

Robust Facial Image Super-Resolution by Kernel Locality-Constrained Coupled-Layer Regression

GUANGWEI GAO, Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, China and Digital Content and Media Sciences Research Division, National Institute of Informatics, Japan

DONG ZHU, College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, China

HUIMIN LU, Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Japan

YI YU, Digital Content and Media Sciences Research Division, National Institute of Informatics, Japan

HEYOU CHANG, Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing XiaoZhuang University, China

DONG YUE, College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, China

Super-resolution methods for facial image via representation learning scheme have become very effective methods due to their efficiency. The key problem for the super-resolution of facial image is to reveal the latent relationship between the **low-resolution (LR)** and the corresponding **high-resolution (HR)** training patch pairs. To simultaneously utilize the contextual information of the target position and the manifold structure of the primitive HR space, in this work, we design a robust context-patch facial image super-resolution scheme via a **kernel locality-constrained coupled-layer regression (KLC2LR)** scheme to obtain the desired HR version from the acquired LR image. Here, KLC2LR proposes to acquire contextual surrounding patches to represent the target patch and adds an HR layer constraint to compensate the detail information.

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

Authors' addresses: G. Gao, Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, No.9 Wenyuan Road, Nanjing City 210023, China; Digital Content and Media Sciences Research Division, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo City 101-8430, Japan; Provincial Key Laboratory for Computer Information Processing Technology, Soochow University, No.1 Shizi Street, Suzhou City 215006, China; email: csggao@gmail.com; D. Zhu, College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, No.9 Wenyuan Road, Nanjing City 210023, China; email: 757419374@qq.com; H. Lu (corresponding author), Department of Mechanical and Control Engineering, Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu City 804-8550, Japan; email: dr.huimin.lu@ieee.org; Y. Yu, Digital Content and Media Sciences Research Division, National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo City 101-8430, Japan; email: yiyu@nii.ac.jp; H. Chang, Key Laboratory of Trusted Cloud Computing and Big Data Analysis, Nanjing XiaoZhuang University, No.3601, Hongjing Road, Nanjing City 211171, China; email: cv_hychang@126.com; D. Yue, College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, No.9 Wenyuan Road, Nanjing City 210023, China; email: medongy@vip.163.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1533-5399/2021/05-ART67 \$15.00

<https://doi.org/10.1145/3418462>

Additionally, KLC2LR desires to acquire more high-frequency information by searching for nearest neighbors in the HR sample space. We also utilize kernel function to map features in original low-dimensional space into a high-dimensional one to obtain potential nonlinear characteristics. Our compared experiments in the noisy and noiseless cases have verified that our suggested methodology performs better than many existing predominant facial image super-resolution methods.

CCS Concepts: • **Computing methodologies** → **Image representations**; **Vision for robotics**; • **Information systems** → **Multimedia content creation**; • **Security and privacy** → *Social aspects of security and privacy*

Additional Key Words and Phrases: Face super-resolution, contextual patches, locality-constrained representation, coupled-layer representation

ACM Reference format:

Guangwei Gao, Dong Zhu, Huimin Lu, Yi Yu, Heyou Chang, and Dong Yue. 2021. Robust Facial Image Super-Resolution by Kernel Locality-Constrained Coupled-Layer Regression. *ACM Trans. Internet Technol.* 21, 3, Article 67 (May 2021), 15 pages. <https://doi.org/10.1145/3418462>

1 INTRODUCTION

Multimedia information plays a very important role in life sciences [8, 13, 35]. Among them, the images obtained from a surveillance camera are very important for social development [3, 10, 12, 30]. However, the photos obtained by the cameras are usually fuzzy and have low quality due to the influence of some external factors such as noise and blur. Facial image super-resolution is proposed to acquire synthesized high-quality faces from those low-quality ones [5, 7, 9, 14, 18, 31].

According to the research ways, the existing face super-resolution approaches could be categorized into learning-based and multi-frame synthesis-based algorithms. The methods based on multi-frame reconstruction obtain consecutive multi-frame low-quality images with the analogical scene [15], then align the images according to datum points, and map the information to the corresponding **high-resolution (HR)** grid to fulfill the HR image reconstruction. The method is sensitive to the factors that affect the accuracy of registration, and the reconstruction results are susceptible to noise, and so forth [2, 26]. Moreover, some scholars have proved that even under ideal conditions such as accurate registration and no noise in the observed **low-resolution (LR)** images, the algorithm still has insufficient complementary information when the reconstructed image is magnified multiple times, even if the image frame is increased. This defect greatly limits the development of multi-frame-based reconstruction methods. Therefore, the learning-based algorithm has become an important solution.

Learning-based face super-resolution approaches use image prior knowledge to constrain or guide the procedure of the super-resolution reconstruction, so as to gain high-quality facial images from the low-quality ones [25, 27]. The face hallucination concept was first put forward by Baker and Kanade [1]. Previous learning-based super-resolution approaches for facial images could be usually classified into the global-based algorithms [17, 34] and the local-based algorithms [20, 32]. The global-based super-resolution methods can well maintain the global structure of the original HR images [4, 33]. Also, the global-based methods have good robustness to noise, but the computation is large and the recovery of details is poor. In order to obtain more detailed information of face features, the local-based methods divide image into blocks according to positions. Super-resolution reconstruction is performed for each face block, and finally HR image is obtained by integrating position blocks.

For instance, Mohamed et al. [29] presented a reconstruction method of facial images by considering the sparse prior. The approach assumed that the test patch can be linearly represented by an over-complete dictionary. A simple yet powerful position-patch face super-resolution algorithm was presented in [28]. To pursue locality and sparsity simultaneously, the authors of [20] designed an effective **locality-constrained representation (LcR)** algorithm, which achieves good performance on image reconstruction. Jiang et al. [19] proposed an iterative version of LcR, which keeps the original data structure of HR image by searching the neighbors in HR training sample space, and updates the HR block iteratively, so as to enhance the image quality. The authors of [21] introduced the smooth sparse priors into the super-resolution process to ensure that analogous training patches naturally have semblable local structure. The authors of [22] presented a context-patch super-resolution scheme for facial images via a thresholding LcR method. This method incorporated contextual information of the position patch, which can have merit of working with large patches. Rajput et al. [32] iteratively utilized locality with sparsity regularization term for more reliable output. The authors of [6] presented a pioneering deep neural network-based super-resolution method. After that, many deep learning-based models have been presented. Li et al. [24] resorted to coarse-to-fine scheme and then proposed a two-branch face super-resolution framework. A multi-scale residual network was utilized in [23] to take full advantage of the image features in diverse scales.

The above-mentioned local-based methods have obtained promising super-resolution reconstruction performance. However, these algorithms still have some drawbacks. For example, the existing context-patch-based face super-resolution methods are all having the hypothesis that the LR feature space has the coincident local structure with that in the corresponding HR space. This will not be held anymore when the input LR images have noise. Also, they may also ignore the underlying nonlinear similarity among samples, which will create advantages for the computing of the more accurate representations. To tackle these problems, we propose a **kernel locality-constrained coupled-layer regression (KLC2LR)** method for context-patch face super-resolution in this article. Particularly, several contributions are made: (i) KLC2LR proposes to utilize the contextual surrounding patches to represent the target patch, which will be severely degraded in practical application scenes. (ii) KLC2LR utilizes the HR error term as a reconstruction constraint and searches the K nearest neighbors from the HR training samples for compensating the lost detailed high-frequency information. (iii) KLC2LR further maps the high-low resolution features into the high-dimensional space to acquire the nonlinear characters among samples. Experimental tests on several facial image databases have confirmed the effectiveness and efficiency of our presented method when compared with some existing facial image super-resolution algorithms.

2 PROPOSED KLC2LR

2.1 Notations

In this article, the HR facial image training set is denoted as $X_H^m (m = 1, 2, \dots, M)$; M counts the whole training set. The corresponding LR training set is denoted as $X_L^m (m = 1, 2, \dots, M)$. Y_L represents a low-quality testing image. According to position information, we divide the LR testing image, HR training samples, and LR training samples into overlapping patches, denoted as $y_L(p)$, $x_H^m(p)$, and $x_L^m(p)$, respectively, where p denotes the patch position. Here, we introduce a contextual framework to our method to add contextual information. We use surrounding blocks to represent the p position block. $X_{HC} = [x_{HC}^1, x_{HC}^2, \dots, x_{HC}^N]$ and $X_{LC} = [x_{LC}^1, x_{LC}^2, \dots, x_{LC}^N]$ are

frequently used to express HR and LR context-patches corresponding to position p in the training set. The constructions of the context-patches are the same as that in [22].

2.2 Problem Formulation

The previous local-based face super-resolution methods usually focus on the position information and ignore the contextual information which can both have advantages of local patches and large patches. Here, we utilize the contextual patches around position p to replace this single patch. We also add HR image information as a constraint to compensate the lost high-frequency details. For simplicity, the position p in the following formula will be omitted. Our model is formulated as follows:

$$J(w) = \|y_L - X_{LC}w\|_2^2 + \|y_H - X_{HC}w\|_2^2 + \lambda \|Dw\|_2^2, \quad (1)$$

where $w = [w^1, w^2, \dots, w^N]^T$, and λ expresses the locality-constraint parameter. D is a diagonal matrix and $D = \text{diag}(d)$, where vector d represents the similarity measures between the LR input patch and those corresponding training ones, and $\text{diag}(\cdot)$ is a diagonalization operator. Since we utilize context-patches to represent an input patch, the computational complexity increases. To tackle this problem, we introduce the thresholding method into this model, which can reduce the algorithm time and also can further improve the super-resolution accuracy. To capture the nonlinear characteristics, kernel function ϕ is utilized to project the features into high-dimensional feature space:

$$J(w) = \|\phi(y_L) - \phi(X_{LC}^K w)\|_2^2 + \|\phi(y_H) - \phi(X_{HC}^K w)\|_2^2 + \lambda \|D^K w\|_2^2, \quad (2)$$

where X_{LC}^K expresses the K nearest LR contextual patches, X_{HC}^K represents the corresponding HR ones, and D^K are the metrics between the LR test patch and the K nearest contextual ones. y_H represents a supporting HR patch, and its initial value can be estimated as follows:

$$J(w_s) = \|y_L - X_{LC}^K w_s\|_2^2 + \lambda \|D_s^K w_s\|_2^2. \quad (3)$$

We search the nearest neighbor value in the original HR image space to retain the original data structure, which can further deal with the problem of neglecting the high-frequency information:

$$D_s^K = \|y_{ini}^H - X_{HC}^K\|_2. \quad (4)$$

Here, the value of the estimated HR image block y_{ini}^H can be acquired by the model in the formula (3), where the similarity measure in the locality-constraint term denotes the Euclidean metric between the acquired LR and the corresponding referenced ones.

2.3 Optimization

The optimal representation coefficient could be solved as follows:

$$\min_w \|\phi(y_L) - \phi(X_{LC}^K w)\|_2^2 + \|\phi(y_H) - \phi(X_{HC}^K w)\|_2^2 + \lambda \|D^K w\|_2^2. \quad (5)$$

By introducing the following auxiliary variables

$$h = \begin{pmatrix} \phi(y_L) \\ \phi(y_H) \end{pmatrix}, G = \begin{pmatrix} \phi(X_{LC}^K) \\ \phi(X_{HC}^K) \end{pmatrix}, \quad (6)$$

the optimization problem (5) can be simplified as

$$\begin{aligned} \min_w & \|h - Gw\|_2^2 + \lambda \|D^K w\|_2^2 \\ \text{s.t.} & \mathbf{1}^T w = 1. \end{aligned} \quad (7)$$

The above Equation (7) is a regularized least square problem, whose proximate solution is

$$w^* = (G + \lambda D^2) \setminus ones(K, 1), \quad (8)$$

where column vector $ones(K, 1)$ has a size of $K \times 1$ with all ones, the matrix G is denoted as $G = Z^T Z$, and

$$Z = \begin{pmatrix} \phi(Y_L) \\ \phi(Y_H) \end{pmatrix} - \begin{pmatrix} \phi(X_{LC}^K) \\ \phi(X_{HC}^K) \end{pmatrix}, \quad (9)$$

where $\phi(Y_L) = [\phi(y_L), \phi(y_L), \dots, \phi(y_L)]$, $\phi(Y_H) = [\phi(y_H), \phi(y_H), \dots, \phi(y_H)]$, $\phi(X_{LC}^K) = [\phi(x_{LC}^{(1)}), \phi(x_{LC}^{(2)}), \dots, \phi(x_{LC}^{(K)})]$, and $\phi(X_{HC}^K) = [\phi(x_{HC}^{(1)}), \phi(x_{HC}^{(2)}), \dots, \phi(x_{HC}^{(K)})]$. By means of some facile algebraic operations, matrix G could be computed by

$$\begin{aligned} G &= \phi(Y_L)^T \phi(Y_L) + \phi(Y_H)^T \phi(Y_H) - \phi(Y_L)^T \phi(X_{LC}^K) \\ &\quad - \phi(Y_H)^T \phi(X_{HC}^K) - \phi(X_{LC}^K)^T \phi(Y_L) - \phi(X_{HC}^K)^T \phi(Y_H) \\ &\quad + \phi(X_{LC}^K)^T \phi(X_{LC}^K) + \phi(X_{HC}^K)^T \phi(X_{HC}^K). \end{aligned} \quad (10)$$

In the high-dimensional feature space, the inter product of samples can be conducted by the kernel function. Each term in formulation (10) has the similar form. For example, term $\phi(Y_L)^T \phi(X_{LC}^K)$ can be denoted as

$$\begin{bmatrix} k(y_L, x_{LC}^{(1)}) & k(y_L, x_{LC}^{(2)}) & \dots & k(y_L, x_{LC}^{(K)}) \\ k(y_L, x_{LC}^{(1)}) & k(y_L, x_{LC}^{(2)}) & \dots & k(y_L, x_{LC}^{(K)}) \\ \vdots & \vdots & \ddots & \vdots \\ k(y_L, x_{LC}^{(1)}) & k(y_L, x_{LC}^{(2)}) & \dots & k(y_L, x_{LC}^{(K)}) \end{bmatrix}, \quad (11)$$

where symbol k denotes the kernel function.

Upon obtaining the optimal representation coefficients w^* , we can acquire the corresponding HR patch as

$$y_H = X_{HC}^K w^*. \quad (12)$$

2.4 Face Super-Resolution via KLC2LR

Let X_L^m and X_H^m ($m = 1, 2, \dots, M$) express the training LR and corresponding HR face set, respectively. The facial image super-resolution aims to acquire the desired HR face Y_H when observing the corresponding LR test Y_L .

First, divide the observed LR test and each of the training samples into overlapped patches $y_L(p)$, $x_L^m(p)$, and $x_H^m(p)$. Denote the K nearest LR and HR contextual patches corresponding to the LR input as $X_{LC}^K = [x_{LC}^{(1)}, x_{LC}^{(2)}, \dots, x_{LC}^{(K)}]$ and $X_{HC}^K = [x_{HC}^{(1)}, x_{HC}^{(2)}, \dots, x_{HC}^{(K)}]$. With regard to each observed LR input $y_L(p)$, it can be represented as a linear weighted sum over the context-patches that surround the LR training patch $x_L^m(p)$ via KLC2LR. By maintaining the representation coefficients, the LR referenced patches are substituted using the related HR ones to acquire the desired HR estimated image patch. Finally, the desired HR face could be acquired by means of combining all the estimated pieces and then averaging the overlapping areas. The whole super-resolution framework for facial image is depicted in Algorithm 1. The flowchart of our method is drawn in Figure 1.

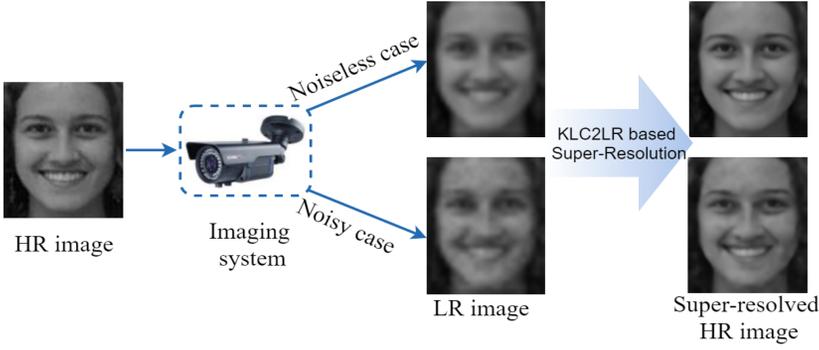


Fig. 1. The flowchart of our method.

ALGORITHM 1: Super-resolve robust facial images via KLC2LR.

- 1: **Input:** Observed LR image Y_L , HR gallery set X_H^m and the corresponding LR set X_L^m ($m = 1, 2, \dots, M$). The parameters: λ , K , and σ .
 - 2: **Output:** The super-resolved HR image Y_H .
 - 3: Divide the HR training samples, LR training samples, and testing image into overlapping patches: $x_H^m(p)$, $x_L^m(p)$, and $y_L(p)$, respectively.
 - 4: **for** each LR input variable y_L **do**
 - 5: (a): Calculate the Euclidean metric between the y_L and all context-patches X_{LC} , and then select the K most similar LR training patches:
 - 6: $D = \|y_L - X_{LC}\|_2^2$.
 - 7: (b): Compute the optimal representation coefficients w^* with regard to the LR input y_L by KLC2LR, and update the representation coefficients according to formula (2).
 - 8: (c): Construct the desired HR patch by
 - 9: $y_H = X_{HC}^K w^*$.
 - 10: **end for**
 - 11: Combine the whole super-resolved HR patches in the light of the corresponding position.
-

3 EXPERIMENTAL EVALUATIONS

3.1 Dataset Description

Experimental results are obtained from the FEI¹ and CAS-PEAL [11] face datasets to show the superiority of our method. The FEI dataset includes 200 objects, and everyone provides two frontal facial images. One is expressionless and the other is smile. 360 face samples are picked as the gallery set and the remaining 40 samples as the query set. The original HR images have a size of 100×120 pixels. By performing smoothing (the filter has a size of 4×4) and then the down-sampling (with a factor of 4) operations on the primordial HR face samples, one can obtain those simulated LR ones. Thus, the LR image has a size of 25×30 . The CAS-PEAL database includes 1,040 objects; we select the frontal images of 40 persons as the testing set and the remaining 1,000 frontal facial images as the training set. The size of the example image in the CAS-PEAL database is 128×112 pixels and the corresponding low-resolution one is 32×28 pixels. Samples from the FEI database and the CAS-PEAL database can be seen in Figure 2.

¹<http://fei.edu.br/~cet/facedatabase.html>.

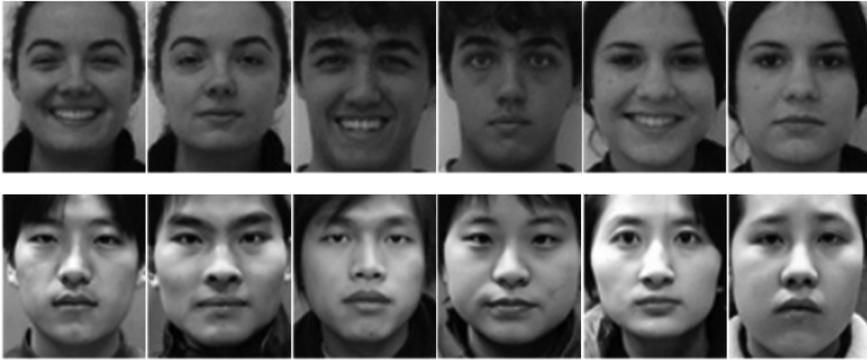


Fig. 2. The first row is examples from the FEI dataset and the second row is examples from the CAS-PEAL dataset.

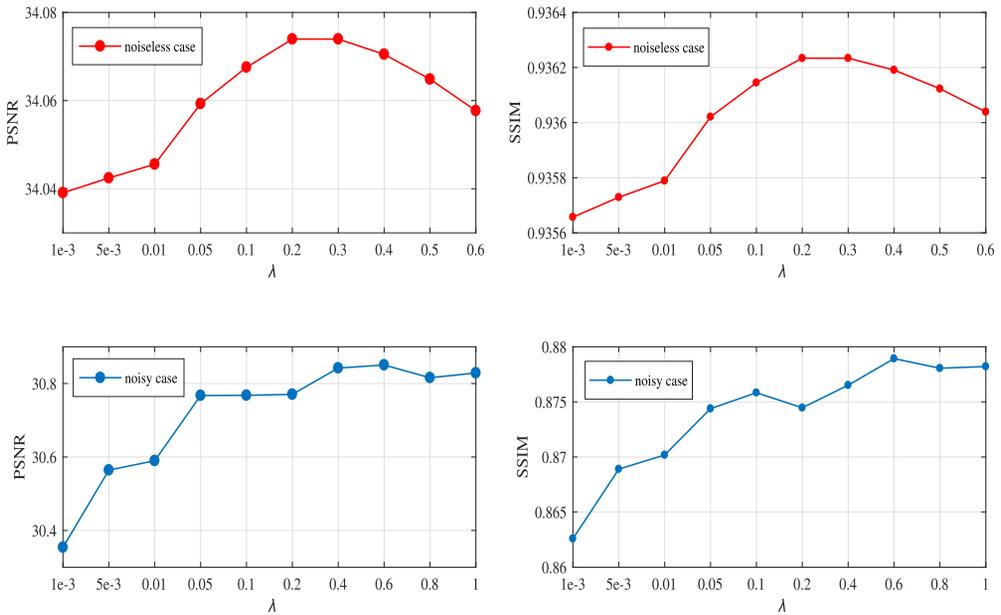


Fig. 3. The effect of the parameter λ on our method.

3.2 Parameter Settings

To obtain the best results of our method, we need to adjust the values of three arguments: the locality-constraint parameter λ , the nearest neighbor size K , and the Gaussian kernel function parameter σ . We choose PSNR and SSIM [16] values as the evaluation standard of our algorithm performance.

3.2.1 The Influence of the Parameter λ . Fix parameters K and σ , adjust the values of λ . Figure 3 depicts the effect (in the form of PSNR and SSIM values) of various values of λ on our method. In Figure 3, we could intuitively notice that the average quantitative indexes can attain the peak when $\lambda = 0.2$ in the noiseless case. When the value of λ exceeds 0.2, the average quantitative results start to decline. Thus, we set the value of λ as 0.2 in the noiseless case. Analogously, we choose the value of λ as 0.6 in the noisy case.

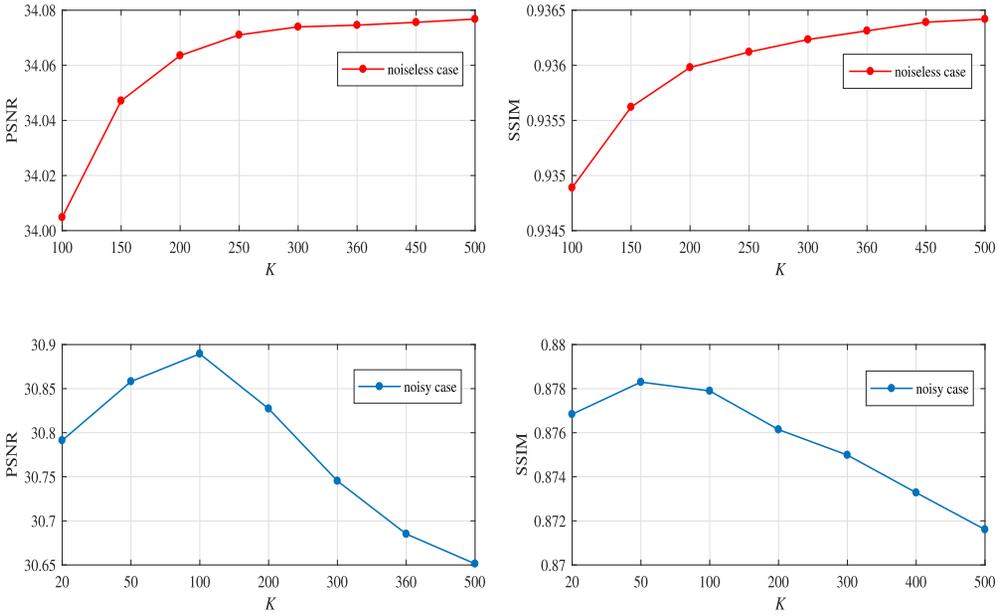


Fig. 4. The effect of the parameter K on our method.

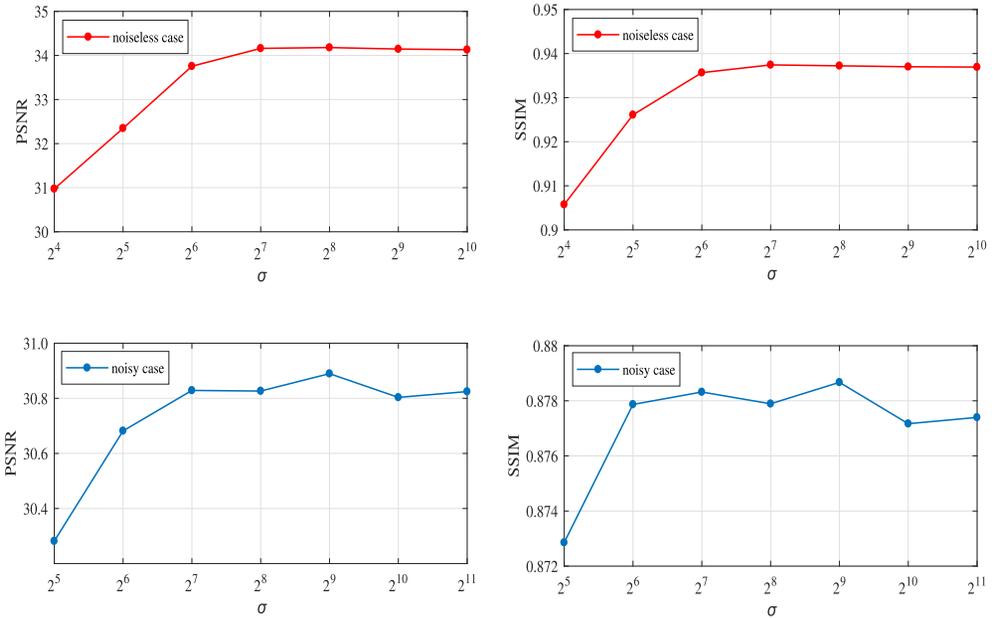


Fig. 5. The effect of the parameter σ on our method.

3.2.2 *The Influence of the Parameter K.* Fix parameters λ and σ , and adjust the value of K . Figure 4 plots the effect of the parameter K on our method. From Figure 4, we can observe that the PSNR value achieves the peak when $K = 300$ in the noiseless case. When the value of K exceeds 300, the quantitative performance improvement tends to be gentle. As more similar patches are selected, the computational complexity increases; we determine the value of K as 300 in the

Table 1. Algorithm Results with Different Kernel Functions

Kernel function	Linear	Polynomial	Gaussian
PSNR (dB)	30.67	30.74	30.88
SSIM	0.8660	0.8754	0.8783

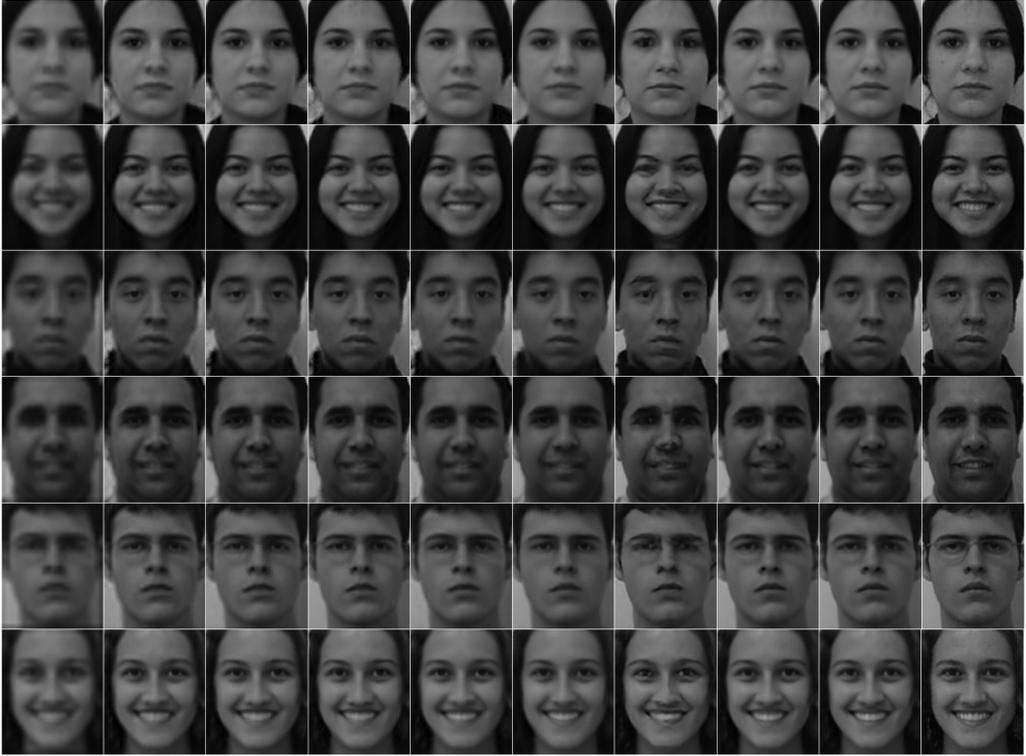


Fig. 6. Some super-resolved examples on the FEI database in the noiseless case. The first to the tenth columns are successively the input LR samples, reconstructed outputs of the LSR [28], LcR [20], LINE [19], SSR [21], ISLcR [32], MSRN [23], TLcR [22], our KLC2LR, and the original HR images.

noiseless case for the consideration of balance between performance and complexity. While in the noisy case, as more similar patches are selected, the PSNR and SSIM values tend to decline, because more irrelevant patches are used. Thus, in the noisy case, we pick the value of K as 100.

3.2.3 The Influence of the Parameter σ . Fix parameters λ and K , and adjust the value of σ . Figure 5 exhibits the effect (in the form of PSNR and SSIM values) of various σ on our approach. From Figure 5, we can see that when the index of σ is greater than 2^8 , the variation of the objective values remains gentle, so we pick the value of σ as 2^8 in the noiseless case. The PSNR and SSIM values fluctuate but decrease as a whole in the noisy case, so we pick the value of σ as 2^9 .

3.2.4 The Influence of the Kernel Function k . Finally, we also show the performance of our method using various kernel functions (Linear, Polynomial, and Gaussian kernel). The experimental settings are the same as the above sections. Table 1 lists the average objective performance of our method with various kernel functions. We can apparently find that our designed regression

Table 2. The Objective Compared Indexes on the FEI Dataset in the Noiseless Case

Methods	PSNR (dB)	SSIM	Year
LSR [28]	31.90	0.9032	2010
LcR [20]	32.75	0.9145	2014
LINE [19]	32.98	0.9176	2014
SSR [21]	32.64	0.9139	2017
ISLcR [32]	32.12	0.9072	2018
MSRN [23]	32.93	0.9174	2018
TLcR [22]	33.86	0.9336	2019
KLC2LR	34.18	0.9372	-



Fig. 7. Some Super-resolved examples on the CAS-PEAL dataset in the noiseless case. The first to the tenth columns are successively the input LR samples, reconstructed outputs of the LSR [28], LcR [20], LINE [19], SSR [21], ISLcR [32], MSRN [23], TLcR [22], our KLC2LR, and the original HR images.

algorithm together with Gaussian kernel function gains better performance than that using Linear and polynomial kernel functions. Thus, we choose Gaussian kernel function in all our experiments.

3.3 Experimental Evaluations

We further perform extensive experiments to assess the superiority of our proposed KLC2LR by comparing it with several existing face super-resolution methods: LSR [28], LcR [20], LINE [19], SSR [21], MSRN [23], TLcR [22], and ISLcR [32].



Fig. 8. Some super-resolved examples on the FEI dataset in the noisy case. The first to the tenth columns are successively the input LR samples, reconstructed outputs of the LSR [28], LcR [20], LINE [19], SSR [21], MSRN [23], TLcR [22], ISLcR [32], our KLC2LR, and the original HR images.

Table 3. The Objective Compared Indexes on the CAS-PEAL Dataset in the Noiseless Case

Methods	PSNR (dB)	SSIM	Year
LSR [28]	28.35	0.8960	2010
LcR [20]	28.96	0.9070	2014
LINE [19]	29.35	0.9115	2014
SSR [21]	29.04	0.9077	2017
ISLcR [32]	28.74	0.9015	2018
MSRN [23]	28.61	0.9017	2018
TLcR [22]	30.20	0.9288	2019
KLC2LR	30.54	0.9329	-

3.3.1 Super-Resolved Results in Noiseless Case. Some super-resolved results in the noiseless case are shown in Figures 6 and 7. The reconstructed details of the LSR, LcR, LINE, SSR, and ISLcR algorithms are satisfying, but the super-resolved images are smoother. In this case, for the MSRN algorithm, the reconstruction of the eye contour and other details are not ideal. The reconstruction details of our algorithm compared with the TLcR and ISLcR algorithms are not very obvious, but the restoration of the local details has been improved. As revealed in Tables 2 and 3, in the FEI face dataset, the superior improvements of our KLC2LR over the TLcR algorithm reach 0.32 dB



Fig. 9. Some super-resolved examples on the CAS-PEAL dataset in the noisy case. The first to the tenth columns are successively the input LR samples, reconstructed outputs of LSR [28], LcR [20], LINE [19], SSR [21], MSRN [23], TLcR [22], ISLcR [32], our KLC2LR, and the original HR images.

(according to the PSNR values) and 0.0036 (according to the SSIM values). In the CAS-PEAL face dataset, the performance improvements reach 0.34 dB (according to the values) and 0.0041 dB (according to the SSIM values). In all, the performance improvement of our KLC2LR scheme over other compared methods is significant.

3.3.2 Super-Resolved Results in Noisy Case. It can be seen from the reconstructed samples in Figures 8 and 9 that our algorithm has better noise immunity than the LSR, LcR, LINE, SSR, MSRN, and TLcR algorithms. Because the MSRN algorithm ignores the prior knowledge of the facial image structure, it is less robust to noise, and the recovered facial image results are poor. From Figures 8 and 9, we can also find that although the synthetic facial images of the ISLcR algorithm have almost smoother face contour, some fine details in the eyes, mouths, and noses around are lost. In contrast, the local details of the edge part of the facial images are better restored by our algorithm. The objective comparison results in Tables 4 and 5 also evaluate the efficiency of our method in the noisy case.

4 CONCLUSIONS

In this work, a KLC2LR method is presented for robust context-patch face super-resolution task. The proposed KLC2LR method not only efficiently utilizes the contextual information, but also uses a high-resolution layer as a constraint to compensate the lost high-frequency detail information. Furthermore, we utilize the kernel trick to project the feature into the high-dimensional variable space to acquire the underlying nonlinear characteristics. Experiments on the two representative

Table 4. The Objective Compared Indexes on the FEI Dataset in the Noisy Case

Methods	PSNR (dB)	SSIM	Year
LSR [28]	29.40	0.8286	2010
LcR [20]	29.98	0.8455	2014
LINE [19]	29.92	0.8428	2014
SSR [21]	30.26	0.8633	2017
MSRN [23]	28.72	0.7778	2018
TLcR [22]	30.12	0.8493	2019
ISLcR [32]	30.45	0.8783	2018
KLC2LR	30.88	0.8799	-

Table 5. The Objective Compared Indexes on the CAS-PEAL Dataset in the Noisy Case

Methods	PSNR (dB)	SSIM	Year
LSR [28]	27.14	0.8465	2010
LcR [20]	27.60	0.8597	2014
LINE [19]	27.80	0.8620	2014
SSR [21]	27.85	0.8721	2017
MSRN [23]	26.42	0.7981	2018
TLcR [22]	28.04	0.8669	2019
ISLcR [32]	27.89	0.8814	2018
KLC2LR	28.44	0.8854	-

face databases under noisy and noiseless conditions have verified that our proposed KLC2LR is robust to noise and also yields obvious superiority over existing face super-resolution approaches.

In practical industrial scenes, the pose and misalignment variations also cannot be overlooked. In our future work, we will introduce some face structure priors into the deep network together with the denoising methods to robustly reconstruct the real-world facial images.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the anonymous reviewers for their critical and constructive comments and suggestions.

REFERENCES

- [1] Simon Baker and Takeo Kanade. 2000. Hallucinating faces. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 83–88.
- [2] Simon Baker and Takeo Kanade. 2002. Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 9 (September 2002), 1167–1183.
- [3] Debarpan Bhattacharya, Sudip Misra, Nidhi Pathak, and Anandarup Mukherjee. 2020. IDeA: IoT-Based autonomous aerial demarcation and path planning for precision agriculture with UAVs. *ACM Trans. Internet Things* 1, 3 (June 2020), 1–21.
- [4] Ayan Chakrabarti, AN Rajagopalan, and Rama Chellappa. 2007. Super-resolution of face images using kernel PCA-based prior. *IEEE Trans. Multimedia* 9, 4 (June 2007), 888–892.
- [5] Berk Dogan, Shuhang Gu, and Radu Timofte. 2019. Exemplar guided face image super-resolution without facial landmarks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1–10.
- [6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 2 (February 2016), 295–307.

- [7] Guangwei Gao and Jian Yang. 2014. A novel sparse representation based framework for face image super-resolution. *Neurocomputing* 134 (June 2014), 92–99.
- [8] Guangwei Gao, Jian Yang, Xiao-Yuan Jing, Fumin Shen, Wankou Yang, and Dong Yue. 2017. Learning robust and discriminative low-rank representations for face recognition with occlusion. *Pattern Recogn.* 66 (June 2017), 129–143.
- [9] Guangwei Gao, Yi Yu, Jin Xie, Jian Yang, Meng Yang, and Jian Zhang. 2020. Constructing multilayer locality-constrained matrix regression framework for noise robust face super-resolution. *Pattern Recogn.* (July 2020), 107539.
- [10] Guangwei Gao, Yi Yu, Meng Yang, Heyou Chang, Pu Huang, and Dong Yue. 2020. Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression. *Inf. Sci.* 506 (January 2020), 19–36.
- [11] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao. 2008. The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans* 38, 1 (January 2008), 149–161.
- [12] Shiming Ge, Shengwei Zhao, Xindi Gao, and Jia Li. 2019. Fewer-shots and lower-resolutions: Towards ultrafast face recognition in the wild. In *Proceedings of the 27th ACM International Conference on Multimedia*. ACM, 229–237.
- [13] Bilal Ghanem, Paolo Rosso, and Francisco Rangel. 2020. An emotional analysis of false information in social media and news articles. *ACM Trans. Internet Technol.* 20, 2 (April 2020), 1–18.
- [14] Klemen Grm, Walter J. Scheirer, and Vitomir Štruc. 2019. Face hallucination using cascaded super-resolution and identity priors. *IEEE Trans. Image Proc.* 29, 1 (October 2019), 2150–2165.
- [15] Hu He and Lisimachos P. Kondi. 2006. An image super-resolution algorithm for different error levels per frame. *IEEE Trans. Image Proc.* 15, 3 (March 2006), 592–603.
- [16] Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *Proceedings of the IEEE International Conference on Pattern Recognition*. IEEE, 2366–2369.
- [17] Xiao Hu, Peirong Ma, Zhuohao Mai, Shaohu Peng, Zhao Yang, and Li Wang. 2019. Face hallucination from low quality images using definition-scalable inference. *Pattern Recogn.* 94 (October 2019), 110–121.
- [18] Kui Jia and Shaogang Gong. 2008. Generalized face super-resolution. *IEEE Trans. Image Proc.* 17, 6 (June 2008), 873–886.
- [19] Junjun Jiang, Ruimin Hu, Zhongyuan Wang, and Zhen Han. 2014. Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning. *IEEE Trans. Image Process.* 23, 10 (August 2014), 4220–4231.
- [20] Junjun Jiang, Ruimin Hu, Zhongyuan Wang, and Zhen Han. 2014. Noise robust face hallucination via locality-constrained representation. *IEEE Trans. Multimedia* 16, 5 (August 2014), 1268–1281.
- [21] Junjun Jiang, Jiayi Ma, Chen Chen, Xinwei Jiang, and Zheng Wang. 2017. Noise robust face image super-resolution through smooth sparse representation. *IEEE Trans. Cybern.* 47, 11 (November 2017), 3991–4002.
- [22] Junjun Jiang, Yi Yu, Suhua Tang, Jiayi Ma, Akiko Aizawa, and Kiyoharu Aizawa. 2019. Context-patch based face hallucination via thresholding locality-constrained representation and reproducing learning. *IEEE Trans. Cybern.* 50, 1 (January 2019), 324–337.
- [23] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. 2018. Multi-scale residual network for image super-resolution. In *Proceedings of the European Conference on Computer Vision*. 517–532.
- [24] Mengyan Li, Yuechuan Sun, Zhaoyu Zhang, and Jun Yu. 2018. A coarse-to-fine face hallucination method by exploiting facial prior knowledge. In *Proceedings of the 25th IEEE International Conference on Image Processing (ICIP '18)*. IEEE, 61–65.
- [25] Xiaoming Li, Ming Liu, Yuting Ye, Wangmeng Zuo, Liang Lin, and Ruigang Yang. 2018. Learning warped guidance for blind face restoration. In *Proceedings of the European Conference on Computer Vision (ECCV '18)*. 272–289.
- [26] Zhouchen Lin and Heung-Young Shum. 2004. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 1 (January 2004), 83–97.
- [27] Tao Lu, Jiaming Wang, Junjun Jiang, and Yanduo Zhang. 2020. Global-local fusion network for face super-resolution. *Neurocomputing* 387 (April 2020), 309–320.
- [28] Xiang Ma, Junping Zhang, and Chun Qi. 2010. Hallucinating face by position-patch. *Pattern Recogn.* 43, 6 (June 2010), 2224–2236.
- [29] Mohamed Naleer Haju Mohamed, Yao Lu, and Feng Lv. 2013. Effective two-step method for face hallucination based on sparse compensation on over-complete patches. *IET Image Process.* 7, 6 (August 2013), 624–632.
- [30] Sivaram Prasad Mudunuri, Shashanka Venkataramanan, and Soma Biswas. 2019. Dictionary alignment with re-ranking for low-resolution NIR-VIS face recognition. *IEEE Trans. Inf. Forensics Secur.* 14, 4 (April 2019), 886–896.
- [31] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. 2003. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.* 20, 3 (May 2003), 21–36.
- [32] Shyam Singh Rajput, K. V. Arya, and Vinay Singh. 2018. Robust face super-resolution via iterative sparsity and locality-constrained representation. *Inf. Sci.* 463 (October 2018), 227–244.

- [33] Ching-Ting Tu and Jang-Ren Luo. 2015. Robust face hallucination using ensemble of feature-based regression functions and classifiers. *Image Vis. Comput.* 44 (December 2015), 59–72.
- [34] Chih-Yuan Yang, Sifei Liu, and Ming-Hsuan Yang. 2013. Structured face hallucination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1099–1106.
- [35] Yi Yu, Suhua Tang, Francisco Raposo, and Lei Chen. 2019. Deep cross-modal correlation learning for audio and lyrics in music retrieval. *ACM Trans. Multim. Comput.* 15, 1 (February 2019), 1–16.

Received December 2019; revised November 2020; accepted March 2021