Face detection of migrating learning based on constrained scene

Xuewen Ding[1,4,*], *Jing* Chen[2], *Yuting* Lei[3], *Panpan* Sun[1], *Xinnan* Cai[1], and *Limei* Chang[1]

[1]School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, China
[2]Tianjin Yunzhitong Technology Co., Ltd, Tianjin, China
[3]Section 4, Multidisciplinary Digital Publishing Institute, Tianjin, China
[4]Tianjin Enterprise Key Laboratory of High-speed Railway Wireless Communication, Tianjin, China

Abstract. Face detection places an important role in face recognition which is a popular choice for biometric systems. To solve the low face detection rate problem of face detection in constrained scenario, an efficient face detection method based on migration learning was proposed in this paper. In the proposed facial detection approach, data-cleaning was firstly used to optimize the face database. Then the Visual Geometry Group 16 (VGG16) deep learning network was improved to realize migration learning by replacing the softmax regression layer with the multi-scale feature detection layer. Finally, the constrained scene face images for testing were detected and labeled by the trained migration learning model. The WIDER FACE dataset was used for experiments. Experiment results showed that the proposed method can successfully perform face detection in the WIDER FACE dataset and obtain more than 90% detection rate.

1 Introduction

As the basis of face recognition, face detection which has been widely used in production and life plays an important role in face recognition [1-3]. However, the accuracy of face detection systems varies according to the different factors like face size, face direction, face features, and also the condition of lightning, image resolution, Illumination, imaging angle, etc. Moreover, the face detection rate will be different according to wearing of accessories, subject's hair, and expressions within various frontal and profile faces. It is still challenging to overcome the above effects in facial detection. Traditional algorithms for face detection include SVM (support vector machines), PCA (principal component analysis) and linear discriminant methods. The training of SVM method requires complexity. The extremely high quadratic programming problem will result in too much computation and difficulty in training. The PCA method only considers low-order statistical information in image data, and does not consider high-order statistical information, which will cause a lot of useful information to be lost during feature extraction. Linear discriminant method is a more common classification method, but Small sample problems and dimensional problems occur. There are other methods such as SIFT[4] and HOG [5],

[*] Corresponding author: dingxw@tute.edu.cn

etc. Because of low robustness and heavy computation, these methods are inconvenient to use. With the rapid development of deep learning, the face detection by neural network came into being [6-7]. For a security-first enterprise such as a bank, face detection is usually done under a constraint scenario. This paper aims to solve the problem of accurate and reliable face detection in restricted scenes by combining data cleaning with deep learning model.

2 Flow chart of face detection

Face recognition proposed in this paper is mainly composed of data cleaning, migration learning model training and real scene prediction, which is shown in Figure 1.

Fig. 1. Flow chart of face detection.

3 Algorithm design

3.1 Database and data preparation

3.1.1 WIDER FACE dataset

The WIDER FACE dataset [8] which is organized based on 61 event classes has a total of about 30,000 pictures and 400,000 faces. The data set features collection of face photos in real and natural complex scenes, including dense faces, painting faces, testing faces and small size faces.These faces have a wide range of changes in scale, posture, illumination, expression, and occlusion. Among them, some labeled faces are in the training set and some faces are in the validation set. Each subset contains 3 levels of detection difficulty: Easy, Medium, and Hard.

3.1.2 Data-cleaning

Data cleaning will keep the data consistent with the front face constraint scene, and filter out the data of side face, incomplete face and so on. WIDER FACE annotation information contains six classification mode information of the current face: size, posture, occlusion, extreme, painting and illumination. Taking these six kinds of annotation parameters as the screening conditions, the occlusion, extreme, painting and other scene data are removed. The data will become less after data filtering, which is not conducive to model training. Therefore, image expansion technology is used to expand the data set. The commonly used expansion methods such as Image Horizontal flip, zoom or shrink, horizontal offset, vertical offset and rotation by angle are included in data cleaning. In order to conform to the hardware memory of computer in this experiment, the image size is finally trimmed to 256×256. The original annotation information is visualized. The cleaned database is shown in figure 2 and then is packaged into Pascal VOC format as subsequent network input.

(a) Multiple faces (b) Single face 1

(c) Single face 2 (d)Single face 3

Fig. 2. Data set after data-cleaning.

3.2 Migration learning

In order to solve the problem of mass labeling data and long training time, migration learning is used. The excellent model selected in this paper is VGG16[9] . VGG16 which combines the advantages of precision and speed is a fast and efficient end-to-end target detection method. The basic structure of the migration learning model in this paper is shown in Figure 3. It can be divided into three parts. The first part is the general target detection backbone network VGG16 which retaining the original convolution kernel weight, ensuring its good object feature extraction function and removing the softmax classification layer. The second part is to change the removed softmax into multi-scale feature detection layer. The multi-scale feature detection layer is composed of convolution kernels of different sizes. Large convolution kernels (38x38) is used for small targets and small convolution kernels (3x3) is used for large targets. By this method, the feature details is better preserved. The third part is feature fusion which combining the features obtained in the first two parts to prepare reducing the loss function value for the next step.

Loss function is the key to the quality of Migration model learning, which is used to measure the difference between the the effect of model and the ground truth. The smaller the loss function is, the better the model works. Sigmoid is used for the loss function in this paper. The overall loss function is as follows:

$$L(x,c,l,g)=\frac{1}{N}(L_{conf}(x,c)+\partial L_{loc}(x,l,g)) \quad (1)$$

In the formula, L_{conf} represents the object type loss function, L_{loc} represents the regression position loss function, and assigns ∂ weight parameter to adjust the focus in different scenarios; N is the number of default boxes that match the object frame, x is the model prediction type, c is the real type, l is the a priori box, and g is the label box. Through the migration learning training, the loss function is continuously optimized by the momentum_optimizer to reduce the difference between the result and the ground truth.

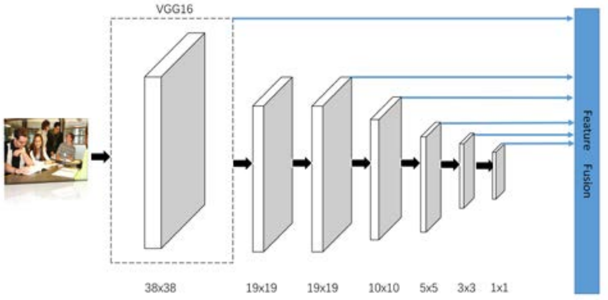

Fig. 3. Migration learning model.

4 Experimental design

Windows 10 system, Tensorflow-gpu-1.12 version and Tensorflow object detection API interface are used for our all experiments by using Intel Core i5 4590 computer with 2GB NVIDIA graphics card and 8GB RAM. Image Pre-processing converted all the images in the database into 256×256. For the training, the learning rate is set to 0.004, the batch size is set to 4, and the training period is 16000.

The test samples by data cleaning from WIDER FACE dataset is used to carry out the experiment and compare our experimental results with others. The corresponding test result overview of our proposed model on WIDER FACE dataset is shown in figure 4. Table 1 shows the appropriateness and comparison of our proposed model results with other models results. As can be discerned from the above table, that was a 90.2% detection rate which is higher than all others. Experiment result show that this method can realize face detection under constraint scenarios.

(a) sample 1 (b) sample 2 (c) sample 3 (d) sample 4

Fig. 4. The experiment result overview on the WIDER FACE dataset.

5 Conclusion

By the effective data cleaning and the migration learning based on deep learning network, the face detection method proposed in this paper solve the low face detection rate problem of face detection in constrained scenario. By data cleaning, the training images of complex scenes are excluded, the images conforming to the scene are screened out and the database is enhanced firstly. Then the training is performed by the deep learning model which is improved for migration learning. Finally, the trained model is used to detect the test face data set. Detection rates as high 90 percent above are obtained on WIDER FACE dataset in the experiment. This proves that our proposed method can be successfully used for detection facial image in the constraint scene, and it is an effective attempt to apply migration learning of deep learning network to detection for human faces.

Table 1. The Comparison of Face detection accuracy with some other methods for the WIDER FACE dataset.

Detection Method	Detection Rate
Two-stage CNN	68.1
Faceness-WIDER	69.1
Multiscale Cascade	71.3
LDCF+	79.0
MTCNN	84.8
ScaleFace	86.8
Our Proposed Method	90.2

This research work is supported by Tianjin Science and technology planning project (No. 20YDTPJC01110) and Tianjin University Science and Technology Development Fund Project (No. 20110710).

References

1. Paul Viola, Michael J. Jones. Robust Real-Time Face Detection. International Journal of Computer Vision, **57**(2):pp.137-154 (2004)
2. E. Osuna, R.M. Freund, F. Girosit. Training support vector machines: an application to face detection. computer vision and pattern recognition, pp.130-136 (1997)
3. Z.Cuiping, S. Guangda. A Survey of Face Recognition Technology. Chinese Journal of Image and Graphics,11:pp.7-16 (2000)
4. D.G.Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, **60**(2):pp.91-110 (2004)
5. N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. computer vision and pattern recognition, pp.886-893 (2005)
6. C. Cortes, V. Vapnik. Support Vector Network. Machine Learning, **20**(3):pp.273-297. (1995)
7. Z. Zhu, P. Luo, X. Wang. Deep Learning Identity-Preserving Face Space. international conference on computer vision, pp.113-120 (2013)
8. S. Yang, P. Luo, C.C. Loy. WIDER FACE: A Face Detection Benchmark. computer vision and pattern recognition, pp. 5525-5533 (2016)
9. W. Liu, D. Anguelov, D.Erhan. SSD: Single Shot MultiBox Detector. european conference on computer vision, pp.21-37 (2016)