Using Synthetic Images with Deep Convolutional Neural Networks for Racial Face Recognition

YEN-LUN CHEN[1], YI-LEH WU[1], CHENG-YUAN TANG[2]

[1]Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taiwan.
[2]Department of Information Management, Huafan University, Taiwan.

Abstract:

In the past, people usually employ the facial feature extraction and shallow learners such as decision trees, SVM, Naive Bayes, etc. to classify faces of different races. Deep learning usually takes lots of time to train. But with the advances in hardware and new algorithm proposed, the training time problem is gradually alleviated. The deep convolutional neural networks have good effect on images classification. In this paper, we use the deep convolutional neural networks to try to solve the problem of classification faces of different racial origin. Because the convolutional neural networks usually require a huge amount of data for training for good performance, such training set of real racial faces is not available to us. As a result of small set of real racial faces, this study proposes to incorporate synthetic facial images in our training set to sufficiently increase the size of the training set. To the best of our knowledge, this study is the first to propose to incorporate synthetic racial faces to train a deep convolutional neural network to classify real racial faces. We compare the performance of only employ synthetic facial images and mixtures of synthetic and real facial images in the training set. Our experiments show that training with only the real facial images (2,500 images) can achieve 91.25% accuracy in classifying faces of three different race origins. However, the classification when training with a mixture of 2,500 real facial images and 15,000 synthetic facial images can be further improved to 98.5% in accuracy.

Keywords: Caffe, Racial Face Classification, Convolution Neural Network, Deep Learning, Synthetic Image

1. Introduction

Human faces contain lots of information of soft biometric attributes, such as race, expression, identity, age, gender, etc. All of which have become a very active inter-disciplinary research area, such as computer vision, machine learning and biometrics. Among all, racial face analysis is one of the hottest topics because it has been widely utilized in practical applications; e.g., security, defense, surveillance and biometric-based identification. Therefore, racial face classification from 2D images is one of the most important problems.

The recognition methods for facial race classification can be usually divided into two main steps: feature extraction and classification. Many methods have been proposed with different features and classifiers, and most of these methods are extract hand-crafted features. Some previous work on racial face classification are listed in in Table 1.

Table 1. Previous work on racial face recognition

Authors and Year	Approach (feature + classifier)	Database	Race	Accu racy
Hosoi et al. [1], 2004	Gabor + SVM	HOIP DB	Asian, African American Caucasian	96%
Lin et al. [2] 2006	Gabor + Adaboost + SVM	FERET DB	Eastern Asian, Caucasian , African American	95%
Lyle et al. [3] 2010	Grayscale and LBP	FRGC (4232 faces)	Asian and Non-Asian	94%
Demirkus et al. [4] 2011	Skin color + Hair color SVM	Online DB (600 subjects)	Asian, African American , Caucasian	94%
Kumar et al.[5] 2011	Color histogram SVM	PubFig DB(58797 faces)	Asian, African, Indian, American , Caucasian	94.6 %
Xie et al. [6] 2012	KFCA and color features	Mugshot DB(5000 Caucasian, 50000	Asian, African American	98%

978-1-7281-2627-2/19/$31.00 ©2019 IEEE
DOI 10.1109/IIAI-AAI.2019.00099

		African, 4000 Asian)	, Caucasian	
Klare et al. [7] 2012	LBP + Gabor PCA + LDA	Operation DB	White, Hispanic, African American	91%
Huang et al.[8] 2014	Boosted local texture and shape features + LCP	FRGC v2.0 and BU-3DFE databases (3,676 textured 3D face models of 418 subjects)	White v.s Asian	97% (Average)
Wang et al. [9] 2016	CNN	MORPH-II (21,060 faces) Multiple large databases given above (240,000 faces) Multiple large databases given above (330,411 faces)	White and African American Chinese and Non-Chinese Han, Uyghur and Non-Chinese	99.7 % 99.85 % 99.6 %

In the past, to obtain more accurate results, researchers employ neural networks deeper and deeper. However, if we want to get more accurate results, our methods need to build on big enough of neural networks and need to spend a lot of time learning and computing. But nowadays, due to the rapid development of graphics processing unit (GPU) every day, the performance of parallel image processing is more than ever before. In the past few years, the GPU performance growth rate exceeds many times of the CPU performance. Figure 1 shows the development of GPU in the recent years. Therefore, deep learning (deep neural network) is a hotter topic than the traditional neural network recently. Especially since the Imagenet competition in 2012, many famous big data companies put more efforts in deep learning; e.g., Google, Facebook, and Baidu. Deep learning is more powerful than many other techniques for many different machine learning problems. Deep neural network is applied widely and has gain attention in recent years.

A deep convolution neural network has become popular in these years because it can get an excellent result in speech analysis and image recognition field. In the past, deep convolution neural network has been used to solve the racial recognition problems [9] and get accurate results. If we want to have a better learning on training convolution neural network model, the first requirement is to obtain a large data set. But how to collect a large number of data is a big problem. In order to resolve it, we propose to incorporate synthetic images in the training dataset to see if using the synthetic images can improve classification results. In this paper, we propose a method to recognize racial faces by employing the deep convolutional neural networks with high precision. We research the employment of synthetic face image of with our method to understand the effects through experiments.

2. Deep learning model

There are two main categories of neural networks, the shallow neural networks and the deep neural networks [10]. In the past, the shallow neural networks are much more popular than the deep neural networks. The shallow neural networks are much more efficient to train due to their simple structures. In recent years, more and more researchers study the deep neural networks by exploiting the improved hardware efficiency. Although the shallow neural networks are theoretically powerful enough, with more and more big data in recent years, they are still limited by their ability. On the other hand, despite the challenge to train and so sophisticated structural decisions, the deep learning networks are powerful undoubtedly.

For example, the LeNet-5 [11] is the most famous hand-writing recognition deep learning model. The LeNet-5 has 7 layers, including the input layer, each layer contains training parameters (connection weights). Especially, the Lenet-5 [11] neural network is a convolutional neural network.

According to the above reasons, more and more researchers are doing researches on deep learning. Deep learning is also called deep structured learning, hierarchical learning, or deep machine learning, and it is a class of machine learning method based on a set of algorithms. By using multiple nonlinear processing layers, the algorithms attempt to model high-level abstractions in data.

In the past, some of researches (e.g., face recognition or facial expression recognition) handcrafted features because image can be represented in many ways such as a vector of intensity values per pixel, or abstract way as a set of edges, regions of particular shape, etc. However, one of the advantages of deep learning is using the efficient algorithms to replace handcrafted features. The efficient algorithms are unsupervised or semi-supervised feature learning and hierarchical feature extraction. To learn from large-scale unlabeled data, people try to make better representations. Further we learn these representations by creating models.

Nowadays, there are huge numbers of variants of deep architectures such as deep convolutional neural networks, deep belief networks, deep recurrent neural networks have been applied on fields like image recognition, video analysis, natural language processing, handwriting recognition, speech recognition, etc. The deep learning networks have been performed to produce state-of-the-art results on various tasks. The convolutional neural networks are highly invariant in translation, scaling, inclination, or the common forms of distortion. The convolutional neural networks have been widely utilized in image and video recognition, recommender systems and natural language processing.

3. Experiments

In this section, we introduce our datasets, the synthetic facial dataset and the real facial dataset. The real face dataset has 3000 images. The dataset includes 1000 Caucasian face images, 1000 African American face images, and 1000 Asian face images. First, we select 200 real face images from each race (total 600 images) in random as the testing data and we call it real_face_testing. Then we decrease the remaining real face images of every race averagely because we want to combine synthetic faces with real faces to find the best combination. The synthetic face dataset has 120000 images. The dataset includes 40000 synthetic Caucasian face images, 40000 synthetic African American face images, and 40000 synthetic Asian face images. We divided this dataset into nine groups because we want to experiment whether the number of synthetic faces will affect the accuracy.

3.1 Using Real face dataset on Alex Net Model

In this section, the result shows 91.25% classification accuracy at 50000 iterations.

Figure 1 shows the results by using the Alex Net Model to train on the real face dataset and tested on the real_face_testing set.

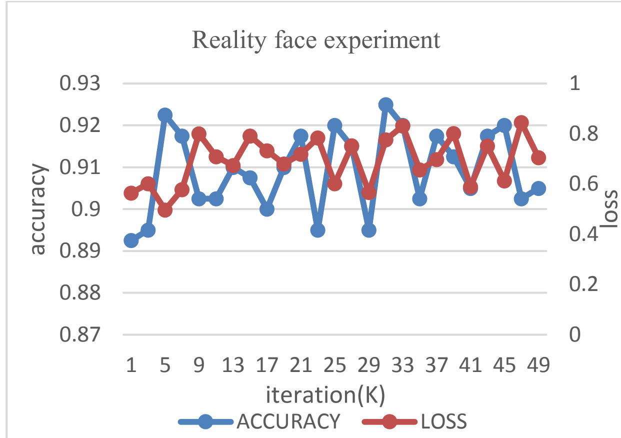

Figure 1. Real face training dataset

3.2 Using Synthetic face dataset on Alex Net Model

In this section, we increase the number of synthetic face images of every races evenly as the training sets. Finally, we conduct eight experiments.

Tabel 2 shows the number of images in the eight training and testing experiments.

We get the classification of accuracy at 50000 iterations with 51.16%, 55.26%, 60.65%, 57.78%, 67.7%, 63.375%, 61.4%, 63.16% in Experiment a, b, c, d, e, f, g, h, respectively. We observe that the number of synthetic faces increase gradually, and the accuracy enhance effectively. But only using the synthetic face dataset produce much inferior classification results than by using the real face dataset, even with much greater number of training images.

Figure 2 shows the results of experiment.

Table 2. Using Synthetic face dataset to experiments

	Asian face	African American face	Caucasian face	Total training	Real test
a	5K	5K	5K	15K	600
b	10K	10K	10K	305K	600
c	15K	15K	15K	45K	600
d	20K	20K	20K	60K	600
e	25K	25K	25K	75K	600
f	30K	30K	30K	90K	600
g	35K	35K	35K	105K	600
h	40K	40K	40K	120K	600

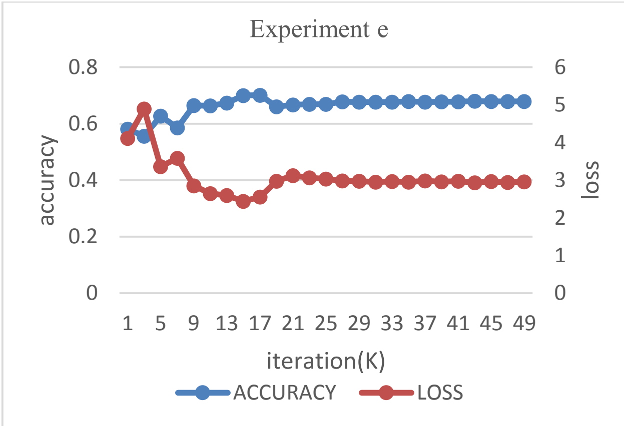

Figure 2. Result of experiment e

3.3 Using Synthetic face dataset on Alex Net Model

In this section, we use both real and synthetic face datasets as our training set. According to the definition of the above experiments, we divide the real face dataset into five groups and divided the synthetic face dataset into eight groups. Next, we use both real and synthetic face dataset with every combination. Therefore, we have forty experiments that need to be done and decide the best combination. Table 3 shows the results of employing the mixture of the real face dataset and the synthetic face dataset.

Table 3 Results of employing the mixture dataset

Synthetic face	Real face				
	500	1000	1500	2000	2400
15K	0.909	0.975	0.971	0.974	**0.985**
30K	0.916	0.955	0.959	0.961	0.981
45K	0.902	0.938	0.944	0.958	0.978
60K	0.875	0.926	0.953	0.943	0.976
75K	0.889	0.938	0.938	0.956	0.952
90K	0.858	0.901	0.930	0.944	0.959
105K	0.871	0.895	0.922	0.941	0.969
120K	0.834	0.880	0.915	0.937	0.939

4. Conclusions and Future Work

In this study, we propose a new method to improve accuracy of racial face classification by employing the deep convolutional neural network and using synthetic human faces. As far as we know, this paper is the first to use synthetic human faces as training set. We use the mixture of real face dataset and synthetic face dataset to train and test the deep convolutional neural networks.

In our experiments, first we can achieve about 91.25% classification accuracy while training only with real face dataset. Second, we show that by only using the synthetic face dataset for training produces much inferior classification accuracy than by using only the real face dataset, even with much greater number of synthetic training images. Finally, we combine 2,500 real faces with 15,000 synthetic faces to achieve 98.85% classification accuracy. However, keep increasing the number of synthetic images in the training set do not further increase the classification accuracy. We conjecture that this may be the result of over-fitting in training.

For future work, we will explore to employ other larger real racial face datasets. We would also explore to classify more finely in similar races, such as Chinese, Japanese, Southeast Asian, etc.

Acknowledgements: This work was partially supported by the Ministry of Science and Technology, Taiwan, under the Grants No. MOST 106-2221-E-011-160, 107-2221-E-011-126, 107-2632-M-211-001.

References

[1] Hosoi S., Erina T., and Masato K., "Ethnicity estimation with facial images," in Proc. IEEE 6th Int. Conf. Autom. Face Gesture Recog., , pp. 195–200, 2004, pp.

[2] Lin H., Lu H. C., and Zhang L. H., "A new automatic recognition system of gender, age and ethnicity," in Proc. World Congr. Intell. Control Autom., pp. 9988–9991, 2006.

[3] Lyle J. R., Miller P. E., Pundlik S. J., and Woodard D. L., "Soft biometric classification using periocular region features," in Proc. IEEE 4th Int. Conf. Biometrics: Theory Appl. Syst., pp. 1–7, 2010.

[4] Demirkus M., Garg K., and Guler S., "Automated person categorization for video surveillance using soft biometrics," in Proc. SPIE, Biometric Technol. Human Identification VII, pp. 76670P, 2010.

[5] Kumar N., Berg A., Belhumeur P., and Nayar S., "Describable visual attributes for face verification and image search," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 10, pp. 1962–1977, Oct. 2011.

[6] Xie Y., Luu K., and Savvides M., "A robust approach to facial ethnicity classification on large scale face databases," in Proc. IEEE Int. Conf. Biometrics: Theory, Appl. Syst., pp. 143– 149, 2012.

[7] Klare B. F., Burge M., Klontz J., Bruegge R. W. V., and Jain A. K., "Face recognition performance: Role of demographic information," IEEE Trans. Inform. Forensics Security, vol. 7, no. 6, pp. 1789–1801, Dec. 2012.

[8] Huang, D., Ding, H., Wang, C., Wang, Y., Zhang, G., Chen, L.: Local circular patterns for multi-modal facial gender and ethnicity classification. Image Vis. Comput. 32(12), 1181–1193, 2014.

[9] Wang W., He F., Zhao Q.,: Facial Ethnicity Classification with Deep

[10] Bengio, Y. Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, *2*(1), 1-127, 2009.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324, 1998.