Iran University of Science and Technology
Deep Learning (Spring 2019)

Face Verification on Blurred and Profile faces

Mohammad hasan Shamgholi
Department of Computer Engineering
Iran University of Science
and Technology
m_shamgholi@comp.iust.ac.ir

Amir Mohammad Salehoof
Department of Computer Engineering
Iran University of Science
and Technology
a_m_salehoof@comp.iust.ac.ir

Abstract

Face verification is one of the most important tasks in the field of machine vision and deeply inspected in research. Deep learning is involved in most of the recent works in the field and they have achieved very good results in standard datasets. However, in practice, we may see some major reduction in their performance on real-world applications. These include the pictures which are not taken in stable situations and have some non-clarity (blurriness), or the picture has some face profiles.

In this work, we use transfer learning and then fine-tuning the base model to solve the above-mentioned problems. The result shows that the achieved model not only works as fine as the base model on standard datasets (like LFW) but also has higher accuracy on harder face-verification cases.

1 Introduction

Face Verification is a sub-task of Face Recognition. Inputs are two faces and the output shows whether faces belong to the same person or not. There are many models for face verification and their accuracy is above 99.5% on popular datasets (1). But when testing in real-world applications like CC-Cameras videos, these models do not perform well. We think that the main reason for this poor performance is the model's training set which is collected in stable conditions and pictures have high quality. But in practice, pictures may be taken in hard situations (e.g. blurriness of any kind, low-quality of the camera, image/video compression, etc).

We introduced a dataset contains about 4200 identities. Each identity has both low and high-quality pictures of their face (Figure 1a 1b). Then we fine-tune a base-line model and show that the resulting model performs better in one of the real-world datasets. As a result of fine-tuning, we may have some accuracy reduction on the main test-set. We evaluate our fine-tuned model on the test set of the base-line model and results shows no change in performance.

2 Related work/Background

The task of face verification and face recognition is well researched. Recent works have a conventional pipeline including face detection, alignment, representation, and classification. Yaniv Taigman et al. (2) works on the alignment and representation step. They use a 3D face modeling to apply an affine transformation and make a face representation out of a nine-layer deep neural network. Fang Hua et al. (3) use the modulation transfer function (MTF) method for measuring the sharpness of a picture and also compared the result with other metrics. They evaluate their method on a dataset made by

(LateX template borrowed from NeurIPS 2019)

(a) low quality

(b) high quality

Figure 1: Sample of low quality and high quality pictures. both pictures belong to same person.

changing of the focal plane across face video sequences during acquisition from the Q-FIRE dataset. Swami Sankaranarayanan et al. (4) focuses on unconstrained face verification task which remains a challenging problem despite recent advances in the face recognition task. The paper proposed an approach that couples a deep CNN-based using triplet probability constraints to address this problem. They claim that their approach is robust to challenges including blurring by performing simple clustering experiments on standard datasets.

3 Proposed method

Voxceleb2 is a dataset usually used in the speech recognition task. It contains 6000 identities and each identity has some videos collected from YouTube in different situations and times. These videos contain both high-quality and low-quality videos of a person. We extract all frames from the videos. For the sake of storage limitation, we saved every 50 frames of videos of about 4200 identities.

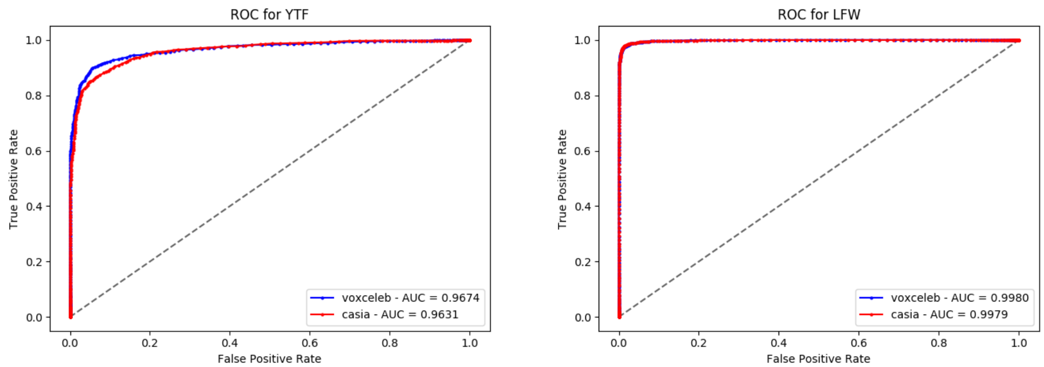
This new dataset contains 25 blurry (low quality) and 25 non-blurry (high quality) picture of faces for each identity. To measure the amount of blurriness, we used an edge detector operator on pictures called Laplacian. This operator calculates a value for each input image. Blurry pictures yield low values with this metric while high values go for quality ones. So, we can use a threshold to distinguish between low quality and high-quality pictures.

We claim that this new dataset will improve models to have better performance in real-world applications. To prove this claim, we chose a popular and simple model named MobileNet (5) and used a pre-trained version which trained on Casia WebFace (6), the second-largest dataset of faces that contains 11 thousand identities with about half million of pictures in total. Before feeding data to the network, we perform face detection and face alignment on each picture in the dataset. Hence pictures only contain faces, with no background. This will prevent the network to pay attention to the background.

Fine-tuning a pre-trained model may have a bad effect when performing the fine-tuned model on its original test set. To examine this fact in our model, we evaluate mobilenet on LFW (1) dataset before and after fine-tuning. LFW is a popular face dataset contains about 5700 identities and 13000 images. In literature, it is common to use this dataset to evaluate the proposed model. This test shows that our approach not only performance as well as baseline model on standard datasets but also has good accuracy on blur faces.

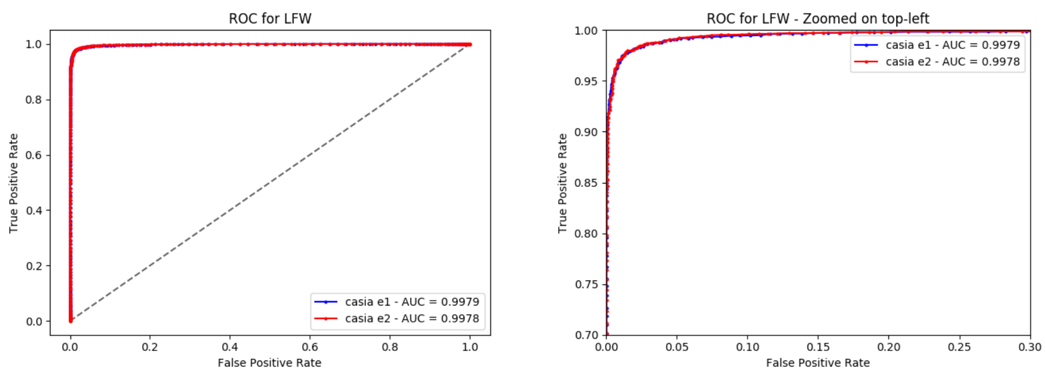
4 Results

To evaluate our work we used another dataset collected from youtube named YouTube Faces (7) (YTF). The dataset contains 3,425 videos of 1,595 different people. YTF introduced many video pairs from its dataset that is suitable for verification tasks such as face verification.



(a) (b)

Figure 2: Results of fine-tuned model and base model on: (a) YTF , (b) LFW datasets



(a) model is going to lose its accuracy on the second epoch

(b) zoomed for better observation

Figure 3: training baseline for two more iterations.

Figure 2a shows fine-tuned model performs better than base model on YTF. Also we evaluate our model on baseline's original test set (LFW) to prove that fine-tuning has no bad effect on base model. This is shown in Figure 2b.

To be sure that baseline does its best performance we continued training for two more iterations and results in Figure 3a shows that model is over-fitting on train set.

5 Discussion

The proposed method achieved some good results on both simple and hard cases on face detection, but the process can be improved by the following works: 1- In the image selection section, we use a simple laplacian filter to find some blur images, but we can do something more advanced; for example, we can pass the images through a deep network and get an embedding for each of them. Then we can use a metric like euclidian distance to find more different faces for an identity. In this way, the fine-tuning section will get some better data and the results should be improved. 2- The face alignment section can be improved by using some more advanced techniques.

References

[1] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in *Advances in face detection and facial image analysis*, pp. 189–248, Springer,

2016.

[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708, 2014.

[3] F. Hua, P. Johnson, N. Sazonova, P. Lopez-Meyer, and S. Schuckers, "Impact of out-of-focus blur on face recognition performance based on modular transfer function," in *2012 5th IAPR International Conference on Biometrics (ICB)*, pp. 85–90, IEEE, 2012.

[4] S. Sankaranarayanan, A. Alavi, C. D. Castillo, and R. Chellappa, "Triplet probabilistic embedding for face verification and clustering," in *2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS)*, pp. 1–8, IEEE, 2016.

[5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[6] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[7] L. Wolf, T. Hassner, and I. Maoz, *Face recognition in unconstrained videos with matched background similarity*. IEEE, 2011.