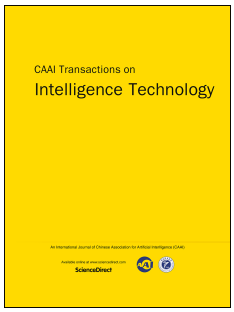
Accepted Manuscript



Face recognition using both visible light image and near-infrared image and a deep network

Kai Guo, Shuai Wu, Yong Xu

PII: S2468-2322(17)30014-8

DOI: 10.1016/j.trit.2017.03.001

Reference: TRIT 41

To appear in: *CAAI Transactions on Intelligence Technology*

Received Date: 30 January 2017

Accepted Date: 28 March 2017

Please cite this article as: K. Guo, S. Wu, Y. Xu, Face recognition using both visible light image and near-infrared image and a deep network, *CAAI Transactions on Intelligence Technology* (2017), doi: 10.1016/j.trit.2017.03.001.

Face Recognition Using Both Visible light Image and Near-infrared image and a Deep Network

Kai Guo, Shuai Wu, Yong Xu
School of Computer Science and Technology
Harbin Institute of Technology Shenzhen Graduate School
Shenzhen, China
guokaiwork@yeah.net

***Abstract*—In recent years，deep networks has achieved outstanding performance in computer vision, especially in the field of face recognition. In terms of the performance for a face recognition model based on deep network, there are two main closely related factors: 1) the structure of the deep neural network, and 2) the number and quality of training data. In real applications, illumination change is one of the most important factors that significantly affect the performance of face recognition algorithms. As for deep network models, only if there is sufficient training data that has various illumination intensity could they achieve expected performance. However, such kind of training data is hard to collect in the real world. In this paper, focusing on the illumination change challenge, we propose a deep network model which takes both visible light image and near-infrared image into account to perform face recognition. Near-infrared image, as we know, is much less sensitive to illuminations. Visible light face image contains abundant texture information which is very useful for face recognition. Thus, we design an adaptive score fusion strategy which hardly has information loss and the nearest neighbor algorithm to conduct the final classification. The experimental results demonstrate that the model is very effective in real-world scenarios and perform much better in terms of illumination change than other state-of-the-art models. The code and resources of this paper are available at http://www.yongxu.org/lunwen.html.**

***Index Terms*—Deep Network, Face recognition, Illumination change, Insufficient training data**

I. INTRODUCTION

Biometrics is one of the most important branches of pattern recognition [1-3]. Face recognition is one of the most attractive biometric techniques. Nevertheless, face recognition in real applications is still a challenging task [4]. The main reason is that the face is a non-rigid object, and it often has different appearance owing to various facial expression, different ages, different angles and more importantly, different illumination intensity. In recent years, deep learning has become more and more prevalent in computer vision. AlexNet [5] which is designed by Alex Krizhevsky got the champion of ILSVRC-2012 competition and outperformed the second place nearly 10 percents.

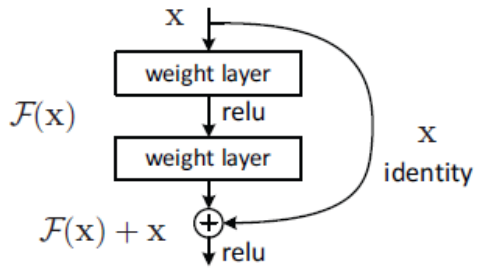


Figure 1: shotcut structure

From then on, researchers realized the powerful analysis ability of deep convolution network. Simonyan [6] proposed a very deep convolutional network, which is called VGG net, and get the second place of ILSVRC-2014 competition. VGGnet could be divided into several blocks and each block contains several convolutional layers which have identical kernel size and channel number. GoogleNet [7] is specially designed for the ILSVRC-2014 competition by google team. It applies a structure called inception to preserve locally compact connection, while the global structure is sparse. GoogleNet won the first place of the ILSVRC-2014 competition whose top-5 error rate is less than 7%. VGGnet and GoogleNet are very deep and have 19 and 22 layers respectively. On account of the great success of VGG and GoogleNet, researchers began to apply different methods to increase the depth of network. However, a degradation problem is exposed: when the network depth increases, accuracy gets saturated and then degrades rapidly. Such degradation, unexpectedly, is not related with overfitting. In order to solve such problem, He [8] designs a shotcut structure which is illustrated in Figure 1. This structure combines the input x and F(x) as the final output, so F(x) is considered as the residual. The network piled up by this structure is called residual network. Residual network could have the depth of 152 layers and the top-5 error is less than 4% on imagenet database.

Face recognition could be considered as a special classification task and the deep network is pretty suitable for face recognition. Deep neural networks have powerful feature extraction ability, and can obtain competitive extractor by using massive training sets. For some public face data sets, such as Labeled Faces in the Wild Home (LFW) [9], the accuracy of deep networks can even reach 99.8%. However, face