Research on Classroom Attendance System Based on Face Recognition

Jingyu Pu, Liang Zhang*, Jinqian Zhang, Ziwei Xue, Xixin Zhang, Jie Yang

Department of Information Engineering, Sichuan Agricultural University
201803707@stu.sicau.edu.cn

Abstract. Deep learning face recognition technology has become one of the most popular technologies. The explosive growth in the use of face recognition has brought a variety of practical needs. Face recognition can be applied to identity authentication, bank security, forensic investigation, face scanning payment, access control system and so on. Existing face recognition methods have achieved remarkable results. With the continuous development of deep learning technology, many depth methods show better accuracy than human recognition. Based on the in-depth study and discussion of the deep learning method, this paper applies the theory to practice, and designs and develops a set of face recognition system for classroom attendance.

1. Introduction

In today's university campus, the traditional campus attendance products, mainly represented by clocking in and swiping, have disadvantages such as replacing clock in, low efficiency, difficulty in statistics, and high cost of management, use and maintenance.

Although some fingerprint clock products have been put into use at present, more than 5% of people are born with shallow fingerprints and cannot be identified by fingerprint. At the same time, they have high requirements on the environment and are sensitive to the humidity and cleanliness of fingers. Dirt, oil and water will not be recognized or affect the identification results.

Every use of fingerprints on fingerprint left the user's fingerprint imprint, and these fingerprints are used to replicate the possibility of fingerprints, peeling, have scars, such as low quality fingerprint identification difficulties, the problem of lower recognition rate, for some hand calluses more manual workers such as part of the registration and identification of special populations would be difficult; It is not universal. By analyzing the above characteristics, we design a face recognition attendance system based on biometric technology and combining face recognition technology with campus data system[1].

Based on deep learning, a face recognition system is constructed by combining face detection method (MTCNN)[2], Open CV affine transformation method and Insight face method.

2. Algorithm analysis

2.1. Insight-face

ArcFace/InsightFace was published by Deng Jiankang et al., Imperial College London on January, 2018. Based on SphereFace, it improved the normalization and additive Angle interval of eigenvectors, improved the separability between classes, and strengthened the intra-class tightness and inter-class difference.
In terms of network structure and setting, high-performance convolutional neural networks, such as ResNet and Inception-ResNet, have better performance than VGG and Inception V1. Different applications balance speed and accuracy, and accuracy is more important for mobile devices.

At an algorithmic level, using the latest Insight-face works better than the relatively early Facenet, as demonstrated by the megaface challenge.

Fig 1. The input-output process of ArcFace

2.2. Loss function

Face recognition is divided into four processes: face detection, face alignment, feature extraction and feature matching.

Feature extraction is the most critical step in face recognition, and the extracted features tend to be the "unique" features of the face, which plays an important role in feature matching. Our network and model bear the heavy responsibility of feature extraction, and excellent network and training strategies make the model more robust.

However, to improve the performance of the face recognition model in the case of excellent Resnet network performance, in addition to optimizing the network structure, modifying the loss function[3] is another option. Optimizing the loss function can enable the model to learn more valuable information from the existing data.

Ever contact the classification problem of a large part using the Softmax loss as the loss of the network layer, experiments show that Softmax loss considering whether samples can be classified correctly, and in expanding class and narrow the distance between the similar between heterogeneous samples within the class of distance between sample has a lot of optimization on the question of space, so we in the process of the construction of the entire model, using Arcface as loss function, in the early stage of the data cleaning work at the same time, the above ways to improve the Resnet network structure to make it more suitable for studying the characteristics of the human face.

(1) Softmax loss

\[ L_1 = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{W_{yi}^T x_i + b_j}}{\sum_{j=1}^{n} e^{W_{ji}^T x_i + b_j}} \]  

(2) Normalized version of Softmax Loss (NSL)

The eigenvectors are normalized, and the weights are normalized

\[ L_2 = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s \cos \theta_j}}{\sum_{j=1, j \neq y_i}^{n} e^{s \cos \theta_j} + \sum_{j=1, j \neq y_i}^{n} e^{s \cos \theta_j}} \]  

(3) ArcFace loss

Adding an Angle interval of m between XI and Wji punishes the Angle between the depth characteristics and their corresponding weights in an additive way, thus simultaneously enhancing in-class tightness and inter-class difference.
L2 normalization corrects the single weight $||W_j||=1$, and also fixes the embedded feature $||x_i||$ through L2 normalization, and rescales it to $s$.

$$L_3 = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{e^{s(cos(\theta_i + m))}}{\sum_{j=1, j \neq i}^{n} e^{s\cos\theta_j}}$$  \hspace{1cm} (3)

Angular space features step by step: ArcFace has more compact feature distribution and more obvious decision boundaries than Softmax. One arc length represents one class.

Angular distance has a more direct effect on the Angle than cosine distance.

Fig.2 Angular spatial characteristic distribution

| Loss Functions       | Decision Boundaries                                                                 |
|----------------------|--------------------------------------------------------------------------------------|
| Softmax              | $(W'_1 - W'_2)\ x + b_1 - b_2 = 0$                                                  |
| W-Norm Softmax       | $||x|| (\cos \theta_1 - \cos \theta_2) = 0$                                         |
| SphereFace           | $||x|| (\cos m \theta_1 - \cos \theta_2) = 0$                                       |
| F-Norm SphereFace    | $s(\cos m \theta_1 - \cos \theta_2) = 0$                                           |
| CosineFace           | $s(\cos \theta_1 - m - \cos \theta_2) = 0$                                         |
| ArcFace              | $s(\cos(\theta_1 + m) - \cos \theta_2) = 0$                                        |

3. Image acquisition and experimental results

3.1. Image acquisition and processing

Labeled Faces in the Wild Home (LFW) data set was constructed for the study of face recognition in unlabeled environment. The dataset consists of more than 13,000 images of human faces, collected on the Internet, each of which is standardized to a person's name. Of those, about 1,680 contained more than two faces.

The CPF (KINETIC in Frontal Profile) dataset was 500 celebrity faces, with 10 Frontal images and 4 Frontal images, because the most challenging Frontal and image images were used as validation (CFP,Frontal - Profile, CFP - FP), AgeDB (Age Database) data set contains 440 people face (including 12240 faces), which contains photos of different people at different ages, the average Age for each object scope for 49 years, like LFW verification method, used in the process of training VGGFACE2, UMD_Faces and project team members using high-definition video cameras recorded 16 students, by
loading the face detector, load the video streaming, draw frame. A total of 1672 multi-pose face data were obtained by face recognition as the training set, and LFW, CPF and AgeDB were used as the verification set.

Image preprocessing consists of two parts: face detection and face alignment. Using the MTCNN module in Insight-Face, the intercepted face and key point coordinates are obtained, and according to the key point coordinates of the face, the face is corrected to a standard position by OpenCV affine transformation.

3.2. Experimental results and analysis

In this paper, VGGFACE2, MS1MV2 and UMD_Faces were trained respectively and verified on LFW, CFP-FP and AGEDB. The accuracy of the training set and the test set was shown in Table 1.

| Datasets | Accuracy rate |
|----------|---------------|
| LFW      | 98.20%        |
| CFP-FP   | 95.10%        |
| AGEDB    | 92.91%        |

During the experiment, five facial key points were generated according to MTCNN to generate normalized facial clipping images (112×112). For embedded networks, we use the widely used architecture, ResNet50. The batch size is set to 512. On MS1MV2, the learning rate starts from 0.1 and is divided by 10 at 20K and 28K iterations. The training process was completed at the time of the 32K iteration. We divided the learning rate into 100K and 160k iterations, and completed the iteration at 180K. Due to equipment reasons, a larger scale of training cannot be completed, and the accuracy obtained on each verification set does not fully reproduce the performance of this algorithm on MegeFace. However, it has certain application value in the field of daily campus classroom attendance clock in.

4. System design

This system studies the Insight-face face recognition algorithm and develops the project based on the actual situation. The development environment adopted by this system is CentOS 7.4 / Opencv / Python 2.7 / Mxnet /. The model is applied to the face recognition system by Python code, and the functional modules are integrated by PythonFlaskWeb framework to develop a back-end server for the face recognition system. In the front part, the Vue. Js code is used to write business logic and interface for the front end, and face registration, recognition, detection and other functions are realized.
According to the function, the system is divided into the following five parts:

1. Face acquisition module[4]. Face images can be sourced from local files, networks or video frames. It is mainly transmitted to the server through the form of frame extraction by the camera, or obtained from the video file and the camera by using the OpenCV function.

2. Face correction module. Since the intercepted face image is multi-pose, in order to improve the recognition accuracy, an affine transformation method provided in OpenCV is used to correct the intercepted face to a standard position by using the detected key points of the face.

3. Face detection module, which is used to preprocess the input image, judge the number of faces in the image, detect the location of the face and coordinates of key points of the face, and intercept the face image from the background.

4. Face recognition module. Face recognition module is generally composed of two parts, namely face feature extraction and face index database establishment. After the pre-processed face image is obtained, the face is input into the pre-trained model to extract multidimensional face features and build an index for it.

5. Face attribute module. First, the face image detected is preprocessed to obtain a standard face image[5]. Then, the face image is input into the pre-trained model, and 128 predicted values are finally obtained, respectively representing whether the face image has this attribute.

5. Conclusion
This paper studies the application of face recognition algorithm in college attendance, features generated and used by traditional face recognition algorithms can be considered as shallow features, and cannot get deeper high semantic features and their deep features from the original image. At the same time, the complexity of face information is high and it is not easy to recognize. Multi-face recognition also has interference problems. This paper proposes a class attendance system based on the Insight-face algorithm. It achieves 98.2% accuracy on LFW. In the application, for the special scene of the classroom, the problem of low face recognition rate still needs to be solved. At the same time, there are many places worth further research and discussion in terms of improving detection accuracy and parameter optimization methods.

Acknowledgments
This research was supported by Research Interests Training Program No.2020637 of Sichuan Agricultural University, and Key Laboratory of Agricultural Information of Sichuan Province.

References
[1] Pranav K B,Manikandan J. Design and Evaluation of a Real-Time Face Recognition System using Convolutional Neural Networks[J]. Procedia Computer Science,2020,171.
[2] Ouyang junlin, Huang jingtao, zhang xiao. Face recognition system based on deep learning [J]. Contemporary education theory and practice,2019,11(06):133-138.(in Chinese).
[3] J. Deng, Y. Zhou, and S. Zafeiriou. Marginal loss for deep face recognition. In CVPR Workshop, 2017.
[4] Wu yao, Qiu weigen. Face recognition based on improved deep convolutional neural network [J]. Computer engineering and design,2017,38(08):2246-2250.(in Chinese).
[5] Zhang shunfeng. Study on Face Recognition Algorithm based on Deep Learning [D]. University of Electronic Science and Technology of China,2018.(in Chinese).