

Discriminative Face Hallucination via Locality-Constrained and Category Embedding Representation

Licheng Liu¹, Member, IEEE, Rushi Lan², and Yaonan Wang³

Abstract—Recent years have witnessed the rapid development of face image hallucination techniques. However, the previous face hallucination methods are unsupervised and ignore the label information of training samples, leading to undesirable results. This article proposes a locality-constrained and category embedding representation (LCER) method to super-resolve face image in a supervised manner by embedding the label information in data representation. The proposed LCER incorporates the locality prior and category information into one unified framework, which aims to learn both the advantages of locality in preserving the true typologic structure of data manifold and the discriminability in exposing the class subspace information. Such strategy allows the LCER not only to preserve more sharpen image details but also to guarantee the face structure pattern be transferred mainly from the same subject in super-resolution reconstruction. Extensive experiments were conducted to evaluate the proposed LCER, and the comparative results demonstrate that it achieved superior face hallucination performance in both the quantitative measurements and visual impressions compared to several state-of-the-art.

Index Terms—Category embedding, discriminative face hallucination, locality-constrained representation (LcR), manifold learning.

I. INTRODUCTION

FACE hallucination, aiming to promote the performance of face recognition [1], is such a domain-specific super-resolution technique that tries to infer high-resolution (HR) face images from the corresponding low-resolution (LR) observations. After the seminal works of [2] and [3], many efforts have been devoted to increasing more detailed face features as much as possible. Generally speaking, the face

hallucination methods in literature can be classified into two main types, namely, global face methods based on statistical models and local face methods based on small patch priors.

The statistical model-based global face methods try to reconstruct HR images on the correlation mapping derived from the LR and HR face image pairs by leveraging the face statistical models. Exemplificative models, including but not limited to the principal component analysis (PCA) [4], the locality preserving projections (LPPs) [5], the canonical correlation analysis (CCA) [6], the singular value decomposition (SVD) [7], [8], the orthogonal procrustes regression (OPR) [9], and so on. Though these algorithms are able to preserve the global structures of human face well, the global-based methods share a common challenge in preserving image details like edge and texture information.

In contrast to global reconstruction, the local patch prior-based face approaches can further preserve the image details via handling small patches rather than holistic images. The neighbor embedding (NE)-based method is the first attempt that studied the correlations from neighbor patches for super-resolution reconstruction [10]. Jiang *et al.* [11] further presented a coupled-layer NE (CLNE) with graph regularization for facial image hallucination, where a more robust NE was achieved by iteratively updating the representation weights. The success of NE-based methods is attributed to the assumption that the LR and HR patch spaces maintain the same topology structure, which is originated from the well-known locally linear embedding (LLE) [12] algorithm. Thus, the projection coefficients learned from the LR patch space can be used directly in the HR patch space to synthesis the desirable HR patch.

Actually, the human face is highly structured, and similar patterns are always trend to present in the same position. Based on this observation, Ma *et al.* [13] introduced a least square regression (LSR)-based position patch method for face hallucination, where the query LR patch was projected on the space spanned by all the training patches seated at the identical location. A flaw of LSR is that it employs no constraint on representation coefficients, leading to unstable solutions. Up to now, various prior knowledge has been exploited to stabilize the linear system, among which the sparsity and locality are the two representative priors.

The sparsity prior assumes that data is inherent sparse and can be approximated by a subset of training samples [14]–[16].

Manuscript received September 3, 2019; revised December 1, 2019; accepted December 25, 2019. Date of publication February 14, 2020; date of current version November 18, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61702169, Grant 61702129, and Grant 61733004, in part by the Natural Science Foundation of Hunan Province under Grant 2018JJ3070, in part by the Fundamental Research Funds for the Central Universities under Grant 531107050878, and in part by the Guangxi Science and Technology Project under Grant 2019GXNSFAA245014 and Grant AD18281079. This article was recommended by Associate Editor Z. Yu. (Corresponding author: Rushi Lan.)

Licheng Liu and Yaonan Wang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: lichenghnu@gmail.com; yaonan@hnu.edu.cn).

Rushi Lan is with the Guangxi Key Laboratory of Image and Graphic Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: rslan2016@163.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TSMC.2020.2965572>.

Digital Object Identifier 10.1109/TSMC.2020.2965572

Motivated by this, Yang *et al.* [17] developed a sparse representation model for face image super-resolution with the assistance of coupled LR and HR dictionaries. Li *et al.* [18] proposed to hallucinate face image by learning the sparse local-pixel structures of the target HR facial images. Wang *et al.* [19] introduced an l_p -norm induced weighted sparse representation model for face hallucination. Jiang *et al.* [20] presented a smooth sparse representation (SSR) to hallucinate face image, which tries to strengthen the smooth structure of the training data manifold. The collaborative representation is involved to face super-resolution too, since it can achieve a comparable performance compared to sparse coding but with a very faster speed [21]. To take advantage of deep architecture, Farrugia and Guillemot [22] proposed to hallucinate the face image via a coupled layer collaborative representation.

In contrast, the locality prior induced by weighted l_2 -norm encourages the training samples similar with the test one to contribute significantly to the final result. Jiang *et al.* [23], [24] introduced the locality-constrained representation (LcR) scheme to model the position patches for face hallucination. The representation capacity of LcR was later enhanced by incorporating with Tikhonov regularization [25]. Shi *et al.* [26], [27] employed the LcR model in high dimensional feature space induced by a kernel function, where the nonlinear characteristics of the original HR space can be captured to enhance the qualities of super-resolution results. The LcR framework was further extended into the quaternion domain [28] and a quaternion locality-constrained coding (QLC) was proposed for color face image super-resolution [29]. Later, Jiang *et al.* [30] presented a thresholding LcR with the strategy of reproducing learning (TLcR-RL) for face hallucination, in which the newly reconstructed face was added back to the training dataset to strengthen the distribution comparability between the query face and the training images. Considering that images are easily corrupted by noise in practice, Liu *et al.* introduced two different reweighting strategies into the locality model and presented a robust locality-constrained bi-layer representation (RLcBR) [31] as well as an iterative relaxed collaborative representation (iRCR) [32] to hallucinate face images affected by impulse noise and Gaussian noise, respectively.

Recent years, the deep learning techniques have been successfully applied for image super-resolution and face hallucination. For example, Dong *et al.* [33] proposed a convolutional neural network (SRCNN) for general image super-resolution. This article is the first attempt to utilize deep models for super-resolution reconstruction by learning an end-to-end mapping between the LR and HR spaces. Liu *et al.* [34] introduced the domain expertise to a cascade sparse coding-based network (CSCN) for super-resolution. Lu *et al.* [35] presented a deep linear mapping learning (DLML) framework for face hallucination. Kim *et al.* [36] proposed a very deep convolutional network for super-resolution (VDSR), in which the cascaded small filters were used to extract context characteristics over large image regions. Ledig *et al.* [37] extended the generative adversarial network (GAN) to SRGAN for

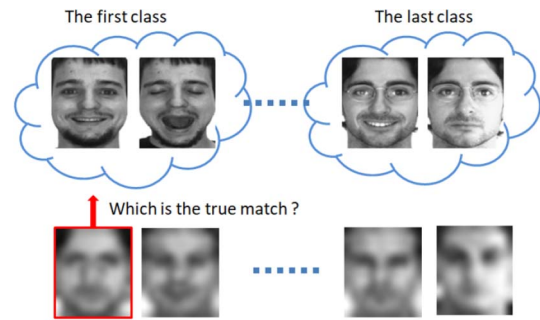


Fig. 1. Discriminative face hallucination problem encountered in the face recognition system, where the training images are partitioned into different classes and the input probe image belongs to one of the classes in the training dataset but with low resolution. The low-resolution problem will severely degrade the performance of the recognition system. Therefore, how to improve the qualities of test face images with the assistant of labeled training samples becomes an emergency issues in the face hallucination field.

single image super-resolution. Yu and Porikli [38] adopted the decoder–encoder–decoder scheme and presented a transformative discriminative autoencoder for face hallucination. Generally, the deep face hallucination trends to utilize certain pretrained deep models and fine tuning with the facial priors to generate temporary faces, which are further used to sharpen the hallucination results from the regression model [39], [40].

Though the previous methods achieved promising performance to a certain extent, one common limitation is that they ignore the discriminative face hallucination problem which is frequently encountered in the face recognition system. Considering that the LR problem commonly exists in many practical face recognition systems, and the degraded input face images will severely decline the system performance (see Fig. 1) [41]. One intuitive way to promote the system performance is to improve the visual quality of input face images by some appropriate face image super-resolution techniques. Therefore, how to design the discriminative face hallucination methods to improve the qualities of test face images with the assistant of labeled training samples in the recognition system becomes an emergency issue in the face hallucination field.

However, the aforementioned face hallucination methods mainly focus on the manifold structure assumption but rarely take into consideration of the discriminative information of training data, which significantly benefits the image representation and classification applications [42], [43]. To address this concern, in this article, we propose a new discriminative face hallucination method named locality-constrained and category embedding representation (LCER) from a different point of view to previous ones. In LCER, the unsupervised locality-constrained coding and supervised category embedding are united into one framework, which benefits our method in twofold. First of all, since the proposed LCER is a position-patch-based approach, it is reasonable to apply the nearest neighbors to represent the test patch. Moreover, for discriminative face super-resolution, since the label information of training samples is known in advance (shown in Fig. 1), it is meaningful to encourage the samples with the same label to

offer more contribution in data representation. The advantages of using locality prior as well as label information admits the proposed LCER method to preserve more image details and desirable face patterns in super-resolution reconstruction. The contributions of this article are summarized as the following threefold.

- 1) The proposed LCER employs label information of training samples in data representation, which guarantees the face structure pattern be mainly transferred from the correct class in HR face synthesis. To the best of our knowledge, this is the first attempt to super-resolved face images in the supervised manner.
- 2) The discriminative term that exploits the local information for super-resolution representation in LCER is interpreted from the Bayes probability perspective. This not only ensures the rationality of the proposed model but also demonstrates the capacity of LCER in preserving the face subspace features from the statistical viewpoint.
- 3) We have carried out extensive experiments on both the standard face datasets and real-word images to verify the efficiency and effectiveness of the proposed method in hallucinating both the clean and noisy face images.

The remainder of this article is organized as follows. The proposed discriminative face hallucination model, as well as the corresponding optimization strategy is introduced in Section II. The experimental results and analysis are presented in Section III. Section IV draws a conclusion.

II. PROPOSED DISCRIMINATIVE FACE HALLUCINATION METHOD

A. LCER Model

In this article, we consider a different scenario, where the training samples contain the face images that belong to the same class (person) but with different illuminations and expressions associated to the input test one. Such scenario is necessary and meaningful, especially, for the video surveillance system (see Fig. 1). Therefore, in such case, the face hallucination becomes into the task that super-resolves the LR observations by the training samples partitioned into different classes, among which the test face image belongs to. Such face image super-resolution problem with labeled training data is called discriminative face hallucination.

Suppose that the training LR and HR datasets are composed of C classes, e.g., $\mathbf{X}^L = \{\mathbf{X}_1^L, \mathbf{X}_2^L, \dots, \mathbf{X}_C^L\} \in R^{N \times K}$ and $\mathbf{X}^H = \{\mathbf{X}_1^H, \mathbf{X}_2^H, \dots, \mathbf{X}_C^H\} \in R^{M \times K}$, where each face image is stretched into an N dimension column vector for the LR face and an M dimension column vector for the HR face, and the i th class includes J_i number of face images, that is, $\mathbf{X}_i^L = \{X_1^L, \dots, X_{J_i}^L\}$ and $\mathbf{X}_i^H = \{X_1^H, \dots, X_{J_i}^H\}$ with $K = \sum_{i=1}^C J_i$ be the total number of training images.

Denote by $\mathbf{y}^L \in R^N$ a vectorized test face image, in the position-patch manner, it is first divided into H small overlapping patches $\{\mathbf{y}^L(i, j) | 1 \leq i \leq P, 1 \leq j \leq Q\} \in R^{n \times K}$ with $H = PQ$, where P is the number of overlapping patches in each row, and Q is the number of overlapping patches in each column. Analogously, the LR and HR training image pairs can be also

divided into small patch sets, $X_k^L(i, j) = \{x_k^L(i, j) | 1 \leq i \leq P, 1 \leq j \leq Q\}_{k=1}^K \in R^{n \times K}$ and $X_k^H(i, j) = \{x_k^H(i, j) | 1 \leq i \leq P, 1 \leq j \leq Q\}_{k=1}^K \in R^{m \times K}$.

For the (i, j) th test sample (image patch) $\mathbf{x}^L(i, j) \in R^n$, basically, it can be represented by a linear combination of the corresponding LR training image patches seated at the same position

$$\mathbf{y}^L(i, j) = X^L(i, j)\mathbf{s}(i, j) \quad (1)$$

where $\mathbf{s}(i, j) = [s_1(i, j), \dots, s_K(i, j)] \in R^K$ is the representation coefficients.

Equation (1) is a linear regression problem, whose solution is [44]

$$\mathbf{s}(i, j) = (X^L(i, j)^T X^L(i, j))^{-1} X^L(i, j)^T \mathbf{y}^L(i, j) \quad (2)$$

in which the superscript T denotes the transportation operator.

Unfortunately, (1) may obtain unstable solutions, especially, when K is larger than n . Thus, additional regularization is needed in the objective function to penalize the representation coefficient as follows:

$$\min_{\mathbf{s}(i, j)} \|\mathbf{y}^L(i, j) - X^L(i, j)\mathbf{s}(i, j)\|_2^2 + \Gamma(\mathbf{s}(i, j)). \quad (3)$$

One vital problem in (3) is the chosen of penalty function $\Gamma(\cdot)$. Different settings of Γ will result in different regularization terms, achieving different representation models. For example, when Γ is set as l_1 -norm, (3) becomes to the well-known sparse representation, while Γ choose to be l_2 -norm, (3) changes into the so-called collaborative representation. Though the work in [45] argues that sparse prior of the coefficient vector $\mathbf{s}(i, j)$ is important, the work in [46] claims that the sparsity may be not so necessary since it is the collaborative representation but not the sparse representation that makes the model efficiently in representation.

For the discriminative face image super-resolution, since the label information is available in the training dataset, it is believed that, similar to the supervised approaches [47], the utilization of label information can help to improve the representation capacities of face hallucination models. Motivated by this, we suggest the following discriminative model for face image super-resolution:

$$\min_{\mathbf{s}(i, j)} \left\| \mathbf{y}^L(i, j) - X^L(i, j)\mathbf{s}(i, j) \right\|_2^2 + \Gamma(\mathbf{s}(i, j)) + \Psi(\mathbf{z}_c^L(i, j) - \mathbf{z}^L(i, j)) \quad (4)$$

where \mathbf{z}_c^L is the estimation of \mathbf{y}^L on the subspace spanned by training samples from c th class and \mathbf{z}^L is the estimation of input LR face \mathbf{y}^L with respect to the whole training samples.

Since both \mathbf{z}_c^L and \mathbf{z}^L are the reconstructions of \mathbf{y}^L associated to X_c and X , respectively, $\Psi(\mathbf{z}_c^L(i, j) - \mathbf{z}^L(i, j))$ is reformulated as $\Psi(X_c^L(i, j)\mathbf{s}_c(i, j), X^L(i, j)\mathbf{s}(i, j))$. To emphasize the discriminative information, we choose to minimize the Euclidean distance between $X_c^L(i, j)\mathbf{s}_c(i, j)$ and $X^L(i, j)\mathbf{s}(i, j)$, therefore, the third term in (10) is specified as

$$\Psi(\mathbf{z}_c^L(i, j), \mathbf{z}^L(i, j)) = \frac{\gamma}{C} \sum_{c=1}^C \left\| X_c^L(i, j)\mathbf{s}_c(i, j) - X^L(i, j)\mathbf{s}(i, j) \right\|_2^2. \quad (5)$$

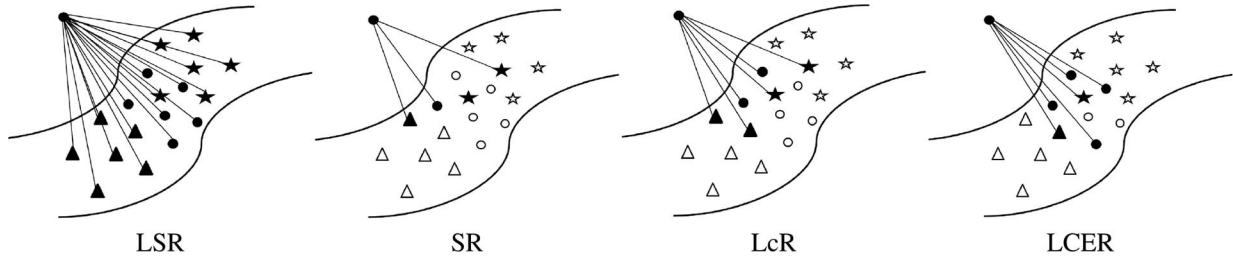


Fig. 2. Visual comparison between LSR, SR, LcR, and LCER, where the training samples with different classes are denoted by different shapes. In the proposed LCER, the majority of samples used for representation of the query patch come from the same class. Moreover, some samples, though from different classes but similar to the test one, are also used for data representation. In contrast, the LSR used all the training samples, while the SR just controls the sparsity of data encoding. The LcR takes into account the similarity prior between the test sample and training samples, but without considering the class label information.

Equation (5) penalizes the summation of the differences between the reconstruction from each class and that from the whole training dataset. By minimizing (5), the input image patch $\mathbf{y}^L(i, j)$ is encouraged to be represented by the training samples with the identical label associated with $\mathbf{y}^L(i, j)$. This can induce discriminative information for representing the query image patch (the visual impression of LCER is shown in Fig. 2).

To further emphasize the contribution of the class in representation, we add a weight into (5) as

$$\Psi(\mathbf{z}_c^L(i, j), \mathbf{z}^L(i, j)) = \frac{\gamma}{C} \sum_{c=1}^C \beta_c \|X_c^L(i, j)\mathbf{s}_c(i, j) - X^L(i, j)\mathbf{s}(i, j)\|_2^2 \quad (6)$$

where β_c is the weight assigned for c th class which is defined as

$$\beta_c = \begin{cases} 1, & \text{if } c \text{ is the correct label for } \mathbf{y}^L(i, j) \\ \frac{1}{C}, & \text{otherwise.} \end{cases} \quad (7)$$

The l_2 -norm is chosen to serve as $\Gamma(\cdot)$ to penalize the coefficients. However, different from collaborative representation, which uses l_2 -norm to regularize the coefficients directly, we prefer to regularize the weighted coefficients as follows:

$$\Gamma(\mathbf{s}(i, j)) = \|\boldsymbol{\tau}(i, j) \odot \mathbf{s}(i, j)\|_2^2 \quad (8)$$

where \odot represents the Hadamard product and $\boldsymbol{\tau}(i, j) = (\tau_1(i, j), \tau_2(i, j), \dots, \tau_K(i, j))^T$ denotes a K -dimensional weight vector that penalizes the distance between the test face image patch and each training patch. The k th entry of $\boldsymbol{\tau}(i, j)$ is defined simply by the Euclidean distance

$$\tau_k(i, j) = \|\mathbf{y}^L(i, j) - \mathbf{x}_k^L(i, j)\|_2^2. \quad (9)$$

As can be seen that, if the k th training sample approaches to the test patch, it is assigned by a small weight, while a large weight is assigned to the training sample, that is, far away from the test one. Therefore, by using the Euclidean distances as the weights, the nearest neighbors of the test patch in training samples will provide more contribution in representation than those non neighborhoods, which can enclose the true topological structure of the patch manifold [48].

Substituting (5) and (8) into (4), we obtain the final LCER framework for discriminative face hallucination

$$\begin{aligned} \min_{\mathbf{s}(i, j)} \quad & \|\mathbf{y}^L(i, j) - X^L(i, j)\mathbf{s}(i, j)\|_2^2 + \lambda \|\boldsymbol{\tau}(i, j) \odot \mathbf{s}(i, j)\|_2^2 \\ & + \frac{\gamma}{C} \sum_{c=1}^C \beta_c \|X_c^L(i, j)\mathbf{s}_c(i, j) - X^L(i, j)\mathbf{s}(i, j)\|_2^2 \\ \text{s.t.} \quad & \mathbf{1}^T \mathbf{s}(i, j) = 1. \end{aligned} \quad (10)$$

The proposed LCER model utilizes the locality prior and the discriminative information for face image hallucination. It is expected to learn both the advantages of locality in preserving the true manifold structure and the discriminant in exposing the class-specific information. On the one side, for discriminative face image super-resolution, since the label information of the test image is known in advance, it is reasonable to utilize the data from the same class to reconstruct the HR face image. On the other side, the proposed framework is a position-patch-based method, in which small position patches are served as basic units to synthesize the target HR patches. Therefore, it is meaningful to use the nearest neighbors not limited to the corresponding class but throughout the whole training dataset to represent the test LR patch, since small similar patterns can exist across different images in different classes [21].

B. Interpretation From Probability Perspective

For a deep observation, the innovation of LCER is mainly contributed to the discriminative term $\gamma \sum_{c=1}^C \beta_c \|X_c^L \mathbf{s}_c - X^L \mathbf{s}\|_2^2$, which admits the proposed LCER significantly differ from LcR or other typical face hallucination models. Actually, the discriminative term in LCER can be interpreted from the probability viewpoint as follows.

For a collection of LR training samples from C classes $X^L = [X_1^L, X_2^L, \dots, X_C^L]$, where X_c^L denotes the data matrix of class c and each column of X_c^L is a sample vector. Thus, the data matrix X^L can be viewed as an expanded class, and we use l_X^L to represent the label set of all the candidate classes in X^L . Denote by \mathcal{S} the linear space spanned by all the training samples in X . Suppose $\mathbf{x}^L \in \mathcal{S}$ is an estimation of the sample \mathbf{y}^L , it can be linearly represented by the training samples with proper regression model, i.e., $\mathbf{x}^L = X^L \mathbf{s} = \sum_{c=1}^C \mathbf{x}_c^L \mathbf{s}_c$, where $\mathbf{s} = [\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_C^T]^T$, and \mathbf{s}_c is the coefficient vector corresponding to \mathbf{x}_c^L . Thus, $\mathbf{x}_c^L = X_c^L \mathbf{s}_c$ is a data point in the subspace associated with class c .

Recall that the discriminative learning is try to approximate \mathbf{y}^L by a proper subspace rather than the holistic one. In other words, it is expected that the label of \mathbf{x}^L should be close to \mathbf{x}_c^L as much as possible to preserve the face subspace characteristics. This can be achieved by maximizing the following probability:

$$P(l(\mathbf{x}^L) = l(\mathbf{x}_c^L) | l(\mathbf{x}_c^L) \in l_X^L). \quad (11)$$

One can adopt the Gaussian kernel, which is a widely used measure to characterize the neighbor-based similarity, to define the probability [42]

$$P(l(\mathbf{x}^L) = l(\mathbf{x}_c^L) | l(\mathbf{x}_c^L) \in l_X^L) \propto \exp(-\gamma\beta_c \|\mathbf{x}^L - \mathbf{x}_c^L\|).$$

To further improve the representation capacity, we choose to maximize the joint probability

$$\begin{aligned} \max P(l(\mathbf{x}^L) = l(\mathbf{x}_1^L), \dots, l(\mathbf{x}^L) = l(\mathbf{x}_C^L) | l(\mathbf{x}_1^L) \in l_X^L \\ \dots, l(\mathbf{x}_C^L) \in l_X^L) \propto \max \prod_c P(l(\mathbf{x}^L) = l(\mathbf{x}_c^L)) \\ \propto \max \exp\left(-\gamma\beta_c \sum_{c=1}^C \|\mathbf{x}^L - \mathbf{x}_c^L\|_2^2\right). \end{aligned} \quad (12)$$

Applying the logarithmic operator and discard the constant terms, (12) can be reformulated as

$$\min \gamma \sum_{c=1}^C \beta_c \|\mathbf{X}^L \mathbf{s} - \mathbf{X}_c^L \mathbf{s}_c\|_2^2 \quad (13)$$

where β_c is the class specific parameter, indicating certain prior knowledge.

As can be seen, (13) is actually the third term of the objective function in LCER.

C. Optimization Strategy

The proposed LCER model admits a closed form solution since it is actually a regularized least square problem. In this section, the indices i and j are omitted without confusion. For convenience, we write the objective function of (10) into subobjectives as follows:

$$F(\mathbf{s}) = F_D(\mathbf{s}) + \lambda F_L(\mathbf{s}) + \frac{\gamma}{C} F_C(\mathbf{s}) \quad (14)$$

with

$$\begin{aligned} F_D(\mathbf{s}) &= \|\mathbf{y}^L - \mathbf{X}^L \mathbf{s}\|_2^2 \\ F_L(\mathbf{s}) &= \|\boldsymbol{\tau} \odot \mathbf{s}\|_2^2 \\ F_C(\mathbf{s}) &= \sum_{c=1}^C \beta_c \|\mathbf{X}_c^L \mathbf{s}_c - \mathbf{X}^L \mathbf{s}\|_2^2 \end{aligned}$$

be the data fidelity term, the locality regularization term, and the category regularization term, respectively.

The second term can be rewritten as

$$F_L(\mathbf{s}) = \|\text{diag}(\boldsymbol{\tau})\mathbf{s}\|_2^2 \quad (15)$$

where $\text{diag}(\cdot)$ is diagonalization operation which transforms a vector into a diagonal matrix with the diagonal be the vector.

To maintain the consistency of the variables, we convert X_c into \dot{X}_c which has the same size as X . The elements of

X_c are assigned to \dot{X}_c at their corresponding locations in X , while the elements located at other locations are set to zeros. Mathematically, \dot{X}_c is defined as

$$\dot{X}_c = [0, \dots, X_c, \dots, 0]. \quad (16)$$

Therefore, the third term in (14) can be reformulated as

$$F_C(\mathbf{s}) = \sum_{c=1}^C \beta_c \|\mathbf{X}^L \mathbf{s} - \dot{X}_c^L \mathbf{s}\|_2^2 = \sum_{c=1}^C \beta_c \|\bar{X}_c^L \mathbf{s}\|_2^2 \quad (17)$$

in which $\bar{X}_c^L = \mathbf{X}^L - \dot{X}_c^L$.

With the constraint $\sum_{k=1}^K s_k = 1$, we rewrite the objective function into the following form:

$$\begin{aligned} F(\mathbf{s}) &= \left\| \mathbf{y} \sum_{k=1}^K s_k - \sum_{k=1}^K (\mathbf{x}_k s_k) \right\|_2^2 \\ &\quad + \lambda \|\text{diag}(\boldsymbol{\tau})\mathbf{s}\|_2^2 + \frac{\gamma}{C} \sum_{c=1}^C \beta_c \|\bar{X}_c^L \mathbf{s}\|_2^2 \\ &= \left\| \sum_{k=1}^K (\mathbf{y} - \mathbf{x}_k) s_k \right\|_2^2 + \lambda \|\text{diag}(\boldsymbol{\tau})\mathbf{s}\|_2^2 + \frac{\gamma}{C} \sum_{c=1}^C \beta_c \|\bar{X}_c^L \mathbf{s}\|_2^2 \\ &= \|(\mathbf{y}\mathbf{1}^T - X)\mathbf{s}\|_2^2 + \lambda \|\text{diag}(\boldsymbol{\tau})\mathbf{s}\|_2^2 + \frac{\gamma}{C} \sum_{c=1}^C \beta_c \|\bar{X}_c^L \mathbf{s}\|_2^2 \\ &= \mathbf{s}^T (\mathbf{y}\mathbf{1}^T - X)^T (\mathbf{y}\mathbf{1}^T - X) \mathbf{s} + \lambda \mathbf{s}^T \text{diag}(\boldsymbol{\tau})^2 \mathbf{s} \\ &\quad + \mathbf{s}^T \left(\frac{\gamma}{C} \sum_{c=1}^C \beta_c (\bar{X}_c^L)^T \bar{X}_c^L \right) \mathbf{s} \\ &= \mathbf{s}^T Q \mathbf{s} \end{aligned} \quad (18)$$

where $Q = (\mathbf{y}\mathbf{1}^T - X)^T (\mathbf{y}\mathbf{1}^T - X) + \lambda \text{diag}(\boldsymbol{\tau})^2 + (\gamma/C) \sum_{c=1}^C \beta_c (\bar{X}_c^L)^T \bar{X}_c^L$.

Finally, the optimization problem in (10) is formulated as follows:

$$\min \mathbf{s}^T Q \mathbf{s}, \quad \text{s.t. } \mathbf{1}^T \mathbf{s} = 1. \quad (19)$$

The Langrange of the objective function is

$$L(\mathbf{s}, \nu) = \mathbf{s}^T Q \mathbf{s} + \nu (\mathbf{1}^T \mathbf{s} - 1) \quad (20)$$

where ν is the Langrange multiplier.

The optimal \mathbf{s} can be obtained by taking the derivative of L with respect to \mathbf{s} and ν , and setting them to zeros, respectively,

$$\begin{cases} \frac{\partial L}{\partial \mathbf{s}} = 2Q\mathbf{s} + \nu = 0 \\ \frac{\partial L}{\partial \nu} = \mathbf{1}^T \mathbf{s} - 1 = 0. \end{cases} \quad (21)$$

From the first equation in (21), one can obtain

$$\mathbf{s} = -\frac{\nu}{2} Q^{-1} \mathbf{1}. \quad (22)$$

Substituting (22) into the second equation in (21), we have

$$\nu = -2 \left(\mathbf{1}^T Q^{-1} \mathbf{1} \right). \quad (23)$$

Finally, by substituting (23) into the first equation in (21), the optimal \mathbf{s} can be obtained

$$\hat{\mathbf{s}} = \frac{Q^{-1}}{\mathbf{1}^T Q^{-1} \mathbf{1}}. \quad (24)$$

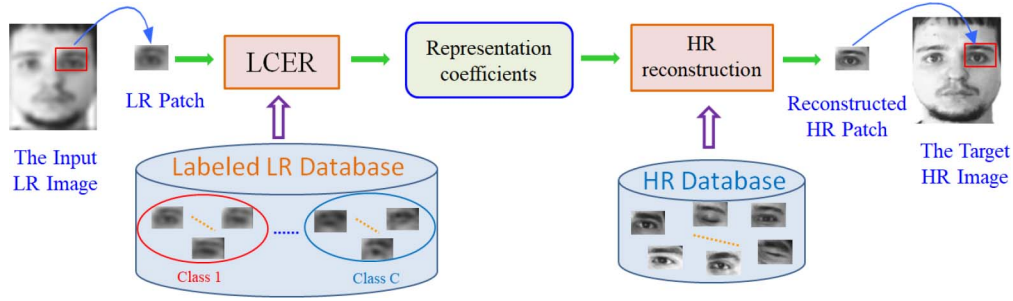


Fig. 3. Flow chart of position patch-based face hallucination via LCER.

Instead, one can also solve the linear system of equations $Qs = 1$ to avoid computing the inverse of Q . Then the obtained coefficient vector \hat{s} need to be normalized so as to meet the constraint $\mathbf{1}^T s = 1$.

D. Face Super-Resolution via LCER

For face image super-resolution, two training sets are needed, namely, the LR training set composed of LR training face images and the HR training set containing the corresponding HR training face images. Denote by $\{X_k^H\}_{k=1}^K$ and $\{X_k^L\}_{k=1}^K$ the HR and LR training sets, respectively, the primary task is to forecast the HR face image Y^H from its LR observation Y^L .

First, the input LR and training face images are all partitioned into small overlapping patches by the separation strategy used in [23]. Each input LR test image patch now can be represented via a linear combination of the LR training ones located at the identical position by using the proposed LCER model. By doing this, a weight vector (representation coefficient) will be achieved associated to the LR training image patches. Analogously, by assuming the HR image patch manifold share the identical topological structure with the LR one [10], the target HR image patch can be reconstructed by transforming the weights from the LR image patch manifold to the corresponding HR ones. The final HR face image can be achieved by concatenating all the synthesized HR image patches to the corresponding positions with the pixel values in the overlapping regions be averaged. The flow chart of position-patch-based face hallucination method via LCER is shown in Fig. 3, and the implementation details are summarized in Algorithm 1.

E. Initial Label Estimation

As discussed previously, the label information of the input LR face is needed in the proposed LCER method. However, in practical application scenarios, e.g., the face recognition system, the label of the input image is actually unknown, which usually needs to be identified. To solve this problem, an intuitive way is employing some well-known classification algorithms to recognize the label of the input image. To this end, various face recognition methods can be chosen. In this article, we estimate the label of input LR face by the LLC algorithm. Mathematically, the input LR face y^L is first represented via LR training samples with the LLC model

$$\min_s \|y^L - X^L s\|_2^2 + \lambda \|\tau \odot s\|_2^2 \quad (25)$$

Algorithm 1 Robust Face Super-Resolution via LCER

Input: The input LR face image Y^L , the LR training images $\{X_k^L\}_{k=1}^K$, the HR training images $\{X_k^H\}_{k=1}^K$, and the regularization parameter γ .

1: Divide the input LR face image into small overlapped image patches $\{y^L(i, j)\}_{i=1, j=1}^{P, Q}$.

2: For the (i, j) th position patch $y^L(i, j)$, **do**

- 1) Extract the LR and HR training samples from the LR and HR training images according to the same location. The extracted LR and HR training samples are used to construct the LR and HR codebooks $X^L(i, j)$ and $X^H(i, j)$ respectively.
- 2) Construct the locality weight vector $\tau(i, j)$ by calculating the distances between the test LR patch and the LR training samples,

$$\tau_k(i, j) = \|y_k^L(i, j) - X_k^L(i, j)\|_2^2 \quad k = 1, \dots, K$$

- 3) Set the weight vector β_c via Eq. (7).
- 4) Compute the optimal representation coefficient vector $s(i, j)$ for the test LR image patch $y(i, j)$ associated with the LR training samples $X^L(i, j)$ via Eq. (10).
- 5) Forecast the HR patch by

$$y^H(i, j) = X^H(i, j)s(i, j).$$

3: **End For.**

4: Synthesis the final target HR face image by concatenating all the above-reconstructed HR image patches with the pixel values averaged in the overlapped positions.

Output: The final hallucinated HR image Y^H .

where X^L is the codebook with each column be a vectorized training image, τ is the correlation adaptor, and s is the corresponding encoding coefficient vector.

Once the optimal coefficient \hat{s} is obtained, the sparse encoding residual associated to each class can be calculated

$$\text{Res}_i(y^L) = \|y^L - X^L \delta_i(\hat{s})\|_2^2 \quad (26)$$

where $\delta_i(\hat{s})$ means selecting the coefficients of \hat{s} corresponding to the i th class.

Finally, the class of the input LR face y^L is determined by the one that achieves the minimal sparse coding residual

$$\text{Class}(y^L) = \min_k \text{Res}_k(y^L). \quad (27)$$

TABLE I
AVERAGED QUANTITATIVE EVALUATION SCORES GENERATED BY ALL COMPARATIVE ALGORITHMS FOR 50 TEST IMAGES IN AR DATASET AND 100 TEST IMAGES IN NUST RWFR DATASET DEGRADED BY DIFFERENT DOWNSAMPLING FACTORS

Comparison Methods	AR Database						NUST RWFR Database					
	PNSR			SSIM(%)			PNSR			SSIM(%)		
	4	8	16	4	8	16	4	8	16	4	8	16
LSR	28.65	24.57	21.82	86.72	74.46	66.32	27.46	23.79	21.23	82.16	65.34	54.71
LcR	29.53	25.51	22.66	88.42	78.09	70.69	27.69	24.02	21.44	82.85	66.67	56.14
LM-CSS	28.27	24.63	21.87	87.07	74.99	66.50	26.85	23.81	21.25	82.56	66.83	55.57
SSR	29.12	23.58	21.86	87.22	73.30	67.62	27.64	23.94	21.31	82.49	66.09	54.96
RLcBR	29.56	25.62	22.84	88.45	78.59	71.69	27.76	24.09	21.52	83.13	67.29	56.85
LCER	29.98	26.74	24.21	89.09	82.86	77.71	27.99	24.55	22.22	83.89	70.24	61.36

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Settings

In our experiments, two public face datasets are chosen to evaluate LCER in face hallucination from different perspectives.

AR Dataset: A subset of 700 face images associated to 50 subjects (persons) are selected for experiments. Thus, each subject holds 14 face images, and we randomly select one image for testing with the remaining 13 ones for training. Thus, all the test face images are absent in the training set. The HR face images are cropped into 120×100 pixels and aligned according to the positions of two eyes.

NUST RWFR Dataset: There are totally 2400 face images with size of 80×80 pixels corresponding to 100 persons (classes) in NUST RWFR dataset. That is, each person holds 24 face images with various expressions and illuminations. Moreover, the face images in NUST RWFR dataset are not well aligned, leading to a more difficult hallucination problem. One face image was randomly chosen from each class for testing. Therefore, 100 face images are used as test images while the remaining 2300 ones are used for training.

The LR face images are obtained by downsampling (by a factor of 4) and blurring (by a 4×4 average smoothing filter) the corresponding HR faces. For the position-patch-based method, small image patches should be first extracted from the holistic image. In order to balance the super-resolution performance and the computational time, the HR patches are recommended with the size of 12×12 pixels, with four pixels (the experiments of patch size and overlapping analysis have been conducted in Section III-F) be overlapped between two adjacent patches. The two parameters λ and γ are empirically set to be 0.9 and 0.005, respectively.

B. Comparison on Standard Datasets

This section evaluates the proposed LCER method by comparing with five state-of-the-art face super-resolution approaches, namely, the LSR [13], the LcR [23], the linear model of coupled sparse support (LM-CSS) [22], the SSR [20], and the RLcBR [31]. The hallucination results

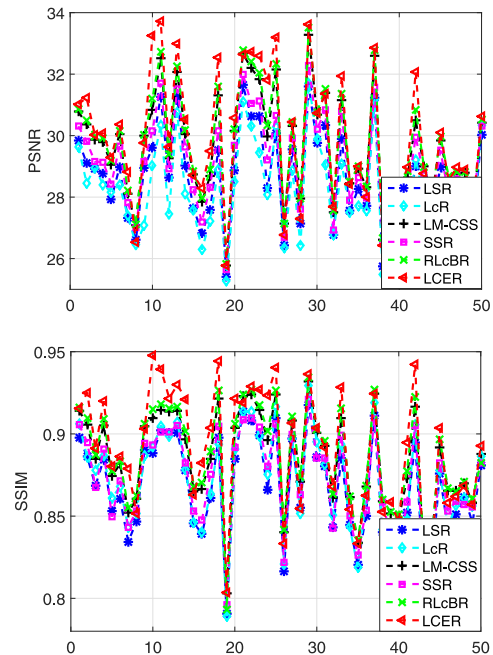


Fig. 4. PSNR and SSIM scores of super-resolved results versus the index of test images in the AR face dataset.

are quantitatively evaluated by two commonly used quality assessment measurements, namely, peak signal to noise ratio (PSNR) and structural similarity (SSIM) [49]. Generally speaking, higher PSNR and SSIM values always indicate better qualities of the hallucination results. The source codes of LcR, LM-CSS, SSR, and RLcBR are provided by the original authors, while the other methods are implemented by ourselves. Fig. 4 plots the PSNR and SSIM scores of the hallucination results for all the 50 test images in the AR dataset generated by these comparison methods, while the averaged PSNR and SSIM values are reported in Table I. One can see that, the proposed LCER method obtains the highest PSNR and SSIM values for almost all the face images in the AR dataset. The quantitative measurement results indicate the proposed LCER method achieved the best face super-resolution performance. This owes to the LCER not only

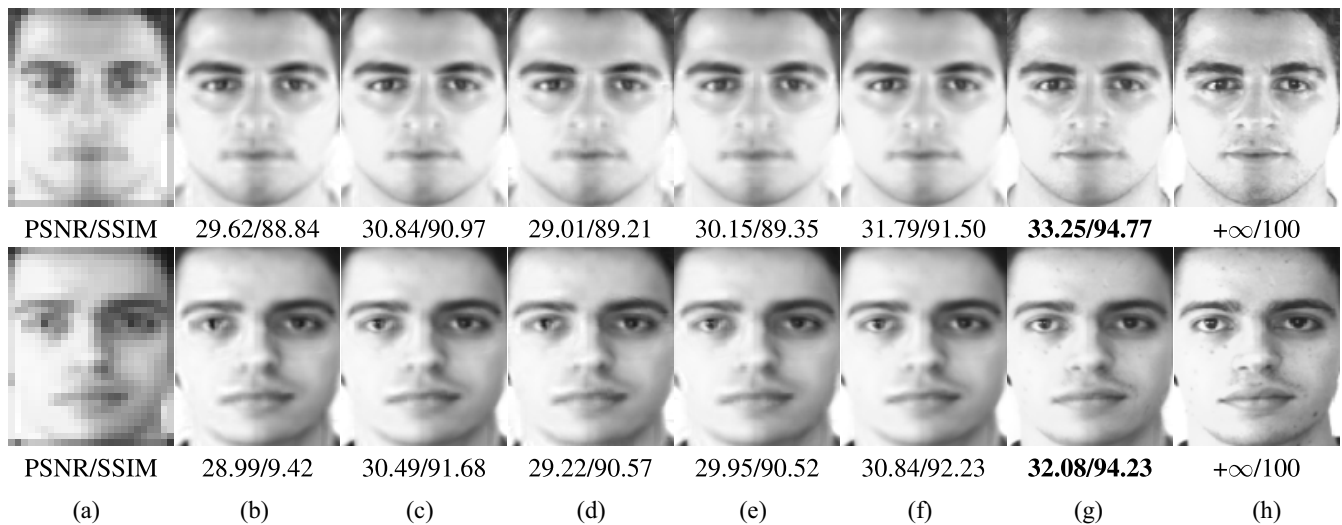


Fig. 5. Visual comparison on the AR face dataset. (a) LR faces. (b) LSR. (c) LcR. (d) LM-CSS. (e) SSR. (f) RLcBR. (g) LCER. (h) Ground truth. We have shown the SSIM values in percentage.

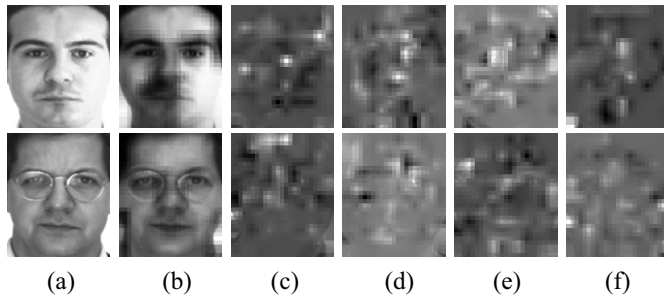


Fig. 6. Reconstruction results generated from the top five most relevant classes. (a) Target HR faces. (b) Reconstructed face from the most related class. (c)–(f) Reconstructed faces from other unrelated classes.

taking into consideration of the locality prior of patch manifold but also the discriminative label information of training data in super-resolution processing.

To give some visual impressions, Fig. 5 shows some face hallucination results of the AR database from all the compared methods. In this figure, the first column exhibits the LR faces, the last column lists the ground truth, and columns 2–6 present the hallucination results of different methods, respectively. From Figs. 4 and 5, the following conclusion can be drawn.

- 1) The locality-constrained-based methods (e.g., LcR and RLcBR) obtained better super-resolution results than the LSR-based method (i.e., LSR). This verifies the locality prior which has been proven more important than sparsity, indeed helpful to reconstruct local detailed patterns in the hallucination results.
- 2) The two-layer-based representation methods (e.g., LM-CSS and RLcBR) achieved relative poor performance. The reasons are as follows. Though it employs two layers to exploit the local geometrical structure on the HR manifold for super-resolution reconstruction, the performance of LM-CSS is limited by the representation capacity of the collaborative representation model. The RLcBR utilizes an HR layer

to compensate the noisy LR layer which is effective for noisy face hallucination, but useless for clean face hallucination.

- 3) The SSR obtains better super-resolution results than the locality prior-based methods (LcR, RLcBR). This is because that the SSR presents the smooth geometrical structure of the training image patch space by encouraging similar training samples holding similar sparse coding coefficients.
- 4) All the previous schemes are unsupervised and do not comprehensively take into account the category characteristics in super-resolution, ignoring the discriminative face patterns. In contrast, the proposed LCER achieves the most satisfactory hallucination performance, in which the face details have been well reconstructed. The reason lies in that the results of LCER benefit from the locality prior knowledge to preserve local geometrical manifold as well as the category embedding to ensure that the discriminative face patterns are mainly transferred from exemplar images belonging to the same subject.

C. Reconstruction Comparison of Each Class

In the proposed LCER model, a discriminative category embedding regularization is used to encourage the training samples from the correct category to play key roles in representation. Such strategy guarantees the reconstructed face structures be mainly transferred from the exemplar face images corresponding to the same subject.

To show the effectiveness of the discriminative term, in what follows, we exhibit the reconstruction results of two face images in the AR dataset from the most related five classes. The reconstruction results are depicted in Fig. 6, where the first column lists the original HR faces, columns 2–6 show the reconstructed faces from most related five classes. As can be seen, the reconstruction results from the correct class (the most relevant class) shown in the second column contain the

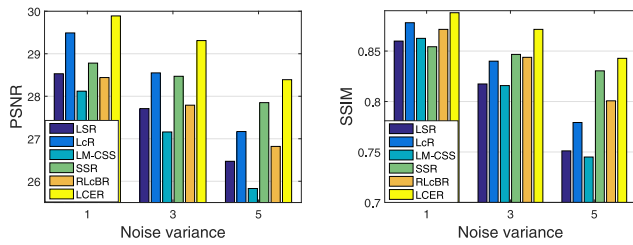


Fig. 7. Comparison of different methods on hallucinating noisy LR faces images.

major patterns of the target HR faces. The reconstructions from other classes hold little patterns but some detail information. This demonstrates that the discriminative term indeed helpful in excavating the inherent characteristics from the same subject for super-resolution reconstruction.

D. Robust Against Noise

In reality, the captured images are inevitably influenced by noise. Thus, a good face hallucination method should be able to not only enhance face details but also robust to noise. In this section, we conduct extensive experiments on the AR dataset to assess the robustness of all the comparison methods by super-resolving noisy face images. The noisy LR faces were generated by adding different levels of Gaussian noise (with standard deviation $\sigma = 1, 3,$ and 5) to the LR faces. The noisy LR faces were then fed directly into each compared method. The results were still evaluated by PSNR and SSIM indexes, and the averaged PSNR and SSIM scores are plotted in Fig. 7. As can be seen, the LCER obtains the highest PSNR and SSIM values for noisy LR face image super-resolution. This demonstrates that the proposed method is still effective in noisy environment.

E. Results on Very LR Face Images

In this section, extensive experiments were carried out to evaluate the efficiency of LCER in hallucinating very LR face images. The very LR faces are generated by blurring and downsampling the corresponding HR face images with factors of 8 and 16, respectively. The very LR face images were then hallucinated by all the comparison methods and the results are assessed by PSNR and SSIM. The averaged PSNR and SSIM values of all the 50 test images in the AR dataset and 100 test images in NUST-RWFR dataset with different magnification factors are tabulated in Table I. From this table, one can see that, the proposed LCER method obtained higher PSNR and SSIM values than other state-of-the-art for all the cases. Interestingly, the larger the magnification factor is, the more PSNR and SSIM gains achieved by the proposed LCER. Fig. 8 shows the super-resolution results of two LR face images downsampled by factors of 8 or 16 from different methods. As can be seen, the performance of all the comparison methods degraded with the increasing of the downsampling factor. This is reasonable since the large the downsampling factor is, the more image details lost in the LR images, resulting in a hard recovery problem. However, the LCER still preserved well the texture details even in high magnification factor environment

TABLE II
AVERAGED QUANTITATIVE EVALUATION SCORES OF THE RESULTS GENERATED BY THE COMPARED ALGORITHMS FOR 50 TEST IMAGES IN AR DATASET DEGRADED BY DIFFERENT DOWNSAMPLING FACTORS

patch size	overlap				
		4	8	12	16
PSNR					
0		26.21	29.04	29.52	29.42
4		—	29.69	29.97	29.68
8		—	—	30.04	29.71
12		—	—	—	29.80
SSIM					
0		79.30	86.84	88.09	87.99
4		—	88.35	89.03	88.64
8		—	—	89.31	88.73
12		—	—	—	89.00

where the structure characteristics are almost lost because of the large downsampling factor.

F. Effect of Different Patch Sizes and Overlappings

In this section, we test the performance of the proposed LCER method with various settings of patch size and overlapping pixels. The experimental settings are as follows. The HR images are set to be with the size of $4 \times 4, 8 \times 8, 12 \times 12,$ and 16×16 pixels with the overlaps between two adjacent patches to be 0, 4, 8, and 12 pixels, respectively. Thus, the patch size of the corresponding LR images is $1 \times 1, 2 \times 2, 3 \times 3,$ and 4×4 pixels with 0, 1, 2, and 3 pixels be overlapped between two adjacent patches. The LCER method with different patch size and overlap settings are applied to the 50 LR face images in the AR dataset. The averaged PSNR and SSIM values are tabulated in Table II. As can be seen, for the same patch size, the performance of LCER becomes better when the overlap becomes larger. This is reasonable since larger overlap brings in more information from adjacent patches to synthesis each pixel in the reconstructed results. For the same overlap, the performance of LCER improves along with the patch size increasing except for the patch size with 16×16 pixels. Basically, image patches with large size contain more patterns than small patches, and can provide more useful information for representation. However, too large patch may cause smoothness and lose some face details in super-resolution reconstruction, leading to undesirable results. Generally speaking, the patch size should be set neither too large nor too small. Too small patch size with little overlapping pixels will lost the human structure patterns in super-resolution reconstruction, while too large patch size with many pixels overlapped will not only lost the face details but also increase the computational complexity dramatically. Therefore, the HR patch size is set to be 12×12 with 4 pixels been overlapped between two adjacent patches to balance the hallucination performance and computational complexity.

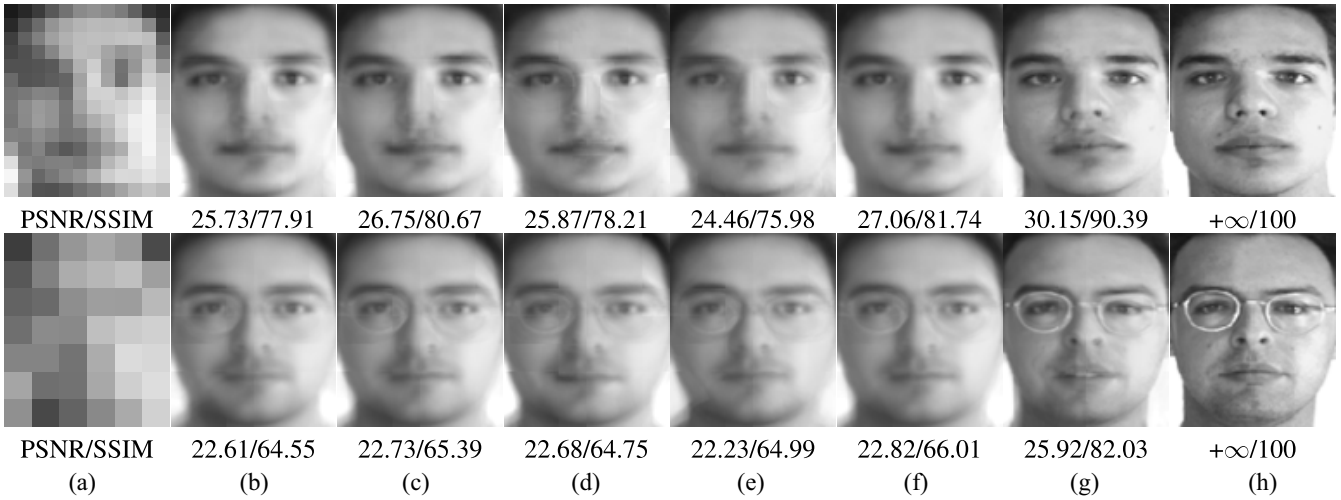


Fig. 8. Comparison of hallucination results of very LR faces in the AR face dataset. The super-resolution reconstruction of LR images downsampled by the magnification factor of 8 is shown in the first two rows, while those of LR images downsampled by magnification factor of 16 are the last two rows. (a) LR faces. (b) LSR. (c) LcR. (d) LM-CSS. (e) SSR. (f) RLcBR. (g) LCER. (h) Ground truth. The SSIM values are shown in percentage.

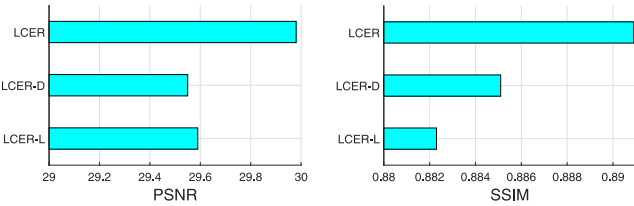


Fig. 9. Comparison of different methods on hallucinating noisy LR faces images.



Fig. 10. Two test samples in CMU + MIT database.

G. Locality Versus Discriminate

The proposed LCER achieves promising face super-resolution performance, attributing to two priors it used, namely, locality and discriminability. The former is used to enclose the inherent manifold structure, while the latter is utilized to enhance the pattern representation. To further verify the roles they played in (10), we implement two variants of LCER. Denote by “LCER-L” the model designed from LCER without the discriminate term, that is, $\gamma = 0$ in (10). The other variant denoted by “LCER-D” is designed from the final model by discarding the locality term, that is, $\lambda = 0$ in (10).

The two variants are then used to hallucinate the LR face images in the AR database and compared with LCER. The comparisons of averaged PSNR and SSIM values are in Fig. 9.

TABLE III
RUNNING TIME (SECOND) OF DIFFERENT METHODS ON ONE RANDOMLY CHOSEN LR FACE IMAGE IN THE AR DATASET

Method	LSR	LcR	LM-CSS	SSR	RLcBR	LCER
Time	4.85	6.17	12.75	86.09	13.62	10.58

From the figure, it can be seen that LCER-L and LCER-D gained competitive performance. This indicates that the locality and discriminate priors play almost equal contribution in representation. In contrast, the LCER obtains the highest scores, which means that both the locality and discriminability are necessary and meaningful in the final face hallucination model, without any of which, the performance of the final model will be degraded.

H. Hallucination of Real World Images

In the experiments mentioned above, the used LR images were generated by conducting downsampling and blurring operators on the corresponding HR images. However, as pointed out in [50], the actual spatial feature correlation between the LR and HR face spaces cannot be correctly determined by the manually adjusted LR face images.

In what follows, we conduct extensive experiments in real-world images to further evaluate the performance of LCER in practical applications. Two images holding the real-life scenes shown in Fig. 10 are chosen from the CMU + MIT database [51] for testing. The two images are with low resolution and low quality since they are captured from a certain distance. To further well simulate the real scenarios, the two image are corrupted by Gaussian noise, where seven face images are manually cropped for testing. The comparison hallucination results are shown in Figs. 11 and 12. As can be observed, the proposed LCER method can still produce high-quality HR face images in the real-word scenarios.

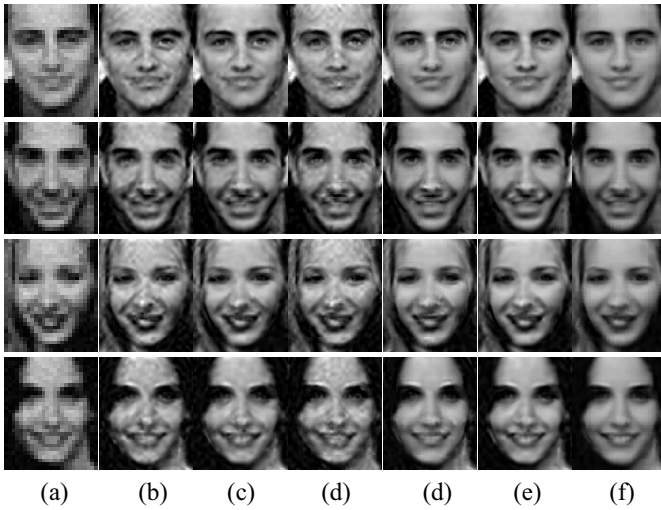


Fig. 11. Comparison of hallucination results of LR face images cropped from the left image in Fig. 10. (a) Cropped LR faces. (b) LSR. (c) LcR. (d) LM-CSS. (e) SSR. (f) RLcBR. (g) LCER.

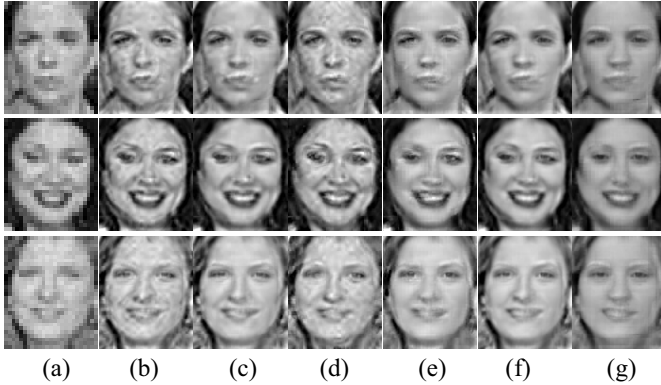


Fig. 12. Comparative hallucination results of LR face images cropped from the right image in Fig. 10. (a) Cropped LR faces. (b) LSR. (c) LcR. (d) LM-CSS. (e) SSR. (f) RLcBR. (g) LCER.

I. Computational Complexity Analysis

The computational cost of our proposed method is mainly come from the calculation of matrix Q , the matrix inverse and product in (24). Basically, it requires $\mathcal{O}(nK^2 + CK^3)$ operators for the calculation of matrix Q . For the matrix inverse and multiplications, $\mathcal{O}(2K^3)$ operators are needed. Thus, the total computational complexity of LCER is $\mathcal{O}(nK^2 + (C+2)K^3)$ for each LR test patch, in which n is the dimension of LR patch, C is the number of classes in training data, and K denotes the total number of training samples.

In addition, we also compared the running time of our proposed method against to that of other methods. One face image in the AR dataset is randomly chosen as the test data. All the experiments are carried out on a PC equipped with 2.60-GHz CPU. The running time of each method is tabulated in Table III. One can see that, the LCER is slightly slower than the single regression methods (LSR, LcR), but faster than these two layer-based methods (LM-CSS, RLcBR) and the smoothness regularized method (SSR).

IV. CONCLUSION

This article presented a novel LCER for discriminative face image super-resolution. In contrast with the previous works that are unsupervised and take into no account the class label information, the proposed LCER super-resolves face images in a supervised manner by learning a category embedding representation. More specifically, a category embedding penalty term is introduced into the objective function, enabling the training samples holding the same label with the query one to offer more contribution in super-resolution reconstruction. This guarantees that the face patterns are mainly transformed from exemplar faces belonging to the same subject. Besides, considering that similar small patches distribute throughout the whole training data space, we employed the locality penalty to force training samples not limited to the same class but close to the test one offer more contribution to the final representation. The union of locality prior and category information admit the hallucination results of LCER to preserve more image details and desirable patterns. The evaluation results demonstrate the superiority of LCER compared with several state-of-the-art schemes in terms of both quantitative measurements and visual impressions.

REFERENCES

- [1] Y. Wang, Y. Y. Tang, L. Li, H. Chen, and J. Pan, "Atomic representation-based classification: Theory, algorithm, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 6–19, Jan. 2019.
- [2] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2000, pp. 83–88.
- [3] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2001, pp. 192–198.
- [4] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 35, no. 3, pp. 425–434, Aug. 2005.
- [5] X. Zhang, S. Peng, and J. Jiang, "An adaptive learning method for face hallucination using locality preserving projections," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–8.
- [6] 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.
- [7] M. Jian, K.-M. Lam, and J. Dong, "A novel face-hallucination scheme based on singular value decomposition," *Pattern Recognit.*, vol. 46, no. 11, pp. 3091–3102, 2013.
- [8] M. Jian and K.-M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1761–1772, Nov. 2015.
- [9] D. Zhu, G. Gao, H. Gao, and H. Lu, "Nuclear norm regularized structural orthogonal procrustes regression for face hallucination with pose," in *Cognitive Internet of Things: Frameworks, Tools and Applications*. Cham, Switzerland: Springer Int., 2020, pp. 159–169.
- [10] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2004, pp. 275–282.
- [11] J. Jiang, R. Hu, Z. Wang, Z. Han, and J. Ma, "Facial image hallucination through coupled-layer neighbor embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1674–1684, Sep. 2016.
- [12] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [13] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch," *Pattern Recognit.*, vol. 43, no. 6, pp. 2224–2236, 2010.
- [14] W. Yang, T. Yuan, W. Wang, F. Zhou, and Q. Liao, "Single-image super-resolution by subdictionary coding and kernel regression," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 9, pp. 2478–2488, Sep. 2017.
- [15] L. Liu, L. Chen, C. L. P. Chen, Y. Y. Tang, and C. M. Pun, "Weighted joint sparse representation for removing mixed noise in image," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 600–611, Mar. 2017.

- [16] V. Venugopal and S. Sundaram, "Modified sparse representation classification framework for online writer identification," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, doi: [10.1109/TSMC.2018.2871267](https://doi.org/10.1109/TSMC.2018.2871267).
- [17] 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.
- [18] Y. Li, C. Cai, G. Qiu, and K.-M. Lam, "Face hallucination based on sparse local-pixel structure," *Pattern Recognit.*, vol. 47, no. 3, pp. 1261–1270, 2014.
- [19] 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.
- [20] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3991–4002, Nov. 2017.
- [21] Y. Zhang *et al.*, "Collaborative representation cascade for single-image super-resolution," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 5, pp. 845–860, May 2019.
- [22] R. A. Farrugia and C. Guillemot, "Face hallucination using linear models of coupled sparse support," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4562–4577, Sep. 2017.
- [23] 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.
- [24] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4220–4231, Oct. 2014.
- [25] J. Jiang, C. Chen, K. Huang, Z. Cai, and R. Hu, "Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation," *Inf. Sci.*, vols. 367–368, pp. 354–372, Nov. 2016.
- [26] J. Shi and C. Qi, "Kernel-based face hallucination via dual regularization priors," *IEEE Signal Process. Lett.*, vol. 22, no. 8, pp. 1189–1193, Aug. 2015.
- [27] J. Shi, X. Liu, Y. Zong, C. Qi, and G. Zhao, "Hallucinating face image by regularization models in high-resolution feature space," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2980–2995, Jun. 2018.
- [28] R. Lan, H. Lu, Y. Zhou, Z. Liu, and X. Luo, "An LBP encoding scheme jointly using quaternionic representation and angular information," *Neural Comput. Appl.*, to be published. [Online]. Available: <https://doi.org/10.1007/s00521-018-03968-y>
- [29] L. Liu, S. Li, and C. L. P. Chen, "Quaternion locality-constrained coding for color face hallucination," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1474–1485, May 2018.
- [30] J. Jiang, Y. Yu, S. Tang, J. Ma, A. Aizawa, and K. Aizawa, "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.
- [31] L. Liu, C. L. 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.
- [32] L. Liu, S. Li, and C. L. P. Chen, "Iterative relaxed collaborative representation with adaptive weights learning for noise robust face hallucination," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 5, pp. 1284–1295, May 2019.
- [33] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Jul. 2016.
- [34] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3194–3207, Jul. 2016.
- [35] T. Lu, L. Pan, J. Jiang, Y. Zhang, and Z. Xiong, "DLML: Deep linear mappings learning for face super-resolution with nonlocal-patch," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2017, pp. 1362–1367.
- [36] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 1646–1654.
- [37] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 105–114.
- [38] 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. (CVPR)*, 2017, pp. 5367–5375.
- [39] J. Jiang, Y. Yu, J. Hu, S. Tang, and J. Ma, "Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination," 2018. [Online]. Available: [arXiv:1806.10726](https://arxiv.org/abs/1806.10726).
- [40] Y. Song *et al.*, "Joint face hallucination and deblurring via structure generation and detail enhancement," *Int. J. Comput. Vis.*, vol. 127, no. 6, pp. 785–800, 2019.
- [41] S. P. Mudunuri and S. Biswas, "Low resolution face recognition across variations in pose and illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 1034–1040, Aug. 2016.
- [42] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2950–2959.
- [43] R. Lan, Y. Zhou, Z. Liu, and X. Luo, "Prior knowledge-based probabilistic collaborative representation for visual recognition," *IEEE Trans. Cybern.*, early access, doi: [10.1109/tcyb.2018.2880290](https://doi.org/10.1109/tcyb.2018.2880290).
- [44] W. Deng, J. Hu, and J. Guo, "Face recognition via collaborative representation: Its discriminant nature and superposed representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2513–2521, Oct. 2018.
- [45] 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.
- [46] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2011, pp. 471–478.
- [47] N. Han *et al.*, "Double relaxed regression for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, early access, doi: [10.1109/tcsvt.2018.2890511](https://doi.org/10.1109/tcsvt.2018.2890511).
- [48] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 3360–3367.
- [49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [50] P. P. Gajjar and M. V. Joshi, "New learning based super-resolution: Use of DWT and IGMRF prior," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1201–1213, May 2010.
- [51] 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.



Licheng Liu (Member, IEEE) received the B.S. degree in information and computing science from the China University of Geosciences, Wuhan, China, in 2010, the M.S. degree in applied mathematics from Hunan University, Changsha, China, in 2012, and the Ph.D. degree in software engineering from the University of Macau, Macau, China, in 2016.

He is currently an Associate Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include image processing, sparse representation, pattern recognition, and machine learning.



Rushi Lan received the B.S. degree in information and computing science and M.S. degree in applied mathematics from the Nanjing University of Information Science and Technology, Nanjing, China, and the Ph.D. degree in software engineering from the University of Macau, Macau, China.

He is currently an Associate Professor with the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China. His research interests include image classification, image denoising, and metric learning.



Yaonan Wang received the Ph.D. degree in electrical engineering from Hunan University, Changsha, China, in 1994.

He was a Postdoctoral Research Fellow with the Normal University of Defence Technology, Changsha, from 1994 to 1995. From 1998 to 2000, he was a Senior Humboldt Fellow in Germany, and from 2001 to 2004, he was a Visiting Professor with the University of Bremen, Bremen, Germany. Since 1995, he has been a Professor with the College of Electrical and Information Engineering, Hunan University. His current research interests include robotics and image processing.