



International Journal of Intellectual Advancements and Research in Engineering Computations

Low resolution face recognition in knowledge stream

I.Kalaimani¹, M.Ishwarya², M.Iswarya²

¹Assistant Professor, Department of ECE, Mahendra Engineering College for Women,
Kumaramangalam, Tamilnadu, India.

²UG Students, Department of ECE, Mahendra Engineering College for Women,
Kumaramangalam, Tamilnadu, India.

ABSTRACT

Typically, the deployment of face recognition models in the wild needs to identify low-resolution faces with extremely low computational cost. To address this problem, a feasible solution is compressing a complex face model to achieve higher speed and lower memory at the cost of minimal performance drop. Inspired by that, this paper proposes a learning approach to recognize low-resolution faces via selective knowledge distillation. In this approach, a two-stream convolutional neural network (CNN) is first initialized to recognize high-resolution faces and resolution-degraded faces with a teacher stream and a student stream, respectively. The teacher stream is represented by a complex CNN for high-accuracy recognition, and the student stream is represented by a much simpler CNN for low-complexity recognition. To avoid significant performance drop at the student stream, we then selectively distil the most informative facial features from the teacher stream by solving a sparse graph optimization problem, which are then used to regularize the fine-tuning process of the student stream. In this way, the student stream is actually trained by simultaneously handling two tasks with limited computational resources: approximating the most informative facial cues via feature regression, and recovering the missing facial cues via low-resolution face classification. Experimental results show that the student stream performs impressively in recognizing low-resolution faces and costs only 0:15MB memory and runs at 418 faces per second on CPU and 9; 433 faces per second on GPU.

Keywords: Face recognition in the wild, Two-stream Architecture, Knowledge Distillation, CNN.

INTRODUCTION

Face, a fundamental attribute that distinguishes one subject from another, needs to be recognized many times every day in modern computer vision and multimedia applications. Among these applications, many well-known face recognition models need to be re-deployed on mobile phones or even smart cameras to meet the real-world requirements that aim to identify low-resolution faces with extremely low computational cost and memory footprint (i.e., face recognition in the wild. Toward this end, it is necessary to explore a feasible solution that can address a key challenge

in face recognition: how to convert an existing complex face model into a more efficient one that still works well on low-resolution faces without remarkable loss of recognition accuracy.

Compared with high-resolution faces, low-resolution faces have their unique visual attributes. Many details are missing in low-resolution faces. However, they are still recognizable for subjects who are familiar with the corresponding high-resolution faces, implying that the neural systems of human beings may have the capability of recovering missing details of familiar faces. Inspired by this fact, many existing low-resolution face models have been proposed, which can be

Author for correspondence:

Department of ECE, Mahendra Engineering College for Women, Kumaramangalam, Tamilnadu, India.

roughly grouped into two categories: the hallucination category and the embedding category [1-5].

Models in the hallucination category propose reconstructing the high-resolution faces before recognition. For example, Kolouri et al. described a transport-based single frame super-resolution method to automatically construct a nonlinear Lagrangian model of high-resolution facial appearance. After that, the low-resolution facial image was enhanced by finding the model parameters that best fit the given low resolution data. Observed that the singular values of a face image at different resolutions have approximately linear relationship. Based on this observation, they first applied singular value decomposition for face representation to learn the mapping function between low-resolution and high resolution face pairs, and then performed both hallucination and recognition of low-resolution faces simultaneously. sentation to perform joint hallucination and recognition, which can synthesize person-specific versions of low-resolution faces with recognition guarantee. Typically, these approaches exhibit impressive performance in recognizing the reconstructed high resolution faces, while the super-resolution operation often brings in additional computational cost and leads to low recognition speed.

RELATED WORK

The approach we proposed in this paper aims to distil knowledge from complex face models for low resolution face recognition. Therefore, we briefly review related work from three aspects, including the general face recognition models, low resolution face recognition and knowledge distillation.

SYSTEM DESIGN

Existing System

Toward this end, it is necessary to explore a feasible solution that can address a key challenge in face recognition: how to convert an existing complex face model into a more efficient one that still works well on low-resolution faces without remarkable loss of recognition accuracy? However,

they are still recognizable for subjects who are familiar with the corresponding high-resolution faces, implying that the neural systems of human beings may have the capability of recovering missing details of familiar faces. Inspired by this fact, many existing low-resolution face models have been proposed, which can be roughly grouped into two categories: the hallucination category and the embedding category. Note that the teacher stream can adopt any architecture of existing deep face models, implying that the proposed approach can convert any existing face model into a much simpler one with higher speed and lower memory at the cost of minimal performance drop. Experimental results on four public datasets show that the student stream performs impressively in recognizing faces at extremely low resolutions. In particular, it uses only 0:15MB memory and runs at about 418 faces per second on a single CPU thread or 9; 433 faces per second on GPU.

PROPOSED SYSTEM

Propose a face model compression method via selective knowledge distillation, which can greatly reduce model size and complexity without remarkable performance drop; Propose graph-based optimization algorithm that can extract the most discriminative facial features from existing face models, which can be used to supervise the training process of low-resolution face models; Conduct comprehensive experiments to show that the compressed model can achieve an extremely high recognition speed with a comparable accuracy with the state-of-the-art high-resolution face models [6-10].

GENERAL FACE RECOGNITION MODEL

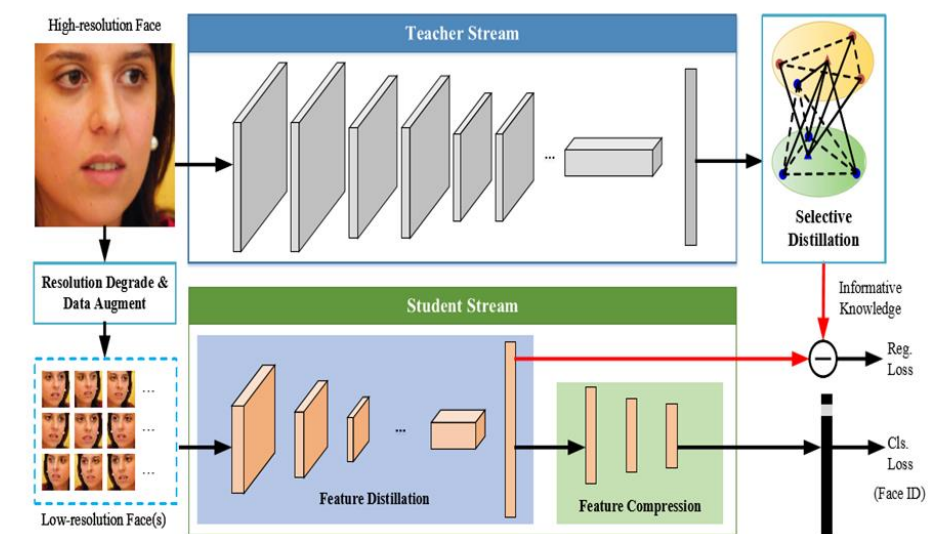
Recently, the general face recognition technique has evolved from the classic shallow frameworks to deep ones with impressive performance improvements. For the deep approaches, a key factor to distinguish them is the loss functions they adopted. For example, Deep Face is an early attempt to ensemble Convolutional Neural Networks (CNNs) by building 3D faces

with identification loss. After that, various loss functions have been proposed for training face recognition CNNs, such as triplet loss, center loss and range loss. In, the tasks of identifying faces and their attributes were simultaneously considered to enhance the recognition performance. For the Deep ID series, several small CNNs using different facial patches were first separately trained, and its subsequent works incorporate face verification signals and change the base networks to increase accuracy. Generally speaking, these deep models have achieved impressive performance in recognizing general faces. As shown in Tab. I, however, many of such generic models have a large amount of parameters, high dimensional feature representations and complex classification function for inference. The complexity faces of these models prevent them from being directly deployed in the wild where the computational resource is limited. Although Deep ID series take low-resolution faces as the input, the unique attributes of low-resolution faces are not explored. To further enhance the performance of low resolution face recognition, it is necessary to explore the missing features during the resolution degradation.

Knowledge Distillation

Instead of mining the knowledge from high-resolution faces, another way to obtain a low

resolution face model (i.e. , the student network) is distilling such knowledge directly from pre-trained complex face model (i.e. , the teacher network). With the development of much deeper and wider networks, such distillation technique has been adopted in many works to compress a complex model (or an ensemble) into a simpler model that is much is a to deploy. Among this works, luo et al. utilized the learned knowledge of a large teacher network or the ensemble of some networks as the supervision to train a compact student network for face recognition. In their approach, the most relevant neurons for face recognition were selected at the higher hidden layers for knowledge transfer. Lopez-Paz et al. Proposed the general distillation framework to combine distillation and learning with privileged information. Su and Maji proposed cross quality distillation to learn models for recognizing low resolution image, non-localized objects and line-drawings by using labeled high-resolution images, labeled localized objects and color images, respective. Radosavovic et al. proposed data distillation to ensemble predictions from multiple transformations of unlabeled data to automatically generate new training annotations.



TRANSACTIONS ON IMAGE PROCESSING

LOW-RESOLUTION FACE RECOGNITION

Typically, there are two ways for low-resolution face recognition. The hallucination category aims to reconstruct high-resolution faces before recognition, while the embedding category proposes extracting features directly from low-resolution faces via the embedding schema. In the hallucination category, Kolouri et al. constructed a nonlinear Lagrangian model of high-resolution facial appearance and then found the model parameters that best fit the low-resolution faces. Jian et al. Proposed a framework based on singular value decomposition and performed face hallucination and recognition simultaneously. In a joint face hallucination and recognition framework was proposed based on sparse representation. This framework can synthesize person-specific low-resolution faces for recognition. In a system was proposed to recognize faces by using sparse representation with the specific dictionary involving many natural and facial images. Moreover, deep models like and can generate extremely realistic high-resolution images from low-resolution faces. However, the speed of such hallucination or super resolution based approaches may be a little slow due to the complex high-resolution face reconstruction process, which hinders their direct deployment in real-world scenarios with limited computational resources.

Instead of reconstructing high-resolution faces, a more direct approach is embedding low-resolution faces into various external contexts to recover the missing features during resolution degradation. Inspired by that, some approaches proposed transforming both high-resolution and low-resolution faces into a unified feature space for matching, while in the multi-scale (multi-resolution) faces were simultaneously analyzed to extract better features. In the multidimensional scaling was adopted to learn a common transformation matrix to simultaneously transform the facial features of low-resolution and high-resolution training images. Shear et al, proposed a joint sparse coding technique for robust recognition at low-resolution, while Wang et al, attempted to solve very low resolution recognition problem using deep learning methods.

In[37], CNNs were adopted with a manifold-based track comparison strategy for low-resolution face recognition in videos.

THE APPROACH

Our two-stream knowledge distillation framework consists of a teacher stream and a student stream. The teacher stream can adopt any complex face recognition neural networks that have been previously trained (and the training data may be no longer available). The distillation process aims to learn a simple and compact student stream that imitates the teacher stream for its practical deployment in real-world scenario.

The learning process consist of three stages: 1) the initialization stage initializes the teacher stream by taking a complex CNN or an ensemble of several CNNs pre-trained on high-resolution face images, and the student stream by classifying low resolution face images with identity labels; 2) the Selection stage extracts the most informative knowledge from the teacher stream where the “right” knowledge is selected while the “wrong” one is wiped out; and 3) the Fine-tuning stage transfers the selective knowledge from teacher and low resolution face images to co-supervise the fine-tuning progress of the student stream by jointly performing feature regression and face identify classification. More details of the three stages are described as follows.

DEFINITION

For the sake of simplicity, we define the key components of the two-stream CNNs as follows: Teacher stream. The teacher stream $\phi_t(F; W_t)$ is a complex CNN (or an ensemble of several complex CNNs) with the set of parameters W_t pre-trained for recognizing a higher solution face F . Here we assume that W_t absorbs the rich knowledge encoded in massive high-resolution face images from a teacher face set D_t , in which each face is labeled by an integer from the identity set L_t . Generally, the number of training face images $|D_t|$ is very large and may be invisible to the student stream (e.g., a CNN model released on the Internet is pre-trained with additional face images from private datasets). Student stream. The student

stream is a much simpler CNN for recognizing a low-resolution face \tilde{F} with parameters W_s . It is learned from the student face set where $|D_s|$ is the number of higher resolution faces. For each high-resolution face F_i , the student face set also contains its N resolution-degraded versions, and the j th resolution-degraded face is denoted as F_{ij} . Note that both the high-resolution face F_i and its degraded versions correspond to the same identity label l_i from the identity set L_s . Here we assume that there are totally C classes of faces in L_s , and the number of high-resolution faces for the c th class is K_c such that $|D_s| = \sum_{c=1}^C K_c$.

Initialization of the two-stream CNNs

As shown is our two-stream CNNs simultaneously conduct high-resolution and low-resolution face recognition with a teacher stream $\phi_t(F; W_t)$ and a student stream $\phi_s(\tilde{F}; W_s)$, respectively. The parameter set W_t of the teacher stream can be initialized by state-of-the-art face recognition models or their ensemble, such as VGG Face with VGG16 architecture, Face Net with Google Net architecture and VGGFace2 with ResNet50 architecture. As a representative example, we use the architecture of VGGFace in the teacher stream and initialize W_t with the author-released model. Note that VGGFace is pre-trained on a massive face image dataset.

CONCLUSION

At present, the problems of large model parameters and high feature dimension widely exist in face recognition models based on deep learning, which hinders their practical deployment

REFERENSES

- [1]. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, 1701–1708.
- [2]. Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, 1891–1898.
- [3]. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 815–823.
- [4]. O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in British Machine Vision Conference (BMVC), 3, 2015, 6.

on resource-restricted applications (e.g., on embedded or mobile devices). To address this problem, this paper proposes a knowledge distillation method, adopting original large model as the teacher network and letting the teacher selectively supervise the training of student networks via designing the multi-task loss function combining regression and classification items. We have accomplished combination of high dimensional deep feature regression and low-resolution facial classification, which achieves the uniform compression of deep network and feature dimension with recognition accuracy rate assured. Experimental results show that the proposed approach can transfer the informative knowledge from the teacher network to student models, leading to compact face recognition models with impressive effectiveness and efficiency.

In our future work, we will tentatively explore the usage of recurrent mechanism that aims to handle the failure cases in the teacher stream. Face attributes such as gender, age and makeup will be incorporated into the multi-task framework to further enhance the performance of the compressed model.

FUTURE WORK

In our future work, we will tentatively explore the usage of recurrent mechanism that aims to handle the failure cases in the teacher stream. Face attributes such as gender, age and makeup will be incorporated into the multi-task framework to further enhance the performance of the compressed model.

- [5]. B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition brary with mobile applications," CMU School of Computer Science, 2016.
- [6]. A. Pentland and T. Choudhury, "Face recognition for smart environments," *Computer*, 33(2), 2000, 50–55.
- [7]. D. Liu, B. Cheng, Z. Wang, H. Zhang, and T. S. Huang, "Enhance visual recognition under adverse conditions via deep networks," arXiv preprint arXiv: 1712.07732, 2017.
- [8]. S. Kolouri and G. K. Rohde, "Transport-based single frame super resolution of very low resolution face images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, 4876–4884.
- [9]. M. Jian and K.-M. Lam, "Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition," *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, vol. 25(11), 2015, 1761–1772.
- [10]. M.-C. Yang, C.-P. Wei, Y.-R. Yeh, and Y.-C. F. Wang, "Recognition at along distance: Very low resolution face recognition and hallucination," in *International Conference on Biometrics (ICB)*, 2015, 237–242.