An Asian Face Dataset and How Race Influences Face Recognition

Zhangyang Xiong[1(✉)], Zhongyuan Wang[1], Changqing Du[2],
Rong Zhu[1], Jing Xiao[1], and Tao Lu[3]

[1] NERCMS, School of Computer, Wuhan University, Wuhan 430072, China
xiong_zhangyang@qq.com
[2] School of Information Engineering, Qujing Normal University,
Qujing 655000, China
[3] Hubei Key Laboratory of Intelligent Robot,
School of Computer Science and Engineering, Wuhan Institute of Technology,
Wuhan 430073, China

Abstract. The face recognition scheme based on deep learning can give the best face recognition performance at present, but this scheme requires a large amount of labeled face data. The currently available large-scale face datasets are mainly Westerners, only containing few Asians. In practice, we have found that models trained using these data sets are lower in accuracy in identifying Asians than Westerners. Therefore, the establishment of a large-scale Asian face dataset is of great value for the development and deployment of face related applications for Asians. In this paper, we propose a simple semi-automatic approach to collect face images from Internet and build a large-scale Asian face dataset (AFD) containing 2019 subjects and 360,000 images. To the best of our knowledge, this is the largest Asian face image dataset proposed so far. To illustrate the quality of AFD, we train 3 different models with the same CNN structure yet by different training datasets (AFD, WebFace, mixed WebFace&AFD) and verify them on one Western and two Asian face testing datasets. Extensive experimental results show that the model by our AFD outperforms counterparts by a large margin for Asian face recognition. We have made the AFD dataset public to facilitate face recognition development for Asians.

Keywords: CNN · Face recognition · Asian face dataset · Race

1 Introduction

In recent years, face recognition has achieved great progress with the help of deep convolutional neural network (CNN). Usually, there are two essential issues in CNN based tasks, such as the structure of CNN and the training dataset. Large face datasets are important for advancing face recognition research, but most public face datasets mainly consist of Western face images, with only a few Asian face images included. For Asian face recognition, the deep learning models trained by these datasets do not provide satisfactory recognition accuracy comparable to that of Westerners.

R. Hong et al. (Eds.): PCM 2018, LNCS 11165, pp. 372–383, 2018.
https://doi.org/10.1007/978-3-030-00767-6_35

The construction of face datasets is tedious because a lot of work must be done to clear large amounts of raw data. To facilitate this task, we have developed a semi-automatic method for constructing a face dataset that detects faces in images returned by public persons retrieved on Internet, and then automatically discards those not belonging to each queried person. We create a collection name indexes and then gather photos from Web based on the collection. Considering that a single photo may contain multiple persons or not only facial images, we detect the individual's face based on the MTCNN method [7]. Finally, because of incorrect labeling or face detection errors, some candidate faces do not match with the actual individuals. We therefore use the Google Face-net CNN model [11] for face matching to eliminate false candidates. Following this way, we offer a face image dataset specific for Asian, including 2019 individuals and total about 360,000 images, called AFD. AFD is currently the largest public Asian face image dataset as we know. This dataset can be applied to training face recognition CNN or other purposes, which is not only used for academic research but also for real applications. AFD can be downloaded at https://github.com/X-zhangyang/AFD-dataset.

To illustrate the quality of AFD, we use different datasets (merely Western faces, merely Asian faces and mixture) to train same model, and then compare model's performance on three different testing datasets (one is Western and the other are Asian). This way, we exploit how training datasets of different races influence experimental results in face recognition. Particularly, we intend to prove that the model trained by Asian faces is able to give better recognition performance on Asian faces than the model trained by Western faces.

The major contributions are highlighted as follows.

(1) Based on the massive Internet photos, we propose a universal semi-automatic method to build new face image datasets with high classification rate.

(2) We create the largest Asian face dataset so far, containing 360,000 face images by 2019 individuals. In contrast, the second largest Asian face dataset CASIA-FaceV5 merely includes 2500 images by 500 individuals.

(3) We experimentally prove that different races influence face recognition performance in terms of the consistency of training and testing datasets. Particularly, our built Asian face dataset largely outperforms Western face datasets for Asian face recognition.

2 Related Work

In face recognition, training data and algorithm are two significant issues. Nowadays, CNN as well as its variants are mainstream algorithms, the structures of which have been becoming deeper and wider. Using deeper and wider structure means we need a larger scale of training dataset. Otherwise, if training data is insufficient, the CNN will become overfit and produces poor recognition performance. Many face data sets have been used in industry and academia which consist of Western face images, have been made public. In this section, we present some popular face image datasets.

Labeled Face in the Wild (LFW) [1]. A dataset of face photographs is designed for studying the problem of unconstrained face recognition. The dataset contains more than 13,000 human facial images collected from Web. Each face has been labeled with the name of the person picture. Among them 1680 faces have two or more distinct photos in the dataset. The only constraint on these faces is that they were detected by the Viola–Jones [13] face detector. More details can be found in website. LFW dataset was very popular in past few years, as the developed algorithm based on it mostly can achieve up to 99% [14]. It also describes a pipeline to build a face image dataset.

FaceScrub [2]. The FaceScrub dataset was created from Internet, followed by manually checking and cleaning the results. It comprises a total of 106,863 face images of male and female 530 celebrities, with about 200 images per person. As such, it is one of the largest public face databases.

CASIA-WebFace [3]. The CASIA-WebFace is one of largest face image datasets, almost from Western race. It includes about 1000 subjects and 494,414 face images. This dataset is widely used in face recognition, especially used to train CNNs. Nevertheless, not all face images are detected and annotated correctly in this dataset. There are no overlapped images between this dataset and LFW. Again, this dataset mainly consists of Western face images.

MegaFace [4]. The MegaFace dataset is the largest publicly available facial recognition dataset with a million faces and their respective bounding boxes. All images were obtained from Flickr (Yahoo's dataset) and licensed under creative commons. MegaFace has become most popular in face recognition.

YouTube Faces [5]. A database of face videos is designed for studying the problem of unconstrained face recognition in videos. The dataset contains 3,425 videos of 1,595 different people. All the videos are downloaded from YouTube. An average of 2.15 video clips are available for each subject, and the average length of a video clip is in 181.3 frames.

FaceDB [6]. FaceDB offers an anonymous "face-to-face" authentication service with high accuracy liveness detection. It supports multimodal biometrics software completely in-house that can be easily integrated in any applications.

CASIA-FaceV5 [15]. This dataset contains 2,500 color facial images of 500 subjects. All images are Asian human face. The number of images is too small to train a CNN because the CNN will become overfitted.

AFD. Our created dataset AFD contains 360,000 color face images from 2019 different individuals. On average, each subject takes about 178 facial images in diversified poses, expressions, deformations, and even lighting conditions. Therefore, this dataset is particularly suitable for training face applications based on deep learning, given that it is not only large-scale, but also rich in poses and lighting. All face images are gathered from Web by Asian names and are also discarded from the wrong labeled ones. To the best of our knowledge, AFD is the largest Asian face dataset proposed so far.

Table 1 tabulates some typical face datasets, where most datasets mainly contain Western human facial images. In order to find out how different races influence face recognition, we manage to build an Asian dataset, termed Asian Face Dataset (AFD). More details will be shown in the next section.

Table 1. Popular facial image datasets.

Datasets	Subjects	Total images	Races
LFW [1]	5,749	13,233	Mainly Western
FaceScrub [2]	530	107,818	Mainly Western
CASIA-WebFace [3]	10,575	494,414	Mainly Western
MegaFace [4]	4,030	4,400,000	Mainly Western
YouTube Faces [5]	1,595	Videos	Mainly Western
FaceDB [6]	23	1521	Mainly Western
CASIA-FaceV5 [15]	500	2500	Only Asian
AFD	2,019	360,000	Only Asian

3 Building AFD Dataset and Training CNN

The previous methods to build a dataset are to collect images from Internet and then tag each image. This paper proposes a new pipeline GDC (Gather-Detect-Classify) to build face image datasets. As most human images are already labeled when uploaded to Internet (usually are tagged with person name, although some images may be mismatched to their labels), we firstly create a name-index set before collecting images and then collect subject images according to the set.

Since the single photo may contain multiple faces and furthermore the faces are not separated from the entire person's image, we need to use face detection method to isolate individual's face. In addition, photos may be incorrectly associated with an individual due to incorrect annotations when people upload photos. For this reason, face recognition techniques need to be used to eliminate false match by classifying matched and mismatched candidates. Moreover, we train face recognition models based on the established dataset to verify its validity.

3.1 GDC Pipeline

A. Gather Raw Images

The main difference of our proposed GDC pipeline from DAR (Detection-Alignment-Recognition) [1] pipeline is that we build a name-index set which is used to collect intended images before gathering raw images. In practice, we create a large name-index set, covering about 2500 individuals. Given that movie star photos are easier to collect from Web, our name-index set contains a large number of Asian stars. According to the constructed name-index set, we search for their photo images one by one, resulting in a raw picture collection about 630,000 images. All images are saved in JEPG format with the largest size being 500 × 500 pixels.

B. Detect Face Images

Although 630,000 pictures were collected from the Internet, what we really care about is facial images. Note that there may appear multiple persons in one picture, or the image may contain entire person's photo rather than face. We therefore use MTCNN [7] to detect faces in each image so as to separate individual's face image. Since MTCNN cannot detect faces with almost 100% accuracy, and some raw images have multiple faces, we end up with more than 910,000 face images totally, taking a variety of mismatched faces with the specific subject. For the total of 2019 individuals, each one takes about 450 faces, including diverse facial poses and expressions as well as mislabeled and morphed faces. After detecting the faces, we crop them horizontally and vertically to 250 × 250 pixels about the center to exclude the noisy background. The ultimate collection can be treated as the preliminary version of AFD dataset.

C. Classify Labeled Face Images

Specifically, we consider collections of face images obtained by running a face detector on images returned from a search engine by person's name. Within each collection for a specific person, some are false positives (e.g., non-faces) found by the imperfect face detector, and a number of them belong to other people appearing in the same image as the queried person or people in images irrelevant to the query. We refer the all faces acquired for a person as candidates, and those mismatched faces as outliers. Our goal is to remove the outliers among the detected candidates for each queried person, so that we obtain faces belonging just to him/her and a cleaned dataset overall. As CNN-based face recognition has given promising results, we use Google Face-net CNN model [2, 11] to select the right face images for each person.

3.2 Discarding Mismatched Candidates

In order to determine whether the collected and detected faces really belong to a person, we need a face image as a benchmark. We manually select a right labeled and clear frontal facial image as the benchmark. By comparing the candidate faces with the benchmark, we decide whether they belong to the same person. To ensure metric accuracy, this similarity comparison is performed in the feature domain instead of the pixel domain. Specifically, we firstly input the benchmark and all candidate face images of the same individual into a trained CNN model to obtain a feature vector for each image, and then compute the Euclidean distance between the benchmark and each candidate.

We observe that images of the same person usually enjoy smaller Euclidean distances, while images belonging to different individuals have larger Euclidean distances. When a Euclidean distance equals to 0, it means these two images are exactly same. Therefore, we need to set a threshold to determine whether to dismiss the candidate face image or not. Since a larger threshold retains more candidates but meanwhile mistakes a certain one, the appropriate threshold directly affects the availability of the dataset.

We sort the distance vector from small to large and then depict it in Fig. 1. We find that it has a steep front and back end and an approximately linear distribution in the middle. Obviously, the element at the steep front should be accepted, and the element at the steep end should be rejected instead. The key is how to determine the attribution of the middle element.

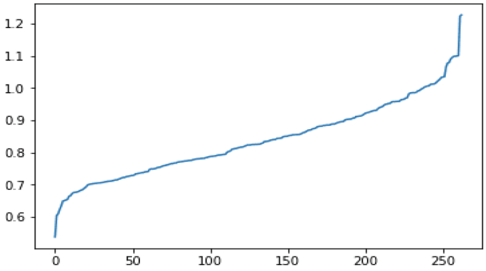

Fig. 1. The distribution of the sorted Euclidean distances by index vs. distance value.

We examine the mean distance of each person, by averaging all the Euclidean distances with

$$D_{mean} = \frac{\sum_{i=0}^{n} D_i}{n} \tag{1}$$

Where D_i represents the Euclidean distance between the i-th candidate face and the benchmark face.

Ranging all D_i from small to large, we find that they are unevenly distributed. For most individuals, their candidate faces are correctly associated. Consequently, most distance values are statistically small and are seldom larger than D_{mean}. In this case, the average distance can be used as a suitable classification criterion. More specifically, the distances less than D_{mean} usually give a correct classification, while those larger than D_{mean} instead give wrong classification.

However, for some individuals, most of the candidates do not match their labels. In this case, D_{mean} is mostly close to the very end D_i. Due to the large number of outliers, the average distance is also very large, so that it is obviously not appropriate to use only the average as the threshold. Even if the distance is less than D_{mean}, the correct classification rate is still very low. Therefore, we need to decrease the threshold against D_{mean} to improve the correct classification rate. In order to eliminate the interference of many outliers, we consider a parameter that is not much affected by high proportions and large values of outliers. Median filtering is a non-linear signal processing technique that can effectively suppress noise and outliers based on sorting statistics theory. We therefore use the median filtering of the distance vector to produce a more robust parameter D_s. In general, for a vector containing many large outliers, its median filtering result is smaller than its average. But for insurance, we still examine the minimum of the two parameters. We choose the minimum one between D_{mean} and D_s as the optimal threshold using

$$D_f = \min(D_{mean}, D_s) \tag{2}$$

For a variety of candidate faces for a particular individual, we exclude candidate faces whose distance values are greater than the threshold, leaving only candidates that are less than the threshold. After dismissing mislabeled candidates, AFD is reduced to 360,000 images from the preliminary 910,000 images, with 178 faces for an individual on average. Subject to the classification accuracy, AFD still takes a very small number of mismatched face images. Actually, the existence of few mismatched images is hopefully to improve the robustness of the model training against data errors and noise. As a concrete example, Fig. 2 shows facial images for 6 individuals under varied poses and expressions, where each row corresponds to the same person.

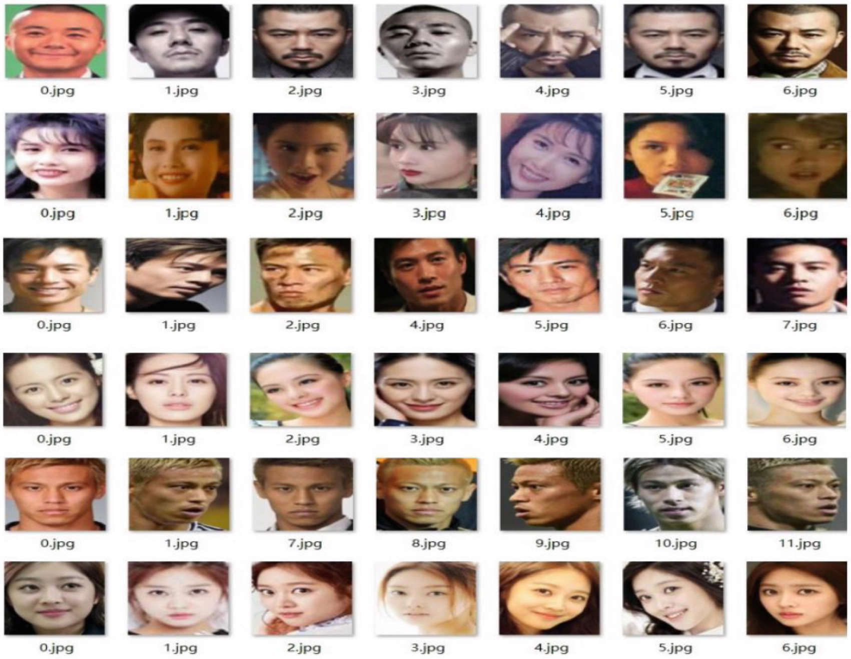

Fig. 2. Six individuals under varied poses or expressions in AFD dataset.

3.3 Training CNN Models

To validate our built face dataset, we need to perform verification on face recognition by training the same CNN model with different training datasets, including ours and publicly available ones. The widespread CNN structures in face recognition include Alex-net [8], VGG [9] and Inception [10]. Considering the comprehensive performance, we implement our CNN structure based on Inception-Resnet [11], using the triplet loss [12] to guide convergence.

4 Experiments

This section experimentally validates the performance of the established Asian face dataset in face recognition. For fair comparison, we use 3 different datasets to train the same CNN structure, such as WebFace, AFD and mixed WebFace&AFD. Few Asian face images are available in WebFace, but AFD contains only Asian face images. Models' details are tabulated in Table 2.

Table 2. Three models trained with the same structure but different training datasets.

Models	Structures	Training datasets	Total Images
Model_1	Inception + Triplet Loss	WebFace (Western)	490,000
Model_2	Inception + Triplet Loss	AFD (Asian)	360,000
Model_3	Inception + Triplet Loss	WebFace&AFD	850,000

The testing images are retrieved from the face datasets and are taken under uncontrolled real-world situations (unconstrained environments). Three trained models will be verified on three different testing datasets, shown in Table 3. Among them, one is for Westerner and two are for Asian.

Table 3. Three different testing datasets, including 1 Western face dataset and 2 Asian face datasets.

Datasets	Positive pairs	Negative pairs	Testing methods	Races
LFW	3000	3000	Cross-validation & ROC	Western
CASIA-FaceV5	3000	3000	Cross-validation & ROC	Asian
RealPhoto	116	7772	Mean Euclidean distance	Asian

Each testing dataset includes a large number of positive pairs (two facial images are from the same identity) and negative pairs (two facial images are from different identities). We took part of faces from public LFW and CASIA-FaceV5 to form the first and second test data sets. To make the test closer to the actual application environment, we build the third testing dataset RealPhoto by ourselves, which contains 68 individuals. Each of them has one Chinese ID card photo and several photos taken under the actual situation. Note that that all face images in training datasets (Table 2) and training datasets (Table 3) are mutually exclusive.

As for experiments on LFW and CASIA-FaceV5, we use cross validation to verify three above models. For visual classification tasks such as face recognition, excellent models should be able to reduce intra-class differences and increase inter-class differences at the same time. For this reason, in the third test, we mainly examine the ability of the model to distinguish between the same person's face and different people's faces. This will be compared by measuring the Euclidean distances of the feature vectors given by the model.

4.1 Results on LFW

The test results on LFW and their corresponding ROC curves are shown in Figs. 3 and 4, respectively. Green line, red line and blue line represent models trained by WebFace, AFD and WebFace&AFD, respectively. The plot illustrates that model trained by mixture dataset achieves the highest accuracy, and meanwhile model trained by WebFace gives a higher accuracy than model trained by AFD. It seems reasonable since WebFace and LFW are both Western face images while AFD contains Asian faces.

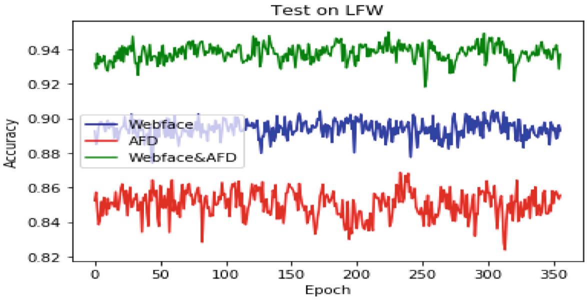

Fig. 3. The results of three models on LFW testing dataset. Blue line by WebFace&AFD tells the best accuracy, about 94%. Green line by WebFace indicates the accuracy of 89%. Red line by AFD shows the lowest mean accuracy, about 85%. (Color figure online)

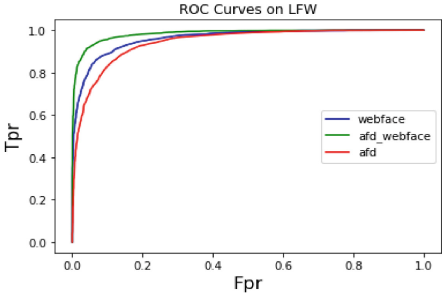

Fig. 4. Three different models' ROC curves on LFW, highly agreeing with the results shown in Fig. 3.

4.2 Results on CASIA-V5

Results on CASIA-V5 and their corresponding ROC curves are shown in Figs. 5 and 6, respectively. The curves illustrate that red line by AFD shows the best performance and green line by WebFace shows the worst instead. Testing dataset CASIA-V5 only includes Asian face images. As we expect, the model trained by Asian face images (AFD) outperforms the models trained by Western faces (WebFace) or mixed races (WebFace&AFD) for Asian face recognition. Once again, it is evident that face recognition performance is best when the face race in the test data set is consistent with the training data set.

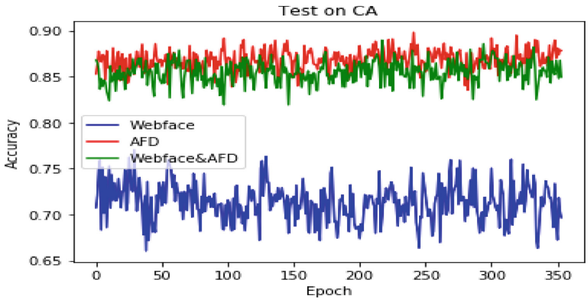

Fig. 5. The results of three models on CASIA-V5 testing dataset. Red line is trained on AFD and green line is trained on WebFace. The mean accuracy of red line is about 87%, and green line seems to be around 72%. (Color figure online)

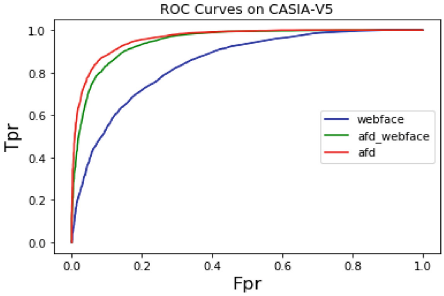

Fig. 6. Three different models' ROC curves on CASIA-V5, highly agreeing with the results shown in Fig. 5.

4.3 Results on RealPhoto

Three models are also verified on a real dataset RealPhoto built by ourselves. This testing dataset includes 68 standard Chinese ID card photos and 116 facial photos taken in real environment. There are the maximum number of pair is 7888 (a pair consists of 1 ID photo and 1 captured photo). Among them, there exist 116 pairs of same identity labels and 7772 pairs of different identity labels. We calculate the mean Euclidean distances with respect to same and different pairs. Same pair means that those two images belong to the same individual, and different pair indicates that they are from different individuals. We examine the difference between the average Euclidean distance of different pair and the average Euclidean distance of the same pair. If the difference is large, there is a large gap between the same pair and different pair, then the corresponding model has a higher accuracy of distinguishing the face image.

Results on the average Euclidean distance of same pair, the average Euclidean distance of different pair and their difference as well are shown in Table 4. Again, in this Asian face testing dataset, the model trained by AFD achieves the largest Euclidean gap 0.58327, while the model trained by WebFace gives the smallest gap 0.24718. This again shows that in Asian face recognition, the model trained with the Asian face dataset works best.

Table 4. Mean Euclidean distances of same and different pairs on testing dataset RealPhoto. The value of the fourth column is equal to the third column minus the second column. The larger the value of the fourth column, the better effect of training with the corresponding data set. The 0.58327 mean AFD dataset achieves best performance.

Training datasets	Mean distances of same pair	Mean distances of different pair	Difference between the two
AFD	0.83953	1.42280	**0.58327**
WebFace	0.69745	0.94463	0.24718
WebFace&AFD	0.72562	1.11575	0.39013

5 Conclusion

In this paper, we propose a semi-automatic way to collect face photos from Internet and build a large scale Asian face dataset containing 2019 subjects and 360,000 images, called AFD. To the best of our knowledge, the size of this dataset rank first in the literature from the perspective of Asian face datasets. Further, we explore how different races of facial image datasets (Western and Asian) affect face recognition performance. In conclusion, the model trained by Asian face images is able to provide better face recognition performance on Asian people than the model trained by Western faces. More generally, when the training dataset agrees with the testing dataset in terms of race, the more promising recognition performance will be achievable. Therefore, for face recognition task under real situations, it is strongly recommended that the agreeable training set be used.

So far, more than 80% facial images in AFD are Chinese. In the future, we would like to collect more face images, such as Japanese and Korean, to build a larger and more comprehensive Asian face image dataset.

Acknowledgements. We'd like to thank Institute of Automation, Chinese Academy of Sciences (CASIA) for offering CASIA-FaceV5 and CASIA-WebFace datasets. The research was supported by National Natural Science Foundation of China (61671332, 61502354, U1736206), Hubei Province Technological Innovation Major Project (2017AAA123), National Key R&D Project (2016YFE0202300), Hubei Provincial Natural Science Fund Key Project (2018CFA024), and Joint Project of Yunnan Provincial Science and Technology Department (2017FH001-060).

References

1. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments, vol. 1, no. 2. Technical Report 07-49, University of Massachusetts, Amherst (2007)
2. Ng, H.W., Winkler, S.: A data-driven approach to cleaning large face datasets. In: IEEE International Conference on Image Processing (ICIP), pp. 343–347 (2014)
3. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch. arXiv preprint arXiv:1411.7923 (2014)
4. Miller, D., Brossard, E., Seitz, S., Kemelmacher-Shlizerman, I.: MegaFace: a million faces for recognition at scale. arXiv preprint arXiv:1505.02108 (2015)
5. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In: CVPR, pp. 529–534. IEEE (2011)
6. Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: Robust face detection using the Hausdorff distance. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 90–95. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45344-X_14
7. Chen, D., Ren, S., Wei, Y., Cao, X., Sun, J.: Joint cascade face detection and alignment. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 109–122. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_8
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: imagenet classification with deep convolutional neural networks. Commun. ACM **60**, 84–90 (2017)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
10. Szegedy, C., et al.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9 (2015)
11. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: AAAI, vol. 4 (2017)
12. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823 (2015)
13. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**, 137–154 (2004)
14. Learned-Miller, E., Huang, G.B., RoyChowdhury, A., Li, H., Hua, G.: Labeled faces in the wild: a survey. In: Kawulok, M., Celebi, M.E., Smolka, B. (eds.) Advances in Face Detection and Facial Image Analysis, pp. 189–248. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-25958-1_8
15. http://biometrics.idealtest.org/findTotalDbByMode.do?mode=Face