Front. Comput. Sci., 2017, 11(2): 208–218
DOI 10.1007/s11704-016-6076-3

RESEARCH ARTICLE

VIPLFaceNet: an open source deep face recognition SDK

Xin LIU[1,2], Meina KAN[1,2], Wanglong WU[1,2], Shiguang SHAN (✉)[1,2], Xilin CHEN[1,2]

1 Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing 100190, China
2 University of Chinese Academy of Sciences, Beijing 100049, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2016

Abstract Robust face representation is imperative to highly accurate face recognition. In this work, we propose an open source face recognition method with deep representation named as VIPLFaceNet, which is a 10-layer deep convolutional neural network with seven convolutional layers and three fully-connected layers. Compared with the well-known AlexNet, our VIPLFaceNet takes only 20% training time and 60% testing time, but achieves 40% drop in error rate on the real-world face recognition benchmark LFW. Our VIPLFaceNet achieves 98.60% mean accuracy on LFW using one single network. An open-source C++ SDK based on VIPLFaceNet is released under BSD license. The SDK takes about 150ms to process one face image in a single thread on an i7 desktop CPU. VIPLFaceNet provides a state-of-the-art start point for both academic and industrial face recognition applications.

Keywords deep learning, face recognition, open source, VIPLFaceNet

1 Introduction

Face recognition, as one of the typical problems in computer vision and machine learning, plays an important role in many applications, such as video surveillance, access control, computer-human interface and mobile entertainments [1]. Generally speaking, a conventional face recognition system consists of four modules, face detection, face alignment, face representation and identity classification. In this pipeline, the key component for accurate face recognition is the third module, i.e., extracting the representation of an input face, which this paper mainly focuses on.

The main challenges of face representation lie in the small inter-person appearance difference caused by similar facial configurations, as well as the large intra-person appearance variations due to large intrinsic variations and diverse extrinsic imaging factors, such as head pose, expression, aging, and illumination. In the past decades, face representation is mostly based on hand-crafted local descriptors [2–8] and shallow learning-based representation models [9–14]. With the development of deep learning technology, it becomes a more potent approach for face representation learning, especially in the real-word scenarios. Compared with the previous hand-crafted routine, deep face representation is learned in a data-driven style which can guarantee better performance as validated in Refs. [15–19]. Taking the de-facto real-world face recognition benchmark LFW as an example, hand-crafted descriptor recorded 95.17% set by high-dimensional LBP [4], while 99.63% accuracy achieved by the latest deep FaceNet in Ref. [19].

In spite of many decades of research and development on face recognition, few open-source face recognition systems are publicly available yet. An open-source SDK with high accuracy in general scenarios is in great need for both academic research and industrial applications. Therefore, in this work, we meet this requirement and propose a deep face recognition model named as VIPLFaceNet, which is released as a BSD-license open source software with detailed implementation of the recognition algorithm. VIPLFaceNet is a powerful deep network for face representation with ten layers including

Received February 2, 2016; accepted August 5, 2016

E-mail: sgshan@ict.ac.cn

seven convolutional layers and three fully-connected layers. As a BSD-license open source software, VIPLFaceNet allows both academic research and industrial face recognition applications in different software and hardware platforms for free.

The contributions of this paper are summarized as follows:

1) We propose and release an open source deep face recognition model, VIPLFaceNet, with high-accuracy and low computational cost, which is a 10-layer deep convolutional neural network that achieves 98.60% mean accuracy on the real-world face recognition benchmark LFW.

2) We investigate the network architecture design and simplification. By careful design, VIPLFaceNet reduces 40% computation cost and cuts down 40% error rate on LFW compared with the AlexNet [20].

3) The VIPLFaceNet SDK code is written in pure C++ code under the BSD license. It is free and easy to be deployed in various software or hardware platforms for both academic research and industrial face recognition applications.

In summary, VIPLFaceNet is an open source deep face recognition SDK with high accuracy in general scenarios, which is built for facilitating the academic and industrial application of various real-world face recognition tasks. The rest of this paper is organized as follows. Section 2 presents the related work on face representation learning and introduces the face recognition benchmarks. Section 3 presents the network architecture design and technical details of our VIPLFaceNet. Section 4 conducts the experimental evaluation with comprehensive discussions and Section 5 concludes this paper.

2 Related work

In this section, we give a brief review of the related work on face representation learning. Moreover, we briefly review the face recognition benchmarks and discuss the performance evolution on the de-facto real-world face recognition benchmark LFW.

2.1 Face representation before deep learning

In the past decades, numerous hand-crafted local features were proposed for face representation, e.g., Gabor wavelets [2], local binary pattern (LBP) [3] and its high dimensional variant [4], scale-invariant feature transform (SIFT) [8], histogram of oriented gradients (HOG) [5], patterns of oriented edge magnitudes (POEM) [6], local quantized pattern (LQP) [7], etc. However, designing an effective local descriptor demands considerable domain specific knowledge and a great deal of efforts.

Besides the hand-crafted local features, learning-based representation is also popular and reports promising accuracy. In Refs. [9, 10], filters are learned to maximize the discriminative power for face recognition. In Ref. [21], faces are represented from their responses to many pre-trained object filters. In Refs. [11–13, 22], codebook learning technologies are utilized for robust face representation. More recently, faces are represented with mid-level or high-level semantic information. For instance, the attributes and simile classifier [23] represent faces by the mid-level face attributes and so-called simile feature. Tom-vs-Pete classifier [14] encodes faces with high-level semantic information by the output scores of a large number of person-pair classifiers. Different from the deep learning approaches, the above methods are still shallow models and mostly rely on hand-crafted local features.

2.2 Deep face representation learning

In recent years, deep learning methods are exploited to learn hierarchical representation and report state-of-the-art performance on LFW [15–19, 24].

DeepFace is an early attempt of applying deep convolutional neural network in real-world face recognition. There are four highlights in DeepFace: 1) a 3D model based face alignment to frontalize facial images with large pose; 2) a very large scale training set with four million face images of 4 000 identities; 3) deep convolutional neural network with the local connected layer that learns separate kernel for each spatial position; and 4) a Siamese network architecture to learn deep metric based on the features of the deep convolutional network.

The DeepID [16], DeepID2 [17], and DeepID2+ [18] are a series of works, which provide a very good example of deep network evolution. In DeepID, 25 CNN networks are trained on each face patch independently. Besides, Joint Bayesian method [25] is applied to learn robust face similarity metric. Finally, an ensemble of 25 deep networks achieve 97.45% mean accuracy on LFW. The DeepID2 introduces the joint identification and verification losses. The performance of DeepID2 on LFW is improved to 99.15%. The DeepID2+ just makes the network deeper and adds auxiliary loss signal on lower layer. Besides, the activation of the feature embedding layer is also studied as sparse, selective and robust. The mean accuracy of DeepID2+ on LFW is 99.47% with 25 CNN models.

Learning face representation from scratch [24] presents a