Shape Driven Kernel Adaptation in Convolutional Neural Network for Robust Facial Trait Recognition

Xuewen Yang

August 7 2018

Abstract

One key challenge of facial trait recognition is the large non-rigid appearance variations due to some irrelevant real world factors, such as viewpoint and expression changes. In this paper, the author explore how the shape information, i.e. facial landmark positions, can be explicitly deployed into the popular Convolutional Neural Network (CNN) architecture to disentangle such irrelevant non-rigid appearance variations. First, instead of using fixed kernels, they propose a kernel adaptation method to dynamically determine the convolutional kernels according to the spatial distribution of facial landmarks, which helps learning more robust features. Second, motivated by the intuition that different local facial regions may demand different adaptation functions, they further propose a tree-structured convolutional architecture to hierarchically fuse multiple local adaptive CNN subnetworks. Comprehensive experiments on WebFace, Morph II and MultiPIE databases well validate the effectiveness of the proposed kernel adaptation method and tree-structured convolutional architecture for facial trait recognition tasks, including identity, age and gender recognition. For all the tasks, the proposed architecture consistently achieves the state-of-the-art performances.

1. Introduction

In the last decade, great progress has been made in developing deep neural network for various computer vision tasks. Among them, Convolutional Neural Network (CNN) has achieved exciting performance on digit recognition [2], traffic sign recognition [2], object recognition [1] and scene labeling.

In this paper, the author also propose to exploit the potential of facial shape information, i.e. a set of facial landmarks, to help CNN based methods learning more powerful and robust face representation. The basic idea is using different convolutional kernel according to the shape information, i.e. distribution of face landmarks, in order that the learned features would become more invariant to appearance variations caused by different viewpoints or expressions. Another potential of this kernel adaptation is that the additional discriminant information contained in the shape may also be directly coded into the learned feature. Specifically, they propose a shape driven kernel adaptation for CNN and use automatically adapted kernels to more efficiently disentangle the mixed factors in each input face image.

2. Kernel Adaptation for CNN

In real world environment, facial appearance may change significantly due to different poses and expressions, one traditional convolutional layer with fixed kernel functions may generate undesirably different responses for the same face. To achieve feature invariance under these complex variations, a mechanism, which can make the convolution kernel automatically adapt to the specific variations of each face instance, will be beneficial.

To this end, they propose a kernel adaptation mechanism for traditional CNN framework. Suppose the input face image is I and the kernel function is f, they hope the kernel function can be automatically adapted for the input face image according to a latent variable S. Then the convolution with kernel adaptation can be formulated as:

$$f = \phi(S, \theta) \tag{1}$$

$$C = \varphi(I * f + b) \tag{2}$$

3. Conclusion

In this paper, the author propose a kernel adaptation method in CNN to exploit shape information for disentangling irrelevant non-rigid facial appearance variations. Since different facial regions have different deformations, to better exert its function, they adopt kernel adaptation in multiple local regions respectively and further propose a tree-structured convolutional architecture to jointly learn features in an end-toend manner. Evaluations on facial trait recognition tasks demonstrate the state-of-the-art performances of the proposed tree-a-CNN model.

Although this network has relatively shallow structure comparing to the state-of-the-art deep convolutional neural networks [1], our method achieves comparable or better performance than AlexNet [1] and DeepFace [3]. These results suggest that kernel adaptation method provides a more compact and effective way to disentangle complex factors in facial images. This is very helpful for an large end-to-end system like deep networks. In future, they will try to deploy this kernel adaptation method into larger and deeper networks to fully explore the potential of shape information for robust feature learning.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012. 1, 2

[2] J. Schmidhuber, U. Meier, and D. Ciresan. Multi-column deep neural networks for image classification. In *CVPR*, 2012. 1

[3] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *CVPR*, 2014. 2