Low Resolution Face Recognition and Reconstruction via Deep Canonical Correlation Analysis

Zhao Zhang[1], Yun-Hao Yuan[1,*], Xiao-Bo Shen[2], Yun Li[1]
[1]Yangzhou University and [2]Nanyang Technological University
yhyuan@yzu.edu.cn

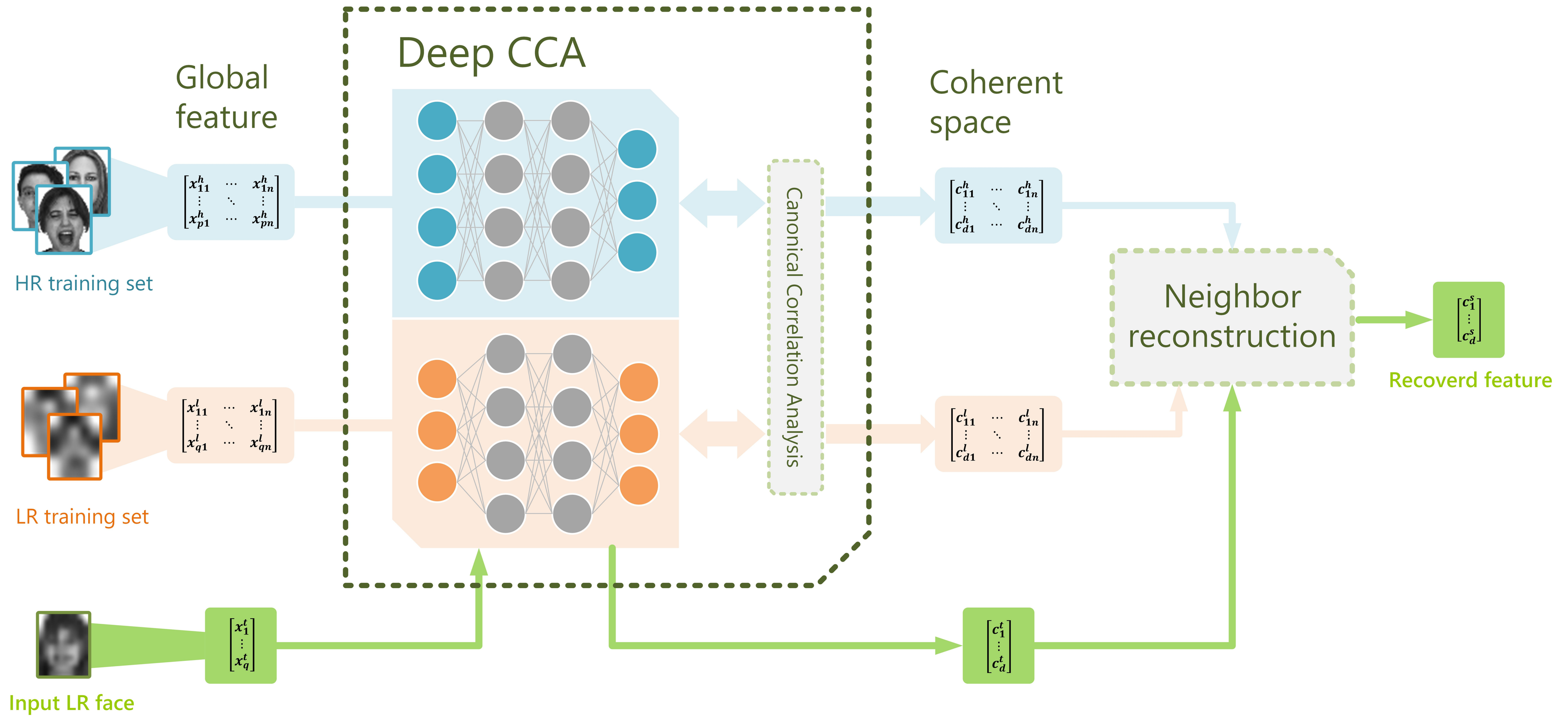

Figure 1: Basic flowchart of our proposed method.

Overview

Background

- Low-resolution (LR) face identification is always a challenge in computer vision.
- Many face super-resolution (FSR) methods, which aim at inferring high-resolution (HR) facial images or recognition features from LR ones, have been proposed to overcome the LR problem.

Motivations

- Canonical correlation analysis (CCA) is a classical but powerful learning method, which seeks basis vectors for two sets of variables, such that their projections onto the basis vectors have a maximized correlation.
- Huang *et al.* [1] first employed CCA in FSR problem, and achieved a promising result.
- However, CCA is a linear learning approach in essence, thus difficult to measure the nonlinear relationships between HR and LR facial images.
- In our previous work [2], we proposed a kernel CCA based LR face recognition method, where the nonlinear correlation between HR and LR face features can be well depicted by two kernel mappings.
- However, the learned nonlinear representation is limited by the fixed kernel. More importantly, projecting original features into kernel spaces is opaque, which makes the mapping process irreversible. Therefore it is difficult to reconstruct the HR face images from the LR one.

Contributions

- We propose a new LR FSR approach based on deep CCA (DCCA). The method can simultaneously recognize and reconstruct LR faces.
- The proposed method can learn flexible nonlinear representations by passing HR and LR facial features via multiple stacked layers of nonlinear transformation.
- We apply a radial basis function (RBF) based neural network to build a regression model to overcome the irreversibility in [2].
- In addition, we also design two residual compensation methods for identification and vision enhancement, respectively.
- A number of experimental results on benchmark datasets have demonstrated the effectiveness and robustness of our method.

Details of Our Method

Our approach employs a two-step framework: in the first step, we carry out facial features/images reconstruction. In the second step, we adopt different residual compensation methods according to the purposes of identification and vision enhancement.

Training

1. HR and LR face set are $I^h = [i_1^h, i_2^h, \dots, i_m^h] \in \mathbb{R}^{p\times m}$ and $I^l = [i_1^l, i_2^l, \dots, i_m^l] \in \mathbb{R}^{q\times m}$.
2. Center the LR and HR training images by $\hat{I}^l = \{I_j^l - \mu^l\}_{j=1}^m$ and $\hat{I}^h = \{I_j^h - \mu^h\}_{j=1}^m$.
3. Employ PCA to extract the global facial features by $X^l = P_l^\top \hat{I}^l$ and $X^h = P_h^\top \hat{I}^h$.
4. Use DCCA to learn flexible nonlinear representations to enhance consistency, as follows:
$$\begin{aligned} \left(\theta_f^*, \theta_g^*, W_f^*, W_g^*\right) &= \underset{\theta_f,\theta_g,W_f,W_g}{\arg\max}\ \mathrm{tr}\left(W_f^\top \Sigma_{fg} W_g\right) \\ s.t.\ W_f^\top \Sigma_{ff} W_f &= W_g^\top \Sigma_{gg} W_g = I, \end{aligned} \tag{1}$$
where $\Sigma_{ff} = F(X^l;\theta_f)F(X^l;\theta_f)^\top + r_f I$, $\Sigma_{gg} = G(X^h;\theta_g)G(X^h;\theta_g)^\top + r_g I$, $\Sigma_{fg} = F(X^l;\theta_f)G(X^h;\theta_g)^\top$, $F(X^l;\theta_f)$ and $G(X^h;\theta_g)$ are the centered outputs of two DNNs, θ_f and θ_g are the vectors containing all parameters of two DNNs, r_f and r_g are two small positive numbers.
5. We get the correlational features $C^l = W_f^{*\top} F(X^l;\theta_f^*)$ and $C^h = W_g^{*\top} G(X^h;\theta_g^*)$.
6. We reestablish the relationship between C^h and X^h by $X^h = W_{RBF}\Phi$. The (i,j)th element in the matrix $\Phi \in \mathbb{R}^{m\times m}$ is calculated by $(\Phi)_{ij} = \exp(-||c_i^h - c_j^h||/2\sigma^2\ )$ with c_i^h as the ith column in C^h and σ as the parameter of RBF. Accordingly, the weighting coefficient matrix can be obtained by $W_{RBF} = X^h\Phi^{-1}$.

Reconstruction

1. Input a new LR face i_t.
2. Compute its principal component feature by $x_t^l = P_l^\top (i_t - \mu_l)$.
3. Transform it to the coherent subspace by $c_t^l = W_f^{*\top} F(x_t^l;\theta_f^*)$.
4. For c_t^l, we find its nearest k neighbors $\{C_{t_j}^l\}_{j=1}^k$ in C^l measured by Euclidean distance. The coefficients $A = \{\alpha_{t_j}\}_{j=1}^k$ are obtained via minimizing $\varepsilon = \left\| c_t^l - \sum_{j=1}^k \alpha_{t_j} C_{t_j}^l \right\|$ $s.t.\ \sum_{j=1}^k \alpha_{t_j} = 1$.
5. The corresponding HR feature $\tilde{c}_t^h$ can be reconstructed by applying A to $\{C_{t_j}^h\}_{j=1}^k$ in C^h. It can be feed into a classifier for recognition.
6. Calculate x_t^h with W_{RBF} according to $\tilde{c}_t^h$.
7. The preliminary reconstruction image can be expressed as $\tilde{i}_t^h = P_h x_t^h + \mu^h$.

Residual Compensation

I. For vision enhancement

1. We use the method in the last section to generate HR face set $\tilde{I}^h$ according to I^l. Then we get HR and LR residual face sets $R^h = I^h - \tilde{I}^h$ and $R^l = I^l - \mathrm{downsample}(\tilde{I}^h)$.
2. Same way as *Reconstruction 4-5*, the HR residual r_t^h can be calculated by keeping the neighborhood relationship from R^l to R^h.
3. The high-quality image we eventually produce is $i_t^h = \tilde{i}_t^h + r_t^h$.

II. For face recognition

1. First we reconstruct feature set $\tilde{C}^h$ according to the LR feature set C^l with method in *Test*, and the corresponding HR set $\tilde{I}^h$ can be generated.
2. Following the down sampling, we get $\tilde{I}^l$. Then we calculate correlational feature $\tilde{C}^l$ of $\tilde{I}^l$.
3. The feature residual sets are expressed as $E^h = C^h - \tilde{C}^h$ and $E^l = C^l - \tilde{C}^l$.
4. We down-sample the reconstructed test image and extract the correlational feature c_t^l.
5. We find feature residual $r_c^l = c_t^l - \tilde{c}_t^l$. Similar to the *Residual Compensation I.2*, we get the corresponding HR feature residual r_c^h.
6. Finally, we obtain the recognition feature $c^h = \tilde{c}_t^h + r_c^h$.

Experimental Results

- We denote the proposed method containing residual compensation as SRDCCA-RC, and the method without residual compensation as SRDCCA.
- We carry out our SRDCCA and SRDCCA-RC methods on the CMU PIE and Yale-B databases to test recognition performance. The nearest neighbor (NN) classifier is used in all the experiments. The results are recorded in Fig.2, Fig.3, and Fig.4.
- We evaluate the effect of our method by reconstructing the image on the CAS-PEAL database. The peak signal to noise ratio (PSNR) result is illustrated in Fig.5.

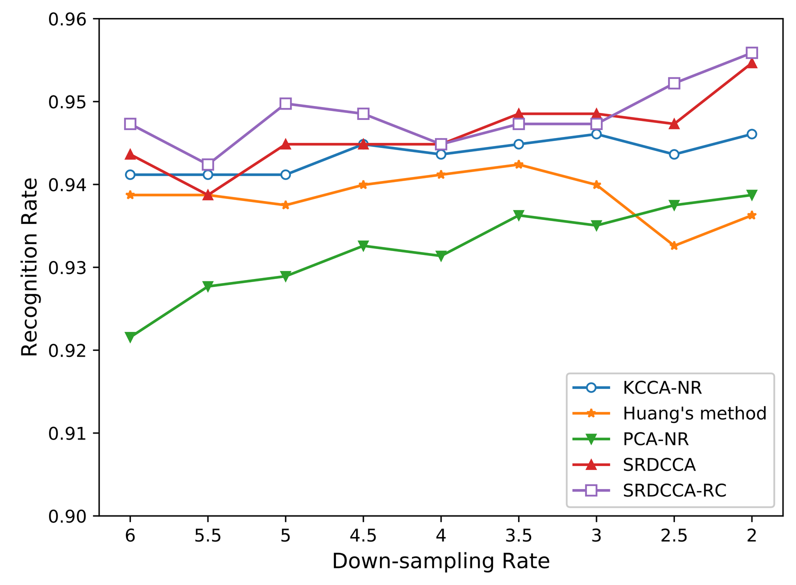

Figure 2: Recognition rate vs down-sampling rate on CMU PIE database. The size of HR images is 64×64.

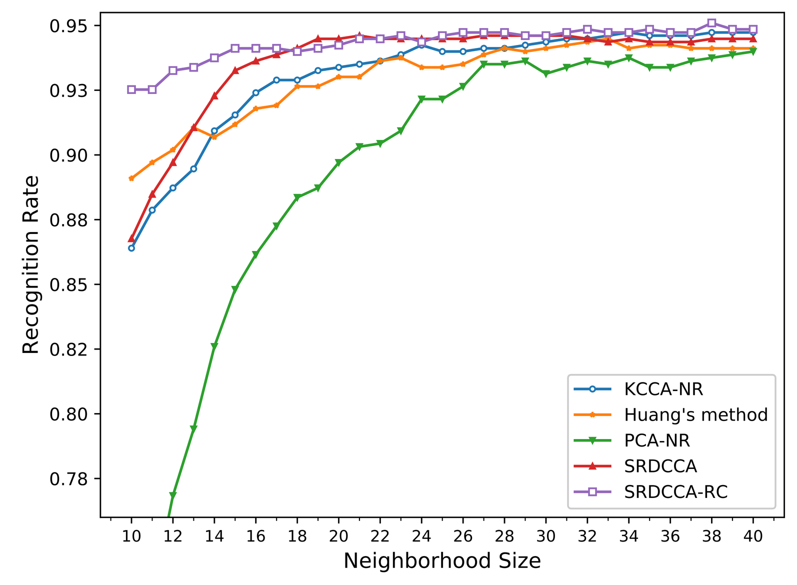

Figure 3: Recognition rate vs neighborhood size on CMU PIE database. The sizes of HR and LR face images are set to 64×64 and 16×16.

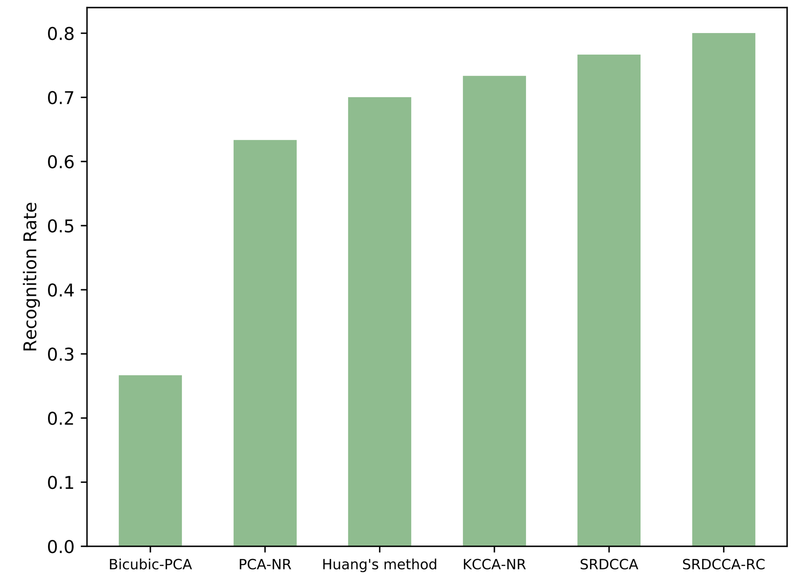

Figure 4: Test on Yale-B database. The HR face image size is still set to 64×64. Differently, LR images present very low resolution: 4×4.

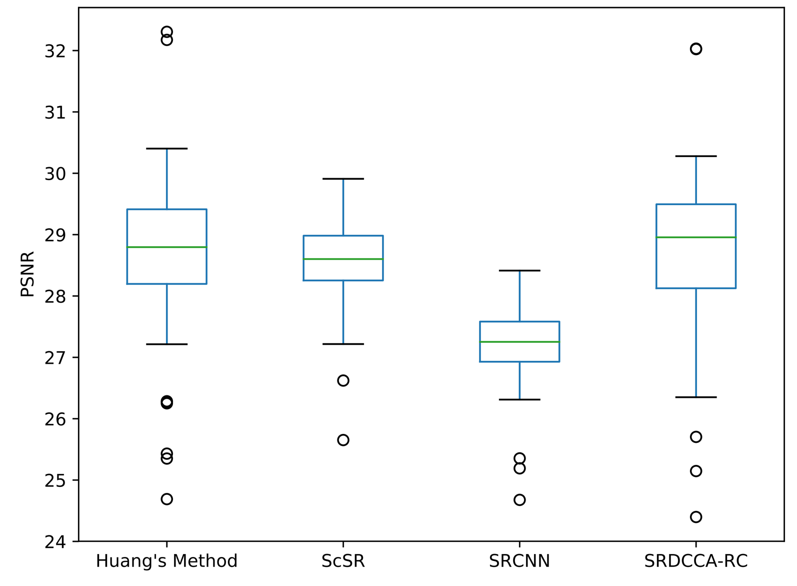

Figure 5: PSNR on CAS-PEAL database.

References

[1] Hua Huang, Huiting He, Xin Fan, and Junping Zhang. Super-resolution of human face image using canonical correlation analysis. *Pattern Recognition*, 43(7):2532–2543, 2010.

[2] Zhao Zhang, Yun-Hao Yuan, Yun Li, Bin Li, and Ji-Peng Qiang. Face hallucination and recognition using kernel canonical correlation analysis. In *ICONIP*, pages 633–641, 2017.