ISBN: 978-981-14-2064-1

The Application of TensorFlow Facial Recognition Technology in Education Assistant System

Wen-yuan Hao[1,a], Jae-woong Kim[2,b] and Jun He[1,c]

[1]Xiamen Academy of Arts & Design,Fuzhou University
[2]Computer Science & Engineering, kongju National University
[1]Xiamen, China, [2]Cheonan, Korea
[a]mainfunction@yeah.net, [b]jykim@kongju.ac.kr, [c]Dr.hejun@qq.com

Abstract

TensorFlow is an open source artificial intelligence system from Google Inc., which was announced in October 2015. With flexibility, high efficiency, and good scalability and portability, it can be applied to a variety of computing environments from smartphones to large computing clusters. It is currently used in many fields.

This study experimentally proposes a technical solution, based on TensorFlow to construct a convolutional neural network model for face recognition, and counts the number of faces identified, so as to quickly perform attendance statistics.

Key words: TensorFlow, Convolutional neural network, Deep learning, Face recognition

Introduction

In recent years, with the continuous development of artificial intelligence technology, it has become a reality that computers can recognize objects and make computers have human-like visual capabilities. Especially, convolutional neural networks are widely used in image classification, target detection, and image semantics. The field of computer vision and image processing has achieved great success.[1] The use of convolutional neural networks for face recognition has achieved great results.[2] which has made face recognition technology widely used in industry, such as access control systems, identification, human-computer interaction, etc. However, the current application in education Not much. This study experimentally applies face recognition technology to the attendance of college classrooms, using TensorFlow as a framework for building a convolutional neural network for face detection, and using face recognition as a basis for attendance in class, Classroom automated attendance.

TensorFlow[3] was officially open sourced as a machine learning library in November 2015. As a second-generation open source system released by Google, it has high flexibility, good portability and cross-platform capabilities, and is the most abundant API in all deep learning frameworks. The implementation of neural networks including a variety of structures and convenient visualization aids provide an excellent research platform for machine learning. The dramatic increase in user volume since the open source is currently the most noteworthy deep learning framework. The core concept of TensorFlow is the calculation graph, which is an abstraction of the computational model. Data is represented by a tensor, and the state is maintained with a variable. The actual calculation needs to be performed in a context called a session. Therefore, the Tensorflow program can generally be divided into two phases: the first phase is the build, which is used to construct the entire calculation process in the form of a graph; the second phase is the execution, and the calculation in the graph is performed by creating a session. The graph consists of nodes, similar to the layers in the convolutional neural network. The nodes of the graph can have zero or more input calculations to produce the output, and the combination of the nodes ultimately constitutes a complete computational flow. In addition, you can also import those already trained open-source models into the current project.

Framework Design of Face Recognition Education Assistant System

Class attendance through face recognition is designed to enable teachers to quickly and accurately grasp student attendance. Compared to real-time positioning attendance, such as by using GPS. Or use other biometrics such as fingerprints, palm prints, voiceprints, irises, etc. Face recognition technology extracts people's facial feature information for identity authentication, which has low application cost, convenient use, high accuracy and convenience. Stronger. Therefore, the application of face recognition technology, image processing technology and database, etc., to form a complete teaching aid system, to help teachers save class time, fast and accurate classroom attendance.

The design idea is that the teacher or the camera shoots multiple classroom photos as data collection, and all the photos are detected by the face and the attendees are identified. Taking into account the conditions such as front and rear occlusion or poor lighting conditions, the unrecognized face is displayed at the same time as the standard photo of the person not included in the time sheet, allowing the teacher to judge the unrecognized face and standard according to the actual situation. Whether the photos are consistent to determine if the unidentified students belong to the classroom. Then, the standard photos of the students who are not included in the attendance are displayed for the teacher to confirm. The specific work flow chart of the system is shown in Figure 1.

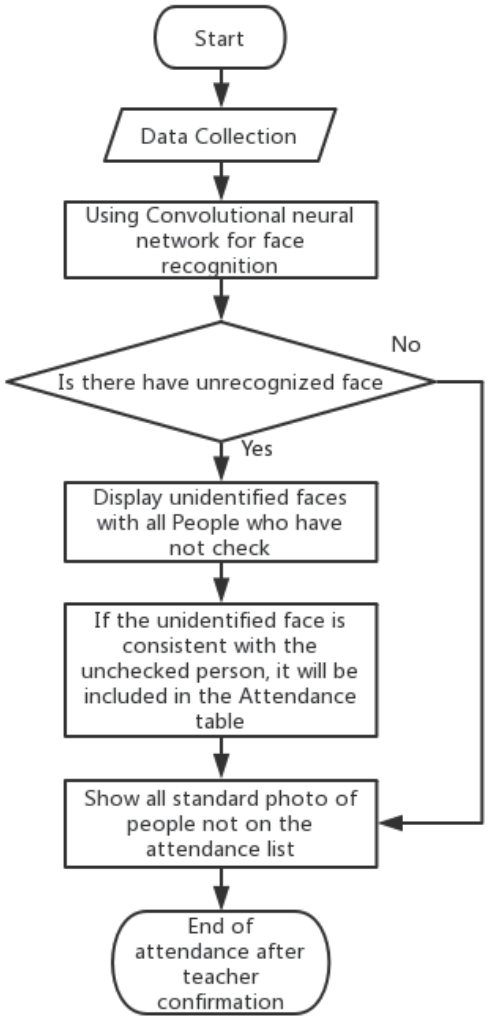

Fig. 1 System flow chart

Convolutional Neural Network

A. The realization of traditional face detection

Before the convolutional neural network has not yet become popular, face recognition mainly extracts the features contained in the face based on the position and size of the face and the relative position between the facial organs, and then compares it with the existing database to realize the identity. Identification. The widely used Haar feature extraction combined with the Adaboost classifier for face detection was proposed by Viola and Jones.[4] This method firstly uses the Haar-like feature to traverse the target image, extracts the features, and then uses the Adaboost to determine the extracted features. Adaboost is divided into three categories: strong classifier, weak classifier and feature node. The core idea is to pass the above Haar feature. The template extracts facial features, and the extracted facial features are quickly calculated to select key features. Finally, it is sent to the cascaded classifier for iterative training. The classifier training uses Adaboost to learn a low detection rate. A multi-layer tree classifier with a rejection rate target.

This method is not effective compared to the current popular use of convolutional neural networks for face detection. The detection data is very demanding, and the face can only be detected well under the condition that the face is front and the vertical light is shining, and the person cannot be detected when the side of the character, the partial occlusion and the light condition are not good.

B. Convolutional neural network algorithm

Convolutional neural networks are similar to fully connected neural networks and are widely used in computer vision. The convolutional neural network consists of an input layer, several convolutional layers, a pooled layer, and a fully connected layer. The layer is the core part of the convolutional neural network, also known as the filter or convolution kernel, the receptive field in animal vision.[5] The convolution kernel is a three-dimensional matrix in mathematics, and the length and width of the matrix are generally specified by hand. The commonly used size is 3 x 3 or 5 x 5. The advantage of convolution is that regardless of the size of the picture, the number of weights to be trained is only related to the size of the convolution kernel and the number of convolution kernels, which reduces the parameters in the network and thus reduces the amount of computation during network training.

The forward propagation algorithm of the convolutional neural network needs to specify the depth of the convolution kernel and the size of the convolution kernel matrix. In order to increase the use of edge pixels, a number of circles 0 are added around the input matrix, and then the calculation is called padding. The size of the pixel distance per movement during the convolution process is defined as the stride. After the forward propagation, the output layer results are obtained. The forward propagation algorithm can be roughly summarized as:

$$a^{l} = \sigma(\sum_{k=1}^{M} a_k^{l-1} * W^{l} + b^{l}) \tag{1}$$

Where 'l' represents the number of layers, '*' represents the convolution, 'W' represents the weight 'b' represents bias, 'σ' as the activation function, and mostly the 'ReLU' function. After obtaining the predicted value of the output layer, the difference between the two is obtained by comparing with the real value, and then the gradient of the loss function for each parameter is calculated by the back propagation algorithm, and then the gradient value is used to correct the weight value and bias value according to the gradient and the learning rate. The formula can be roughly summarized as:

$$\begin{aligned} W_{i+1} &= W_{i+1} - \eta \frac{\partial L}{\partial W_i} \\ b_{i+1} &= b_{i+1} - \eta \frac{\partial L}{\partial b_{i+1}} \\ W_i &= W_i - \eta \frac{\partial L}{\partial f_{i+1}} \frac{\partial f_{i+1}}{\partial f_i} \frac{\partial f_i}{\partial W_i} \\ b_i &= b_i - \eta \frac{\partial L}{\partial f_{i+1}} \frac{\partial f_{i+1}}{\partial f_i} \frac{\partial f_i}{\partial b_i} \end{aligned} \tag{1}$$

Where 'L' is the loss function and 'η' is the learning rate. If the learning rate is not set properly, the model will not converge when it is trained.

Convolutional neural networks are now widely used in the field of computer vision. The face detection and face recognition used in this paper are all implemented using a convolutional neural network model.

C. Face detection using MTCNN

Face detection is an indispensable part of the face recognition system. It is the paving work for applications such as face recognition and key point positioning. Considering the complexity of the actual environment in the specific application of classroom lighting and students' front and rear occlusion, this study chose to use the deep learning MTCNN[6] based cascade neural network face detection proposed by Kaipeng Zhang et al. in order to balance the accuracy and rate. method. MTCNN can also get better detection results under different angles of light environment and facial expressions.

MTCNN adopts a three-level network structure Proposal Network (P-Net), which is mainly used to obtain the regression window of the candidate window and the bounding box of the face region. The boundary box is used for regression to calibrate the candidate window.

Refine Network (R-Net), the network structure still removes those false-positive areas through bounding box regression and NMS.

Output Network (O-Net), get more detailed processing results. The effect is roughly the same as the R-Net layer. However, this layer has more supervision on the face area and also outputs 5 landmarks. Network structure shown in Figure 2.

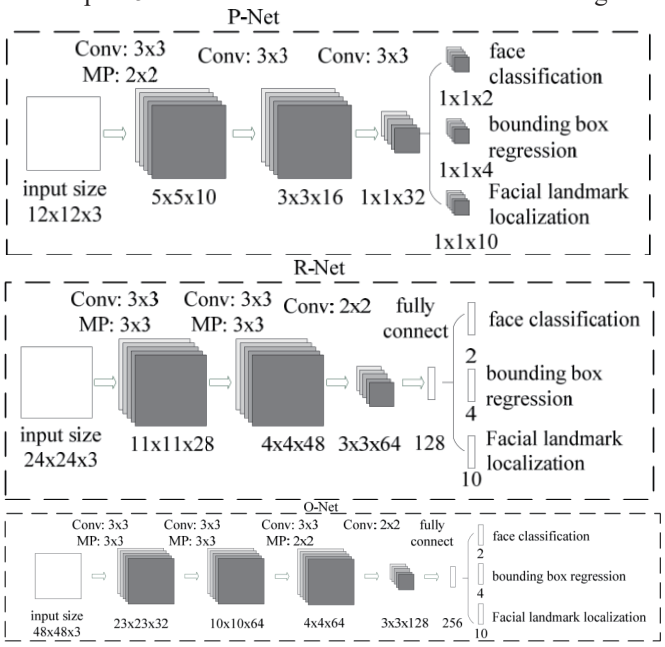

Fig. 2 The architectures of P-Net, R-Net, and O-Net, where "MP" means max pooling and "Conv" means convolution. The step size in convolution and pooling is 1 and 2, respectively

The MTCNN feature descriptor mainly consists of three parts, face/non-face classifier, bounding box regression, and landmark positioning. This article uses GitHub open source MTCNN trained model for face detection to get a good detection effect, as shown in Figure 3.

Fig. 2 Face detection effect

Use FaceNet for Face Recognition

FaceNet[7] is a deep learning model that Google published on CVPR in 2015 to extract facial features. The network is mainly composed of a batch input layer and a deep convolutional neural network, and then the predicted values are normalized to output facial features. The network uses the Triplet Loss method to calculate the loss function. When training, the model aims to continuously reduce the distance between different people's faces while increasing the distance between different people to ensure network extraction. The singularity and consistency of the facial features.

FaceNet can directly map face images to Euclidean space, and the distance of space represents the similarity of face images. As long as the mapping space is generated, tasks such as face recognition, verification, and clustering can be easily accomplished. The method is based on a deep convolutional neural network with an accuracy of 0.9963 on the LFW dataset and 0.9512 on the YouTube Faces DB dataset. Therefore, this study uses Kaiyuan's FaceNet model as a module for face recognition, and achieved a good recognition effect, as shown in Figure 4.

ISBN: 978-981-14-2064-1

Fig. 4 The Face recognition effect

Conclusion

How to promote the effective application of artificial intelligence in the field of education At present, the hot issues in the field of education, combined with the development status of deep learning technology, the experimental design of this research has realized the use of face recognition technology for fast and accurate examination and attendance assisted classroom teaching . After the actual application test can effectively save the time of the class name, the attendance can be performed multiple times in the classroom to prevent the students from leaving early. Basically achieved the design goals.

In order to further explore the application of computer vision in education and improve the level of education and teaching intelligence, the teaching aid system can be further developed. The behavior analysis will pay attention to the performance of each classmate in the classroom, and the classroom performance will be included in the comprehensive evaluation.

References

[1] Szegedy C , Liu W , Jia Y , et al. “Going Deeper with Convolutions,” 2014.

[2] Ahonen T , Hadid A , Pietikainen M . “Face Description with Local Binary Patterns: Application to Face Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28(12), pp. 2037-2041, 2006.

[3] Abadi M, Agarwal A, Barham P, et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” arXiv:1603.04467, 2016.

[4] Viola P , Jones M J . “Robust Real-Time Face Detection,” International Journal of Computer Vision, vol. 57(2), pp. 137-154, 2004.

[5] Hubel D H , Wiesel T N . “Receptive fields, binocular interaction and functional architecture in the cat\"s visual cortex,” The Journal of Physiology, vol. 160(1), pp. 106-154, 1962.

[6] Zhang K , Zhang Z , Li Z , et al. “Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks,” IEEE Signal Processing Letters, vol. 23(10), pp.1499-1503, 2016.

[7] Schroff F, Kalenichenko D, Philbin J. “FaceNet: A unified embedding for face recognition and clustering,” 2015.