity, and local structural regularity into a unified iterative framework. Here, the NLM was used to learn a non-local prior. This algorithm achieved a good reconstruction effort, but the burden of calculation was rather heavy, and the running speed was rather slow. In this paper, NLM is regarded as prior information and tries to acquire the initial value of the MAP estimation in the learning process. The desire of the proposed algorithm is to accelerate the convergence process and obtain better quality of reconstruction and time performance.

Apart from the above image priors, sparse-representation models have been widely used in single-image SR tasks. Yang *et al.* [9] first proposed an image SR reconstruction scheme based on sparse representation. An image patch was regarded as the sparse representation for an over-complete dictionary. The sparsity of an image patch was used as *a priori* information with which to regularize the SR reconstruction problems. The algorithm was improved in their later paper [10], not directly using the HR and LR image-patch pairs as dictionaries, but using a sparse-coding algorithm to learn more compact dictionary pairs, which greatly improved the calculation speed. Zeyde *et al.* [11] applied the k singular value decomposition (K-SVD) dictionary training procedure for the LR training patches, and orthogonal matching pursuit (OMP) algorithm to solve parse-coding. Experiments results showed that the algorithm can make both visual and peak signal-to-noise ratio (PSNR) improvements. Huang and Dragotti [12] proposed a deep dictionary model in which the dictionaries consist of L layers, with the first $L-1$ layers being analysis dictionaries and the last layer a synthesis dictionary, and the dictionary was updated iteratively in a backward fashion.

In sparse-representation-based SR, the construction of dictionaries is especially important for the performance of the reconstruction model. In general, these dictionary construction methods can be divided into two categories: analytical and learning based methods [13]. Dictionaries created from analytical approaches are "implicit" dictionaries described by algorithms such as wavelet, complex wavelet, contourlet, and bandelet algorithms. In contrast, learning-based dictionaries offer more convenience and power to handle specific signal data. These learning-based approaches usually use machine-learning techniques to infer dictionaries from training datasets, which are typically represented by concrete matrices. Dictionaries obtained via learning can fit the data better, but they make the system more complex and require a high calculation cost. Complexity constraints limit the learning of the dictionaries, especially the size of the atom. From a more practical standpoint, it would be desirable to effectively combine dictionaries of both types. A double-sparsity dictionary algorithm is proposed by Rubinstein *et al.* [14] that used both the analysis-based and learning-based methods. This algorithm had not only the efficiency of an analytical dictionary but also the adaptability of a learning-based dictionary. The structure was based on a sparsity model of the dictionary atoms over a known base dictionary. Ai *et al.* [15] used wavelet coefficients of LR images as the futures to train dictionary, so the trained dictionary pair had the property of double sparsity, and the training dataset can be small. In their later work [16], the bootstrapping method was used to construct training set, and four wavelet sub-bands of the two difference images were used as extra information to train the dictionary.

The high time efficiency of sparsity dictionaries meets the requirements for fast learning, so they are introduced here into the SR reconstruction, which can help obtain better time performance while ensuring reconstruction quality. Our double-sparsity dictionaries are different from those of Ai *et al.* [15], and the double-sparsity dictionaries are trained directly, while Ai *et al.* first performed a wavelet transform on the image, and the dictionaries were trained on the wavelet coefficients to achieve "double sparsity."

In this paper, an image SR reconstruction algorithm based on double-sparsity dictionaries is proposed. The algorithm combines both the advantages of analytical and learning-based dictionaries, which can ensure the quality of reconstruction and improve speed. The algorithm is suitable for applications with high-time-performance requirements. On this basis, non-local self-similar constraints are used in the regularization of SR reconstruction, which further improves the quality of reconstructed images while keeping runtimes within an acceptable range.

The main contributions of this paper are the following.

(1) Combining MAP estimation and NLM in SR reconstruction problems.

(2) Use of double-sparsity dictionaries in the image SR reconstruction process, which combines the efficiency of an analytical dictionary and the adaptability of a learned dictionary, and acquires better time performance while ensuring the reconstruction quality.

(3) Use of the non-local self-similar constraints in the reconstruction process, which further improves the quality of reconstructed images while keeping runtime within an acceptable range. The algorithm offers a desirable compromise between low computational complexity and high reconstruction quality.

The rest of this paper is organized as follows. In Section II, a novel image SR reconstruction algorithm based on double-sparsity dictionaries is proposed, which considers both the sparsity prior and non-local self-similarity prior. The theory of sparse representation of signals is discussed in Section II.*A* and the theory of a double-sparsity dictionary in Section II.*B*. In Section II.*C*, a SR based on double-sparsity dictionaries (DSD-SR) algorithm is proposed that can be divided into an offline training-data pre-processing part and an online testing-data SR reconstruction part. The proposed DSD-SR algorithm is then developed with NLM constraints (NLMDSD-SR) in Section II.*D*. Experimental results and analysis are shown in Section III.

II. Proposed Approach

A. Sparse Representation of Signals

Scientists have put a significant amount of effort on sparse representation of signals over redundant dictionaries. They suppose natural signals can be represented and estimated by a sparse linear combination of pre-determined atom signals, chosen from a group of data called a dictionary.

For a given signal $\mathbf{z} \in \mathbf{R}^{n\times 1}$, the sparse-representation model can be expressed as $\mathbf{Z}=\mathbf{D}\boldsymbol{\alpha}$, where $\mathbf{D} \in \mathbf{R}^{n\times K}$ is a dictionary of *K* atoms of length *n*, and $\boldsymbol{\alpha} \in \mathbf{R}^{K\times 1}$ is the sparse coefficient. The sparse representation of the signal $\boldsymbol{z}$ can be estimated by the following formula:

$$\min \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \quad \|\mathbf{z}-\mathbf{D}\boldsymbol{\alpha}\|_2^2 \le \varepsilon , \qquad (1)$$

where $\|\boldsymbol{\alpha}\|_0$ is the L_0-norm which represents a count of the number of non-zeros in the vector, and ε is the error allowed for accuracy.

Given the signal set $\{\mathbf{z}_i\}_{i=1}^N$, $\mathbf{Z}=\{\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_N\}$, the learning problem can be described as

$$\min_{\mathbf{D},\mathbf{A}} \left\{\|\mathbf{Z}-\mathbf{D}\mathbf{A}\|_F^2\right\} \quad \text{subject to} \ \forall \mathrm{i}, \|\boldsymbol{\alpha}_i\|_0 \le T_0 , \qquad (2)$$

or

$$\min_{\mathbf{D},\mathbf{A}} \sum_i \|\boldsymbol{\alpha}_i\|_0 \quad \text{subject to} \ \|\mathbf{Z}-\mathbf{D}\mathbf{A}\|_F^2 \le \varepsilon , \qquad (3)$$

where $\boldsymbol{\alpha}_i$ is the column i of $\mathbf{A}$ indicating the coefficient corresponding to the signal $\mathbf{z}_i$, and T_0 is a sparse constraint.

B. Double-Sparsity Dictionaries

SR reconstruction is an ill-posed problem since, for a given LR input image **Y**, there are many HR images **X** that satisfy the above reconstruction constraint. The sparsity of image patches is used as prior information to regularize the ill-posed problem in reconstruction. Supposing that $\mathbf{x}_i$ and $\mathbf{y}_i$ are the HR and LR image patches, respectively, which have the same sparse representation under the HR dictionary $\mathbf{D}_h$ and LR dictionary $\mathbf{D}_l$, the problem can be regularized via the sparse-representation prior on small patches *x* of **X**: The patches *x* of the HR image **X** can be represented as a sparse linear combination in a dictionary $\mathbf{D}_h$ of HR patches sampled from training images:

$$\mathbf{x} \approx \mathbf{D}_h\boldsymbol{\alpha} \quad \text{for some } \boldsymbol{\alpha} \in \mathbf{R}^K \text{ with } \|\boldsymbol{\alpha}\|_0 = K , \qquad (4)$$

and $\mathbf{x}_i$ can be reconstructed by $\mathbf{y}_i$ under the sparse priors:

$$\min_{\boldsymbol{\alpha},\mathbf{x}_i} \|\boldsymbol{\alpha}\|_0 \quad \text{subject to} \begin{cases} \|\mathbf{F}\mathbf{D}_l\boldsymbol{\alpha}-\mathbf{F}\mathbf{y}_i\|_2^2 \le \varepsilon_1 \\ \|\mathbf{F}\mathbf{D}_h\boldsymbol{\alpha}-\mathbf{F}\mathbf{x}_i\|_2^2 \le \varepsilon_2 \\ \|\mathbf{SHX}-\mathbf{Y}\|_2^2 \le \varepsilon_3 \end{cases} , \qquad (5)$$

where **F** is a (linear) feature-extraction operator.

Double-sparsity dictionaries combine the efficiency of analytical dictionaries and adaptability of learning-based dictionaries. The structure is based on the sparse model of the dictionary atom on the selected base dictionary:

$$\mathbf{D} = \boldsymbol{\Phi}\mathbf{W} , \qquad (6)$$

where Φ is the base dictionary, and **W** the atomic representation matrix, which is assumed to be sparse, and each column of which has a fixed number of non-zero values, $\|\mathbf{w}_i\|_0 \le P_0$. The selection of the base dictionary Φ has great influence on the performance of the double-sparsity dictionary model. Usually, analytical dictionaries with fast implementation are selected, such as wavelets, complex wavelets, and contourlets.

Compared with analytical dictionaries, the double-sparsity dictionary model provides adaptability through the modification of **W**. In addition, one can use different types of dictionaries as **Φ**. Compared to learning-based dictionaries, the sparse structure of double-sparsity dictionaries is more efficient, which mainly depends on the choice of **Φ**.

The training of double-sparsity dictionaries can be described as solving the following optimization problem:

$$\min_{\mathbf{W},\mathbf{A}} \{\|\mathbf{Z}-\boldsymbol{\Phi}\mathbf{W}\mathbf{A}\|_F^2\} \ \text{subject to} \begin{cases} \forall i, \|\boldsymbol{\alpha}_i\|_0 \le T_0 \\ \forall j, \|\mathbf{w}_j\|_0 \le P_0, \ \|\boldsymbol{\Phi}\mathbf{w}_j\|_2 = 1 \end{cases} , \qquad (7)$$

where **Z** is the signal to be represented, and **A** the sparse coefficient.

C. Super-Resolution based on Double-Sparsity Dictionaries

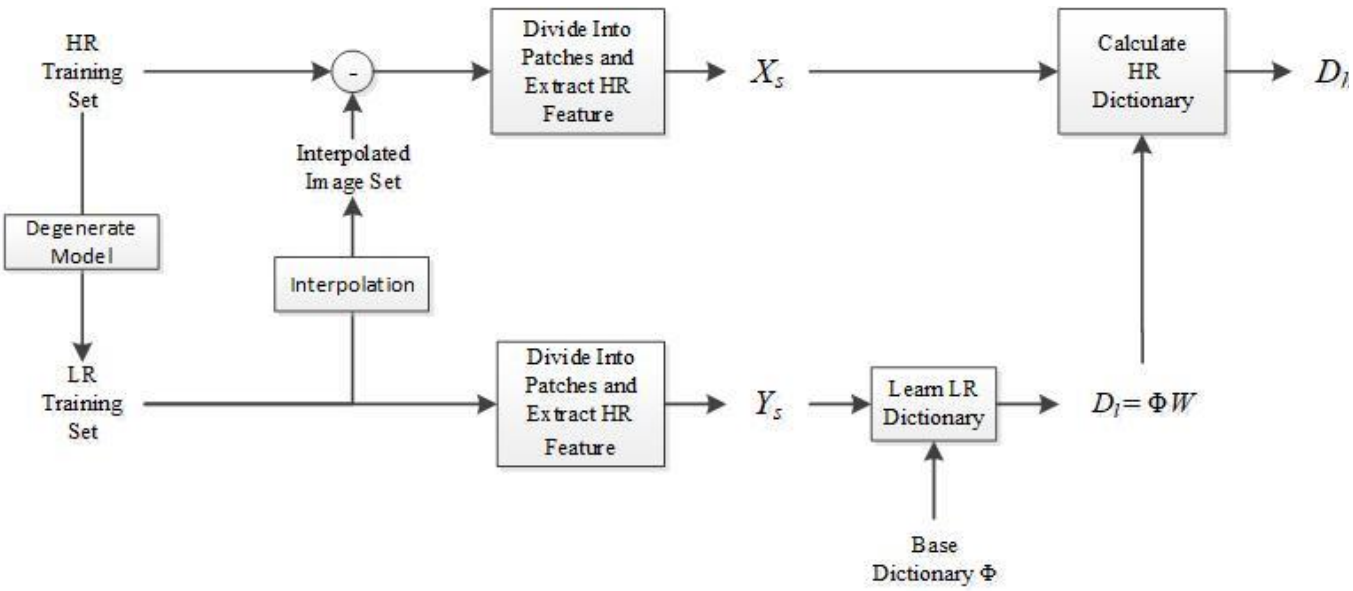

Fig. 1 DSD-SR part 1 framework

Zeyde *et al.* [11] applied K-SVD dictionary training and OMP sparse-coding to obtain good SR performance on learned dictionaries. Rubinstein *et al.* [14] were the first to introduce double-sparsity dictionaries. They proposed a sparse dictionary based on a sparsity model of the dictionary atoms over a base dictionary and demonstrated the advantages of the double-sparsity dictionary structure for computed tomography (CT) denoising. Inspired by these works, an image SR reconstruction algorithm based on double-sparsity dictionaries is proposed herein. The algorithm is divided into two parts: training-data pre-processing and testing-data SR reconstruction. The first part can be done offline.

Part 1 Training-Data Pre-processing

(1) The HR image in the training set is blurred and down-sampled according to the degraded model to obtain a corresponding LR image.

(2) LR feature vectors $\mathbf{Y}_s$ are extracted, and the LR image is divided into $n \times n$ patches and features extracted. A total of four one-dimensional filters are used in this step:

$$\begin{aligned} f_1 &= [-1,1], & f_2 &= f_1^T \\ f_3 &= [1,-2,1]/2, & f_4 &= f_3^T \end{aligned}, \quad (8)$$

These four filters are applied to LR images, and four eigenvectors are obtained for each image patch, which are concatenated as a characteristic representation of an image patch. Therefore, the length of the feature vector corresponding to each LR image patch is $4n^2$, which is denoted by $\{\mathbf{y}_i\}_{i=1}^N$, $\mathbf{Y}_s = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N\}$.

(3) LR dictionaries D_l are trained, and an overcomplete discrete cosine transform (DCT) dictionary is selected as the base dictionary $\mathbf{\Phi}$. That is to solve

$$\min_{\mathbf{W},\mathbf{A}} \{\|\mathbf{Y}_s - \mathbf{\Phi}\mathbf{W}\mathbf{A}\|_F^2\}$$
$$\text{subject to } \begin{cases} \forall i, \|\boldsymbol{\alpha}_i\|_0 \le T_0 \\ \forall j, \|\mathbf{w}_j\|_0 \le P_0, \ \|\mathbf{\Phi}\mathbf{w}_j\|_2 = 1 \end{cases}, \quad (9)$$

where the LR dictionary $\mathbf{D}_l = \mathbf{\Phi}\mathbf{W}$.

(4) HR feature vectors $\mathbf{X}_s$ are extracted. After interpolating LR images to the size of HR images (interpolation of the images), the corresponding interpolated image is subtracted from the HR image, and the high-frequency part can be obtained. The high-frequency image is divided into $(R_n) \times (R_n)$ image patches and the patches converted to vectors, which are regarded as the feature vectors of HR image patches $\{\mathbf{x}_i\}_{i=1}^N$, $\mathbf{X}_s = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$. R is the SR zoom factor.

(5) The HR dictionary $\mathbf{D}_h$ is calculated. Suppose that HR-LR image patches have the same sparse-representation coefficient A under the HR-LR dictionaries pair, then $\mathbf{D}_h$ can be calculated by minimizing the following approximation:

$$\mathbf{D}_h = \arg\min_{\mathbf{D}_h} \|\mathbf{X}_s - \mathbf{D}_h\mathbf{A}\|_F^2 \ , \quad (10)$$

Using the pseudo-inverse solution,

$$\mathbf{D}_h = \mathbf{X}_s\mathbf{A}^+ = \mathbf{X}_s\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \ , \quad (11)$$

where "+" represents pseudo-inverse operator.

The framework of Part 1 of DSD-SR is illutrated in Figure 1.

Part 2 Testing-Data SR Reconstruction

(1) The LR test image $\mathbf{Y}$ is interpolated to the size of the target HR image.

(2) $\mathbf{Y}$ is divided into $n \times n$ image patches and features extracted. For each feature vector $\mathbf{y}_t$, the following steps are

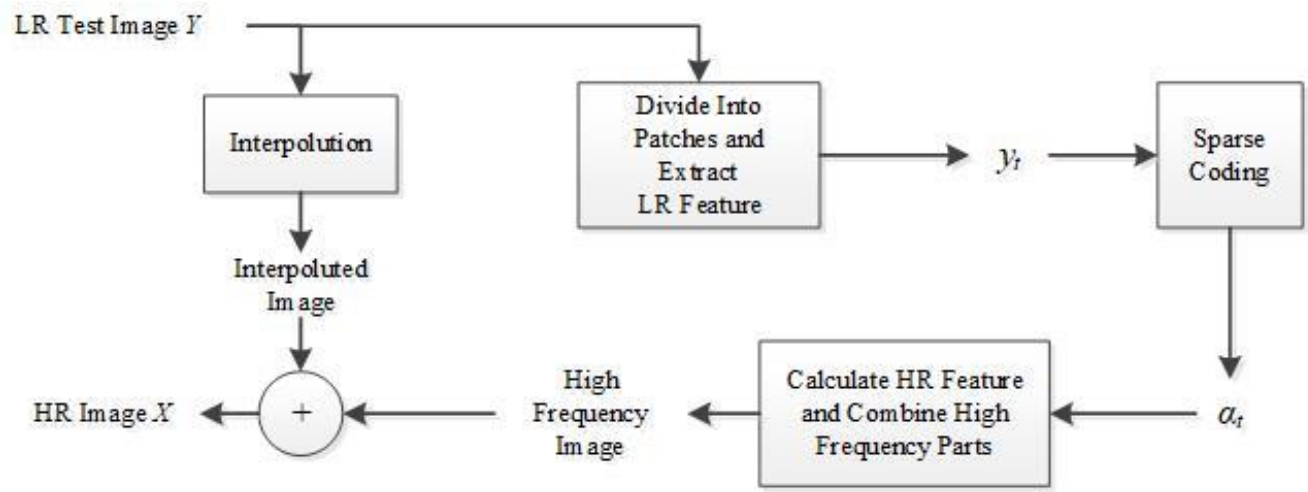

Fig. 2 DSD-SR part 2 framework

performed.

Step 1: The representation vector $\boldsymbol{\alpha}_t$ (sparse coding) is calculated as

$$\begin{aligned} &\boldsymbol{\alpha}_t = \arg\min_{\boldsymbol{\alpha}_t} \{\|\mathbf{y}_t - \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\alpha}_t\|_2^2\} \\ &\text{subject to } \|\boldsymbol{\alpha}_t\|_0 \le T_0 \end{aligned} \quad . \tag{12}$$

Step 2: The feature vector of HR patches $\mathbf{x}_t$ (high-frequency part) is calculated:

$$\mathbf{x}_t = \mathbf{D}_h \boldsymbol{\alpha}_t \quad . \tag{13}$$

(3) When all the HR features are obtained, the high-frequency HR image can be constructed by enforcing local compatibility and smoothness constraints between adjacent patches. The high-frequency HR image is then added to the interpolated image to obtain the target HR image **X**.

The framework of Part 2 of DSD-SR is illustrated in Figure 2.

D. DSD-SR with NLM Constraint

The image-degradation model $\mathbf{Y} = \mathbf{SHX} + \mathbf{V}$ shows that image SR reconstruction is an inverse problem and ill-posed. Many HR images **X** may satisfy the model obtained from a given LR image **Y**. To obtain a well-defined solution, it is important to introduce an effective prior or constraint (regularization) in the reconstruction process:

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \{\|\mathbf{Y} - \mathbf{SHX}\|_2^2 + \lambda \mathbf{R}(\mathbf{X})\}, \tag{14}$$

where the first item $\|\mathbf{Y} - \mathbf{SHX}\|_2^2$ is the data-fidelity item, indicating that the estimated HR image **X** must be consistent with the LR test image **Y** after degradation. The second item **R**(**X**) is the regularization item: the Lagrangian factor λ determines the strength of regularization constraints.

The self-similarity of the local patch mode is an important feature of natural images. Based on this observation, small image patches are redundantly reproduced in the same scale and between different scales. For each local image patch $\mathbf{x}_i$, the patches that are similar to it throughout the whole image are searched for (actually the searching is done in a large enough area around $\mathbf{x}_i$). The top L most similiar patches $\mathbf{x}_i^l$ for $\mathbf{x}_i$ are selected. Letting $\mathbf{p}_i$ and $\mathbf{p}_i^l$ be the center pixels of $\mathbf{x}_i$ and $\mathbf{x}_i^l$, respectively, the weighted average of $\mathbf{p}_i^l$ can be used to predict $\mathbf{p}_i$:

$$\mathbf{p}_i = \sum_{l=1}^{L} \mathbf{w}_i^l \mathbf{p}_i^l \quad , \tag{15}$$

where the weight $\mathbf{w}_i^l$ is defined as

$$\mathbf{w}_i^l = \frac{\exp(-e_i^l / h)}{\sum_{l=1}^{L} \exp(-e_i^l / h)}, \quad e_i^l = \|\mathbf{x}_i - \mathbf{x}_i^l\|_2^2, \tag{16}$$

and h is the global filter parameter, controlling the decline of exponential expression. Considering the non-local redundancy of natural images, the expected estimation error $\left\|\mathbf{p}_i - \sum_{l=1}^{L} \mathbf{w}_i^l \mathbf{p}_i^l\right\|_2^2$ is very small. Letting ω_i be the column vector containing all weights $\mathbf{w}_i^l$, $\boldsymbol{\beta}_i$ is the column vector containing all center pixels. Introducing non-local self-similarity regularization terms into formula (14), one obtains

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \{\|\mathbf{Y} - \mathbf{SHX}\|_2^2 + \lambda \sum_{x_i} \|\mathbf{x}_i - \boldsymbol{\omega}_i^T \boldsymbol{\beta}_i\|_2^2\}, \tag{17}$$

where $\mathbf{X}^*$ is solved using the gradient-descent method.

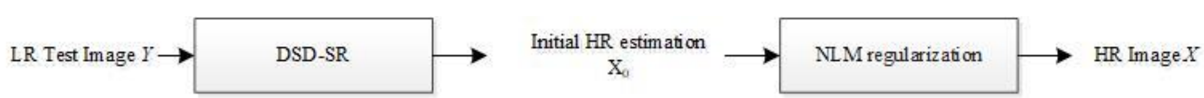

Fig. 3 NLMDSD-SR framework

Inspired by the work of Zhang *et al.* [8] and Rubinstein *et al.* [14], an image SR reconstruction algorithm based on double-sparsity dictionaries and NLM constraint is proposed (denoted NLMDSD-SR). The diagram of the algorithm is shown in Figure 3.

III. Experimental Results and Analysis

Several benchmarks, including Set14 [11], Set5 [17] and B100 [18], were used as the testing set in the present work. Approximately 100,000 training patch pairs were collected from the training image set used in [10].

In the experiments, the size of the LR image patches was 3 × 3, and the number of filters was four, so the length of the feature vectors was 3 × 3 × 4 = 36. An overcomplete DCT dictionary was selected as the base dictionary $\mathbf{\Phi}$. Supposing the number of atoms in the dictionary is K, then the scale of $\mathbf{\Phi}$ is $36 \times K$; the scale of the dictionary representation matrix $\mathbf{W}$ is $K \times K$. Sparse constraint P_0 was set to be 6, and the sparse constraint T_0 of the signal representation matrix $\mathbf{A}$ was set to 5. Experimental results show that the algorithm is not sensitive to the choice of K, which was finally set to be 576. All the simulations were conducted in MatLab R2016a (MathWorks, USA) on a PC with an Intel(R) Core(TM) i7 processor running at 3.6 GHz with 4GB of RAM.

A. Proposed Double-Sparsity Dictionary versus Learned Dictionary

The results of SR reconstruction with the proposed double-sparsity dictionaries (denoted DSD-SR) were evalated was compared with a learning-based dictionary in [11] (denoted LD-SR). To determine the contribution of the double-sparsity dictionary independently, NLM was not applied to enforce the global reconstruction constraint in the experiments for this part.

Seven images were selected from Set14 as the testing set. For each test image, the PSNR between the reconstructed image and original one was calculated, and the reconstruction time given, as shown in Table I (magnification ×2) and Table II (magnification ×4). As can be seen, the proposed DSD-based method has a slightly higher average PSNR than the LD-based method in [11], but the time-performance improvements are greater, i.e., approximately 60% less than the average reconstruction time of [11].

TABLE I
MAGNIFICATION ×2 PERFORMANCE IN TERMS OF PSNR(DB) AND RUNNING TIME(S)

Images	LD-SR		DSD-SR	
	PSNR(dB)	**Time(s)**	**PSNR(dB)**	**Time(s)**
Baboon	22.609	4.064	22.722	1.828
Barbara	26.564	7.303	26.703	2.966
Bird	34.916	1.355	34.942	0.578
Bridge	26.961	4.373	27.081	1.876
Face	31.400	1.238	31.488	0.526
Lenna	32.552	4.631	32.631	1.858
Pepper	30.662	4.366	30.767	1.859
Zebra	30.923	3.735	30.940	1.585
Average	**29.573**	**3.883**	**29.659**	**1.635**

TABLE II
MAGNIFICATION ×4 PERFORMANCE IN TERMS OF PSNR(DB) AND RUNNING TIME(S)

Images	LD-SR		DSD-SR	
	PSNR(dB)	**Time(s)**	**PSNR(dB)**	**Time(s)**
Baboon	20.210	1.463	20.291	0.628
Barbara	23.593	2.613	23.692	0.962
Bird	27.828	0.497	27.885	0.191
Bridge	23.288	1.618	23.308	0.595
Face	28.821	0.451	28.831	0.175
Lenna	28.518	1.611	28.488	0.584
Pepper	27.265	1.626	27.266	0.588
Zebra	23.490	1.359	23.527	0.500
Average	**25.377**	**1.405**	**25.411**	**0.528**

B. Proposed versus Other Algorithms

In this experiment, 2× magnification and 4× magnification of SR reconstruction were performed on the test images. The proposed DSD-SR and NLMDSD-SR methods were compared with the bi-cubic interpolation method [19], Kim's method using sparse regression and a natural image prior [20], Zhang's SR with non-local means and steering kernel regression [8], and Yang's method by simple functions [21]. The results were compared by visual quality subjectively and by numerical measurements of PSNR, mean structural similarity (MSSIM) [22], and feature similarity (FSIM) [23]. The higher the value PSNR, MSSIM, and FSIM, the more similar the reconstructed image is to the original image.

The results of 2× magnification are shown in Table III (PSNR), Table IV (MSSIM), Table V(FSIM).

TABLE III
PSNR(DB) OF THE RECONSTRUCTED IMAGES
BY DIFFERENT METHODS (2×MAGNIFICATION)
(1) BICUBIC[19] (2) KIM[20] (3)ZHANG[8]
(4) YANG[21] (5) DSD-SR (6) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)	(6)
Baboon	21.811	22.452	22.826	22.470	22.722	**22.961**
Barbara	25.528	26.312	26.713	26.457	26.703	**26.982**
Bird	31.914	33.560	35.483	33.771	34.942	**35.485**
Bridge	25.316	26.377	27.250	26.713	27.081	**27.468**
Face	30.287	30.705	31.573	30.676	31.488	**31.653**
Lenna	30.646	31.627	32.865	31.736	32.631	**32.932**
Pepper	29.469	30.249	30.665	29.661	30.767	**30.877**
Zebra	26.868	29.034	31.375	29.861	30.940	**31.721**
Average	27.730	28.790	29.844	28.918	29.659	**30.010**

TABLE IV
MSSIM OF THE RECONSTRUCTED IMAGES
BY DIFFERENT METHODS (2×MAGNIFICATION)
(1) BICUBIC[19] (2) KIM[20] (3)ZHANG[8]
(4) YANG[21] (5) DSD-SR (6) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)	(6)
Baboon	0.602	0.671	0.728	0.700	0.717	**0.743**
Barbara	0.787	0.827	0.849	0.836	0.846	**0.860**
Bird	0.948	0.969	**0.978**	0.974	0.974	0.977
Bridge	0.725	0.785	0.831	0.811	0.822	**0.840**
Face	0.809	0.834	0.857	0.845	0.854	**0.863**
Lenna	0.869	0.888	0.905	0.896	0.902	**0.907**
Pepper	0.870	0.883	0.894	0.889	0.893	**0.895**
Zebra	0.848	0.895	0.929	0.917	0.923	**0.933**
Average	0.807	0.844	0.871	0.858	0.866	**0.877**

TABLE V
FSIM OF THE RECONSTRUCTED IMAGES BY DIFFERENT METHODS (2×MAGNIFICATION)
(1) BICUBIC[19] (2) KIM[20] (3)ZHANG[8]
(4) YANG[21] (5) DSD-SR (6) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)	(6)
Baboon	0.954	0.971	0.985	0.979	0.985	**0.986**
Barbara	0.966	0.977	0.985	0.982	0.984	**0.986**
Bird	0.944	0.964	**0.979**	0.967	0.975	**0.979**
Bridge	0.963	0.977	**0.992**	0.985	0.991	**0.992**
Face	0.897	0.915	0.942	0.923	0.940	**0.947**
Lenna	0.978	0.984	0.996	0.987	0.996	**0.996**
Pepper	0.984	0.989	**0.996**	0.991	0.996	**0.996**
Zebra	0.963	0.980	**0.996**	0.990	0.995	**0.996**
Average	0.956	0.969	0.984	0.975	0.983	**0.985**

The results of 4× magnification are shown in Table VI (PSNR), Table VII (MSSIM), Table VIII (FSIM).

TABLE VI
PSNR(DB) OF THE RECONSTRUCTED IMAGES BY DIFFERENT METHODS (4×MAGNIFICATION)
(1) BICUBIC[19] (2) KIM[20] (3)ZHANG[8]
(4) YANG[21] (5) DSD-SR (6) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)	(6)
Baboon	19.890	20.113	20.351	20.032	20.291	**20.391**
Barbara	22.957	23.476	**23.969**	23.610	23.692	23.960
Bird	25.877	26.726	28.171	26.885	27.885	**28.283**
Bridge	22.219	22.719	23.428	22.974	23.308	**23.559**
Face	27.539	27.780	28.993	27.015	28.831	**29.063**
Lenna	26.720	27.487	28.799	27.104	28.488	**28.831**
Pepper	25.965	26.604	27.419	25.487	27.266	**27.464**
Zebra	21.144	22.610	23.942	22.907	23.527	**24.140**
Average	24.039	24.690	25.634	24.502	25.411	**25.711**

TABLE VII
MSSIM OF THE RECONSTRUCTED IMAGES BY DIFFERENT METHODS (4×MAGNIFICATION)
(1) BICUBIC[19] (2) KIM[20] (3)ZHANG[8]
(4) YANG[21] (5) DSD-SR (6) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)	(6)
Baboon	0.369	0.402	0.461	0.454	0.441	**0.467**
Barbara	0.625	0.662	0.697	0.689	0.678	**0.698**
Bird	0.806	0.844	0.876	0.864	0.867	**0.882**
Bridge	0.473	0.513	0.576	0.573	0.558	**0.585**
Face	0.684	0.701	0.733	0.723	0.725	**0.737**
Lenna	0.752	0.779	0.805	0.794	0.795	**0.806**
Pepper	0.778	0.803	0.818	0.808	0.812	**0.819**
Zebra	0.570	0.632	0.702	0.694	0.684	**0.714**
Average	0.632	0.667	0.709	0.700	0.695	**0.714**

TABLE VIII
FSIM OF THE RECONSTRUCTED IMAGES BY DIFFERENT METHODS (4×MAGNIFICATION)
(1) BICUBIC[19] (2) KIM[20] (3)ZHANG[8]
(4) YANG[21] (5) DSD-SR (6) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)	(6)
Baboon	0.804	0.840	0.897	0.880	0.886	**0.899**
Barbara	0.865	0.898	0.927	0.924	0.918	**0.929**
Bird	0.843	0.867	0.894	0.883	0.889	**0.900**
Bridge	0.832	0.864	0.908	0.901	0.900	**0.912**
Face	0.797	0.806	0.860	0.839	0.850	**0.864**
Lenna	0.904	0.923	**0.957**	0.927	0.951	**0.957**
Pepper	0.915	0.938	**0.958**	0.940	0.952	**0.958**
Zebra	0.800	0.857	0.911	0.888	0.893	**0.911**
Average	0.845	0.874	0.914	0.898	0.905	**0.916**

From the perspective of numerical measures, the NLMDSD-SR algorithm has the best SR reconstruction quality, Zhang's method is second-best, and the DSD-SR algorithm is third-best.

Table IX shows the 2× magnification reconstruction time used to investigate the time complexity, and Table X shows the 4× reults. As can be seen in Table IX, the proposed DSD-SR method has the best time efficiency, with an average time of 1.635s, which is approximately 7 times faster than the average speed of the Kim's method. Next is the NLMDSD-SR method, which improves the reconstructed-image quality, while the time performance remains within an acceptable range.

TABLE IX
RECONSTRUCTION TIME (S) OF DIFFERENT METHODS (2×MAGNIFICATION)
(1) KIM[20] (2)ZHANG[8] (3) YANG[21]
(4) DSD-SR (5) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)
Baboon	15.301	72.230	62.312	1.828	21.880
Barbara	18.565	144.084	83.168	2.966	35.286
Bird	4.541	32.575	19.878	0.578	6.892
Bridge	18.707	86.917	70.919	1.876	23.522
Face	3.481	22.316	16.546	0.526	6.281
Lenna	10.642	77.643	47.254	1.858	21.951
Pepper	9.940	77.046	46.725	1.859	22.039
Zebra	23.652	67.501	67.062	1.585	19.506
Average	13.104	72.539	51.733	**1.635**	19.669

TABLE X
RECONSTRUCTION TIME (S) OF DIFFERENT METHODS (4×MAGNIFICATION)
(1) KIM[20] (2)ZHANG[8] (3) YANG[21]
(4) DSD-SR (5) NLMDSD-SR

Images	(1)	(2)	(3)	(4)	(5)
Baboon	22.710	69.278	17.858	0.628	20.503
Barbara	32.811	119.252	29.005	0.962	32.319
Bird	8.921	23.737	5.546	0.191	6.416
Bridge	33.881	76.952	18.619	0.595	21.809
Face	3.282	21.375	4.767	0.175	5.770
Lenna	18.593	74.657	15.344	0.584	20.013
Pepper	20.521	74.752	15.408	0.588	20.152
Zebra	33.971	65.284	15.665	0.500	18.387
Average	21.836	65.661	15.277	**0.528**	18.171

TABLE XI
AVERAGE PSNR(DB) OF THE RECONSTRUCTED IMAGES OF SET 5, SET14 AND B100

Methods	Set14[11]	Set5[17]	B100[18]
Bicubic	23.098	24.746	23.516
Kim[20]	23.816	25.620	24.021
Zhang[8]	23.894	27.299	24.591
Yang[21]	23.627	25.409	23.574
DSD-SR	24.483	26.918	24.761
NLMDSD-SR	**24.805**	**27.390**	**25.027**

TABLE XII
AVERAGE MSSIM OF THE RECONSTRUCTED IMAGES OF SET 5, SET14 AND B100

Methods	Set14[11]	Set5[17]	B100[18]
Bicubic[19]	0.630	0.742	0.597
Kim[20]	0.667	0.787	0.628
Zhang[8]	0.690	0.811	0.665
Yang[21]	0.696	0.803	0.661
DSD-SR	0.691	0.798	0.656
NLMDSD-SR	**0.708**	**0.814**	**0.673**

TABLE XIII
AVERAGE FSIM OF THE RECONSTRUCTED IMAGES OF SET 5, SET14 AND B100

Methods	Set14[11]	Set5[17]	B100[18]
Bicubic[19]	0.804	0.810	0.706
Kim[20]	0.836	0.847	0.731
Zhang[8]	0.864	0.865	0.796
Yang[21]	0.862	0.851	0.740
DSD-SR	0.867	0.856	0.780
NLMDSD-SR	**0.880**	**0.868**	**0.801**

Table XI (PSNR), Table XII (MSSIM) and Table XIII (FSIM) shows the numerical metrics for reconstructing images by different methods. The experiment was performed on 4× magnification of SR reconstruction on benchmarks, including Set14, Set5, and B100.

From the experimental results, the average performance of the NLMDSD-SR method is the best on all datasets, followed by the DSD-SR and Zhang's methods.The reconstruction results on Set14 and B100 of DSD-SR are better than those of Zhang's method, and the results on Set5 of Zhang's method are better than those of DSD-SR.

Considering all factors, including subjective visual quality, objective assessment, and time complexity, the proposed NLMDSD-SR method obtains good performance for image SR reconstruction. In addition, the proposed DSD-SR algorithm significantly improves reconstruction speed while guaranteeing reconstruction quality, so it is suitable for applications with high-time-performance requirements.

IV. CONCLUSIONS

In this paper, an image SR reconstruction algorithm based on double-sparsity dictionaries is proposed. The algorithm combines both the advantages of analytical and learning-based dictionaries, which can ensure the quality of reconstruction and improve speed. The algorithm is suitable for applications with high-time-performance requirements. On this basis, non-local self-similar constraints are used in regularization of SR reconstruction (donoted as NLMDSD-SR), which further improve the quality of reconstructed images while keeping the runtime within an acceptable range. The algorithm offers a desirable compromise between low computational complexity and reconstruction quality.

An overcomplete DCT dictionary is chosen as a base dictionary. In fact, there are a variety of base dictionaries to choose from, such as wavelets, complex wavelets, contourlets, and bandelets. How to choose a more appropriate base dictionary is one of our planned future research directions.

The non-local self-similarity constraint improves the quality of SR reconstruction, but the time complexity of the regularization reconstruction is relatively high. Finding the regularization constraint that reduces the time complexity and has certain anti-noise performance under the premise of ensuring the reconstruction quality is another planned research direction.

REFERENCES

[1] R. R. Schultz, R. L. Stevenson, "A bayesian approach to image expansion for improved definition," IEEE Transactions on Image Processing, 1994, 3(3): 233-242.

[2] S. P. Belekos, N. P. Galatsanos, A. K. Katsaggelos, "Maximum a posteriori video super-resolution using a new multichannel image prior,"IEEE Transactions on Image Processing, vol. 19, no. 6, 2010, pp. 1451-1464.

[3] M. Mofidi, H. Hajghassem,A. Afifi, "Adaptive image super-resolution via controlled weighting coefficients of a maximum-a-posteriori estimator," Journal of Electronic Imaging, 2018, 27(4): 043031.

[4] W. Li, M. Niu, Y. Zhang, et al, "Super-Resolution Imaging of Real-Beam Scanning Radar Base on Accelerated Maximum a Posteriori Algorithm,"in Proceedings of the IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 3173-3176.

[5] L. Zhang, W. Zuo, "Image Restoration: From Sparse and Low-Rank Priors to Deep Priors, "IEEE Signal Processing Magazine, vol. 34, no. 5, 2017, pp. 172-179.

[6] A. Buades, B. Coll and J. M. Morel, "A non-local algorithm for image denoising," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, vol. 2, 2005, pp. 60-65.

[7] W. Dong, G. Shi, L. Zhang and X. Wu, "Super-resolution with nonlocal regularized sparse representation," in Proceedings of the International Society for Optical Engineering Visual Communications and Image Processing (VCIP), Huangshan, China, 2010.

[8] K. Zhang, X. Gao, D. Tao D and X. Li, "Single image super-resolution with non-local means and steering kernel regression," IEEE Transactions on Image Processing, vol. 21, no. 11, 2012, pp. 4544-4556.

[9] J. Yang, J. Wright, Y. Ma and T. Huang, "Image super-resolution as sparse representation of raw image patches," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, 2008, pp.1-8.

[10] J. Yang, J. Wright, T. Huang and Y. Ma, "Image super-resolution via sparse representation," IEEE Transactions on Image Processing, vol. 19, no. 11, 2010, pp. 2861-2873.

[11] R. Zeyde, M. Elad, and M. Protter. "On single image scale-up using sparse-representations," in Proceedings of the International Conferences on Curves and Surfaces, Tønsberg, Norway, 2010, pp. 711-730.

[12] J. Huang and P. L. Dragotti, "A deep dictionary model for image super-resolution," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, Alberta, Canada, 2018.

[13] R. Rubinstein, A. M. Bruckstein and M. Elad, "Dictionaries for sparse representation modeling," Proceedings of the IEEE, vol. 98, no. 6, 2010, pp. 1045-1057.

[14] R. Rubinstein, M. Zibulevsky and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," IEEE Transactions on Signal Processing, vol. 58, no. 3, 2010, pp. 1553-1564.

[15] N. Ai, J. Peng, X. Zhu and X. Feng, "SISR via trained double sparsity dictionaries," Multimedia Tools and Applications, vol. 74, no. 6, 2015, pp. 1997-2007.

[16] N. Ai, J. Peng, J. Wang, L. Wang and J. Qi, "Single image super-resolution by learned double sparsity dictionaries combining bootstrapping method," in Proceedings of the International Conference on Artificial Neural Networks, Alghero, Sardinia, Italy, 2017, pp. 565-573.

[17] M. Bevilacqua, A. Roumy, C. Guillemot, et al., "Low Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding," in Proceedings of the British Machine Vision Conference, 2012, pp. 135.1-135.10.

[18] R. Timofte, V. DeSmet, L. VanGool, "A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution," In: Cremers, D., Reid, I., Saito, H., Yang, M. H. (Eds.): Computer Vision – ACCV 2014. Springer, Cham, Lecture Notes in Computer Science, Vol. 9006, 2014, pp. 111–126.

[19] Liu, Jing, Zongliang Gan, and Xiuchang Zhu, "Directional Bicubic Interpolation—A New Method of Image Super-Resolution," In Proceedings of the International Conference on Multimedia Technology, Atlantis Press, 2013, pp.463-470.

[20] K. Kim ,Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior ," IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 6, 2010, pp. 1127-1133.

[21] C. Y. Yang and M. H. Yang, "Fast direct super-resolution by simple functions," in Proceedings of the IEEE International Conference on Computer Vision, IEEE Computer Society, Sydney, Australia, 2013, pp. 561-568.

[22] Z. Wang Z, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol.13, no.4, 2004, pp. 600-612.

[23] L. Zhang, L. Zhang, X. Mou and D. Zhang, “FSIM: A Feature similarity index for image quality assessment,” IEEE Transactions on Image Processing, vol. 20, no. 8, 2011, pp. 2378-2386.

Xiuxiu Liao was born in Longhui, HuNan, China, in 1983.

She is a lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China. She received her PhD degree in Computer Application Technology from South China University of Technology in 2013.

Her research interests include image processing, compressive sensing and machine learning.

Kejia Bai was born in GuiYang, HuNan, China, in 1974.

He received the PhD in system engineering from South China University of Technology, Guangzhou, China, in 2009. He is currently an associate professor in Guangdong Polytechnic Normal University, Guangzhou, China.

His research interests include image processing, pattern recognition, artificial intelligence, super-resolution, and target tracking.

Qian Zhang was born in Zibo, ShanDong, China, in 1982.

She is a lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China. She received her PhD degree in Computer Application Technology from South China University of Technology in 2013.

Her research interests include image processing and deep learning.

Xiping Jia was born in Lantian, Shanxi, China, in 1976.

He is an associate professor in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China. He received his PhD degree in Computer Application Technology from South China University of Technology in 2008.

His research interests include image processing, data mining and machine learning.

Jia Ouyang was born in XinHua, HuNan, China, in 1986.

He is a lecturer in the School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou, China. He received his PhD degree in Computer Application Technology from SUN YAT-SEN UNIVERSITY in 2014.

His research interests include differential privacy and machine learning.

Yunzhi Jiang was born in Fuyang, AnHui, China, in 1982.

He is an associate professor in the School of Mathematics and Systems Science, Guangdong Polytechnic Normal University, Guangzhou, China. He received his PhD degree in Computer Application Technology from South China University of Technology in 2012.

His research interests include theoretical foundation and application of evolutionary algorithms, image segmentation, etc.