






# Context-Patch Representation Learning With Adaptive Neighbor Embedding for Robust Face Image Super-Resolution

Guangwei Gao , Senior Member, IEEE, Yi Yu , Senior Member, IEEE, Huimin Lu , Senior Member, IEEE, Jian Yang , Member, IEEE, and Dong Yue , Fellow, IEEE

**Abstract**—Representation learning steered robust face image super-resolution (FSR) methods have attracted extensive attention in the past few decades. Most previous methods were devoted to exploiting the local position patches in the training set for FSR. However, they usually overlooked the sufficient usage of the contextual information around the testing patches, which are useful for stable representation learning. In this article, we attempt to utilize the context-patch around the testing patch and propose a method named context-patch representation learning with adaptive neighbor embedding (CRL-ANE) for FSR. On one hand, we simultaneously use the testing position patch and its adjacent ones for stable representation weight learning. This contextual information can compensate for recovering missing details in the target patch. On the other hand, for each input patch set, due to its inherent facial structural properties, we design an adaptive neighbor embedding strategy to elaborately and adaptively choose primary candidates for more accurate reconstruction. These two improvements enable the proposed method to achieve better SR performance than some of the other methods. Qualitative and quantitative experiments on some benchmarks have validated the superiority of the proposed method over some state-of-the-art methods.

**Index Terms**—Adaptive neighbor embedding, contextual information, face super-resolution, representation learning.

Manuscript received 14 July 2021; revised 7 March 2022, 24 May 2022, and 15 June 2022; accepted 17 July 2022. Date of publication 20 July 2022; date of current version 7 June 2023. This work was supported in part by the National Key Research and Development Program of China under Grants 2018AAA0100102 and 2018AAA0100100, in part by the National Natural Science Foundation of China under Grants 61972212, 61772568, and 62076139, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20190089, and in part by Six Talent Peaks Project in Jiangsu Province under Grant RJFW-011. The guest editor coordinating the review of this manuscript and approving it for publication was Professor David Crandall. (*Corresponding author: Yi Yu.*)

Guangwei Gao is with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China, and also with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: csggao@gmail.com).

Yi Yu is with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: yiyu@nii.ac.jp).

Huimin Lu is with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Kitakyushu 804-8550, Japan (e-mail: dr.huimin.lu@ieee.org).

Jian Yang is with the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: csjyang@njust.edu.cn).

Dong Yue is with the College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: medongyue@vip.163.com).

Digital Object Identifier 10.1109/TMM.2022.3192769

## I. INTRODUCTION

THE aim of image super-resolution (SR) is to reconstruct a high-quality image with more details from its observed low-quality image (s). With the aid of the internet, large-scale data can be collected to train robust models. However, learning the effective potential priors for more robust and accurate image reconstruction is still a crucial challenge. Numerous algorithms have been designed for general image SR problems [1]–[3] and domain-specific SR problems [4], [5]. Face image super-resolution (FSR), as a specific case of general image SR, is a technique to acquire high-resolution (HR) facial images from observed low-resolution (LR) facial images [6]. As the preprocessing step, it has been widely used in various face-related applications, such as face tracking, face detection, face editing, face reconstruction, and face recognition [7]–[9].

In previous decades, tremendous achievements have been made in the SR solution of LR facial images. Previous models mainly focused on the global faces. Some representative manifold learning methods are used to derive the HR facial images based on the correlation mapping learned from the LR/HR face pairs [10]. Liu *et al.* [11] attempted to model the relationship between LR/HR face images using a non-parametric Markov random field model to further compensate for the local details. Jia *et al.* [12] designed a global image-based tensor to describe the mappings across multiple modalities. A local patch-based multi-resolution tensor is utilized to generate HR facial details. The performance of these methods still needs to be improved when the number of the training examples is small or the observed LR images possess pose or noise variations.

In recent years, the part-based methods have received attractive attention since the local patch-steered algorithms could maintain well the facial details well. Baker *et al.* [13] suggested that reconstruction based on the image patches can improve the SR performance. Following this idea, Yang *et al.* [14] extracted patches from the given HR images and then trained a couple of LR and HR dictionaries to super-resolve the desired HR images. Jiang *et al.* [15] applied smooth regression to learn the relationship between HR and LR patches with the assistance of the local structure prior. Zeng *et al.* [16] generated a better training set to enhance the quality of the FSR. Jiang *et al.* [17] further exploited the contextual information and designed a thresholding locality-constrained representation scheme.

Although the above approaches have attained promising performances on FSR, there are still some drawbacks that need to be addressed. On one hand, only the context-patches around the training patch are used, and the contextual information around the testing patch is not well-utilized. On the other hand, previous methods exploited the fixed-size training patches for representation learning, ignoring the inherent structural properties included in the faces. To dispose of the aforementioned issues, we propose an approach called context-patch representation learning with adaptive neighbor embedding (CRL-ANE) for FSR in this paper. In particular, several contributions in this work are made:

- To avoid the tedious and ambiguous parameter search, based on the local inherent facial structural properties, we design a parameter-free adaptive neighbor embedding solution to adaptively select primary candidates for more accurate and stable reconstruction.
- To make full use of the context-patches around the testing patch for stable representation weight learning, we design an effective matrix set steered learning scheme to directly adopt the original form of the contextual patch for robust representation learning. Through our experiments, we find that these context-patches can compensate for the recovery of facial details when the observed LR inputs contain noise.

The rest of this study is as follows. We describe two categories of the related works in Section II. Our motivations are given in detail in Section III. Our devised solution is presented in Section IV. The experimental evaluations and discussions are provided in Section V. In Section VI, we give the conclusion of our work.

## II. RELEVANT WORK

We simply mention two categories of work related to the FSR problem in this section. The conventional methods are mainly based on the statistical geometric structure of the faces. To super-resolve HR face images with more details, researchers have focused on two kinds of approaches: mapping function-based approaches and prior knowledge-based approaches.

**Mapping function-based approaches** are devoted to investigating new models to find the inherent transformation between LR/HR face pairs. Ma *et al.* [18] designed the least square representation (LSR)-steered approach by exploiting the position constraint to restrain the reconstruction process. Jiang *et al.* [19] designed a locality constrained representation (LcR) solution to simultaneously achieve locality and sparsity in the representation learning process. Later, Liu *et al.* [20] introduced this idea into the quaternion space to hallucinate color face images. Shi *et al.* [21] proposed to train a series of adaptive kernel regression functions for high-frequency information prediction. Recently, deep convolutional neural network (CNN) based models have been widely used to learn end-to-end mapping functions. Yu and Porikli [22] introduced a transformative autoencoder to reconstruct very LR unaligned and noisy faces. Cao *et al.* [23] presented an attention-aware framework, which performed effective facial part enhancement via a deep reinforcement learning scheme. Song *et al.* [24] designed a two-step FSR network, which first used the deep models to generate coarse HR faces and then enhanced the facial details via facial component matching.

Zhang *et al.* [25] proposed a copy and paste generative adversarial network (GAN) to super-resolve LR faces under normal illumination conditions.

**Prior knowledge-based approaches** mainly focus on existing reasonable priors for guiding the design of regularization terms [26]–[29]. To resolve the unstable solution of LSR [18], Jung *et al.* [30] adaptively selected primal training patches for efficient reconstruction by introducing the sparsity constraint. Wang *et al.* [31] presented a weighted adaptive sparse regularization method for accurate, stable, and robust face image reconstruction. Jiang *et al.* [32] assigned different model parameters for different patches based on the facial structure prior. Rajput *et al.* [33] presented an iterative sparsity and locality-constrained representation approach for robust FSR. Liu *et al.* [34] proposed a robust locality-constrained bi-layer representation (RLcBR) method to perform FSR and noise removal simultaneously. Chen *et al.* [35], [36] formulated the FSR procedure as a contextual joint representation model to restore HR facial images under noisy LR scenarios. For deep CNN-based methods, facial structural priors are usually introduced into the model for better performance. Zhu *et al.* [37] presented a gated deep bi-network to recover the textural details by exploiting the facial spatial priors. Chen *et al.* [38] introduced the facial parsing and landmark maps into the training phase to super-resolve better results. Yu *et al.* [39] estimated the facial component heatmaps in the network to guide the super-resolution procedure. Recently, they further considered the facial attribute priors and developed an attribute embedding method to hallucinate very low-quality faces [40]. Zhang *et al.* [41] utilized the facial identity information and presented a super-identity convolutional neural network for FSR. Hsu *et al.* [42] leveraged the similar prior for the identity-preserving FSR task. Recently, Ma *et al.* [43] proposed an FSR method with iterative collaboration between facial image recovery and landmark estimation.

In comparison with those previous methods, we elaborately exploit the central position patch and its adjacent patches in the input for more stable representation weight learning. Meanwhile, by taking the inherent facial structural properties into consideration, we design an adaptive neighbor embedding strategy to adaptively choose reliable candidates in the training set for more accurate reconstruction. These two considerations promote the context-patch representation capability, thus making the proposed method achieve better recovery performance.

## III. MOTIVATION

Our approach belongs to the prior knowledge steered ones. In comparison with the existing approaches, in our solution, we propose to fully utilize the adaptive neighbor structure and the contextual information for robust representation learning. We argue that the key to a robust FSR model lies in the robust and effective representation learning with the help of the horizontal neighbors and vertical contextual information. We provide the reasons for the above observations as follows:

- It is known that human faces usually contain abundant structure priors, which can make a great contribution to the final face reconstruction task. For instance, the forehead and cheeks are smooth, while the eyes, mouth, and nose may

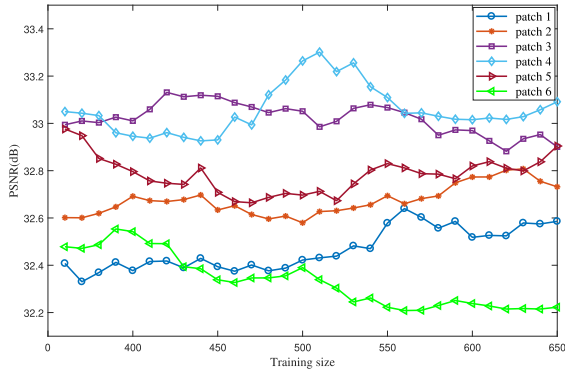


Fig. 1. The PSNR values of different test position patches with different training sizes. For the illustration, we only list 6 patches here.

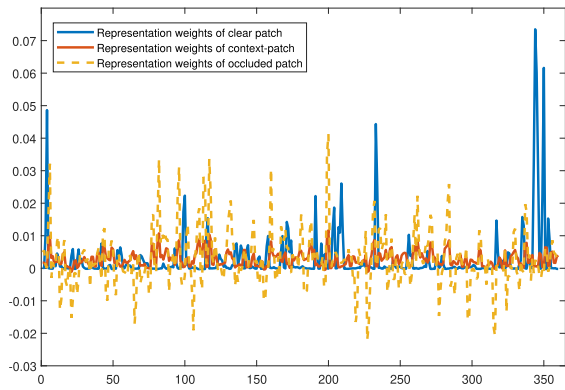


Fig. 2. Representation weights of the clear patch, the occluded patch, and the related context-patch using the LcR method [19], respectively.

contain rich textural information. Generally, in conventional neighbor embedding-based methods, an appropriate training size should be preset to simultaneously capture the facial details and ease the computational costs. Nevertheless, it is usually unrealistic to allot a uniform neighbor for distinct facial parts. In Fig. 1, we give a plot of the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [44] values versus different training sizes with different position patches. It was found that different position patches achieve their best reconstruction performance with different training sizes. This phenomenon motivates us to design an adaptive neighbor embedding scheme for each patch to obtain its best reconstruction performance.

- The key of the learning steered methods is the desired representation weights of the acquired patch over a given training set. When the observed patch is occluded, then the computation of the optimal representation weights becomes intractable, which leads to unsatisfactory facial reconstruction performance. Fortunately, if we resort to its contextual patches, which are visible and may have a similar manifold structure to that of its central patch, it is likely to attain more accurate representation weights for more accurate reconstruction. We provide an example in Fig. 2. With the help of the facial contextual counterparts, the effect of the occlusion in the local central patch can be resolved to some extent.

TABLE I  
NOTATIONS USED IN THIS ARTICLE

Symbol	Meaning
$x_t^i, x_m^i$	the $i$ -th patch of a face image
$S_t^i, S_m^i$	the contextual patch sets
$J_C, J_t$	the constraint terms
$P_t^i, C_k^i$	the cascaded patch sets
$A_t^i, B_k^i$	the diagonal block matrix
$\lambda, \eta, Z$	the auxiliary Lagrange multipliers
$\tau, \epsilon$	the parameters

Based on the above observations, we introduce the contextual prior and the adaptive neighbor embedding scheme into our proposed model to simultaneously mitigate the computational burden and provide complementary information for robust representation learning.

#### IV. PROPOSED METHOD

##### A. Main Model

Our notations are summarized in Table I. In learning steered face hallucination approaches, pairs of LR and HR patches form the training set. Suppose we have  $M$  LR/HR face pairs for training, denoted as  $X = \{X_m\}_{m=1}^M$  and  $Y = \{Y_m\}_{m=1}^M$ , where  $X_m$  and  $Y_m$  represent the paired LR/HR training examples, respectively. The main goal of the face image super-resolution is to hallucinate the potential high-quality candidate  $Y_o$  from its low-quality observed  $X_t$  by using the LR and HR training set,  $\{X_m, Y_m\}_{m=1}^M$ .

For patch-based methods, all the paired faces in the training set are divided into lapped small patches, denoted as  $\{x_m^i\}_{m=1}^M$  and  $\{y_m^i\}_{m=1}^M$ , where  $x_m^i$  and  $y_m^i$  denote the  $i$ -th patch of the  $m$ -th LR/HR training face, respectively. For the LR testing face image  $X_t$ , its  $i$ -th patch is represented as  $x_t^i$ . Our goal is to obtain the optimal representation weights  $w_m^i$  of  $x_t^i$  over the LR training patches  $\{x_m^i\}_{m=1}^M$ .

As in [17], we utilize all the contextual candidates within a reliable window treated position  $p$  as the center. For the input patch  $x_t^i$ , its contextual patches are denoted as  $x_t^{i,a}$ , where  $a = 1, 2, \dots, c$ . Here,  $c$  denotes the number of contextual patches within a window and can be acquired by the step size ( $ss$ ), the patch size ( $ps$ ), and the window size ( $ws$ ):  $c = (1 + \frac{ws-ps}{ss})^2$ . As in [17], we also fix the step size as 2 in this paper. Then, the contextual patch set  $S_t^i$  of  $x_t^i$  can be denoted as follows:

$$S_t^i = [x_t^{i,1}, x_t^{i,2}, \dots, x_t^{i,c}]. \quad (1)$$

The contextual patch set of the training faces can be constructed in the same manner. For the  $i$ -th patch of the  $m$ -th training face  $x_m^i$ , its contextual patch set  $S_m^i$  can be obtained as follows:

$$S_m^i = [x_m^{i,1}, x_m^{i,2}, \dots, x_m^{i,c}]. \quad (2)$$

Considering that the contextual patches can provide complementary information for discriminative representation learning, especially when the LR input encounters noise, these contextual

patches are represented jointly by sharing the same representation weights as follows:

$$J_c = \sum_{a=1}^c \left\| x_t^{i,a} - \sum_{m=1}^M w_m^i x_m^{i,a} \right\|_{l_p}, \quad (3)$$

where  $w_m^i$  is the target optimal representation weights,  $\|\cdot\|_{l_p}$  denotes the  $l_p$ -norm of a matrix or a vector.

In addition to the above contextual fidelity term, the central patch within a window should also be represented accurately. As a result, we have the following term:

$$J_t = \left\| x_t^i - \sum_{m=1}^M w_m^i x_m^i \right\|_{l_p}. \quad (4)$$

By considering the contextual constraint term  $J_c$  and central representation term  $J_t$  together, we use the minimization function to obtain the optimal weights  $w^i$  as follows:

$$\min_{w^i} \{J_t + J_c + \tau \Omega(w^i)\}, \quad (5)$$

where the third term  $\Omega(w^i)$  denotes the prior about the combination weights,  $\tau$  is the regularization parameter.

### B. Adaptive Neighbor Embedding

In the previous section, the whole training set is used to represent the given patch. However, it is worth noting that when the training sample size is large, the computational cost of function (5) increases dramatically. The previous popular methods usually selected  $K$  nearest neighbors to reduce the computational complexity. Nevertheless, it is usually unreasonable and intractable to preset a reliable value of  $K$  in real-world applications. Additionally, it is unrealistic to designate the uniform  $K$  to individual distinct facial parts.

To avoid the difficulty of conducting parameter selection, we advise a parameter-free adaptive neighbor search strategy here to adaptively embed similar patches for better reconstruction. Let  $N(x_t^i)$  represent the set of neighbors of the input patch  $x_t^i$ , then we define  $N(x_t^i)$  as follows:

$$N(x_t^i) = \left\{ x_b^i \mid \text{if } d(x_t^i, x_b^i) \leq \frac{1}{M} \sum_{m=1}^M d(x_t^i, x_m^i) \right\}, \quad (6)$$

where  $b = 1, \dots, M$ ,  $d(x_t^i, x_m^i) = \|x_t^i - x_m^i\|_2^2$ , and  $\frac{1}{M} \sum_{m=1}^M d(x_t^i, x_m^i)$  is actually the mean of all  $d(x_t^i, x_m^i)$ . For each input LR patch, our strategy could adaptively select primary patches from the training set for more accurate and efficient reconstruction.

An example is shown in Fig. 3, from which we can observe that individual faces have distinct neighbor maps. It was also found that the forehead and cheek regions have a large number of neighbors, while nose, eyes, mouth, and face contours have a small number of neighbors. This further validates that different facial regions have a distinct number of neighbors. In [32], the authors also proposed an adaptive parameter setting strategy. The differences are twofold. On one hand, the method in [32] first performed parameter map learning in a training set, and then used this learned map for HR face prediction. In our method, we

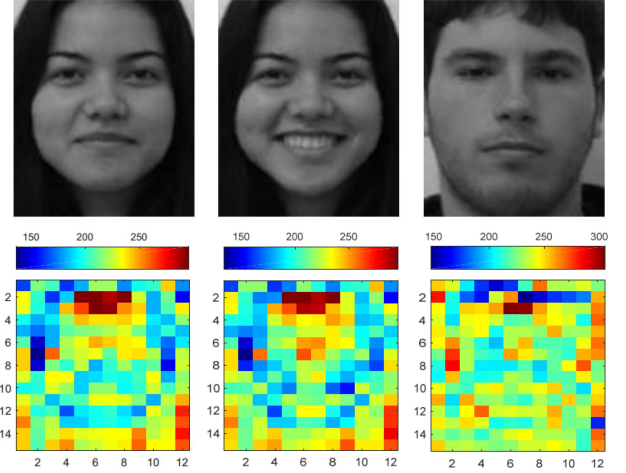


Fig. 3. Visualization of the adaptively embedded neighbors.

directly perform a neighbor search in the testing phase. On the other hand, the method in [32] assigned the same parameter for the patches from the same position of a different subject, while we relaxed this consumption and made these patches adaptively choose their neighbors.

### C. Vector Set-Based CRL

Similar to previous methods, in this part, we use the vector form to elaborate each patch. In this case, the contextual patch set in (1) and (2) can be denoted as  $S_t^i = [x_t^{i,1}, x_t^{i,2}, \dots, x_t^{i,c}] \in \mathfrak{R}^{d \times c}$  and  $S_m^i = [x_m^{i,1}, x_m^{i,2}, \dots, x_m^{i,c}] \in \mathfrak{R}^{d \times c}$ , where  $d$  is the measure of the patch. By considering the aforementioned adaptive neighbor search strategy simultaneously, (3) and (4) can be rewritten as follows:

$$J_t = \left\| x_t^i - \sum_{k \in N(x_t^i)} w_k^i x_k^i \right\|_2^2, \quad (7)$$

$$J_c = \sum_{a=1}^c \left\| x_t^{i,a} - \sum_{k \in N(x_t^i)} w_k^i x_k^{i,a} \right\|_2^2.$$

Researchers [19], [34] have made extensive efforts on exploring effective constraints and advised that the locality constraint is superior to the sparsity constraint in exposing the inherent structure of the nonlinear manifold. To make full use of the merits of both locality and sparsity, in this paper, we wish to incorporate the locality and sparsity constraint into the priority of  $w^i$ . (5) can be rewritten as follows:

$$\min_{w^i} \left\| P_t^i - \sum_{k=1}^K w_k^i G_k^i \right\|_2^2 + \tau \|Dw^i\|_1, \quad (8)$$

where the index  $k$  is in the set  $N(x_t^i)$ ,  $D$  is a diagonal weight matrix with elements  $d_k = \|P_t^i - G_k^i\|_2^2$ ,  $k = 1, 2, \dots, K$  ( $K$  denotes the length of set  $N(x_t^i)$ ). The cascaded testing patch set  $P_t^i$  is denoted as follows:

$$P_t^i = [x_t^i, x_t^{i,1}, x_t^{i,2}, \dots, x_t^{i,c}]. \quad (9)$$

**Algorithm 1:** The algorithm for Solving (11).

**Input:** The  $i$ -th cascading contextual patch set  $P$  from an LR observed face, the corresponding adaptive selected neighbor set  $G = [G_1, G_2, \dots, G_K]$  from the LR training set.

**Parameter:** The locality parameters  $\tau$ , and the maximal iteration steps  $max\_num$ .

**Initialize:**  $z^{(0)} = w^{(0)} = \lambda^{(0)} = 0$ .

**while**  $t < max\_num$  **do**

1. Update  $w$  by (14);
2. Update  $z$  by (15);
3. Update Lagrange multiplier  $\lambda$  by (16);
4.  $t \leftarrow t + 1$ .

**end while**

**Output:** The desired representation weight vector  $w$ .

The cascaded training patch set  $G_k^i$  is similarly denoted as follows:

$$G_k^i = [x_k^i; x_k^{i,1}; x_k^{i,2}; \dots, x_k^{i,c}]. \quad (10)$$

In the next text, we detail the optimization procedure of function (8). For simplicity, we leave out the indexes  $i$  and  $t$ . Directly solving (8) is difficult, and we transform it into the following equivalent formulation:

$$\min_{w,z} \|P - Gw\|_2^2 + \tau \|z\|_1, \quad s.t. \quad z = Dw, \quad (11)$$

where  $w$  is the target optimal representation weight vector consisting of  $K$  values of  $w_k$  and  $G$  is the matrix cascaded by  $K$  vectors of  $G_k$ ,  $k = 1, 2, \dots, K$ . The alternating minimization method can be efficiently used to solve the above multiple variable optimization problem [45]. The corresponding augmented Lagrange function of (11) is denoted as follows:

$$L(w, z, \lambda) = \|P - Gw\|_2^2 + \tau \|z\|_1 + \lambda^T (z - Dw) + \frac{\mu}{2} \|z - Dw\|_2^2, \quad (12)$$

where  $\lambda$  denotes the auxiliary Lagrange multiplier,  $\mu > 0$  is a positive penalty constant. Then  $w$  and  $z$  can be optimized alternatively.

*Updating  $w$ :* By fixing others, the optimal solution of  $w$  can be attained as follows:

$$\begin{aligned} w^{(t+1)} &= \arg \min_w L(w, z^{(t)}, \lambda^{(t)}) \\ &= \arg \min_w \|P - Gw\|_2^2 + \frac{\mu}{2} \|e^{(t)} - Dw\|_2^2, \end{aligned} \quad (13)$$

where  $e^{(t)} = z^{(t)} + \lambda^{(t)}/\mu$ . By considering the derivative of the function  $L$  related to the variable  $w$ , and equating the result to zero, we can attain the following:

$$w^{(t+1)} = (2G^T G + \mu D^T D)^{-1} (2G^T P + \mu D^T e^{(t)}). \quad (14)$$

*Updating  $z$ :* By fixing others and using the soft-thresholding operator [46], the optimal solution of  $z$  can be attained as follows:

**Algorithm 2:** Optimizing (19) via ADMM.

**Input:** The  $i$ -th LR testing diagonal block matrix  $A$ , the corresponding adaptive selected training diagonal block matrix set  $B = [B_1, B_2, \dots, B_K]$ .

**Parameter:** The model parameter  $\tau$ , and the parameter  $\epsilon$  in the termination condition.

**Initialize:**  $E^{(0)} = Z^{(0)} = 0$ ,  $w^{(0)} = z^{(0)} = \gamma^{(0)} = 0$ .

**while not converged do**

1. Update  $E$  by (22);
2. Update  $w$  by (24);
3. Update  $z$  by (25);
4. Update Lagrange multipliers  $\gamma$  and  $Z$  by (26);
5. Check the convergence condition by (27);
6.  $t \leftarrow t + 1$ .

**end while**

**Output:** The desired representation weight vector  $w$ .

$$\begin{aligned} z^{(t+1)} &= \arg \min_z L(w^{(t+1)}, z, \lambda^{(t)}) \\ &= \arg \min_z \frac{\tau}{\mu} \|z\|_1 + \frac{1}{2} \left\| z - \left( Dw^{(t+1)} - \frac{1}{\mu} \lambda^{(t)} \right) \right\|_2^2 \\ &= \text{shrink} \left( Dw^{(t+1)} - \frac{1}{\mu} \lambda^{(t)}, \frac{\tau}{\mu} \right), \end{aligned} \quad (15)$$

where the operator shrink is defined as  $\text{shrink}(x, \sigma) = \text{sign}(x) \cdot \max(|x| - \sigma, 0)$  in a scalar way.

*Updating  $\lambda$ :* Once  $z^{(t+1)}$  and  $w^{(t+1)}$  are updated, the assisted Lagrange multiplier  $\lambda$  can be updated as follows:

$$\lambda^{(t+1)} = \lambda^{(t)} + \mu (z^{(t+1)} - Dw^{(t+1)}). \quad (16)$$

The detailed process to solve (11) is listed in Algorithm 1.

**D. Matrix Set-Based CRL**

Many pioneering works [28], [47] have discussed that the nuclear norm-based constraint can be more appropriate for maintaining the inherent geometry of the reconstruction error. In contrast to the foregoing subsection where each contextual patch is reshaped as a vector, here we directly utilize the 2D form of the contextual patch for robust representation learning. Then the patch set in (1) and (2) can be denoted as  $S_t^i = [x_t^{i,1}, x_t^{i,2}, \dots, x_t^{i,c}] \in \mathfrak{R}^{p \times q \times c}$  and  $S_m^i = [x_m^{i,1}, x_m^{i,2}, \dots, x_m^{i,c}] \in \mathfrak{R}^{p \times q \times c}$ , where  $p$  and  $q$  represent the measure of the observed patch. In this case, (3) and (4) can be rewritten as follows:

$$\begin{aligned} J_t &= \left\| x_t^i - \sum_{k \in N(x_t^i)} w_k^i x_k^i \right\|_*, \\ J_c &= \sum_{a=1}^c \left\| x_t^{i,a} - \sum_{k \in N(x_t^i)} w_k^i x_k^{i,a} \right\|_*, \end{aligned} \quad (17)$$

where  $\|\cdot\|_*$  denotes the nuclear norm of a matrix.

By some explicit algebraic steps, (5) can be formulated as follows:

$$\min_{w^i} \|A_t^i - B^i(w^i)\|_* + \tau \|Dw^i\|_1, \quad (18)$$

where  $B^i(w^i) = w_1^i B_1^i + w_2^i B_2^i + \dots + w_K^i B_K^i$ ,  $A_t^i = \begin{bmatrix} x_t^i & & & & \\ & x_t^{i,1} & & & \\ & & \ddots & & \\ & & & x_t^{i,c} & \\ & & & & \ddots \end{bmatrix}$  is the diagonal block matrix of  $x_t^i$ ,  $B_k^i = \begin{bmatrix} x_k^i & & & & \\ & x_k^{i,1} & & & \\ & & \ddots & & \\ & & & x_k^{i,c} & \\ & & & & \ddots \end{bmatrix}$  is the diagonal block matrix of  $x_k^i$ ,  $k = 1, 2, \dots, K$ .

The alternating direction method of multipliers (ADMM) can be utilized to tackle the previous optimization task. By introducing some auxiliary variables, we transform (18) into the next formulation as follows:

$$\min_{E, w, z} \|E\|_* + \tau \|z\|_1 \quad (19)$$

$$s.t. A - B(w) = E, z = Dw$$

Here, for the sake of description, we also omit the indexes  $i, k$  and  $t$ . The augmented Lagrange function of (19) can be written as follows:

$$L(w, z, \eta, Z, E) = \|E\|_* + \tau \|z\|_1 + \langle Z, A - B(w) - E \rangle$$

$$+ \langle \eta, z - Dw \rangle + \frac{\mu}{2} (\|A - B(w) - E\|_F^2 + \|z - Dw\|_2^2), \quad (20)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product, both  $\eta$  and  $Z$  are the auxiliary Lagrange multipliers, and  $\mu > 0$  is a positive penalty constant. Thus,  $E, w$  and  $z$  can be optimized iteratively.

*Updating E:* By fixing the others, the optimal  $E$  can be gained by optimizing the following problem:

$$E^{(t+1)} = \arg \min_E L(w^{(t)}, z^{(t)}, \eta^{(t)}, Z^{(t)}, E)$$

$$= \arg \min_E \frac{1}{\mu} \|E\|_* + \frac{1}{2} \|E - C\|_F^2, \quad (21)$$

where  $C = A - B(w^{(t)}) + Z^{(t)}/\mu$ . Its solution is

$$E^{(t+1)} = UT_{\frac{1}{\mu}}[Q]V, \quad (22)$$

where  $(U, Q, V^T) = \text{svd}(C)$ ,  $T_{\frac{1}{\mu}}[Q] = \text{diag}(\{\max(0, q_j - \frac{1}{\mu})\}_{1 \leq j \leq r})$ ,  $q_1, \dots, q_r$  denotes the positive singular values, and  $r$  denotes the rank of matrix  $Q$ .

*Updating w:* By fixing the others, the optimal solution of  $w$  can be obtained as follows:

$$w^{(t+1)} = \arg \min_w L(w, z^{(t)}, \eta^{(t)}, Z^{(t)}, E^{(t+1)})$$

$$= \arg \min_w \|b^{(t+1)} - Hw\|_F^2 + \|g^{(t)} - Dw\|_2^2, \quad (23)$$

where  $H = [\text{vec}(B_1), \text{vec}(B_2), \dots, \text{vec}(B_K)]$ ,  $b^{(t+1)} = \text{vec}(A - E^{(t+1)} + \frac{1}{\mu} Z^{(t)})$  and  $g^{(t)} = z^{(t)} + \frac{1}{\mu} \eta^{(t)}$ . The closed solution of  $w$  is given as follows:

$$w^{(t+1)} = (H^T H + D^T D)^{-1} (H^T b^{(t+1)} + D^T g^{(t)}). \quad (24)$$

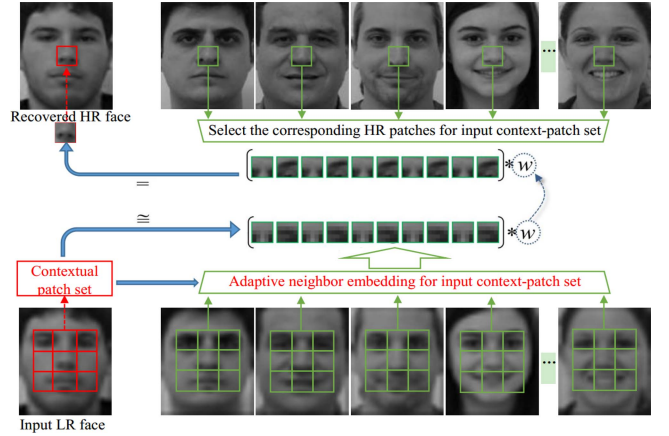


Fig. 4. Flowchart of our algorithm. Given an LR patch, we first choose its adjacent patches as the input contextual patch set, which are marked by red boxes. Then, we try to adaptively select its similar neighbors (green patch set) to perform representation learning and attain the optimal representation weights  $w$ . The required HR patch can be obtained by using the same representation weights over the relevant HR patches.

*Updating z:* By fixing others,  $z$  can be updated as follows:

$$z^{(t+1)} = \arg \min_z L(w^{(t+1)}, z, \eta^{(t)}, Z^{(t)}, E^{(t+1)})$$

$$= \arg \min_z \frac{\tau}{\mu} \|z\|_1 + \frac{1}{2} \left\| z - \left( Dw^{(t+1)} - \frac{1}{\mu} \eta^{(t)} \right) \right\|_2^2$$

$$= \text{shrink} \left( Dw^{(t+1)} - \frac{1}{\mu} \eta^{(t)}, \frac{\tau}{\mu} \right), \quad (25)$$

where the operator shrink has the same definition as that in the previous section.

*Updating Z and  $\eta$ :* Once  $w^{(t+1)}$ ,  $z^{(t+1)}$ , and  $E^{(t+1)}$  are acquired, the aided multipliers,  $\eta$  and  $Z$  can be updated as follows:

$$\eta^{(t+1)} = \eta^{(t)} + \mu \left( z^{(t+1)} - Dw^{(t+1)} \right),$$

$$Z^{(t+1)} = Z^{(t)} + \mu \left( A - B(w^{(t+1)}) - E^{(t+1)} \right). \quad (26)$$

In this work, we use the following termination conditions:

$$\|z^{(t+1)} - Dw^{(t+1)}\|_{\infty} \leq \epsilon,$$

$$\|A - B(w^{(t+1)}) - E^{(t+1)}\|_{\infty} \leq \epsilon, \quad (27)$$

where  $\epsilon$  is a given termination condition tolerance.

The detailed procedure for solving (19) is summarized in the form of Algorithm 2.

After obtaining the optimal representation weight  $w^i$  of the  $i$ -th patch, we would gain the desired HR patch by  $y_o^i = \sum w_k^i y_k^i$ . The target HR face  $Y_o$  can be yielded by integrating all the target HR patches  $y_o^i$  and averaging overlapped pixel values in terms of their respective positions. The whole architecture of our method is displayed in Fig. 4.

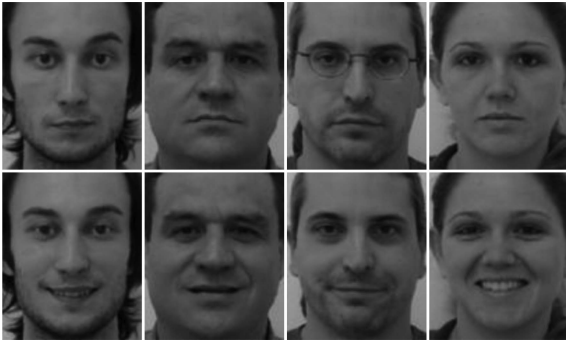


Fig. 5. Some example faces from the FEI face database [48]. Each column denotes the faces from one person.

TABLE II  
ABLATION STUDY RESULTS OF EACH MODULE IN OUR METHOD

Models	VCRL-NE	VPRL-ANE	VCRL-ANE	MCRL-ANE
PSNR (dB)	25.8896	25.7947	26.1050	26.3431
SSIM	0.8119	0.8021	0.8219	0.8284

## V. EXPERIMENTS AND DISCUSSIONS

### A. Dataset Description

In this part, we perform evaluations on the publicly available FEI face dataset [48], which collects 400 frontal faces from 200 subjects. Therefore, each subject possesses two examples: one with a smiling expression and the other with a neutral expression. All the face areas are cropped to have a size of  $120 \times 100$ . We randomly select 360 faces for training and the remaining 40 faces for testing. The HR faces are smoothed (with a window size of  $4 \times 4$ ) and then downsampled by a scale factor of 4 to generate the corresponding LR counterparts with a size of  $30 \times 25$ . The size of the patch and the overlap among adjacent patches in both the training and testing faces are  $12 \times 12$  pixels and 4 pixels, respectively. Some face examples from the FEI face database are depicted in Fig. 5.

### B. Ablation Study

We evaluate the effectiveness of each module in our method. Our method based on the vector and matrix patch is denoted as VCRL-ANE and MCRL-ANE, respectively. Compared to VCRL-ANE, VPRL-ANE replaces the context-patch with the position-patch (i.e., set the window size to 12), and VCRL-NE utilizes the selected training patches for representation learning. The compared results in terms of PSNR (dB) and SSIM are given in Table II. It can be seen that VCRL-ANE outperforms VPRL-ANE, indicating that the contextual information indeed compensates for the recovery of facial details in the target patch. The adaptive neighbor embedding strategy is also important in our method since VCRL-ANE obtains a better performance than VCRL-NE, which reveals that the adaptive neighbor selection strategy can lead to more stable and accurate reconstruction. The gain of MCRL-ANE over VCRL-ANE also validates the effectiveness of matrix regression used in our method.

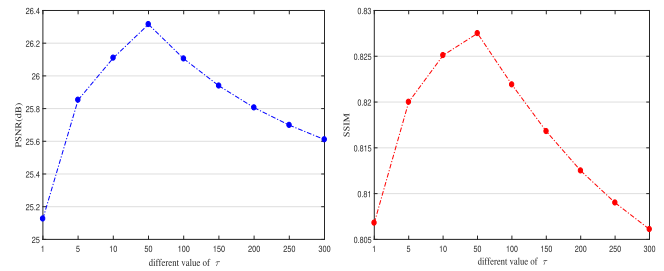


Fig. 6. The performance of our MCRL-ANE method with various indexes of  $\tau$  in terms of the average PSNR (dB) and SSIM values.

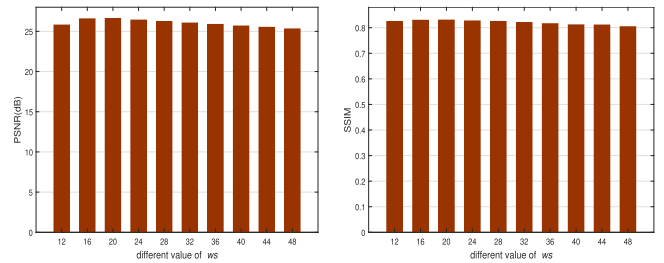


Fig. 7. The performance of our MCRL-ANE method with various indexes of  $ws$  in terms of the average PSNR (dB) and SSIM values.

### C. Parameter Discussions

In this part, we study the effect of our method with various parameter settings of the regularization parameter  $\tau$  and window size parameter  $ws$ .

To observe the effect of the regularization parameter, we conduct face image super-resolution experiments with various values of  $\tau$ . All the input face images are corrupted by a square “baboon” block image and downsampled to be used as the noisy LR test data. In Fig. 6, we draw the PSNR (dB) and SSIM [44] values of our MCRL-ANE approach with various values of  $\tau$ . From Fig. 6, it can be observed that as  $\tau$  grows, the super-resolution performance of MCRL-ANE first rises and then decreases. Values of  $\tau$  that are too large or too small provide no improvement to the reconstruction performance. When the values of  $\tau$  are set to approximately 50, our method can achieve stable performance.

In Fig. 7, we show the effectiveness of our method with different parameter settings of window size  $ws$ . It should be noted that when the patch size and the window size are all  $12 \times 12$  pixels, our method tends to be the position-patch-based approach. By considering more contextual patches (e.g., set window size to 16) when performing representation learning, the performance of our MCRL-ANE approach has a distinct improvement. When the size of the window is larger than  $16 \times 16$  pixels, the performance first increases and then tends to decrease. In our next evaluations, we configure the window size as  $16 \times 16$  to make a desired trade-off between the computational cost and the performance.

### D. Comparison With State-of-the-arts

In this part, we compare our method with some comparative approaches, including several deep CNN-based methods (i.e., SICNN [41], FSRNet [38], DICNet [43], and SPARNet [27]) and

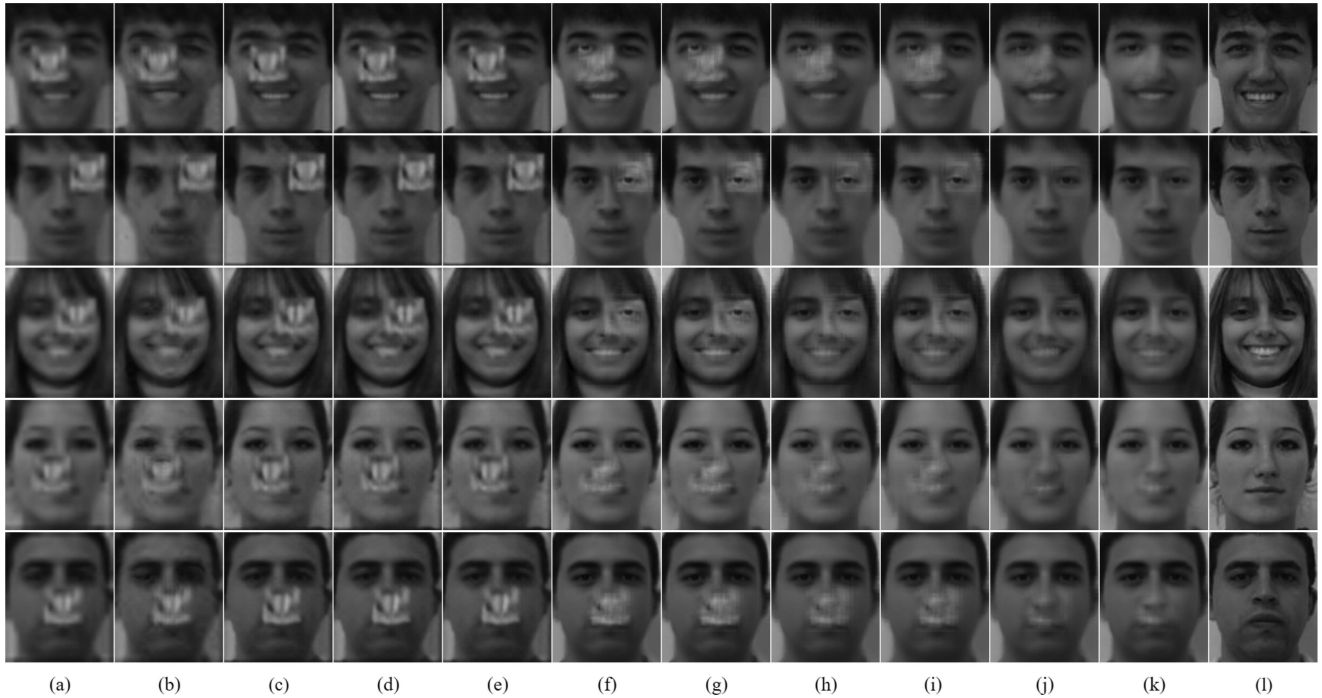


Fig. 8. Hallucinated results of respective methods for corrupted LR faces with block noises. From left to right are (a) the low-quality observations and the hallucinated outputs of (b) SICNN [41], (c) FSRNet [38], (d) DICNet [43], (e) SPARNet [27], (f) LcR [19], (g) RLcBR [34], (h) TLcR [17], (i) PRGFC [36], (j) our VCRL-ANE, (k) our MCRL-ANE, and (l) the original HR faces.

TABLE III

THE QUANTITATIVE COMPARISONS FOR LOW-QUALITY FACES CORRUPTED BY BLOCK NOISE

Methods	PSNR(dB)	SSIM
SICNN [41]	22.4933	0.7452
FSRNet [38]	22.7031	0.7861
DICNet [43]	22.9736	0.7977
SPARNet [27]	23.0327	0.7991
LcR [19]	24.4348	0.8134
RLcBR [34]	24.9246	0.8210
TLcR [17]	25.3728	0.8200
PRGFC [36]	25.6064	0.8199
<b>VCRL-ANE</b>	<b>26.1050</b>	<b>0.8219</b>
<b>MCRL-ANE</b>	<b>26.4431</b>	<b>0.8284</b>

TABLE IV

THE QUANTITATIVE COMPARISONS FOR LR FACES CORRUPTED BY MIXTURE NOISES

Methods	PSNR(dB)	SSIM
SICNN [41]	21.4075	0.6263
FSRNet [38]	21.1536	0.5390
DICNet [43]	21.3730	0.6289
SPARNet [27]	21.5064	0.6351
LcR [19]	23.2248	0.7431
RLcBR [34]	23.7746	0.7739
TLcR [17]	24.2761	0.7712
PRGFC [36]	24.4906	0.7752
<b>VCRL-ANE</b>	<b>24.8022</b>	<b>0.7782</b>
<b>MCRL-ANE</b>	<b>25.0113</b>	<b>0.7817</b>

several position-patch-based methods (i.e., LcR [19], TLcR [17], RLcBR [34], and PRGFC [36]). All the comparative approaches are tuned to attain their best performance.

The performance of each method is evaluated in super-resolving face images corrupted by a square “baboon” image block and mixture noises (e.g., Gaussian noise and block occlusion), respectively. Objectively, the values of PSNR and SSIM are also used to quantitatively investigate the reconstruction performance of each method. Tables III and IV depict the average PSNR (dB) and SSIM values of the respective methods. It can be seen that our method yields the best performance in terms of both PSNR and SSIM. When compared with the recently presented

contextual patch steered PRGFC [36] (i.e., the second-best competing method in the experiments), our method can still have considerable gain.

Some hallucinated faces of each method are listed in Figs. 8 and 9 for further qualitative comparisons. The deep CNN-based methods (i.e., SICNN [41], FSRNet [38], DICNet [43], and SPARNet [27]) cannot attain satisfactory performance when the LR faces contain noise. The reason may be that they do not take into consideration the highly structured position prior and noise prior, which have a crucial impact in robust face image super-resolution tasks. The hallucinated faces of LcR [19] and RLcBR [34] have better visual effects. TLcR [17] and PRGFC [36] were recently proposed as effective methods that





Fig. 9. Hallucinated results of respective methods for corrupted LR faces with mixture noises (Gaussian and block noises). From left to right are (a) the low-quality observations and the hallucinated outputs of (b) SICNN [41], (c) FSRNet [38], (d) DICNet [43], (e) SPARNet [27], (f) LcR [19], (g) RLcBR [34], (h) TLcR [17], (i) PRGFC [36], (j) our VCRL-ANE, (k) our MCRL-ANE, and (l) the original HR faces.

can preserve more facial details. By considering the contextual information around the testing patch and adaptively embedding primary training samples for more accurate and reasonable reconstruction, the hallucinated faces of our method look more similar to the ground truth.

### E. Comparisons on Real-World Faces

In all the above evaluations, the observed low-quality input face images stem from their related original high-quality counterparts. In real application scenes, it is unreasonable and difficult to simulate the process of image degradation. Thus, in this part, we perform experiments to testify the effectiveness of our approach on real-world low-quality face images.

We manually extract the low-quality face images from the CMU+MIT dataset [49] and slightly align them to the samples in the FEI dataset for better reconstruction. Then, these natively low-quality faces are resized to have a size of  $30 \times 25$ . Fig. 10 shows the visual results of the respective methods on several real low-quality images with block or mixture noise. Compared with other methods, our VCRL-ANE and MCRL-ANE can yield the best visual performance. We can see that compared with the results in Figs. 8 and 9, our method can still generate some ghosting effects around the occlusion area, which further illustrates the difficulty of hallucinating faces in real-world applications.

## VI. SUMMARIES AND FUTURE WORK

In this work, an approach named context-patch representation learning with adaptive neighbor embedding (CRL-ANE) was proposed for face image super-resolution. To obtain stable and robust representation weights, we utilized the context-patch

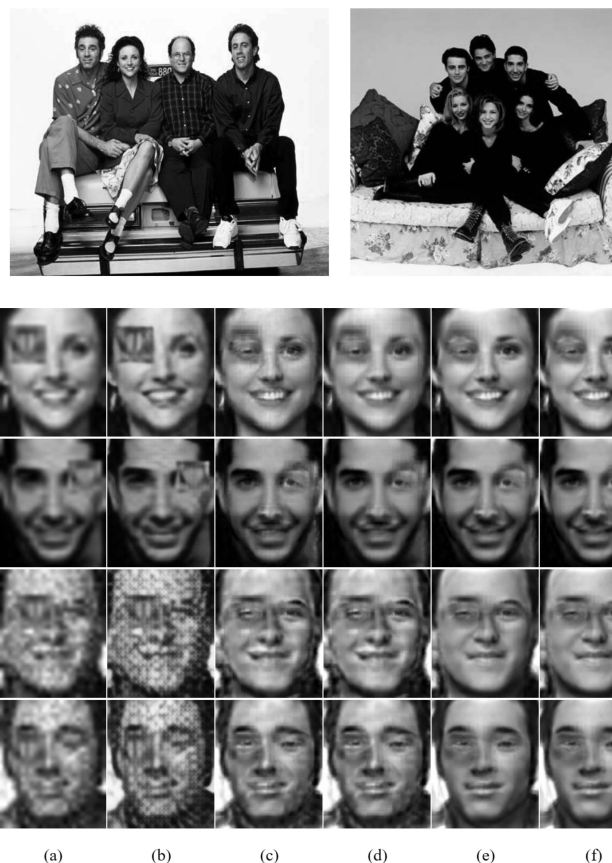


Fig. 10. Hallucinated results for LR faces extracted from the real-world dataset. From left to right: (a) the input LR noisy faces, (b) the results of FSRNet [38], (c) the results of RLcBR [34], (d) the results of PRGFC [36], (e) the results of our VCRL-ANE, and (f) the results of our MCRL-ANE.

around the testing patch for representation learning. Additionally, based on the inherent facial structural properties, we design an adaptive neighbor embedding strategy for each input patch set. By combining the two strategies, the proposed framework can perform stable representation learning and accurate reconstruction. Experiments on the public face dataset have shown the efficiency and effectiveness of the proposed method over some state-of-the-arts.

In real-world surveillance scenes, the pose and misalignment variations cannot be also ignored. For such degraded faces, learning the robust feature representation should also be well-studied. Furthermore, incorporating more highly structured facial priors (e.g., position and noise priors) into deep models to handle the diverse noisy face image super-resolution task is also part of our future work.

## REFERENCES

- [1] J. Li, Z. Pei, and T. Zeng, "From beginner to master: A survey for deep learning-based single-image super-resolution," 2021, *arXiv:2109.14335*.
- [2] G. Gao et al., "Feature distillation interaction weighting network for lightweight image super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, 2022, vol. 36, pp. 661–669.
- [3] G. Gao et al., "Lightweight bimodal network for single-image super-resolution via symmetric cnn and recursive transformer," 2022, *arXiv:2204.13286*.
- [4] X. Jing et al., "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1363–1378, Mar. 2017.
- [5] L. Wang, D. Li, Y. Zhu, L. Tian, and Y. Shan, "Dual super-resolution learning for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3774–3783.
- [6] J. Jiang, C. Wang, X. Liu, and J. Ma, "Deep learning-based face super-resolution: A survey," *ACM Comput. Surv.*, vol. 55, no. 1, pp. 1–36, 2021.
- [7] G. Gao et al., "Learning robust and discriminative low-rank representations for face recognition with occlusion," *Pattern Recognit.*, vol. 66, pp. 129–143, 2017.
- [8] J. Li et al., "Toward a comprehensive face detector in the wild," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 104–114, Jan. 2019.
- [9] G. Gao, Y. Yu, J. Yang, G.-J. Qi, and M. Yang, "Hierarchical deep CNN feature set-based representation learning for robust cross-resolution face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 5, pp. 2550–2560, May 2022.
- [10] H. Huang, H. He, X. Fan, and J. Zhang, "Super-resolution of human face image using canonical correlation analysis," *Pattern Recognit.*, vol. 43, no. 7, pp. 2532–2543, 2010.
- [11] C. Liu, H.-Y. Shum, and W. T. Freeman, "Face hallucination: Theory and practice," *Int. J. Comput. Vis.*, vol. 75, no. 1, pp. 115–134, 2007.
- [12] K. Jia and S. Gong, "Generalized face super-resolution," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 873–886, Jun. 2008.
- [13] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.
- [14] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
- [15] J. Jiang et al., "SRLSP: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 27–40, Jan. 2017.
- [16] X. Zeng, H. Huang, and C. Qi, "Expanding training data for facial image super-resolution," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 716–729, Feb. 2018.
- [17] J. Jiang et al., "Context-patch face hallucination based on thresholding locality-constrained representation and reproducing learning," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 324–337, Jan. 2020.
- [18] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognit.*, vol. 43, no. 6, pp. 2224–2236, Jun. 2010.
- [19] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Noise robust face hallucination via locality-constrained representation," *IEEE Trans. Multimedia*, vol. 16, no. 5, pp. 1268–1281, Aug. 2014.
- [20] L. Liu, S. Li, and C. P. Chen, "Quaternion locality-constrained coding for color face hallucination," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1474–1485, May 2018.
- [21] J. Shi and G. Zhao, "Face hallucination via coarse-to-fine recursive kernel regression structure," *IEEE Trans. Multimedia*, vol. 21, no. 9, pp. 2223–2236, Sep. 2019.
- [22] X. Yu and F. Porikli, "Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3760–3768.
- [23] Y. Shi, L. Guanbin, Q. Cao, K. Wang, and L. Lin, "Face hallucination by attentive sequence optimization with reinforcement learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2809–2824, Nov. 2020.
- [24] Y. Song et al., "Joint face hallucination and deblurring via structure generation and detail enhancement," *Int. J. Comput. Vis.*, vol. 127, no. 67, pp. 785–800, 2019.
- [25] Y. Zhang et al., "Copy and paste GAN: Face hallucination from shaded thumbnails," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 7355–7364.
- [26] Z.-S. Liu, W.-C. Siu, and Y.-L. Chan, "Features guided face super-resolution via hybrid model of deep learning and random forests," *IEEE Trans. Image Process.*, vol. 30, pp. 4157–4170, 2021.
- [27] C. Chen, D. Gong, H. Wang, Z. Li, and K.-Y. K. Wong, "Learning spatial attention for face super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 1219–1231, 2021.
- [28] G. Gao et al., "Constructing multilayer locality-constrained matrix regression framework for noise robust face super-resolution," *Pattern Recognit.*, vol. 110, 2021, Art. no. 107539.
- [29] M. Li, Z. Zhang, J. Yu, and C. W. Chen, "Learning face image super-resolution through facial semantic attribute transformation and self-attentive structure enhancement," *IEEE Trans. Multimedia*, vol. 23, pp. 468–483, 2021.
- [30] C. Jung, L. Jiao, B. Liu, and M. Gong, "Position-patch based face hallucination using convex optimization," *IEEE Signal Process. Lett.*, vol. 18, no. 6, pp. 367–370, Jun. 2011.
- [31] Z. Wang, R. Hu, S. Wang, and J. Jiang, "Face hallucination via weighted adaptive sparse regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 802–813, May 2014.
- [32] J. Jiang, J. Ma, S. Tang, Y. Yu, and K. Aizawa, "Face hallucination through differential evolution parameter map learning with facial structure prior," *Inf. Sci.*, vol. 481, pp. 174–188, May 2019.
- [33] S. S. Rajput, K. Arya, and V. Singh, "Robust face super-resolution via iterative sparsity and locality-constrained representation," *Inf. Sci.*, vol. 463, pp. 227–244, Oct. 2018.
- [34] L. Liu, C. P. Chen, S. Li, Y. Y. Tang, and L. Chen, "Robust face hallucination via locality-constrained Bi-layer representation," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1189–1201, Apr. 2018.
- [35] L. Chen, J. Pan, and Q. Li, "Robust face image super-resolution via joint learning of subdivided contextual model," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5897–5909, Dec. 2019.
- [36] L. Chen, J. Pan, J. Jiang, J. Zhang, and Y. Wu, "Robust face super-resolution via position relation model based on global face context," *IEEE Trans. Image Process.*, vol. 29, pp. 9002–9016, 2020.
- [37] S. Zhu, S. Liu, C. C. Loy, and X. Tang, "Deep Cascaded Bi-Network for Face Hallucination," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 614–630.
- [38] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "FSRNet: End-to-end learning face super-resolution with facial priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2492–2501.
- [39] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, "Face super-resolution guided by facial component heatmaps," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 217–233.
- [40] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Semantic face hallucination: Super-resolving very low-resolution face images with supplementary attributes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2926–2943, Nov., 2020.
- [41] K. Zhang et al., "Super-identity convolutional neural network for face hallucination," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 183–198.
- [42] C.-C. Hsu, C.-W. Lin, W.-T. Su, and G. Cheung, "SiGAN: Siamese generative adversarial network for identity-preserving face hallucination," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6225–6236, Dec. 2019.
- [43] C. Ma, Z. Jiang, Y. Rao, J. Lu, and J. Zhou, "Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5569–5578.

- [44] Z. Wang et al., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [45] A. Gunawardana and W. Byrne, "Convergence theorems for generalized alternating minimization procedures," *J. Mach. Learn. Res.*, vol. 6, no. 12, pp. 2049–2073, Dec. 2005.
- [46] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for  $l_1$ -minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008.
- [47] H. Zhang et al., "Low-rank matrix recovery via modified Schatten- $p$  norm minimization with convergence guarantees," *IEEE Trans. Image Process.*, vol. 29, pp. 3132–3142, 2020.
- [48] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, Jun. 2010.
- [49] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, Jan. 1998.



**Guangwei Gao** (Senior Member, IEEE) received the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, Nanjing, China, in 2014. He was a Visiting Student with Computing, The Hong Kong Polytechnic University, Hong Kong, in 2011 and 2013, respectively. He was also a Project Researcher with the National Institute of Informatics, Tokyo, Japan, in 2019. He is currently an Associate Professor with the Nanjing University of Posts and Telecommunications. He has authored or

coauthored more than 60 scientific papers in IEEE/ACM/AAAI venues, including IEEE TIP/TCSVT/TITS/TMM/TIFS, ACM TOIT/TOMM, AAAI, IJ-CAI, PR, and was a reviewer for journals and conferences including IEEE TPAMI/TMM/TCSVT/TNNLS/TYCB, CVPR, ICCV, ECCV, AAAI. His research interests include pattern recognition and computer vision.



**Yi Yu** (Senior Member, IEEE) received the Ph.D. degree in information and computer science from Nara Women's University, Nara, Japan. She is currently an Assistant Professor with the National Institute of Informatics (NII), Tokyo, Japan. Before joining NII, she was a Senior Research Fellow with the School of Computing, National University of Singapore, Singapore. Her research interests include machine learning and deep learning for multimedia understanding, and knowledge discovery. She was the recipient of the Best Paper Award from the IEEE ISM 2012, the 2nd

prize in Yahoo Flickr Grand Challenge 2015, the 2nd place (out of 29 teams) in ACM SIGSPATIAL GIS Cup 2013, the Best Paper Runner-Up in APWeb-WAIM 2017, and were recognized as finalist of the World's FIRST 10 K Best Paper Award in ICME 2017.



**Huimin Lu** (Senior Member, IEEE) received the B.S. degree in electronics information science and technology from Yangzhou University, Yangzhou, China, in 2008, the M.S. degree in electrical engineering from the Kyushu Institute of Technology, Kitakyushu, Japan, and Yangzhou University in 2011, and the Ph.D. degree in electrical engineering from the Kyushu Institute of Technology in 2014. From 2013 to 2016, he was a JSPS Research Fellow (DC2, PD, and FPD) with the Kyushu Institute of Technology. He is currently an Associate Professor with

the Kyushu Institute of Technology and an Excellent Young Researcher of MEXT-Japan. His research interests include computer vision, robotics, artificial intelligence, and ocean observing.



**Jian Yang** (Member, IEEE) received the Ph.D. degree from the Nanjing University of Science and Technology (NUST), Nanjing, China, on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a Postdoctoral Researcher with the University of Zaragoza, Zaragoza, Spain. From 2004 to 2006, he was a Postdoctoral Fellow with Biometrics Centre of Hong Kong Polytechnic University, Hong Kong. From 2006 to 2007, he was a Postdoctoral Fellow with the Department of Computer Science of New Jersey Institute of Technology, Newark, NJ, USA. He

is currently a Chang-Jiang Professor with 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 with the Scholar Google. His research interests include pattern recognition, computer vision and machine learning. He is/was currently an Associate Editor for the *Pattern Recognition Letters*, *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, and *Neurocomputing*. He is a Fellow of IAPR.



**Dong Yue** (Fellow, IEEE) received the Ph.D. degree in engineering from the South China University of Technology, Guangzhou, China, in 1995. He is currently a Professor and the Dean of the Institute of Advanced Technology and College of Automation & AI, Nanjing University of Posts and Telecommunications, Nanjing, China. He was the Associate Editor for the *Journal of the Franklin Institute* and *International Journal of Systems Sciences* and the Guest Editor of the Special Issue on New Trends in Energy Internet: Artificial Intelligence-based Control, Network Security and Management. Up to now, he has authored or coauthored more than 250

papers in international journals and two books. He holds more than 50 patents. His research interests include analysis and synthesis of networked control systems, multiagent systems, optimal control of power systems, and Internet of Things. Prof. Yue was the Associate Editor for the *IEEE Industrial Electronics Magazine*, *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, *IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS*, *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*.