ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 133

2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE2016)

Cascaded Hallucination-Classification Deep Network for Low-Resolution Face Recognition in the Wild

Zheyu Zhang[1,*] and Peter Cheung[2]

[1]Department of Computer Science and Technology, Tsinghua University, China

[2]Department of Electrical & Electronic Engineering, Imperial College London, UK

*Corresponding author

***Abstract*—Low-resolution face recognition (LR FR) has become an active research subarea due to its significances for real applications. Conventional low-resolution face recognition approaches meet challenges like noise affection and lack of effective features with LR faces. In this paper, we propose a deep learning method for LR FR. Our convolutional neural network (CNN) model directly learns an end-to-end classification on LR faces. Different from normal CNN for high-resolution (HR) face recognition, ours integrates a lightweight hallucination network mapping LR images into HR ones. Furthermore, we concatenate the hallucination and classification networks so that the training propagation is operated in one model, which largely boosts the performance over basic CNN and separate two-step models. Besides, our model is robust to varying poses and illuminations in the wild, and also portable to embedded system for its memory- and energy-saving features.**

Keywords-deep convolutional neural network; face-hallucination; low-resolution face recognition

I. INTRODUCTION

Face recognition has received much attention due to its wide applications. Although the performance of face recognition in controlled environment is satisfactory, the same thing does not apply to real applications with faces from images with small size, poor quality, and wide variations. Therefore, low-resolution face recognition in the wild has become one of the most challenging and crucial topics in computer vision.

Even though many methods have been proposed, LR FR in the wild still remains unsolved due to following challenges:

- Loss of information. Most traditional methods based on HR face images cannot perform well with low resolution, mainly because of the loss of local information and lack of effective representation under LR circumstances.
- Wide Variations. Non-canonical view faces vary largely on poses, angles, expressions, and illuminations. The algorithm has to be robust enough to handle these variations.
- Background Noises. Faces in the wild tend to have complicated background that can affect the recognition performance.
- Random Blur. Face images captured in the wild are often blurred with unknown kernels.

In this paper, we propose a deep convolutional neural network that outperforms normal CNN on LR FR problem. Convolutional neural network is proved to be successful in vision tasks including face recognition. The multilayer model can provide representations from varying faces that are much more robust than hand-crafted features. We apply the well-known VGG-Net [1] to face recognition and it performs robustly with HR face images. We add a three-layer hallucination network cascaded with VGG-Net hallucinating LR faces. Our hallucination model, inspired by a related work of image super-resolution [7], is extended with larger filters and different dataset that restrain the model for face hallucination. Finally, we integrate the hallucination network and classification network to form our model for optimization as a whole. The comparison between our cascaded network and other methods shows that our design outperforms others on the result (see Experiment Section).

The proposed cascaded network has several appealing properties. First, the hallucination structure is lightweight and simple, yet has good performance on LR face images. Second, the classification structure is deep enough to guarantee the robustness for faces in the wild, which contain variations and blur. Third, it is an end-to-end model that takes LR face images as input and directly output the classification result. Fourth, despite the deep structure, the deployed model for recognition is memory-saving and fast in speed, which is ideal for real embedded applications.

The main contribution of this paper is that we introduce a novel model for LR FR by combining face hallucination and classification approaches into deep neural network architecture. It excels most conventional recognition methods in aspects of robustness and performance, while still enjoying simplicity and light weight.

II. RELATED WORKS

A. Low-Resolution Face Recognition

This subarea has been researched for decades and methods can be classified into two categories: indirect methods and direct methods. Indirect methods mainly refer to those who first generate HR face images by super-resolution or face

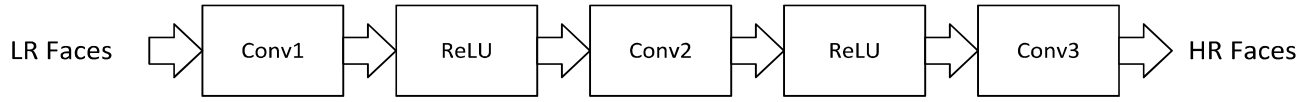


FIGURE I. FACE HALLUCINATION CNN STRUCTURE

hallucination, and then apply traditional HR FR approach for recognition. Significant works include hallucination [2] and S^2R^2 [3]. Direct methods refer to those who extract resolution-robust representations directly from LR face images and do the recognition. Landmarks include color feature [4] and CLPMs [5]. The model proposed in this paper is a mixed method of above two.

B. Convolutional Neural Network

CNN has received explosive popularity recently due to its success in computer vision tasks including image classification. The network can be trained to handle very complex vision problems, and our model benefits from the convolutional architecture to extract robust face representations. VGG-Net [1] is a powerful CNN that can be deployed for multiple applications including face recognition. And there are also eminent works that use CNN for face hallucination [6,7,8], which inspire our work to a large extent.

III. APPROACH

Our cascaded deep network is comprised of hallucination step and classification step. However, unlike normal indirect LR FR methods, our model integrates two parts into one neural network, which means that propagations can be operated in one network instead of two separate ones. More specifically, initially there are two CNNs separately trained for hallucination and classification. After separate models converge to stability, they will be cascaded to establish an integrated model and carry out some other training cycles to fine-tune the parameters until final convergence.

A. Face Hallucination

Our hallucination model is inspired by a related work of image super-resolution [7]. Based on their proposed CNN model, we extend the model into three-channel color-image with a larger hidden layer filter and train the network with face database to restrain the generalized super-resolution into face hallucination.

The network structure is demonstrated in Figure I. The low-resolution images should be first resized by bicubic interpolation, which is the only preprocess required. We denote the interpolated faces as LR faces I_L, and I_L propagate through the network with three convolution layers. Let $F_i(I_L)$ denote the feature maps of convolution layer i, and $F_0(I_L) = I_L$. Let (W_i, B_i) denote the weights and biases of the convolution layer. Then the operation of convolution layers is

$$F_i'(I_L) = W_i * F_{i-1}(I_L) + B_i, \mathrm{i} = 1,2,3 \quad (1)$$

The operation of ReLU (Rectified Linear Unit) layer is

$$F_i(I_L) = \max(0, F_i'(I_L)), \mathrm{i} = 1,2,3 \quad (2)$$

Learning the mapping function F requires the optimization of parameters (W_i, B_i), which is achieved through minimizing the loss function defined below

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\|F(Y_i; \Theta) - X_i\| \quad (3)$$

where Θ is the set of all parameters, n is the number of training samples, and $\{X_i\}$ denotes ground-truth high-resolution training images. This loss function benefits the hallucinated faces in PSNR level, which is a widely-used metric for image restoration, but is not necessarily optimal for vision or classification. That is why we propose the cascaded model.

layer	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
type	input	conv	relu	conv	relu	mpool	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv
name	–	conv1_1	relu1_1	conv1_2	relu1_2	pool1	conv2_1	relu2_1	conv2_2	relu2_2	pool2	conv3_1	relu3_1	conv3_2	relu3_2	conv3_3	relu3_3	pool3	conv4_1
support	–	3	1	3	1	2	3	1	3	1	2	3	1	3	1	3	1	2	3
filt dim	–	3	–	64	–	–	64	–	128	–	–	128	–	256	–	256	–	–	256
num filts	–	64	–	64	–	–	128	–	128	–	–	256	–	256	–	256	–	–	512
stride	–	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	1
pad	–	1	0	1	0	0	1	0	1	0	0	1	0	1	0	1	0	0	1

layer	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
type	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	relu	mpool	conv	relu	conv	relu	conv	softmx
name	relu4_1	conv4_2	relu4_2	conv4_3	relu4_3	pool4	conv5_1	relu5_1	conv5_2	relu5_2	conv5_3	relu5_3	pool5	fc6	relu6	fc7	relu7	fc8	prob
support	1	3	1	3	1	2	3	1	3	1	3	1	2	7	1	1	1	1	1
filt dim	–	512	–	512	–	–	512	–	512	–	512	–	–	512	–	4096	–	4096	–
num filts	–	512	–	512	–	–	512	–	512	–	512	–	–	4096	–	4096	–	2622	–
stride	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1
pad	0	1	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0

FIGURE II. STRUCTURE AND CONFIGURATION OF VGG-NET

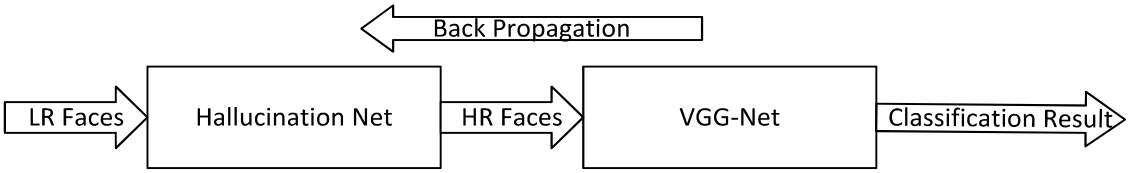


FIGURE III. CASCADED DEEP NETWORK

B. *Face Classification*

We use a pre-trained VGG-Net for classification. Since the model can be fine-tuned with small dataset, the training cycles can be short. Figure II is the structure and configuration of VGG-Net. When applied to different dataset, the only configuration to adjust is the number of filters in fc8 layer. Change the output vector length of the final layer fc8 to the number of individuals in the recognition task and the filter of this layer will be trained from scratch, while others can transfer filters from pre-trained model.

C. *Cascaded Deep Network*

Assume we have already trained the hallucination net and classification net separately. Now we form the cascaded deep network by concatenate the last convolution layer of hallucination net and the input data layer of VGG-Net (Figure III). Then the whole network should be trained with LR images training set for further fine-tuning, which can eventually boost the performance of end-to-end LR classification.

Our proposed model has obvious advantages against the method of directly inputting the hallucinated faces into VGG-Net, for we are optimizing the hallucination part for better classification performance rather than mere PSNR performance mentioned in previous section. In the meantime, the function of hallucination net will be preserved, for we only fine-tune with lower learning rate on the pre-trained mapping parameters. This leads to a boost of accuracy in classification results. More details are discussed in experiment section.

IV. EXPERIMENT

In this section we first describe the dataset for training and testing. Then we demonstrate further details on implementation. Finally, the results compared with other method are provided. Note that all our experiments are implemented in Caffe, Python and Matlab.

TABLE I. CONFIGURATION OF HALLUCINATION NET

Parameters	**Layers**		
	Conv1	***Conv2***	***Conv3***
Filter Dim	9	3	5
Filter Num	64	32	3
Pad	0	0	0
Std	0.001	0.001	0.001
Lr_mult_W	1	1	0.1
Lr_mult_b	0.1	0.1	0.1

TABLE II. EXPERIMENT RESULTS

Evaluation	**Method**		
	Basic CNN	***Separate Training***	***Cascaded Deep Network***
Classification Accuracy	0.9188	0.9225	0.9353
PSNR	33.148	34.258	-

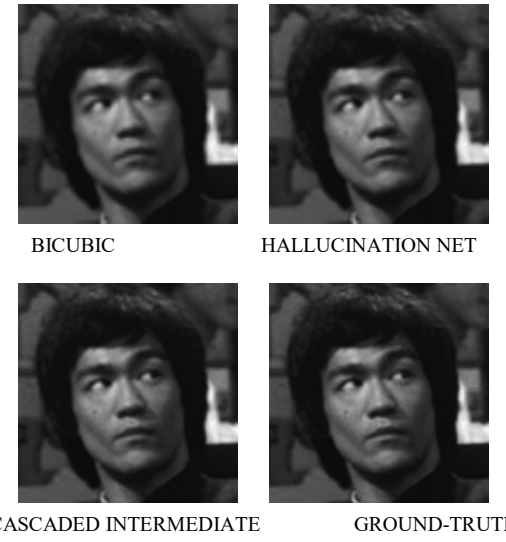
A. *Dataset*

Our dataset is a subset of a very large face database called CASIA-WebFace-Database [9]. The database has 10575 subjects and 494414 images, and 1200 face images of 8 individuals are randomly picked from it. We divide our dataset into training set of 800 faces, validation set of 200 and testing set of 200. The original image size is 250*250, and we downscale them to 224*224 as ground-truth images and 64*64 as low-resolution images.

B. *Implementation Details*

First we perform bicubic interpolation as the only data pre-processing to the LR face images, upscaling them to 224*224. For training and validation set, the interpolated images are cropped into 35*35 patches and ground-truth images into 21*21 patches. These cropped patches are used for separate training for hallucination net. Table I shows the configuration of hallucination net.

For classification net, either interpolated faces or ground-truth faces can be used for the separate training. Pre-trained



BICUBIC HALLUCINATION NET

CASCADED INTERMEDIATE GROUND-TRUTH

FIGURE IV. CASCADED DEEP NETWORK

model of VGG-Net on large-scale face database is available on their webpage. Therefore, with our small dataset the very deep network can be easily fine-tuned and converge in a short period. In our experiment we test on 8 individuals, so the output vector length of fc8 layer is changed to 8. After convergence, build the cascaded model as described in previous section. Note that the parameters from previous training must be preserved for fine-tuning. Then input the interpolated LR face images as a whole (not in patches) and train until reaching stability.

C. Results

The result comparison is demonstrated in Table II. As is shown in the table, hallucination net alone boosts the PSNR, but cascaded network further promotes the accuracy of classification. Figure IV shows the hallucination effect of different methods. But anyway, the visual quality is not what we pursue in this paper since we are not doing face hallucination task. The recognition accuracy is the evaluation of the performance of LR FR approaches, and ours turns out to be superior.

V. SUMMARY

This paper proposes a deep learning approach for low-resolution face recognition. We propose a novel deep neural network structure based on previous related works of face hallucination and recognition. The cascade of two CNNs suggests an effective way of better optimization over conventional indirect methods of LR FR. By optimizing a joint hallucination-classification model, we demonstrate its advantage over conventional ones. Further study can be conducted to explore a more efficient structure to possibly promote the accuracy or accelerate the model.

ACKNOWLEDGMENT

This research was part of the undergraduate IROP (International Research Opportunities Program) of Imperial College London and Tsinghua University.

REFERENCES

[1] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. Proceedings of the British Machine Vision, 1(3):6, 2015

[2] Baker, S., Kanade, T.: Hallucianting faces. In: Proc. IEEE 4th Int. Conf. on Automatic Face and Gesture Recognition (FG), Grenoble, France, Mar. 2000, pp. 83–88 (2000)

[3] Hennings-Yeomans, P.H., Baker, S., Vijaya Kumar, B.V.K.: Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), Alaska, USA, June 2008, pp. 1–8 (2008)

[4] Choi, J.Y., Ro, Y.M., Plataniotis, K.N.: Color face recognition for degraded face images. IEEE Trans. Syst. Man Cybern., Part B, Cybern. 39(5), 1217–1230 (2009)

[5] Li, B., Chang, H., Shan, S.G., Chen, X.L.: Low-resolution face recognition via coupled locality preserving mappings. IEEE Signal Process. Lett. 17(1), 20–23 (2010)

[6] Zhou, E., Fan, H., Cao, Z., Jiang, Y., Yin, Q.: Learning face hallucination in the wild. In: Proc. AAAI Conf. Artificial Intelligence. (2015)

[7] C. Dong, C. Loy, K. He and X. Tang. Learning a deep convolutional network for image super-resolution. European Conference on Computer Vision. Springer. 184–199, 2014.

[8] Zhu, S., Liu, S., Loy, C.C., Tang, X.: Deep cascaded bi-network for face hallucination. In: ECCV. (2016)

[9] Dong Yi, Zhen Lei, Shengcai Liao and Stan Z. Li, "Learning Face Representation from Scratch". arXiv preprint arXiv:1411.7923. 2014.

[10] C. Liu, H. Shum, and C. Zhang. A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model. Proc. of IEEE International Conference