Research on Image Recognition Based on Deep Learning Technology

Fangfang Li

June 20, 2018

Abstract

Nowadays image recognition technology is widely used, and plays a very important in various fields. Deep learning technology uses multilayer structure to analyze and deal with image features, which can improve the performance of image recognition. The popular models of deep learning contain AutoEncoder, Restricted Boltzmann Machine, Deep Belief Network, Convolutional Neural Network, Recurrent Neural Network and other improved methods. The applications of image recognition based on deep learning technology including image classification, facial recognition, image search, object detection, pedestrian detection, video analysis. We believe that in the future deep learning will develop rapidly in theory, algorithm, and application and they will make our lives more intelligent.

1. Introduction

In 2016, in the deep learning field, the most sensational event is that in March AlphaGo defeats Li Shishi, who is a professional nine player and the world chess champion, with the total score 4:1. This is AlphaGo's second victory. In October 2015, it became the first Computer Go program to defeat the European Go champion Fan Hui, a 2-dan (out of 9 dan possible) professional, five to zero. The AlphaGo system was developed by British computer company DeepMind which was bought by Google in 2014. AlphaGo's algorithm uses a Monte Carlo tree search to find its moves based on knowledge previously "learned" by deep learning, both from human and computer play [1].Deep learning simulates biological neural brain to process precise and complex calculating, like people recognizing objects.

Image recognition is an important application and requirement in life and industry. With the need of social practice, there are more and more kinds of things to be classified and identified, and the content of the object is more and more complex. For example, in the traffic management system, the car license automatic recognition based on image recognition technology can record the illegal behaviors of vehicles. So image recognition technology plays a very important role in the fields of industry and our daily life. Therefore, it is of great significance to improve the performance of image recognition technology by using advanced deep learning algorithms.

2. Definition of Deep Learning

Deep learning is a new field in the research of machine learning, which is a kind of unsupervised learning. The motivation of deep learning is to simulate the human brain to establish the neural network, which can interpret the data, such as images, sound and text, like human brain. In the neural network, attribute categories or features are represented by combining low-level features to form more abstract high-level features, in order to discover the distributed feature representation of data. One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction.

The relation between points q on the image and points on the template Q is given by

$$q = sRQ + t. \tag{1}$$

To recover s, R, and t, we first subtract the centroid from both point sets to get $p = q - \overline{q}$ and $P = Q - \overline{Q}$, then estimate a 2×3 linear transformation $A = pP^T(PP^T)^{-1}$ and translation $t = \overline{q} - A\overline{Q}$. To recover an estimate of the rotation and scale we let the third row of A be the cross product between the first two rows and by taking its SVD, $A' = UDV^T$, we estimate the closest rotation in terms of Frobenius norm $R = UV^T$. Two of the singular values of A' are identical [3], and this is our estimate of scale. We thenes timate the yaw, pitch and roll angles from the rotation matrix. Given the estimated pose we transform the template to the orientation of the face in the image, the image is back-projected onto the shape, and then a frontal view of the face is rendered [2]. This results in a collection

of faces where every face is in approximately frontal position as can be seen in Fig. 1

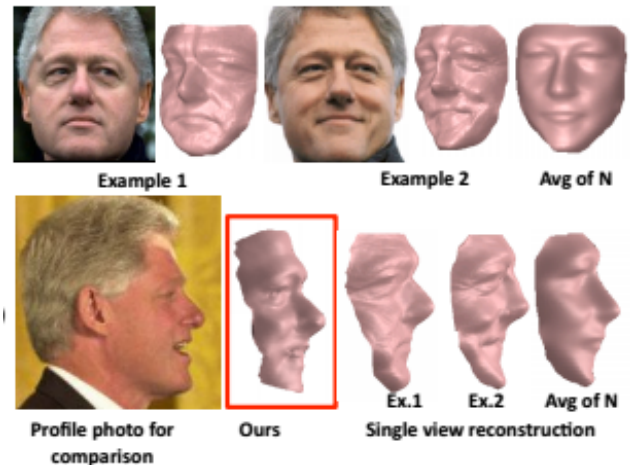

Figure 1. Comparison to single view reconstructions.

Compared with traditional machine learning methods, deep learning methods are particularly suitable for dealing with large data. The deep learning methods can reduce model bias to improve statistical estimation accuracy with more complex models. In addition, deep learning is almost the only end to end machine learning system, which abandons the intermediate step of artificial rules and applies the prior knowledge of the data structure to the new model structure. These advantages make deep learning methods very suitable for image recognition.

References

[1] S. H. Amin and D. Gillies. Analysis of 3D face reconstruction. In *ICIAP*, pages 413–418, 2007. 1

[2] F. Liu, D. Zeng, Q. Zhao, and X. Liu. Joint face alignment and 3D face reconstruction. In *ECCV*, pages 545–560, 2016. 1

[3] Y. Zheng, J. Chang, Z. Zheng, and Z. Wang. 3D face reconstruction from stereo: A model based approach. In *ICIP*, pages 65–68, 2007. 1