



# Constructing multilayer locality-constrained matrix regression framework for noise robust face super-resolution

Guangwei Gao<sup>a,b,c,\*</sup>, Yi Yu<sup>b</sup>, Jin Xie<sup>d</sup>, Jian Yang<sup>d</sup>, Meng Yang<sup>e</sup>, Jian Zhang<sup>f</sup>

<sup>a</sup>Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China

<sup>b</sup>Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo, Japan

<sup>c</sup>Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, Suzhou, China

<sup>d</sup>School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

<sup>e</sup>School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China

<sup>f</sup>Global Big Data Technologies Center, University of Technology Sydney, Ultimo, NSW, Australia

## ARTICLE INFO

### Article history:

Received 16 October 2019

Revised 22 May 2020

Accepted 3 July 2020

Available online 4 July 2020

### Keywords:

Nuclear norm

Matrix based regression

Face super-resolution

Position-patch

## ABSTRACT

Representation learning methods have attracted considerable attention for learning-based face super-resolution in recent years. Conventional methods perform local models learning on low-resolution (LR) manifold and face reconstruction on high-resolution (HR) manifold respectively, leading to unsatisfactory reconstruction performance when the acquired LR face images are severely degraded (e.g., noisy, blurred). To tackle this issue, this paper proposes an efficient multilayer locality-constrained matrix regression (MLCMR) framework to learn the representation of the input LR patch and meanwhile preserve the manifold of the original HR space. Particularly, MLCMR uses nuclear norm regularization to capture the structural characteristic of the representation residual and applies an adaptive neighborhood selection scheme to find the HR patches that are compatible with its neighbors. Also, MLCMR iteratively applies the manifold structure of the desired HR space to induce the representation weights learning in the LR space, aims at reducing the inconsistency gap between different manifolds. Experimental results on widely used FEI database and real-world faces have demonstrated that compared with several state-of-the-art face super-resolution approaches, our proposed approach has the capability of obtaining better results both in objective metrics and visual quality.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

The details of facial features play crucial roles on distinguishing subjects in surveillance applications. However, the captured facial region of interest generally has low-quality because of the restraints of hardware storage, long distance to the interested object and other constraints in electric imaging system. Thus, the limited discriminative details extracted from these low quality faces significantly affect the system performance of face recognition [1–3]. Face super-resolution, also called face hallucination, is the solution that predict high-quality images from low-quality query images to provide rich facial features in the recognition system. Learning-based face super-resolution has been a very active research topic, which motivates us to work on it with more advanced algorithms.

The earliest learning-based face super-resolution work was developed by Freeman et al. [4] to predict the potential relation-

ship between low-resolution (LR) and high-resolution (HR) patches via a Markov network. After that, some machine learning-based approaches have been exploited. Shi et al. [5] designed a unified framework that combines global consistency, local sparsity, and pixel correlation to super-resolve the desired faces. A two-dimensional manifold learning based technology was presented in [6] to maintain the latent relation between the LR and HR faces in their original 2D form. By using domain knowledge and sparse recovery algorithms, Abiantun et al. [7] proposed to super-resolve faces with very low resolutions.

Although these aforementioned global models have achieved satisfactory results in most cases, they will fail in preserving some distinct individual details of the query LR faces sometimes. Ma et al. [8] presented a least square-based face hallucination method through position patch. Then, many researchers integrated the position prior to face image reconstruction [9–11] based on a fact that human faces have distinct structures. To alleviate the trouble that the least square representation may generate a nonunique solution in case that the training dictionary has bigger size, Jung et al. [12] desired to automatically choose principal training patches to

\* Corresponding author at: Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, No. 9 Wenyuan Road, Nanjing, 210023, China.

E-mail addresses: [csggao@gmail.com](mailto:csggao@gmail.com), [tsy\\_1314520@163.com](mailto:tsy_1314520@163.com) (G. Gao).

improve the reconstruction results. Jiang et al. [13] proposed a local patch-based model using neighbor embedding (NE) scheme to restore more facial details. Li et al. [14] introduced the sparse prior, which is then adopted to guide the reconstruction procedure. Jiang et al. [15,16] added the locality-constrained regularization and then obtained state-of-the-art hallucination results. Later, the similar idea is further introduced into quaternion space to hallucinate color faces [17]. Two robust locality-constrained representation models are presented by Liu et al. [18,19] to acquire the target HR faces and eliminate noise simultaneously. Shi et al. [20,21] learned the reconstruction coefficients in the kernel space and synthesized the HR image patches in the spatial domain. By exploiting the subdivided contextual sub-patches, Chen et al. [22] proposed a joint learning framework for face super-resolution.

Recently, deep learning-based models have demonstrated remarkable performance in image super-resolution applications. Dong et al. [23] did the pioneer work to introduce the convolutional neural network into image super-resolution (SRCNN) task. Cao et al. [24] resorted to deep reinforcement learning and then proposed an attention-aware face super-resolution framework. Motivated by deep CNN denoiser, Jiang et al. [25] presented a two-step face hallucination method. Yu et al. [26] developed an attribute-embedded upsampling network to reduce the ambiguity in face image super-resolution. These aforementioned deep based approaches does not take the highly structured facial prior into consideration and may be failed in face super-resolution with noise.

In previous local patch-based approaches, the core procedure can be summarized into two aspects: firstly, the LR input patch is coded as a weighted (linear or nonlinear) representation over the same position LR training patches; the desired HR patch is then rendered by integrating the corresponding HR training patches with the same reconstruction weights. The basic assumption is that the latent embedding geometry between the target HR patch and its LR counterpart is consistent in respective image space. However, this manifold assumption will not be true when the acquired LR face images are severely degraded (e.g., noisy, blurred). Furthermore, previous methods stacked the representation error into a vector and characterized them in the pixel level, neglecting the structural characteristic of the error image.

To tackle above issues, we present a multilayer locality-constrained matrix regression (MLCMR) framework in this work to capture the intrinsic structural characteristic of the representation error and keep the geometrical consistency between LR and HR manifolds simultaneously. In brief, the main contributions of our method are highlighted in the following:

- (1) We design an adaptive neighborhood selection scheme, which aims at adaptively exploiting the intrinsic geometry (similarity) of the target HR manifold to regularize the more accurate reconstruction coefficients learning in the LR manifold.
- (2) Distinguishing from existing methods imposing pixel level constraint (e.g.,  $l_1$  or  $l_2$ -norm) on the representation error, MLCMR applies the image level regularization (i.e., nuclear norm) scheme to reveal the intrinsic structure of the error, to achieve more robust results.
- (3) MLCMR iteratively updates the LR training images to simultaneously learn the representation of the LR input patch and preserve the manifold structure of the primitive HR space, with the goal of achieving the super-resolution of the desired HR patch in a much more consistent space.
- (4) To better evaluate the efficiency of our MLCMR approach, besides comparable results on controlled face images, we also conduct experiments and give some in-depth analysis on real-world faces.

This paper is a further extension of conference version [27]. In this extension, we provide more experimental comparisons and more deep-going analysis. The rest of this paper is organized as follows. Several related approaches are briefly reviewed in Section 2. In Section 3, we detail our multilayer locality-constrained matrix regression framework. Furthermore, the computational complexity and convergence analysis of the MLCMR method are also provided. Experiments performed on controlled face images and real-world face images are shown in Section 4. Section 5 shows the conclusion and future work.

## 2. Related work

Before formulating our proposed approach, in this part, we will first briefly introduce some related position-patch based face super-resolution approaches.

Denote by  $Y_L$ ,  $A_L = \{A_L^1, A_L^2, \dots, A_L^M\}$  and  $A_H = \{A_H^1, A_H^2, \dots, A_H^M\}$  the query LR image, LR and corresponding HR training sets, respectively, where  $M$  represents the scale of the training set. For each LR input, its overlapped patches located at position  $(i, j)$  are represented as  $\{Y_L(i, j) | 1 \leq i \leq R, 1 \leq j \leq C\}$ . The same size overlapped patches are also extracted from  $A_L$  and  $A_H$ , and denoted as  $\{A_L^m(i, j) | 1 \leq i \leq R, 1 \leq j \leq C\}$  and  $\{A_H^m(i, j) | 1 \leq i \leq R, 1 \leq j \leq C\}$ , respectively, where  $R$  and  $C$  denote the patch numbers. Different works convert the query patch vector into a reconstruction weights to hallucinate the target HR patch.

### 2.1. Least square estimation

Considering the structure prior in human faces, Ma et al. [8] designed a simple yet efficient least square model to collaboratively represent each query patch using all training images at the same position. The optimal representation weights affiliated to the input  $Y_L(i, j)$  can be computed as

$$\begin{aligned} x^*(i, j) &= \arg \min_{x(i, j)} \left\| Y_L(i, j) - \sum_{m=1}^M A_L^m(i, j) x_m(i, j) \right\|_2^2 \\ \text{s.t. } \sum_{m=1}^M x_m(i, j) &= 1. \end{aligned} \quad (1)$$

The analytical solution of above formula can be acquired by means of a Gram matrix.

### 2.2. Sparse representation

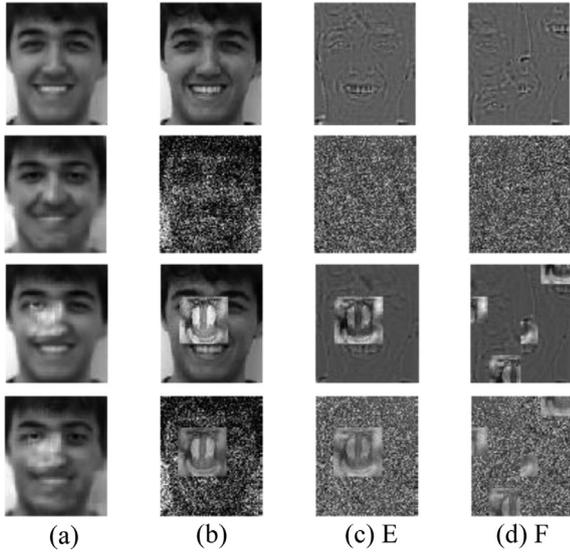
On account of the nonunique of the least square estimation, Jung et al. [12] imposed sparse regularization on the solution vector, proposed to code the input patch by adaptively selected the most relevant training patches. They propose to solve the following  $l_1$ -norm optimization problem:

$$\min_x \|x(i, j)\|_1 \quad \text{s.t.} \quad \left\| Y_L(i, j) - \sum_{m=1}^M A_L^m(i, j) x_m(i, j) \right\|_2^2 \leq \varepsilon. \quad (2)$$

Here, the  $l_1$ -norm accumulates the absolute values of a vector. Some convex optimization methods such as  $l_1$ -ls [28] can be used to solve Eq. (2). It should be pointed out that the  $l_1$ -norm constraint could stabilize the solution, as well as reveal the latent geometry of the training examples.

### 2.3. Locality-constrained representation

The sparse regularization method [12] emphasizes that sparsity of the representation weights is critical to represent the query patch, while locality property of the reconstruction vector is neglected, which is proved to be more useful than sparsity in capturing the inherent manifold geometry. Inspired by the impressive



**Fig. 1.** (a) Recovered images; (b) input images; (c) error images; (d) recombined error images.

**Table 1**  
Norm values of error image and recombined images.

Types of norm	$l_1$ -norm	$l_2$ -norm	nuclear norm	
	E and F	E and F	E	F
Noise-free	171.81	2.48	15.61	16.80
Gauss noise	1988.88	22.29	195.03	196.17
Block noise	534.95	9.96	38.59	47.86
Mixture noise	1939.81	21.69	181.44	186.09

performance of the local coordinate coding scheme [29], a locality-constrained representation [15] model was presented to maintain the sparsity and locality of training images simultaneously. The formulation of the problem can be given as follows:

$$\min_{x(i,j)} \left\| Y_L(i,j) - \sum_{m=1}^M A_L^m(i,j)x_m(i,j) \right\|_2^2 + \lambda \sum_{m=1}^M [d_m(i,j)x_m(i,j)]^2 \quad (3)$$

$$\text{s.t.} \quad \sum_{m=1}^M x_m(i,j) = 1,$$

where parameter  $\lambda$  controls the contribution of locality and the reconstruction error. Each  $d_m(i,j)$  describes the similarity between query  $Y_L(i,j)$  and the  $m$ -th atom  $A_L^m(i,j)$ . By resorting to the regularized least-squares estimator, one can easily gain the analytical solution of problem (3).

### 3. Multilayer locality-constrained matrix regression framework

#### 3.1. Main motivation

All above methods denoted the input patch as a column vector, and then usually imposed the  $l_2$  or  $l_1$  norm (pixel level) constraint on the representation error, thus neglected the structural characteristic (image level) of the error. For better illustration, we give an example in Fig. 1. Images (b) denote the input images, while images (a) are reconstructed ones from (b). Images (c) depicts the error E between images (b) and (a). We have created the error F from the error E by re-ordering the pixel positions (i.e., deranging the structural characteristics). The values of  $l_2$ ,  $l_1$  and nuclear norms between errors (c) and (d) are exhibited in Table 1.

From this table, we can find that, while the error E and the error F have always the same norm values in the cases of  $l_1$ -norm and  $l_2$ -norm, they exhibit different norm values in the case of nuclear norm, especially for images with block and mixture noises.

As we all know, the  $l_2$  and  $l_1$  norms are pixel based values, and they cannot reveal the structural characteristic of the error. This example shows that the nuclear norm of error could depict structural characteristic more effectually than  $l_2$  or  $l_1$  norm. Thus, it is feasible for us to impose nuclear norm regularization on the reconstruction errors.

#### 3.2. Locality-constrained matrix regression

We represent all the image patches in two-dimensional form in this part. Denote by  $Y_L(i,j) \in \mathbb{R}^{p \times q}$  and  $A_L^m(i,j) \in \mathbb{R}^{p \times q}$  the observed LR image patches and collected training image patches, respectively, where the size of the patch is denoted as  $p$  and  $q$ . We leave out the sign  $(i,j)$  in the next part for simplicity. As in [15], the local manifold constraint is also added to acquire more meaningful representation weights. We formulate the problem by

$$\min_x \|A_L(x) - Y_L\|_* + \alpha \|d \otimes x\|_2^2, \quad (4)$$

where  $A_L(x) = x_1 A_L^1 + x_2 A_L^2 + \dots + x_M A_L^M$ ,  $x = [x_1, \dots, x_M] \in \mathbb{R}^M \times 1$ ,  $\|\cdot\|_*$  denotes the nuclear norm (the sum of the singular values of a variable),  $\alpha$  is a parameter to control the contribution of the locality manifold constraint,  $d$  is an  $M$ -dimensional vector used to describe the similarity between  $Y_L$  and each training patch  $A_L^m$ , and  $\otimes$  represents point-wise vector product. As mentioned above, we define  $d$  as follows:

$$d_m = \|Y_L - A_L^m\|_*, \quad 1 \leq m \leq M. \quad (5)$$

Model (4) can be directly used to calculate the desired representation weights. Thus, the basic assumption is that the geometry structure between HR manifold and the LR manifold are similar. Due to the noisy, blurry degradations in real-world application scenarios and the one-to-many correspondence between the LR query patch and corresponding HR one, this hypothesis will not be held anymore. To this end, we propose to learn the representation of the LR input and meanwhile preserve the geometry structure of original HR manifold.

Particularly, from the perspective of graph learning theory, with regard to any two patches from the original HR manifold, if their local distance is small, it is more likely that they may exhibit the close representation weights. Next, we will utilize the intrinsic geometry  $G$  from the HR manifold to induce the representations learning in the observed LR patch manifold.

At first, for each target HR patch  $Y_H$ , we should obtain its neighbor structure in the original HR manifold. The most commonly used  $K$  nearest-neighbor search strategy is generally applied to determine the fixed neighborhoods. Nevertheless, it is usually intractable to predefine the appropriate parameter  $K$  in practical application scenarios. It is also unreasonable to assign the same  $K$  to represent various face structures. To avoid the difficulty to conduct parameter selection and adaptively choose similar patches for reconstruction, we introduce a parameter-free adaptive neighborhood selection strategy as follows. Denote by  $N(Y_H)$  the set of near-est neighbors of  $Y_H$ , then we define  $N(Y_H)$  as

$$N(Y_H) = \left\{ A_H^b \mid \text{if } d(Y_H, A_H^b) \leq \frac{1}{M} \sum_{j=1}^M d(Y_H, A_H^j), b = 1, \dots, M \right\}, \quad (6)$$

where  $d(Y_H, A_H^j)$  denotes the distance [defined in Eq. (5)] between  $Y_H$  and  $A_H^j$ , and  $(1/M) \sum_{j=1}^M d(Y_H, A_H^j)$  is actually the mean of all  $d(Y_H, A_H^j)$  ( $j = 1, \dots, M$ ). For better illustration, we draw the distributions of  $d$  and the number of neighbors [calculated by Eq. (6)] for different target patches in Fig. 2, from which we can observe that for each target HR patch, our strategy could adaptively select similar patches for more accurate reconstruction.

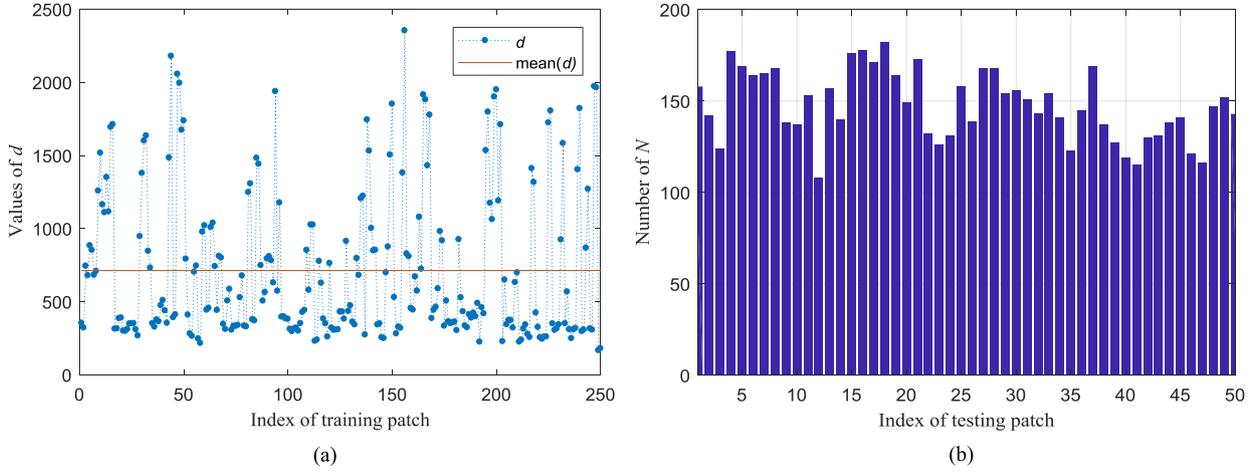


Fig. 2. The distributions of (a) the distance  $d$  and (b) the number of neighbors.

With the above adaptive neighborhood selection strategy, the penalty weighting matrix  $P = (p_{ij})_{K \times K}$  is defined as follows ( $K$  is the number of neighbors in  $N(Y_H)$ ):

$$p_{ij} = \exp\left(-\frac{\|A_H^i - A_H^j\|_2^2}{\sigma_1 \sigma_2}\right), \quad A_H^i, A_H^j \in N(Y_H). \quad (7)$$

Here,  $\sigma_1 = (1/K) \sum_{l=1}^K \|A_H^i - A_H^l\|_2^2$  and  $\sigma_2 = (1/K) \sum_{l=1}^K \|A_H^j - A_H^l\|_2^2$  ( $l = 1, \dots, K$ ). In our setting of symmetric weighting matrix  $P$ , a heavy penalty will be induced if the weights of  $A_H^i$  and  $A_H^j$  are very different. Therefore, minimizing it aims to encourage that if  $A_H^i$  and  $A_H^j$  are close to each other in the HR patch manifold, then so should be  $x^i$  and  $x^j$ . With simple formulation, we have:

$$\frac{1}{2} \sum_{i,j=1}^K (x_i - x_j)^2 p_{ij} = x^T (D - P)x = x^T Lx, \quad (8)$$

where diagonal matrix  $D$  has the entries of being row sums of  $P$ ,  $D_{ii} = \sum_j p_{ij}$  (or column because  $P$  is symmetric), and variable  $L = D - P$  denotes the Laplacian matrix [10].

Without loss of generality, we also use  $A_L$  to denote the adaptively selected neighbor patches set of  $Y_L$  (It should be noted that different  $Y_L$  have different  $K$ ). So far, our locality-constrained matrix regression (LCMR) model can be formulated as

$$\min_x \|A_L(x) - Y_L\|_* + \alpha \|d \otimes x\|_2^2 + \beta x^T Lx. \quad (9)$$

Here, parameter  $\beta$  is used to balance the contributions of the compatible neighbors.

### 3.3. Optimization

The minimization problem (9) could be rewritten as

$$\begin{aligned} \min_x & \|E\|_* + \alpha \|d \otimes x\|_2^2 + \beta x^T Lx \\ \text{s.t.} & A_L(x) - Y_L = E \end{aligned} \quad (10)$$

The alternating direction method of multipliers (ADMM) method [30–33] can be exploited to solve the above formulation. To solve problem (10), let us define the next augmented Lagrange function:

$$\begin{aligned} L_\mu(x, E) = & \|E\|_* + \alpha \|d \otimes x\|_2^2 + \beta x^T Lx \\ & + \text{tr}(Z^T (A_L(x) - Y_L - E)) + \frac{\mu}{2} \|A_L(x) - Y_L - E\|_F^2, \end{aligned} \quad (11)$$

where  $\mu > 0$  is a penalty variable,  $\alpha$  and  $\beta$  are above mentioned regularization parameters,  $\text{tr}(\bullet)$  is the trace operator, while  $Z$  is the

Lagrange multiplier. Following a series of simple algebraic derivation, we can rewrite optimization problem (11) as

$$\begin{aligned} L_\mu(x, E) = & \|E\|_* + \alpha \|d \otimes x\|_2^2 + \beta x^T Lx \\ & + \frac{\mu}{2} \|A_L(x) - Y_L - E + \frac{1}{\mu} Z\|_F^2 - \frac{1}{2\mu} \|Z\|_2^2. \end{aligned} \quad (12)$$

The above problem is unconstrained. So we can optimize the variable one by one by fixing other variables until pre-set terminal conditions are reached.

#### Updating $x$

Given  $E$  together with other variables, our optimization problem can be rewritten as

$$L_\mu(x) = \frac{2\alpha}{\mu} \|d \otimes x\|_2^2 + \frac{2\beta}{\mu} x^T Lx + \left\| Y_L + E - \frac{1}{\mu} Z - A_L(x) \right\|_F^2. \quad (13)$$

Inspired by the description in [8], we can acquire the closed solution of the above regularized least square problem via

$$x^{k+1} = (G + \tau D^2 + \delta L) \backslash \text{ones}(K, 1), \quad (14)$$

Where  $k$  denotes the iteration index,  $\tau = 2\alpha/\mu$ ,  $\delta = 2\beta/\mu$ , column vector  $\text{ones}(K, 1)$  has a size of  $K \times 1$  with the entries of all ones, diagonal matrix  $D$  has a size of  $K \times K$  with the entries of  $D_{mm} = d_m$ , the symbol “ $\backslash$ ” represents the left matrix division, and the covariance matrix  $G = C^T C$  together with

$$C = \text{Vec}\left(Y_L + E - \frac{1}{\mu} Z\right) \text{ones}(M, 1)^T - H. \quad (15)$$

Here, matrix  $H = [\text{Vec}(A_L^1), \dots, \text{Vec}(A_L^K)]$ , operator  $\text{Vec}(\cdot)$  transforms a two-dimensional matrix into a column vector. Then the analytical solution of  $x$  can be obtained by rescaling it to satisfy  $\sum_{m=1}^K x_m = 1$ .

#### Updating $E$

Given  $x$  together with other variables, we can rewrite our optimization problem as

$$E^{k+1} = \arg \min_E \left( \frac{1}{\mu} \|E\|_* + \frac{1}{2} \left\| E - \left( A_L(x) - Y_L + \frac{1}{\mu} Z \right) \right\|_F^2 \right). \quad (16)$$

Its optimal solution can be acquired by [34]:

$$E^{k+1} = U \frac{1}{\mu} [S] V, \quad (17)$$

where  $(U, S, V^T) = \text{svd}(A_L(x) - Y_L + \frac{1}{\mu} Z)$ , operation  $\text{svd}$  denotes the singular value decomposition.

**Algorithm 1**

ADMM algorithm to solve problem (10).

**Input:** Training patch matrices  $A_L^1, \dots, A_L^M \in \mathbb{R}^{p \times q}$  and observed patch matrix  $Y_L \in \mathbb{R}^{p \times q}$ , parameters  $\alpha$  and  $\beta$ , and the terminal condition parameter  $\varepsilon$ .

**Initialization:**  $x = 0, E = Z = 0, \varepsilon = 10^{-6}$

1: Step1: Fix others and update  $x$  according to (14);

2: Step2: Fix others and update  $E$  according to (17);

3: Update multiplier  $Z$  according to (19);

4: Check for convergence:

$$\|A_L(x) - Y_L - E\|_\infty > \varepsilon$$

Go to step 1;

5. **Output:** Optimal representation weights  $x^{k+1}$

The singular value thresholding operator  $T(\bullet)$  is denoted as

$$T_{\frac{1}{\mu}}[S] = \text{diag}\left(\left\{\max\left(0, s_{j,j} - \frac{1}{\mu}\right)\right\}_{1 \leq j \leq r}\right), \quad (18)$$

here  $r$  count the rank of  $S$ .

Finally, the Lagrange multiplier  $Z$  can be updated by

$$Z^{k+1} = Z^k + \mu(A_L(x^{k+1}) - Y_L - E^{k+1}), \quad (19)$$

where  $\mu > 0$  is the penalty variable.

The detailed procedure to solve problem (10) via ADMM is listed in Algorithm 1.

**3.4. Face super-resolution via LCMR**

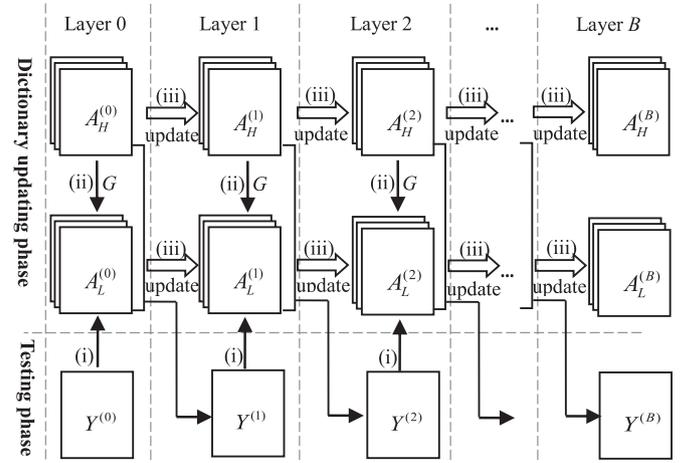
Regard to the super-resolution problem, the training sets consist of HR and corresponding LR pairs. Let  $A_H = \{A_H^1, A_H^2, \dots, A_H^M\}$  the HR training face images and  $A_L = \{A_L^1, A_L^2, \dots, A_L^M\}$  the LR counterparts, the task of face super-resolution is to recover the desired HR candidate  $Y_H$  from observed LR query  $Y_L$ .

As in [15], we first utilize the same dividing strategy to split the query LR input and each face pair in the training set into overlapped patches, which are represented as  $Y_L(i,j)$ ,  $A_L^m(i,j)$  and  $A_H^m(i,j)$ , respectively. With regard to each patch  $Y_L(i,j)$  in LR input image, we first calculate its penalty weighting matrix from the same position training HR patches by means of the desired  $Y_H(i,j)$ , with the aim of exploiting the geometry structure of the target HR manifold to guide the representation weights learning in the LR manifold. Secondly, by using LCMR,  $Y_L(i,j)$  is approximated as a weighted combination over the same position adaptively selected neighbor LR training patches  $A_L^m(i,j)$  ( $m = 1, \dots, K$ ). Then, by employing the same representation weights on the related HR training patch matrices, we would gain the target HR patch  $Y_H(i,j)$ . Finally, we can acquire the desired HR face via integrating all the target HR patches and averaging overlapped pixel values according to their corresponding positions.

**3.5. Constructing MLCMR for efficient face super-resolution**

The representation weights and the desired HR patch are obtained separately in the aforementioned super-resolution scheme. It only takes into account the LR manifold while neglecting the use of the geometry information hidden in the original HR manifold, which is unaffected by the image degradation process [4]. Inspired by multilayer [13] and deep learning [24, 35] solutions, we also propose to construct a multilayer locality-constrained matrix regression framework to update the representation weights and the LR training set step by step. The flowchart of our proposed method is drawn in Fig. 3, where the reconstruction weights and the LR training set are updated alternately.

**Weights updating:** Without loss of generality, the number of layers is defined as  $B$ . Therefore, for the  $b$ -th layer, the optimal re-



**Fig. 3.** Flow diagram of MLCMR method. (i) Calculating the desired representation weights  $x$ ; (ii) Preserving the HR manifold geometry  $G$  by penalty weighting matrix; (iii) Updating the LR and HR training sets.

**Algorithm 2**

Robust face super-resolution via MLCMR.

**Input:** HR training set  $A_H = \{A_H^1, \dots, A_H^M\}$ , corresponding LR one  $A_L = \{A_L^1, \dots, A_L^M\}$ , input LR image  $Y_L$ , layer number  $B$ .

1: The LR input and each training face are portioned into  $N$  overlapped blocks;

2: **For** each layer:  $b = 0$  to  $B$ ;

3: **For** each input patch:

a) Calculate the similarity  $d^{(b)}$  and the penalty weighting matrix  $P^{(b)}$  for the LR input image patch matrix by formulate (5) and (7) respectively;

b) Calculate the optimal representation weights  $x^{(b)}$  of the LR input patch over the same position LR dictionary  $A_L^{(b)}$  by formulation (20) using Algorithm 1;

c) Construct the intermediate HR patch  $Y_H$  by  $Y_H^{(b+1)} = A_H^{(b)}(x^{(b)})$  and as an input of the next layer;

d) Update the next layer dictionaries  $A_L^{(b+1)}$  using *leave-one-out* scheme.

4: **End for**

5: **End for**  $b$

6: Integrate all the super-resolved HR patch matrices to form the desired HR image  $Y_H$ ;

**Output:** The super-resolved HR face image  $Y_H$ .

construction coefficients of input patch at position  $(i,j)$  can be obtained using the following formulation:

$$x^{(b)} = \arg \min_{x^{(b)}} \|A_L^{(b)}(x^{(b)}) - Y_H^{(b)}\|_* + \alpha \|d^{(b)} \otimes x^{(b)}\|_2^2 + \beta (x^{(b)})^T L(x^{(b)}). \quad (20)$$

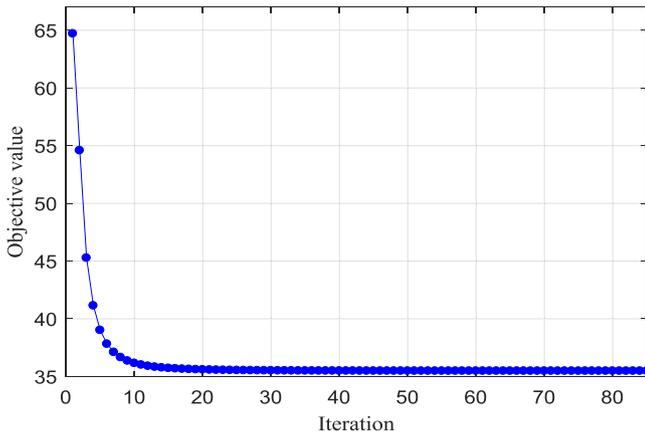
It should be noted that  $Y_H^{(0)} = Y_L$ ,  $A_L^{(0)} = A_L$ . After that, the desired HR patch can be updated by  $Y_H^{(b+1)} = A_H^{(b)}(x^{(b)})$ , where  $A_H^{(b)}(x^{(b)}) = A_H^1(i,j)x_1^{(b)}(i,j) + A_H^2(i,j)x_2^{(b)}(i,j) + \dots + A_H^K(i,j)x_K^{(b)}(i,j)$ .

**Training sets updating:** The LR training set in the next layer is updated using the method in Section 3.3 by a so-called *leave-one-out* scheme: one image from the LR training set is selected as input in turn, while the remaining is regarded as the dictionary. Then, the LR training set is updated by traversing the whole database. In this way, the inconsistency between the HR and LR manifolds is gradually reduced, thus the HR and LR weight pairs get more coupled as well.

By alternately updating the representation weights and LR dictionary, we can construct multilayer LCMR to boost the super-resolution performance. Algorithm 2 details the entire face super-resolution algorithm.

**Table 2**  
Ablation study of each module in the proposed method.

Methods	Metric	MLCMR_NL	MLCMR_NC	MLCMR_NN	MLCMR
Noise-	PSNR	32.25	32.58	32.92	<b>33.09</b>
	SSIM	0.9035	0.9164	0.9212	<b>0.9291</b>
Gauss noise	PSNR	26.28	26.72	26.90	<b>27.01</b>
	SSIM	0.7846	0.7956	0.8084	<b>0.8191</b>
Block noise	PSNR	24.34	25.07	25.27	<b>25.78</b>
	SSIM	0.8195	0.8298	0.8188	<b>0.8484</b>
Mixture noise	PSNR	23.14	23.88	24.14	<b>24.69</b>
	SSIM	0.7513	0.7635	0.7736	<b>0.7852</b>



**Fig. 4.** Convergence analysis of Algorithm 1.

### 3.6. Complexity and convergence analysis

We desire to perform the complexity analysis on the proposed MLCMR method here. Because the dictionary updating process can be conducted offline, thus, we only analyze the operational cost of the online super-resolution procedure in our proposed approach. The major time complexity of Algorithm 1 is taken on Step 1 and Step 2, which involves three parts: (i) adaptive neighborhood selection; (ii) representation weights calculating, and (iii) the SVD operation. It can be found that there are totally four primary elements influencing the time cost in parts (i), (ii) and (iii): the training size  $M$ , the patch numbers  $N$  in one image, the number of the neighbors  $K$ , and the dimension  $d^2$  of one patch.

Inspired by [13], the adaptive neighborhood selection step costs  $O(Kd^2M)$  and the representation weights calculating step costs  $O(d^2K^3)$ . Thus, Step 1 in Algorithm 1 costs  $O(d^2K^3 + Kd^2M)$ . For a matrix in  $\mathbb{R}^d \times d$ , its exact SVD costs  $O(d^3)$ . Therefore, for each patch, it costs  $O(d^3 + d^2K^3 + Kd^2M)$  in Algorithm 1. Finally, by considering the maximum iteration number  $maxIter$ , layer number  $B$  and patch number  $N$ , the total computational complexity of MLCMR framework is about  $O(maxIter(d^3 + d^2K^3 + Kd^2M)NB)$ .

For the convex problem, the convergence properties of the ADMM can be guaranteed. One can refer to [36] for more details. Since it is difficult to guarantee that matrix  $L$  is positive definite, our formulation may non-convex. In this case, it would be difficult to guarantee the convergence of Algorithm 1 in theory. However, throughout our tests, we observe that Algorithm 1 can converge asymptotically. Fig. 4 shows a convergence example of our method, where the objective value tends to be stable after nearly 20 steps.

## 4. Experimental results and discussions

We compare our MLCMR method with some state-of-the-art face super-resolution approaches in this part: sparse representation (SR) model [12], locality-constrained representation

(LCR) model [15], locality-constrained iterative neighbor embedding (LINE) model [13], locality-constrained bi-layer representation (LCBR) model [18]. Furthermore, deep reinforcement learning (DRL) model [24] and deep CNN (DCNN) model [25] are also used for comparison.

To evaluate the effectiveness and efficiency of our proposed algorithm, extensive experiments are performed on controlled face databases (the FEI database [37]) and real-world face databases (the CMU+MIT face database [38]). Besides the qualitative comparisons of each method, the values of Structural SIMilarity (SSIM) and peak signal-to-noise ratio (PSNR) [39] are also utilized to quantitatively investigate the qualities of super-resolved faces.

### 4.1. Dataset description

We first conduct simulation experiments on the FEI face database. Each face is manually aligned based on the positions of three points: center of the mouth, centers of the right and left eyeballs (Fig. 5 shows some example images). For the FEI database, it gathers 400 frontal samples from 200 persons. Thus, each person has two samples: one with a neutral expression while the other is smiling. All the face regions are cropped and normalized to the size of  $120 \times 100$ . We randomly pick 40 face samples for testing in this paper, and the other 250 face samples for training. The HR faces are first smoothed (with window size  $4 \times 4$ ) and then resized by a down-sample factor of 4 to form their LR counterparts, who have size of  $30 \times 25$ .

### 4.2. Ablation study

In this part, we study the effects of each module in our method. Each test image has four degradations: without noise; with Gauss, block or mixture noise (see Section 4.3 for details). Compared to MLCMR,

MLCMR\_NL removes locality constraint, MLCMR\_NC removes consistency constraint and MLCMR\_NN replaces nuclear norm with the  $l_2$  norm. Table 2 presents the average PSNR and SSIM results. One can observe that locality constraint is important since MLCMR outperforms MLCMR\_NL. Moreover, MLCMR has better performance than MLCMR\_NC, indicating that the adaptively selected target HR manifold indeed compensates useful information in the proposed method. Further, MLCMR obtains better performance than MLCMR\_NN, especially in face of block and mixture noise, which reveals that the nuclear norm regularization is also useful to capture the inherent structure of the reconstruction error.

### 4.3. Result comparisons

For fair evaluations, we tune the parameters in comparative approaches to gain their possible best performance. The size of the HR patch is  $12 \times 12$  pixels, and there are 4 columns (or rows) between adjacent patches. Thus, the size of the related LR patch is

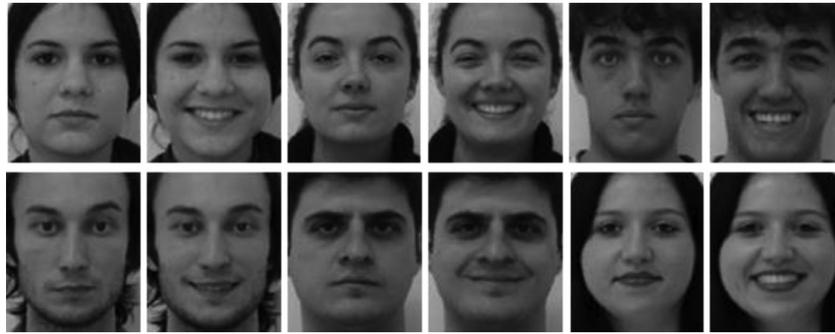


Fig. 5. Some face images from FEI face database.



Fig. 6. Comparisons of different face super-resolution approaches in noise-free case. The first to the ninth columns are successively the query LR images, the results of SR method, LCR method, LINE method, DRL method, DCNN method, LCBR method, the proposed MLCMR, and the referenced HR images.

$3 \times 3$  pixels with one column (or row) overlapped with its adjacencies. For the SR method, when seeking an optimal solution, we set the error tolerance as 0.001. For the LCR method, for best performance, the suitable balance variable is set to 0.04. For the LINE method, the suitable locality parameter, the iteration number, and layer number are set to  $1e-5$ , 4 and 3, respectively. For the DRL method, 320 face samples are used as training set, 40 samples used as validation set and 40 samples used as the testing set. For the RLCBR method, the corresponding parameters are set to 1 and 0.04.

1) *Experiments without noise*: Several example super-resolved faces by compared approaches are shown in Fig. 6. From left to right, the first column denotes the acquired LR images, the second to the seventh columns are the super-resolved HR images by six comparable methods, while the last column is the ground truth references. Some “ghosting” effects are obtained by SR method on locations around margins of the mouth and face contours. LCR and LINE methods have similar subjective visual performance. Multi-layer based methods (DRL, RLCBR and our MLCMR) yield better subjective performance than previous single-layer based ones. Es-



Fig. 7. Comparisons of different face super-resolution approaches for LR faces corrupted by the Gaussian noise. The first to the ninth columns are successively the query LR images, the results of SR method, LCR method, LINE method, DRL method, DCNN method, LCBR method, the proposed MLCMR, and the referenced HR images.

Table 3

The objective compared results of respective approaches with various noises on the FEI database.

Methods	Metric	SR	LCR	LINE	DRL	DCNN	LCBR	MLCMR
Noise-free	PSNR	31.89	32.38	32.61	32.67	32.70	32.78	<b>33.09</b>
	SSIM	0.9087	0.9114	0.9158	0.9235	0.9176	0.9186	<b>0.9291</b>
Gauss noise	PSNR	24.80	26.49	26.64	25.01	26.70	26.79	<b>27.01</b>
	SSIM	0.6414	0.7788	0.8036	0.7481	0.8055	0.8065	<b>0.8191</b>
Block noise	PSNR	22.98	24.43	24.89	22.44	24.90	24.92	<b>25.78</b>
	SSIM	0.7985	0.8234	0.8073	0.7327	0.8210	0.8310	<b>0.8494</b>
Mixture noise	PSNR	21.14	23.22	23.66	21.62	23.70	23.77	<b>24.69</b>
	SSIM	0.5775	0.7431	0.7563	0.6856	0.7640	0.7739	<b>0.7852</b>

pecially, our proposed MLCMR method can obtain more reasonable performance and have more similar visual impressions with the target HR images. Furthermore, the average objective compared results of all the 40 testing images are tabulated in Table 3. Our MLCMR method can obtain the best quantitative results since it incorporates the matrix regression and takes the manifold structure of the desired HR patch space into account.

2) *Robust to noise*: Previous experiments simply assume that the acquired LR faces are noiseless. However, due to the complicated imaging environment, the influence of noise inevitably cannot be ignored in the observed LR faces. Thus, we verify the robustness of MLCMR on noise situations in this part. Experiments

are conducted on three cases: test images are corrupted by zero-mean Gaussian noise with deviations 0.05; test images are occluded by a square “baboon” image block in randomly located position with an occlusion level of 10%; test images are corrupted by a mixture noises (gauss noise and block occlusion). For a detailed comparison of different methods, Fig. 7–9 show several super-resolved face images to depict the visual image quality. Table 3 also tabulates the average evaluation measures obtained by respective methods. We can see that in noisy situations, the qualities of all these super-resolved images are reduced drastically. The reason may be that in the noise scenario, SR seeks the most similar training patches other than alleviating the influence of noise. Thus, the



**Fig. 8.** Comparisons of different approaches for LR faces corrupted by block occlusion. The first to the ninth columns are successively the query LR images, the results of SR method, LCR method, LINE method, DRL method, DCNN method, LCBR method, the proposed MLCMR, and the referenced HR images.

SR based method does not achieve satisfying performance. The LCR and LINE methods obtain more reliable results with less ghosting effect than the SR method. This is because, in LCR and LINE methods, the manifold structure is incorporated, which can enhance the discriminative ability of representation weights. Moreover, the deep based DRL and DCNN do not take the highly structured noise prior into consideration, which has been verified to be crucial in face super-resolution tasks. The RLCBR method can obtain smoother results with less noise effect by improving the locality-constrained model with a weight vector. Compared with other methods, by applying the nuclear norm to regularize the reconstruction error and iteratively using the manifold structure of the desired HR space to induce the reconstruction weights learning in the LR space, the desired faces generated by our MLCMR can reveal more feature details around the mouth, eyes and face contour, together with best visual quality and highest quantified results.

#### 4.4. Compared results on real-world images

In aforementioned experiments, the input smoothed and down-sampled LR testing face images come from its original HR counterparts. However, due to the difficulties of fitting the image degradation process in real-world applications, the real spatial feature cor-

relation between the degraded LR image and the HR one cannot be simply described by only the input LR inputs. Therefore, similar experiments are performed on the real-world CMU+MIT face dataset [38] to further verify the efficiency of our method.

All the testing samples are manually extracted from the CMU+MIT face database and aligned to the training images in the FEI database in accordance with the center points of the two eyes. Then, by simply using Bicubic interpolation, these aligned images are resized to have size  $30 \times 25$  (to have the same size as the training samples). Some extracted LR samples are exhibited in Fig. 10. Unlike those down-sampled LR images by the known degradation process, due to the complexity imaging conditions, these images in this experiment natively have low-resolution. Fig. 11 lists some reconstructed HR images by different methods in face of four degradation modes (i.e., without noise; with Gauss, block or mixture noise). In the noise-free case, all methods obtain similar performance. However, in noisy cases, we can observe that our method yields the best performance with less ghosting effect around the mouth, eyes and face contour. To better evaluate the efficiency of our method, we treat the super-resolved faces obtained by our method in the noise-free case as the “ground truth”, and show the referenced quantitative compared results in Table 4.



**Fig. 9.** Comparisons of different approaches for LR faces corrupted by mixture noise. The first to the ninth columns are successively the query LR images, the results of SR method, LCR method, LINE method, DRL method, DCNN method, LCBR method, the proposed MLCMR, and the referenced HR images.

**Table 4**

The objective compared results of respective approaches with various noises on the CMU+MIT database.

Methods	Metric	SR	LCR	LINE	DRL	DCNN	LCBR	MLCMR
Gauss noise	PSNR	20.58	24.43	25.19	23.93	25.30	25.36	<b>25.60</b>
	SSIM	0.5236	0.7325	0.8048	0.7253	0.8104	0.8203	<b>0.8231</b>
Block noise	PSNR	22.54	23.33	23.64	22.28	23.75	23.81	<b>24.36</b>
	SSIM	0.7914	0.8403	0.8592	0.7606	0.8640	0.8684	<b>0.8786</b>
Mixture noise	PSNR	18.90	21.07	21.43	20.86	21.50	21.57	<b>22.28</b>
	SSIM	0.4660	0.6584	0.7378	0.6149	0.7421	0.7432	<b>0.7579</b>

#### 4.5. Parameter analysis

We survey the effect of the locality parameter  $\alpha$ , consistency parameter  $\beta$  and layer number  $B$  in this part.

1) *The effect of the regularization parameters:* The performance of our method using different parameter settings is evaluated to further test the effect of parameters. In these tests, we just adjust one parameter while given the other one. In this experiment, we use the *leave-one-out* strategy: one image from the training set is selected as input in turn, and the rest of the training set is treated as the dictionary. The LR input is occluded by a square “baboon” image block in randomly located position with an occlusion level of 10%. Fig. 12 plots the aver-

age PSNR and SSIM results. From the results, one can find that as  $\alpha$  grows, the performance of MLCMR first increases and next decreases. In our experiments, we set  $\alpha$  around 10. When the values of consistency parameter  $\beta$  vary from 0.5 to 1, our proposed method can always obtain stable performance.

2) *The effect of layer number:* Fig. 13 lists several super-resolved results of our method using different layer numbers. For comparison, the result of Bicubic Interpolation (BI) is used as the baseline. L1 and L2 respectively denote the single-layer MLCMR and the two-layer MLCMR, and so on. From the visual results, we can observe that the results of BI have blurred details while our model can recover more facial details with more layer numbers. We also tabulate the average quantitative values versus differ-



**Fig. 10.** Some extracted LR images on the real-world CMU+MIT face dataset. For each example, the extracted and aligned LR input is at left, and the super-resolved result of our method is at right.

**Table 5**

The average quantitative values with different layer numbers.

Layer number	BI	L1	L2	L3	L4
PSNR	27.42	32.51	32.97	33.09	33.10
SSIM	0.8411	0.9210	0.9254	0.9291	0.9300

ent layer numbers in Table 5 for comparison. The superiority of our MLCMR approach becomes distinct when the layer number increases. Nevertheless, more layer numbers need more computational cost. We set layer number as 3 in the following testing.

#### 4.6. Computational time

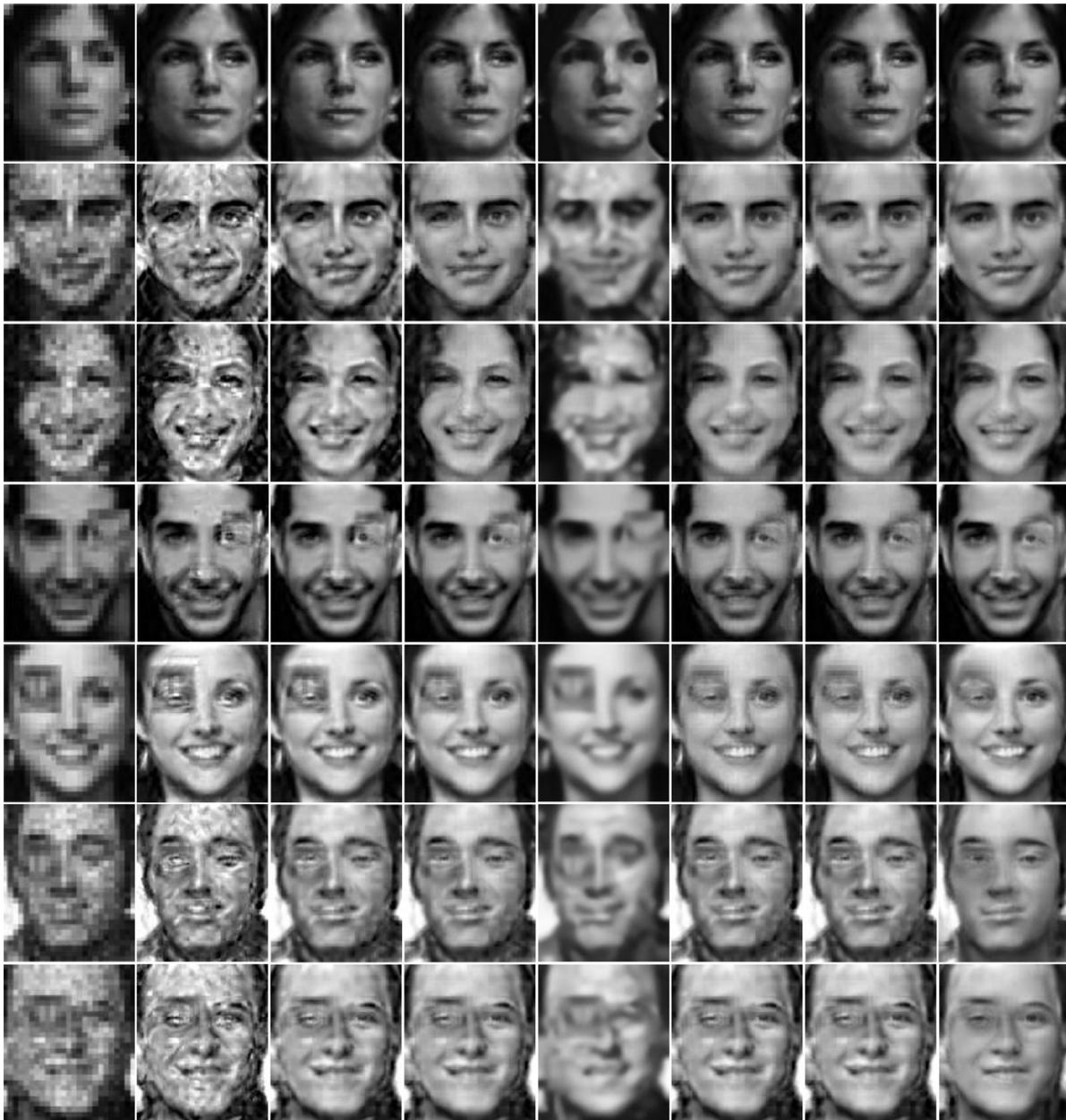
In this section, we evaluate the computational time of respective approaches. The experiments are conducted with the following configurations: the CPU is Intel Core i7 -6700 and the RAM is 16 GBytes. For simplicity, we only perform the comparisons on one face sample from the FEI database. The computational cost of respective approaches is tabulated in Table 6. We can observe that the DRL method has the least computational consumption once the network is pre-trained. Due to the iterative updating scheme, LINE and RLCBR methods require similar computational time. The LCR method is faster than other position-patch based methods since it only requires a few matrix multiplications and additions steps.

On account of the iterative strategy in representation learning, our proposed MLCMR requires much more time in comparison with other approaches.

#### 4.7. Recognition tests

We investigate the recognition performance using the super-resolved HR faces. The extended Yale B face dataset [40], which contains 38 humans, is used here. The 14 images of a subject with normal lighting conditions are selected. With regard to each subject, we pick 7 samples for training and the rest without or with noise (i.e., Gauss, block or mixture noise) are used for testing. Each HR face has size of  $96 \times 84$ , and its corresponding LR one is  $24 \times 21$ . In this part, each query LR face is super-resolved by the *leave-one-out* strategy, and those samples from the testing subject are excluded from the training faces. For simplicity, the matrix regression-based classifier [30] is exploited to recognize super-resolved faces.

Table 7 tabulates the recognition rates (RAs) of different super-resolution approaches in four degradation modes. From Table 7, we can find that the super-resolved faces by our MLCMR method can achieve higher RAs than other methods in block and mixture noise cases. Our proposed MLCMR can retain more feature details around the mouth, eyes and face contour via using the nuclear norm regularization and iteratively applying the manifold structure of the desired HR space to regularize weights learning in the LR space.



**Fig. 11.** Comparisons of different approaches for LR images extracted from real-world CMU+MIT face dataset. The first to the eighth columns are successively the query LR images, the results of SR method, LCR method, LINE method, DRL method, DCNN method, LCBR method, the proposed MLCMR.

**Table 6**

Comparison of computational time (seconds) on the FEI face database.

Methods	SR	LCR	LINE	DRL	DCNN	LCBR	MLCMR
Time	3.2548	1.2331	2.4652	0.3387	1.1543	2.5875	6.5482

**Table 7**

Recognition rates of respective approaches on the Extended Yale B face database.

Methods	SR	LCR	LINE	DRL	DCNN	LCBR	MLCMR
Noise-free	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Gauss noise	0.88	0.95	0.96	0.90	0.97	0.97	0.98
Block noise	0.85	0.92	0.93	0.86	0.93	0.94	0.97
Mixture noise	0.80	0.87	0.90	0.84	0.92	0.92	0.95

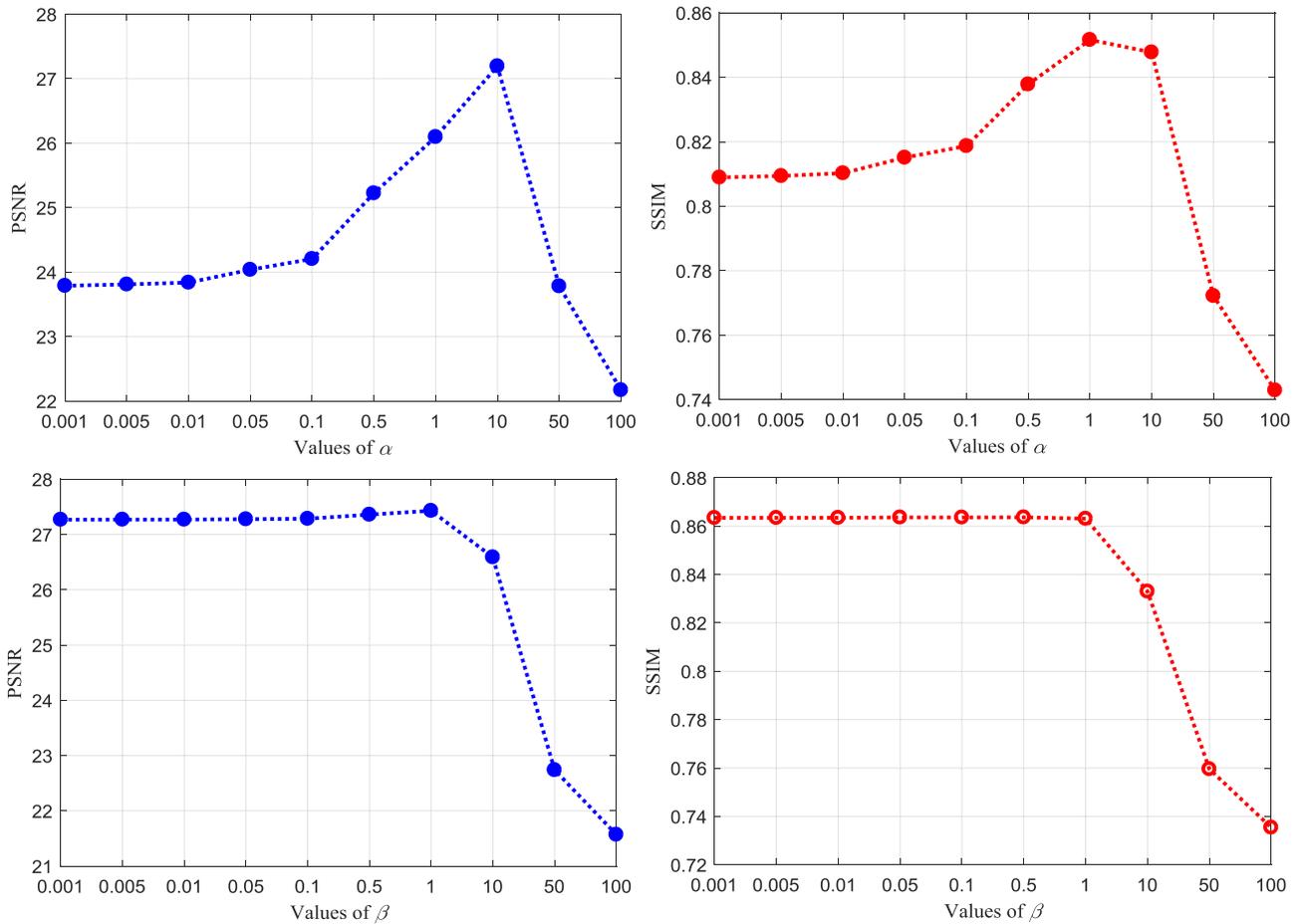


Fig. 12. The averaged PSNR and SSIM values of MLCMR utilizing different indexes of  $\alpha$  (first row) and  $\beta$  (second row).



Fig. 13. Face super-resolution with different layer numbers. The first to the seventh columns are successively the query LR inputs, the super-resolved results of B1's, L1's, L2's, L3's, L4's and the referenced HR images.

## 5. Conclusions and future work

We design a novel noise robust face image super-resolution model called multilayer locality-constrained matrix regression (MLCMR) in this paper. MLCMR uses nuclear norm regularization to capture the structural characteristic of the representation residual. Also, MLCMR is proposed to apply an adaptive neighborhood

selection scheme to find the HR counterpart that is compatible with its neighbors. Furthermore, our proposed MLCMR framework iteratively utilizes the manifold structure of the primordial HR space to guide the representation weights learning in the degraded LR space. Experiments on both controlled and the real-world faces have shown the effectiveness of our method.

In practical surveillance applications, the pose and misalignment variations cannot be neglected. As for those seriously degraded LR images, the deep feature representation of them should be well investigated. In addition, how to incorporate the facial structure prior to deep models is also our future work.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Project nos. 2018AAA0100102 and 2018AAA0100100; the National Natural Science Foundation of China under Grant nos. 61972212, 61772568, 61972213, 61806098 and 61833011; the Six Talent Peaks Project in Jiangsu Province under Grant no. RJFW-011; the Natural Science Foundation of Jiangsu Province under Grant no. BK20190089; and the Open Fund Project of Provincial Key Laboratory for Computer Information Processing Technology (Soochow University) (no. KJS1840).

### References

- [1] X.-Y. Jing, X. Zhu, F. Wu, R. Hu, X. You, Y. Wang, H. Feng, J.-Y. Yang, Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning, *IEEE Trans. Image Process.* 26 (Jan 2017) 1363–1378.
- [2] X. Luo, Y. Xu, J. Yang, Multi-resolution dictionary learning for face recognition, *Pattern Recognit.* 93 (Sept 2019) 283–292.
- [3] X. Wei, Y. Li, H. Shen, W. Xiang, Y.L. Murphey, Joint learning sparsifying linear transformation for low-resolution image synthesis and recognition, *Pattern Recognit.* 66 (June 2017) 412–424.
- [4] W.T. Freeman, E.C. Pasztor, O.T. Carmichael, Learning low-level vision, *Int. J. Comput. Vis.* 40 (Oct 2000) 25–47.
- [5] J. Shi, X. Liu, C. Qi, Global consistency, local sparsity and pixel correlation: a unified framework for face hallucination, *Pattern Recognit.* 47 (Nov 2014) 3520–3534.
- [6] L. An, B. Bhanu, Face image super-resolution using 2D CCA, *Signal Process.* 103 (Oct 2014) 184–194.
- [7] R. Abiantun, F. Juefei-Xu, U. Prabhu, M. Savvides, SSR2: sparse signal recovery for single-image super-resolution on faces with extreme low resolutions, *Pattern Recognit.* 90 (June 2019) 308–324.
- [8] X. Ma, J. Zhang, C. Qi, Hallucinating face by position-patch, *Pattern Recognit.* 43 (Jun 2010) 2224–2236.
- [9] Z. Wang, R. Hu, S. Wang, J. Jiang, Face hallucination via weighted adaptive sparse regularization, *IEEE Trans. Circuits Syst. Video Technol.* 24 (May 2014) 802–813.
- [10] J. Jiang, R. Hu, Z. Wang, Z. Han, J. Ma, Facial image hallucination through coupled-layer neighbor embedding, *IEEE Trans. Circuits Syst. Video Technol.* 26 (Sept 2016) 1674–1684.
- [11] S.S. Rajput, K. Arya, V. Singh, Robust face super-resolution via iterative sparsity and locality-constrained representation, *Inf Sci* 463 (Oct 2018) 227–244.
- [12] C. Jung, L. Jiao, B. Liu, M. Gong, Position-patch based face hallucination using convex optimization, *IEEE Signal Process. Lett.* 18 (Jun 2011) 367–370.
- [13] J. Jiang, R. Hu, Z. Wang, Z. Han, Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning, *IEEE Trans. Image Process.* 23 (Oct 2014) 4220–4231.
- [14] Y. Li, C. Cai, G. Qiu, K.-M. Lam, Face hallucination based on sparse local-pixel structure, *Pattern Recognit.* 47 (Mar 2014) 1261–1270.
- [15] J. Jiang, R. Hu, Z. Wang, Z. Han, Noise robust face hallucination via locality-constrained representation, *IEEE Trans. Multimedia* 16 (Aug 2014) 1268–1281.
- [16] J. Jiang, Y. Yu, S. Tang, J. Ma, A. Aizawa, K. Aizawa, Context-patch face hallucination based on thresholding locality-constrained representation and reproducing learning, *IEEE Trans. Cybern.* 50 (Jan. 2020) 324–337.
- [17] L. Liu, S. Li, C.P. Chen, Quaternion locality-constrained coding for color face hallucination, *IEEE Trans. Cybern.* 48 (May 2018) 1474–1485.
- [18] L. Liu, C.P. Chen, S. Li, Y.Y. Tang, L. Chen, Robust face hallucination via locality-constrained bi-layer representation, *IEEE Trans. Cybern.* 48 (Apr. 2018) 1189–1201.
- [19] L. Liu, S. Li, C.P. Chen, Iterative relaxed collaborative representation with adaptive weights learning for noise robust face hallucination, *IEEE Trans. Circuits Syst. Video Technol.* 29 (May 2019) 1284–1295.
- [20] J. Shi, X. Liu, Y. Zong, C. Qi, G. Zhao, Hallucinating face image by regularization models in high-resolution feature space, *IEEE Trans. Image Process.* 27 (June 2018) 2980–2995.

- [21] J. Shi, G. Zhao, Face hallucination via coarse-to-fine recursive kernel regression structure, *IEEE Trans. Multimedia* 21 (Sept 2019) 2223–2236.
- [22] L. Chen, J. Pan, Q. Li, Robust face image super-resolution via joint learning of subdivided contextual model, *IEEE Trans. Image Process.* 28 (Dec 2019) 5897–5909.
- [23] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (Feb 2016) 295–307.
- [24] Q. Cao, L. Lin, Y. Shi, X. Liang, G. Li, Attention-aware face hallucination via deep reinforcement learning, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1656–1664.
- [25] J. Jiang, Y. Yu, J. Hu, S. Tang, J. Ma, Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination, in: *Proceedings of International Joint Conference on Artificial Intelligence*, 2018, pp. 771–778.
- [26] X. Yu, B. Fernando, R. Hartley, F. Porikli, Super-resolving very low-resolution face images with supplementary attributes, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 908–917.
- [27] G. Gao, X. Jing, Q. Zhou, S. Wu, D. Yue, Locality-constrained matrix regression for position-patch based face hallucination, in: *IEEE International Conference on Image Processing*, 2016, pp. 419–423.
- [28] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, An interior-point method for large-scale  $l_1$ -regularized least squares, *IEEE J. Sel. Top. Signal Process.* 1 (Dec 2007) 606–617.
- [29] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, in: *Proceedings of Advances in Neural Information Processing Systems*, 2009, pp. 2223–2231.
- [30] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (Jan 2017) 156–171.
- [31] J. Chen, J. Yang, L. Luo, J. Qian, W. Xu, Matrix variate distribution-induced sparse representation for robust image classification, *IEEE Trans. Neural Netw. Learn. Syst.* 26 (Oct 2015) 2291–2300.
- [32] J. Yang, Y. Zhu, K. Li, J. Yang, C. Hou, Tensor completion from structurally-missing entries by low-tt-rankness and fiber-wise sparsity, *IEEE J. Sel. Top. Signal Process.* 12 (Dec 2018) 1420–1434.
- [33] Y. Chen, H. Xu, C. Caramanis, S. Sanghavi, Robust matrix completion and corrupted columns, in: *Proceedings of International Conference on Machine Learning*, 2011, pp. 873–880.
- [34] J.F. Cai, E.J. Candes, Z.W. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.* 20 (2010) 1956–1982.
- [35] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2790–2798.
- [36] Z. Lin, R. Liu, Z. Su, Linearized alternating direction method with adaptive penalty for low-rank representation, in: *Proceedings of Advances in Neural Information Processing Systems*, 2011, pp. 612–620.
- [37] C.E. Thomaz, G.A. Giraldo, A new ranking method for principal components analysis and its application to face image analysis, *Image Vis. Comput.* 28 (Jun 2010) 902–913.
- [38] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (1998) 23–38.
- [39] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (Apr 2004) 600–612.
- [40] K.-C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (May 2005) 684–698.

**Guangwei Gao** received the B.S. degree in information and computation science from Nanjing Normal University, Nanjing, China, in 2009, and the Ph.D. degree in pattern recognition and intelligence systems from Nanjing University of Science and Technology, Nanjing, China, in 2014. From March 2011 to September 2011 and February 2013 to August 2013, he was an exchange student of Department of Computing, Hong Kong Polytechnic University. Now, he is an associate professor in the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications. His research interests include pattern recognition, and computer vision. He has served as reviewer for IEEE TNNLS/TIP/TMM/TCYB, Pattern Recognition, Neurocomputing, Pattern Recognition Letter and AAAI/ICPR/ICIP etc.

**Yi Yu** received the Ph.D. degree in information and computer science from Nara Women's University, Japan. He is currently an Assistant Professor with the National Institute of Informatics (NII), Japan. Before joining NII, she was a Senior Research Fellow with the School of Computing, National University of Singapore. Her research covers large-scale multimedia data mining and pattern analysis, location-based mobile media service and social media analysis. She and her team received the best Paper Award from the IEEE ISM 2012, the 2nd prize in Yahoo Flickr Grand Challenge 2015, were in the top winners (out of 29 teams) from ACM SIGSPATIAL GIS Cup 2013, and the Best Paper Runner-Up in APWeb-WAIM 2017, recognized as finalist of the World's FIRST 10K Best Paper Award in ICME 2017.

**Jin Xie** received the Ph.D. degree from the Department of Computing, Hong Kong Polytechnic University in 2012. From 2013 to 2017, he was a Research Scientist with New York University Abu Dhabi, Abu Dhabi, UAE. Now, he is a professor in the School of Computer Science and Engineering of Nanjing University of Science and Technology. His current research interests include image forensics, computer

vision, machine learning, 3-D computer vision with the convex optimization, and deep learning methods.

**Jian Yang** received the PhD degree from Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a Chang-jiang professor in the School of Computer Science and Engineering of NUST. He is the author of more than 100 scientific papers in pattern recognition and computer vision. His papers have been cited more than 4000 times in the Web of Science, and 9000 times in the Scholar Google. His research interests include pattern recognition, computer vision and machine learning. Currently, he is/was an Associate Editor of Pattern Recognition Letters, IEEE Trans. Neural Networks and Learning Systems, and Neurocomputing. He is a Fellow of IAPR.

**Meng Yang** was now an associate professor at School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China. He received his Ph.D. degree from The Hong Kong Polytechnic University in 2012. Before joining Sun Yat-Sen University, he has been working as Postdoctoral fellow in the Computer Vision Lab of

ETH Zurich. His research interest includes sparse coding, dictionary learning, object recognition and machine learning. He has published 11 ICML/AAAI/PR/ICCV/ECCV papers and several IJCV, IEEE TNNLS and TIP journal papers. Now his Google citation is over 3300, and his homepage is [www.yangmeng.org.cn](http://www.yangmeng.org.cn).

**Jian Zhang** received the B.Sc. degree from the East China Normal University, Shanghai, China, in 1982, the M.Sc. degree in computer science from Flinders University, Adelaide, SA, Australia, in 1994, and the Ph.D. degree in electrical engineering from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 1999. He is currently an Associate Professor with the Advanced Analytics Institute and School of Software, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia. His current research interests include multimedia processing and communications, image and video processing, machine learning, pattern recognition, media and social media visual information retrieval and mining, human-computer interaction and intelligent video surveillance systems. He was the General Chair and chaired the International Conference on Multimedia and Expo in 2012. He is an Associated Editor for the IEEE Trans. Multimedia and the EURASIP Journal on Image and Video Processing.