2021 IEEE 10th Data Driven Control and Learning Systems Conference
May 14-16, 2021, Suzhou, China

Face Verification Technology Based on FaceNet Similarity Recognition Network

Fengwei Gu[1,2], Jun Lu[1,2*], Guihua Xia[1,2], Zhiguang Feng[1]

1. College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001,China

2. Key laboratory of Intelligent Technology and Application of Marine Equipment(Harbin Engineering University), Ministry of Education, Harbin, 150001, China
E-mail: lujun0260@sina.com

Abstract: The background complexity of face images is high in actual scenes, and there are a series of problems such as illumination and occlusion, which greatly reduces the performance of the face verification model. This paper proposes a face verification algorithm FaceNetSRM based on the FaceNet similarity recognition network to improve the performance of the face verification model and the accuracy of Chinese face verification. Firstly, the deep convolutional neural network framework in FaceNet is determined, and the similarity recognition module is used to replace the Euclidean distance module in FaceNet. Then, the CASIA-WebFace face dataset and the self-made face dataset C-facev1 are used to train the face verification algorithm of this article. Finally, the trained model is tested and evaluated on the face dataset LFW and CASIA-FaceV5 to show the effectiveness of the face verification method in this article, and the face verification effect of the algorithm is compared with the face verification effect of FaceNet. The experimental results show that the face verification accuracy rate of the FaceNetSRM algorithm in this paper is 1.5% higher than that of FaceNet, and the accuracy rate of Chinese face verification is improved by 2.8%. The algorithm has good robustness and generalization ability, which can be applied in face verification systems.

Key Words: Face Verification, FaceNet, Similarity Recognition Module, Face Dataset

2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS) | 978-1-6654-2423-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/DDCLS52934.2021.9455715

1 Introduction

Face verification technology is a research hotspot in the field of computer vision, which is widely used in the fields of identity verification, financial payment and human-computer interaction. The general process of face verification is as follows: Firstly, the target face in the image is obtained through face detection technology. Then, the face recognition model is used to extract facial features. Finally, the obtained facial features are measured to complete face verification[1].

The existing face recognition methods can be divided into traditional methods and methods based on depth features. Traditional face recognition methods mainly rely on the application of artificial design features and machine learning technology. The representative methods include Histogram of Oriented Gradient (HOG)[4], Principal Component Analysis (PCA)[5], Support Vector Machine (SVM)[6] and Linear Discriminant Analysis (LDA)[7] and so on.

The traditional face recognition method based on artificial design features and machine learning technology can obtain rich face information in the image, which performs well in a specified environment, but it is difficult to ensure the robustness and generalization ability of face detection and recognition in an unconstrained environment. With the rapid development of deep neural networks, traditional face recognition methods have been gradually replaced by deep convolutional neural networks(DCNN)[8] in recent years. DCNN has powerful characterization capabilities to obtain the depth features of images, and is widely used in many fields such as target detection, tracking and image generation. Therefore, DCNN is gradually applied in the field of face detection and recognition.

CNN was introduced to the face recognition task for the first time by DeepFace[9],First of all, DeepFace performed face detection by means of reference points. Then, 3D modeling was performed on the face in the image, and the face alignment was completed. Finally, multiple convolutional layers were used to extract the depth features of the face for classification to complete face recognition. N. P. Ramaiah et al.[10] used deep convolutional neural networks to solve the problem of face recognition under non-uniform lighting conditions and achieved good results. Yang et al.[11] used convolutional neural networks to solve the problem of inaccurate target position in face recognition, and the accuracy rate was significantly improved. Song AP et al. [12] proposed an attention mechanism on convolutional neural networks, which solved the problem of face similarity in face recognition and achieved good experimental results.

The above-mentioned face recognition methods based on depth features have been significantly improved in accuracy, but there are various interference factors such as illumination and occlusion in the actual scene, which increase the difficulty of face detection and recognition. This paper proposes a face verification algorithm FaceNetSRM based on the FaceNet [13] similarity recognition network to ensure the robustness and generalization ability of the face verification model and improve the accuracy of Chinese face verification.

*This work is supported by Natural Science Foundation of Heilongjiang Province of China under Grant No.F201123

Authorized licensed use limited to: Sri Sivasubramanya Nadar College of Engineering. Downloaded on October 25,2023 at 16:05:09 UTC from IEEE Xplore. Restrictions apply.

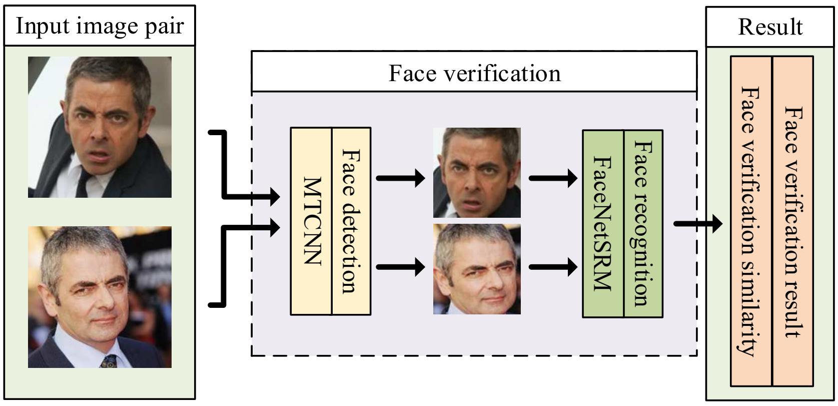

Fig. 1: System flow chart of face verification

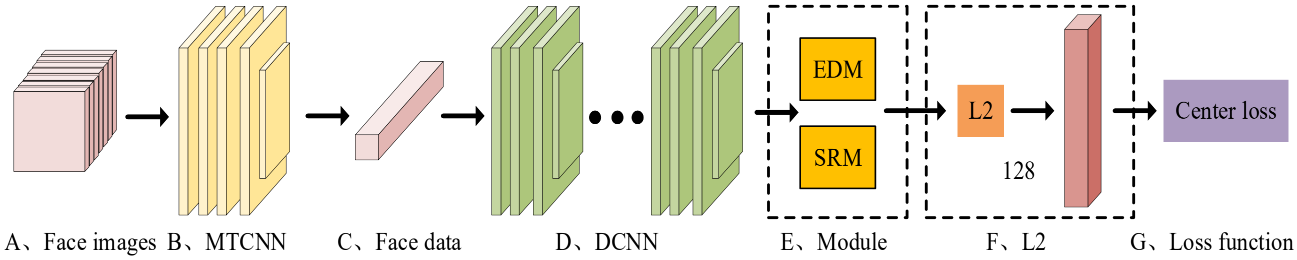

Fig. 2: Network framework of FaceNetSRM

The algorithm uses multi-task cascaded convolutional neural network to complete face detection, and uses similarity recognition module to replace the Euclidean distance module in FaceNet to realize the recognition of high-dimensional face feature vectors, thereby improving the accuracy of face verification. The CASIA-WebFace face data set and the self-made C-facev1 data set are used to train the face verification algorithm in this paper to improve the robustness and generalization ability of the model. To verify the effectiveness of this method, the face dataset LFW and CASIA-FaceV5 [14] are used to test and evaluate the proposed FaceNetSRM, and the algorithm is compared and analyzed with FaceNet.

2 Face Verification Model

The main purpose of face verification is to determine whether two face images are from the same person. This paper mainly uses multi-task cascaded convolutional neural network to detect face information in images, and uses FaceNetSRM to obtain facial features and complete feature recognition to achieve face verification. The flow chart of the face verification system is shown in Figure 1. The flow is as follows: Firstly, the face image is input into the face verification system. Then, MTCNN [15] is used for face detection. Finally, FaceNetSRM is used to extract the depth feature vector of the face, and the similarity recognition is completed to show face verification result.

2.1 Algorithm Framework

Based on the FaceNet network framework, this paper inserts the similarity recognition module into the FaceNet network framework, and replaces the Euclidean distance module with the similarity recognition module to obtain the FaceNetSRM face verification algorithm. The network framework of FaceNetSRM is shown in Figure 2. A is face image. B is MTCNN. C is face data. D is deep convolutional neural network. E is metric module (Euclidean distance module and similarity recognition module). F is the normalization process, and G is the loss function. When the metric module is the Euclidean distance module, FaceNetSRM degenerates to FaceNet. The greater the distance obtained by FaceNet via the Euclidean distance module, the less likely the two face images are from the same person. However, the greater the similarity value obtained from the similarity recognition module in FaceNetSRM, the greater the possibility that the two face images are from the same person. The verification results of the two show a negative correlation trend.

FaceNetSRM takes the output of MTCNN as the network input. The face detection results of MTCNN are mapped to a high-dimensional feature space using a deep convolutional neural network, and high-dimensional face feature vector information is extracted. After L2 normalization, a 128-dimensional feature vector is obtained. The central loss function is used to optimize the network model. The similarity recognition of high-dimensional face vectors is performed through the similarity recognition module, which completes the task of face verification.

The deep convolutional neural network is the core structure in the network framework of FaceNetSRM. The main function is to map the face image to the high-dimensional feature space to obtain the depth features

Authorized licensed use limited to: Sri Sivasubramanya Nadar College of Engineering. Downloaded on October 25,2023 at 16:05:09 UTC from IEEE Xplore. Restrictions apply.

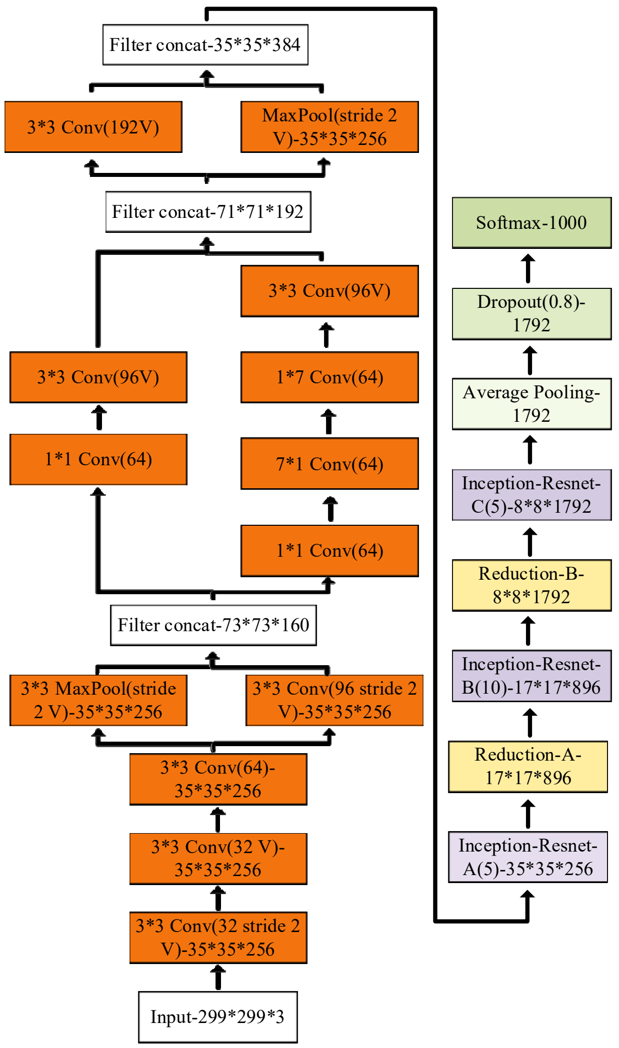

Fig. 3: Network structure of Inception-ResNet-v2

of the face image, and use the high-dimensional depth features to realize the face verification task . In the process of extracting face features, there are disturbances from various environmental factors such as illumination and occlusion, which requires deep convolutional neural networks to have sufficient feature extraction capabilities. Therefore, this article chooses Inception-ResNet-v2 [16] deep convolutional neural network as the backbone network of FaceNetSRM. The network structure of Inception-ResNet-v2 is shown in Figure 3. Inception-ResNet-v2 combines the inception network and residual ideas, which can effectively extract the depth features of the face while ensuring the feasibility of model training.

2.2 Loss Function of FaceNetSRM

In the training process of the FaceNetSRM network model, this paper uses a combination of softmax multi-class loss [17] and center loss [18] to optimize the network model. Using the combined loss function method can not only solve the problem of poor accuracy of the single loss training model, but also improve the training effect of the overall model.The loss function of FaceNetSRM is shown in formula (1):

$$L_{FaceNetSRM} = \lambda L_{center} + L_{soft\max} \qquad (1)$$

Where λ is the scale factor. L_{center} is the center loss, and L_{softmax} is the softmax loss.

The center loss and softmax loss are shown in equations (2) and (3) respectively:

$$L_{center} = \frac{1}{2}\sum_{i=1}^{N} \left\| g(x_i) - c_{\mathrm{ji}} \right\|_2^2 \qquad (2)$$

$$L_{soft\max} = -\sum_{i=1}^{\mathrm{n}} \log \frac{e^{w_{ji}^T g(x_i)+b_{ji}}}{\sum_{j=1}^{n} e^{w_k^T g(x_i)+b_k}} \qquad (3)$$

Where N is the number of face samples. x_i is the ith face sample. $g(x_i)$ is the corresponding face feature. y_i is the face sample category. c_{yi} is the category feature center. n is the number of face sample categories.w_{ji} is the weight of the jth category. b_{ji} is the corresponding offset. w_k is the kth weight, and b_k is the corresponding offset.

2.3 Similarity Recognition Module

After FaceNetSRM extracts the high-dimensional feature vector of the face, it needs to recognize the similarity of the vector. In this paper, SRM is used to replace the Euclidean distance module in FaceNet to complete the similarity recognition of vectors, so as to realize face verification. Now we assume that there are two k-dimensional face feature vectors M and N. At this time $M=[M_1, M_2,..., M_k]^{\mathrm{T}}$, $N=[N_1, N_2,..., N_k]^{\mathrm{T}}$, the similarity calculation method of SRM is shown in formula (4):

$$\cos\langle M, N\rangle_{SRM} = \frac{\sum_{i=1}^{m} M_i \times N_i}{\sqrt{\sum_{i=1}^{m}(M_i)^2} \times \sqrt{\sum_{i=1}^{m}(N_i)^2}} \qquad (4)$$

In this way, the similarity $\cos\langle M, N\rangle_{SRM}$ between the face feature vectors M and N is obtained. When $\cos\langle M, N\rangle_{SRM}$ is greater than the threshold, the face verification result determines that the face image is from the same person, otherwise the determination result is opposite.

3 Experiments

The algorithm in this paper is tested under the framework of Windows10 and tensorflow2.0, by using GPU

Authorized licensed use limited to: Sri Sivasubramanya Nadar College of Engineering. Downloaded on October 25,2023 at 16:05:09 UTC from IEEE Xplore. Restrictions apply.

Fig. 4: Part of the face image of C-facev1

Fig. 5: Comparison of face verification in F1 group

acceleration. The experimental environment is as follows. The CPU configuration is Intel Core i7-10700k 3.8GHz 16GB, and the GPU configuration is Nvidia RTX 2080Ti GPU 11GB. The Adam optimizer is used to optimize the network model in the process of model training, and the learning rate can be dynamically adjusted to reduce loss. To improve the robustness and generalization ability of the algorithm, the self-made C-facev1 face dataset is incorporated into the CASIA-WebFace face dataset to train the network model. The method of dataset combination can improve the accuracy of face verification to a certain extent.

In the test phase, the face dataset LFW and CASIA-FaceV5 are used to verify and evaluate the face verification algorithm FaceNetSRM in this paper. The dataset C-facev1 and CASIA-FaceV5 only contain Chinese face images, and the part of the C-facev1 face dataset is shown in Figure 4.

3.1 Qualitative analysis

In order to demonstrate the effectiveness of the face verification algorithm in this paper, 10 groups of face images with different states or backgrounds were selected from LFW and CASIA-FaceV5 respectively in the experiment to perform qualitative analysis. These images are divided into two groups, F1 and F2. And the image pairs in each group are sequentially numbered NO.01-NO.10, and the number in the label indicates the similarity under the corresponding algorithm. "Y/N" means that the verification result is correct/incorrect.The verification results of F1 and F2 image pairs are shown in Figure 5 and Figure 6, respectively. It can be seen from Figure 5 that the image pairs in the F1 group mainly have some problems such as face pose changes, expression changes, occlusion, and light interference. Among them, NO.01, NO.06, and NO.07 mainly have face posture changes. NO.02, NO.05, NO.07, and NO.08 mainly have occlusion problems. NO.09 and NO.10 mainly have light interference. The image pairs of NO.04, NO.06, and NO.09 mainly have the problems of exaggerated expression changes and illumination. These reasons will increase the difficulty of face verification. However, the verification results of these images under FaceNetSRM are correct, but some of the verification results under FaceNet are wrong, as shown by the verification results of NO.04, NO.06, and NO.09 images. The above verification results show that the FaceNetSRM algorithm has better robustness and generalization ability, as well as better verification accuracy.

The image pairs in the F2 group mainly have lots of problems such as illumination, face pose and occlusion. No. 01, No. 06, No. 07, No. 08, No. 10 image pairs have face pose changes. No.02, No.03, No.04, No.05, No.07 image pairs have lighting problems. NO.01, NO.04, and NO.06

Authorized licensed use limited to: Sri Sivasubramanya Nadar College of Engineering. Downloaded on October 25,2023 at 16:05:09 UTC from IEEE Xplore. Restrictions apply.

Table 1: Test accuracy

Method / Dataset		FaceNet（%）	FaceNetSRM（%）
LFW	D1	97.56	98.12
	D2	98.42	98.86
CASIA-FaceV5	D1	96.16	96.60
	D2	97.84	98.35

Fig. 6: Comparison of face verification in F2 group

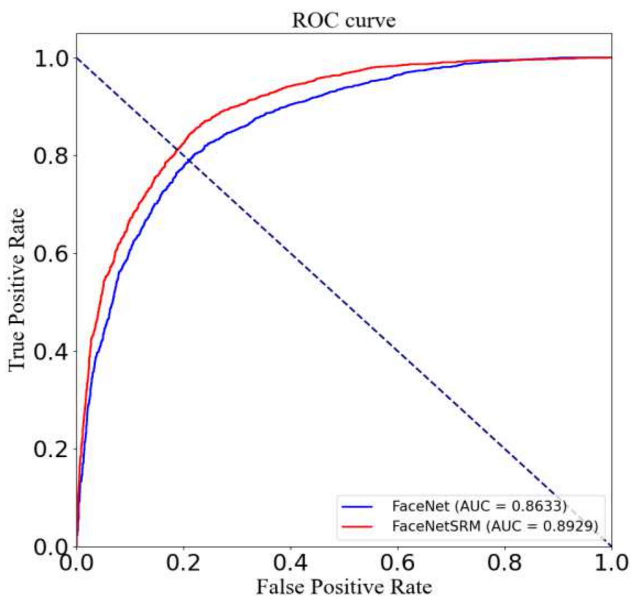

Fig. 7: Comparison of ROC curves

image pairs have occlusion and illumination problems caused by glasses. All images are correct in the FaceNetSRM algorithm verification results, only NO.04 is verified incorrectly under the FaceNet algorithm, which is caused by the inability to effectively extract facial features due to the influence of illumination on the NO.04 image. Although the two algorithms performed well for the F2 group, the FaceNetSRM algorithm performed more prominently.

3.2 Quantitative analysis

The face verification algorithm in this paper is tested on the LFW and CASIA-FaceV5 datasets as a whole, and the CASIA-WebFace dataset and the combination of CASIA-WebFace and C-facev1 datasets are used to train the algorithm in this paper. The CASIA-WebFace dataset is now referred to as D1, and the combined dataset of CASIA-WebFace and C-facev1 is referred to as D2. The test accuracy rate is shown in Table 1. It can be seen from Table 1 that the verification accuracy rate of FaceNetSRM is better than that of FaceNet on the whole, and the test accuracy rate of the model obtained in the D2 dataset is higher than that of the D1 dataset. Specifically, the verification accuracy of FaceNetSRM is improved by 1.30% on the LFW dataset and 2.19% on the CASIA-FaceV5 dataset.

Therefore, the FaceNetSRM algorithm not only improves the overall face verification accuracy rate, but also improves the accuracy rate of Chinese face verification by more than two percentage points. The verification accuracy rate of the FaceNetSRM algorithm is significantly improved.

Authorized licensed use limited to: Sri Sivasubramanya Nadar College of Engineering. Downloaded on October 25,2023 at 16:05:09 UTC from IEEE Xplore. Restrictions apply.

This paper also draws the ROC [19] curve as shown in Figure 7 to evaluate the overall performance of the model. It can be seen from Figure 7 that the abscissa of the ROC curve is False Positive Rate, and the ordinate of the ROC curve is True Positive Rate. The ordinate value of the ROC curve becomes larger as the abscissa value increases. We hope that the True Positive Rate can be as large as possible while the False Positive Rate is as small as possible. From the change trend of the ROC curve of FaceNetSRM and FaceNet in Fig. 7, we hope that the intersection of the ROC curve and y = -x+1 is as close as possible to the point (0,1). In other words, the larger the area under the ROC curve is, the better the performance of the model corresponding to the ROC curve is. As shown in Figure 7, the area under the ROC curve corresponding to the FaceNetSRM algorithm is $AUC_{FaceNetSRM}$=0.8929, and the area under the ROC curve corresponding to the FaceNet algorithm is $AUC_{FaceNet}$=0.8633. Obviously the value of $AUC_{FaceNetSRM}$ is larger, so the model performance of the FaceNetSRM algorithm is better than that of FaceNet.

4 Conclusions

Aiming at the problems of image complexity, illumination and occlusion in actual scenes, in order to improve the accuracy of Chinese face verification, this paper proposes a face verification algorithm FaceNetSRM based on the FaceNet similarity recognition network. The algorithm uses the similarity recognition module to replace the Euclidean distance module in FaceNet to realize the recognition of face feature vectors. CASIA-WebFace and self-made C-facev1 are used to train the face verification algorithm of this article together, and the trained model is evaluated on the LFW and CASIA-FaceV5. The experimental results show that FaceNetSRM has good robustness and generalization ability through the comparative analysis of FaceNetSRM and FaceNet, and the algorithm significantly improves the accuracy of Chinese face verification, which can be applied to actual face verification systems.

Future work may include: 1) The network structure is studied to further improve the feature extraction ability; 2) The structure of the data set is improved to make the data more comprehensive; 3) The model is pruned under the condition of ensuring the accuracy.

References

[1] Wang F, Cheng J, Liu W, et al. Additive margin softmax for face verification[J]. IEEE Signal Processing Letters, 2018, 25(7): 926-930.

[2] Sajjad M, Nasir M, Muhammad K, et al. Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities[J]. Future Generation Computer Systems, 2020, 108: 995-1007.

[3] Cai W, Ma B, Zhang L, et al. A pointer meter recognition method based on virtual sample generation technology[J]. Measurement, 2020, 163: 107962.

[4] Nigam S, Singh R, Misra A K. Efficient facial expression recognition using histogram of oriented gradients in wavelet domain[J]. Multimedia tools and applications, 2018, 77(21): 28725-28747.

[5] Vyanza V E, Setianingsih C, Irawan B. Design of smart door system for live face recognition based on image processing using principal component analysis and template matching correlation methods[C]//2017 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob). IEEE, 2017: 23-29.

[6] Zhu Y, Gao W, Guo Z, et al. Liver tissue classification of en face images by fractal dimension-based support vector machine[J]. Journal of biophotonics, 2020, 13(4): e201960154.

[7] Ouyang A, Liu Y, Pei S, et al. A hybrid improved kernel LDA and PNN algorithm for efficient face recognition[J]. Neurocomputing, 2020, 393: 214-222.

[8] Zangeneh E, Rahmati M, Mohsenzadeh Y. Low resolution face recognition using a two-branch deep convolutional neural network architecture[J]. Expert Systems with Applications, 2020, 139: 112854.

[9] Taigman Y, Yang M, Ranzato M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification[C]// Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014.

[10] N. P. Ramaiah, E. P. Ijjina and C. K. Mohan, "Illumination invariant face recognition using convolutional neural networks," 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Kozhikode, India, 2015, pp. 1-4, doi: 10.1109/SPICES.2015.7091490.

[11] Yang Y X, Wen C, Xie K, et al. Face recognition using the SR-CNN model[J]. Sensors, 2018, 18(12): 4237.

[12] Song A P, Hu Q, Ding X H, et al. Similar face recognition using the IE-CNN model[J]. IEEE Access, 2020, 8: 45244-45253.

[13] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015.

[14] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, Debin Zhao. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Trans. on System Man, and Cybernetics (Part A), vol.38, no.1, pp149-161. 2008.1

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

[16] Thomas A, Harikrishnan P M, Palanisamy P, et al. Moving vehicle candidate recognition and classification using inception-resnet-v2[C]//2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, 2020: 467-472.

[17] Li X, Wang W. Learning discriminative features via weights-biased softmax loss[J]. Pattern Recognition, 2020, 107: 107405.

[18] Farzaneh A H, Qi X. Discriminant Distribution-Agnostic Loss for Facial Expression Recognition in the Wild[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 406-407.

[19] Albiero V, Bowyer K, Vangara K, et al. Does face recognition accuracy get better with age? deep face matchers say no[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020: 261-269.

Authorized licensed use limited to: Sri Sivasubramanya Nadar College of Engineering. Downloaded on October 25,2023 at 16:05:09 UTC from IEEE Xplore. Restrictions apply.