Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230740

The COVERED FACE: FaceNet-based partially occluded person recognition - on attention mechanism

Xiaopeng Feng[1,3,+] and Boyun Zhou[2,+]

[1]School of Information Science and Technology Northwest University, Xi'an, Shaanxi, China
[2]School of Information and Electronic Engineering Zhejiang Gongshang University, Hangzhou, Zhejiang, China

[3]183455667@qq.com
[+]These authors contributed equally to this work and should be considered co-first authors.

Abstract. This paper investigates the problem of improving the recognition accuracy of partially occluded faces. We aim to insert the attention mechanism blocks into the FaceNet to improve its recognition capability. FaceNet has excellent performance as a general-purpose face recognition system. However, recognizing faces with incomplete facial features is a challenge for feature extraction networks. To overcome this weakness and improve the recognition range of FaceNet, this paper proposes to incorporate the attention mechanism in the feature extraction network so that the attention of the model will be focused on the face features that are not occluded. The addition of the attention mechanism makes the application scenario of FaceNet more flexible and improves the anti-interference ability and robustness of the model. The results show that FaceNet trained by non-specific target dataset can improve the performance after adding the attention mechanism. In detail, the improvement in recognition for regular datasets is small. Nevertheless, the performance of the model is significantly improved for certain datasets with extreme conditions, such as face images with occlusions. In general, inserting the attention mechanism blocks in FaceNet can improve its value and expand its usage environment.

Keywords: facenet, face recognition, attention mechanism, deep learning, CNN.

1. Introduction

Face recognition is a kind of biometric technology, which generally identifies people by collecting their facial features. Face recognition system can automatically detect human face in the image, and then perform further processing on the detected face, such as face tracking, face verification, face recognition and face clustering. Face recognition has wide commercial value. In addition, after 30 years of research, face recognition technology has become mature and stable. It is limited, however, by the conditions imposed by a number of practical applications [1].

The World Health Organization declared a pandemic influenza state on March 12, 2020 due to the hyper-transmissibility of syndrome coronavirus 2 (SARS-CoV-2) and the high lethality of COVID-19. The countries of the world have so far suffered heavy losses in this global pandemic, including human deaths, economic stagnation or even regression, and increased poverty [2]. In the global COVID-19

DOI: 10.54254/2755-2721/6/20230740

pandemic environment, people have to wear masks to prevent the disease, especially in crowded public places that include airports, hospitals, schools [3]. Face recognition is considered as a contactless and non-invasive method of authentication which makes it safer in this situation [4]. But problems also arise. Wearing a mask can have an impact on face recognition, for instance by reducing the accuracy of the recognition, as the mask can obscure some of the facial features. On the other hand, removing the mask for face recognition in this environment will undoubtedly increase the risk of infection.

Therefore, this paper considers that it is necessary to improve the accuracy of face recognition with masks. We improve the recognition accuracy of FaceNet in extreme environments, such as when most people are wearing masks, by inserting attention mechanism blocks into its feature extraction network. This paper first describes the equipment and data set used to train the model. Then the experimental method is described and the experimental results are presented.

2. The development of face recognition

Research on face recognition systems started in the 1960s, and improved after the 1980s with the development of computer technology and optical imaging technology. However, it entered the primary application stage in the late 1990s. The period from 1965 to 1990 was the initial stage of face recognition research, which focused on face recognition methods based on Euclidean geometric features. The research in this stage was ultimately limited by theory and had few practical applications. The period from 1991 to 1997 was an active period of face recognition research. During this period the famous and important Eigenface method was proposed by Turk and Pentland at MIT, which uses the statistical principal component analysis (PCA) method [5]. Brunelli and Poggio in 1992 compared structural feature-based and template-matching-based methods and determined that the latter was superior to the former. Since 1998, researchers started to investigate for face recognition under non-ideal conditions. Multi-conditional problems such as consideration of lighting and pose were also introduced, and 3D model-based face modeling and recognition methods emerged.

With the development of deep learning, face detection technology based on deep learning has achieved great success in recent years [6]. During this period, excellent face recognition models have been continuously proposed, such as DeepFace, FaceNet, DeepID, which make face recognition increasingly correct and faster, and also continue to work in unrestricted environments. As one of the most successful applications in the field of image analysis and computer vision, face recognition technology has received attention from business, industry, law and government, and continues to work with different industries [7].

3. FaceNet and CBAM

We wish to train a model that can be used in an unconstrained environment, so FaceNet is chosen. The FaceNet is a face recognition system that can choose the type of embedded CNN neural network to map face images to a 128-dimensional Euclidean space and judge the degree of similarity based on the Euclidean distance of two face photos. The performance of FaceNet is excellent. The model trained with 8 million people and over 200 million photos achieved an accuracy of 99.63% tested on the LFW dataset and 95.12% on the YouTube Faces DB dataset [8]. However, we learned from our tests that the Euclidean distance between the new image and the original image increases after covering the face photo with some occlusions, such as a mask. When different photos of the same person are masked, the Euclidean distance between some of the images will exceed the threshold, which will lead to recognition errors. In extreme environments, such as when everyone wears a mask, the recognition correct rate drops dramatically.

Convolutional Block Attention Module (CBAM) represents the attention mechanism module of the convolutional module. The overview of CBAM is shown in Figure 1. Compared with SENet, CBAM adds the spatial attention mechanism module, while SENet only focuses on the channel attention mechanism [9]. CBAM is chosen in this paper due to its better performance.

DOI: 10.54254/2755-2721/6/20230740

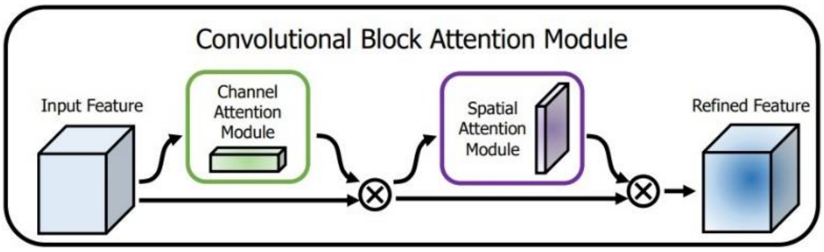

Figure 1. The overview of CBAM [9].

4. The dataset

To adequately test the model performance, several datasets are used for training and testing in this paper. The CASIA-WebFace dataset is chosen as the training set in order to implement the model training process. Because it has a large enough amount of data to ensure the completeness of the model training process. The Labeled Faces in the Wild (LFW) dataset and the Face Mask Lite Dataset are chosen for the test set.

LFW provides images of faces from natural scenes in life that are difficult to recognize due to many factors such as pose, lighting, expression, age, and occlusion. Even for the same person, different photos in the dataset can be very different. Sometimes there are distracting factors in the photos, or even more than one face. Figure 2 shows some of the images in LFW dataset. The processing of this dataset is very complete, and the images are classified and have specific names. Each image is 250 × 250 and most of them are in color, but there are still some black and white images mixed in. In summary, the LFW dataset is widely used for its good attribute and classification processing, and the performance of the model is highly demanded.

Figure 2. Some images in LFW dataset [10].

DOI: 10.54254/2755-2721/6/20230740

The Face Mask Lite Dataset is a simulated face mask dataset that is used to simulate the performance of the model in extreme environments. The authors of this dataset used Style GAN-2 to generate masks at the right locations in 10,000 different face photos and placed the original photos and the generated photos in two separate folders [11]. Figure 3 shows some of the images in this dataset. We used MTCNN to crop the Face Mask Lite Dataset to obtain our test set, denoted as STD. The effect of the attention mechanism on the model recognition ability can be obtained by comparing the performance of different models on the two test sets.

Figure 3. Some images in Face Mask Lite Dataset [11].

5. Experimental results

We wish to obtain the relationship between attention mechanisms and face recognition models. This paper designs two sets of scenarios and control the models to run different training sets to evaluate their performance. As a model trained by two experimental groups separately, we can also ensure that the final optimized solution is repeatable and not a result of chance.

This paper separately tested the performance of the model trained by the original FaceNet and the model trained by the FaceNet with the attention mechanism on the LFW test. Figure 4 shows the overview of the original network structure and the network structure with CBAM. On the left is the original InceptionResnetV1 structure, and on the right is the InceptionResnetV1 structure with the CBAM added. The attention mechanism is added behind the stem layer and Inception Resnet C layer.

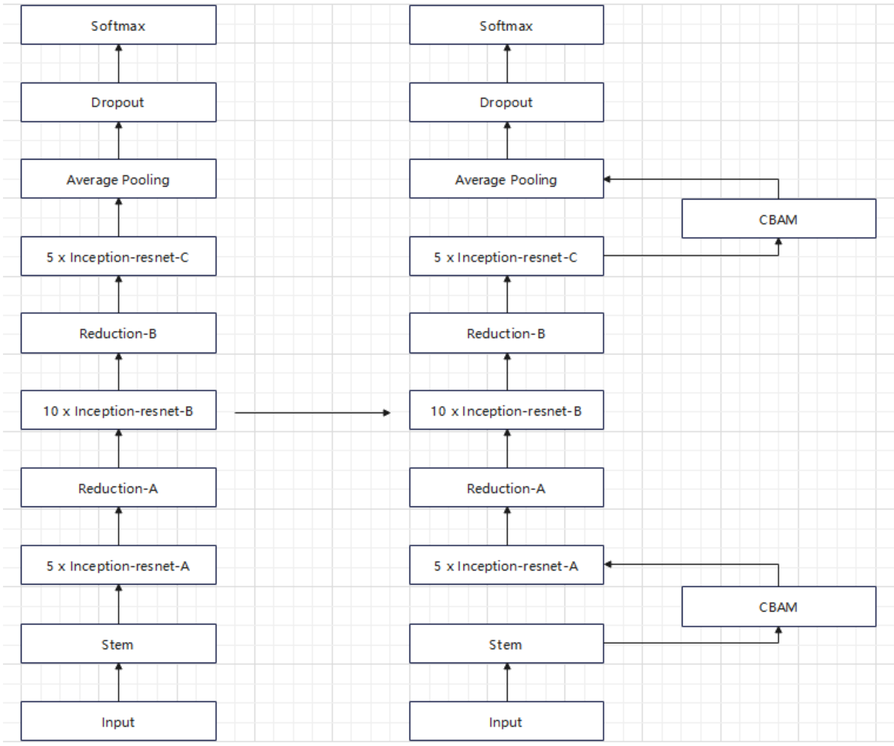

Figure 4. Overview of the original network structure and the network structure with CBAM.

We selected different test sets for the model to test its performance. Table 1 shows the test results of the two models on LFW. The original model trained by FaceNet achieves 97.54% correctness on LFW. Meanwhile, the model trained by FaceNet with the attention mechanism achieves 97.56% correctness on LFW. This shows that the attention mechanism does not improve the general recognition much. Then we separately tested the performance of the model trained by the original FaceNet and the model trained by the FaceNet with the attention mechanism on the STD test set.

Table1. Test results of the two models on LFW.

Model	Backbone Network	Traning Set	Test Set	Accuracy
Ordinary FaceNet	Inception Resnet V1	CASIA WebFace	LFW	97.54%
FaceNet with Attention Mechanism	Inception Resnet V1	CASIA WebFace	LFW	97.56%

Table 2 shows the test results of the two models on STD. The original model trained by FaceNet achieves 81.34% correctness on STD. Meanwhile, the model trained by FaceNet with the attention mechanism achieves 93.71% correctness on STD. In this round of experiments, it can be concluded that inserting the attention mechanism in FaceNet can improve the performance on the STD dataset. Nevertheless, the accuracy of the model cannot reach the accuracy on LFW. The attentional mechanism has limited enhancement to the model.

Table2. Test results of the two models on STD.

Model	Backbone Network	Traning Set	Test Set	Accuracy
Ordinary FaceNet	Inception Resnet V1	CASIA WebFace	STD	81.34%
FaceNet with Attention Mechanism	Inception Resnet V1	CASIA WebFace	STD	93.71%

6. Conclusion

In summary, this paper argued that the attention mechanism is a plug-and-play module that can improve the performance of the original model. Inserting the attention mechanism module into FaceNet can improve the recognition of obscured faces. Meanwhile, the improvement in the ability to recognize regular datasets by inserting the attention mechanism into the model is minimal. In contrast, the enhancement of the ability of the attention mechanism for models dealing with data sets of extreme environments, such as recognizing a large number of occluded face images, is significant. Nevertheless, we found using datasets from general recognition environments to train models, the attention mechanism has very limited enhancements to the models. Future research should consider the potential effects of training sets more carefully, for example, creating a dataset containing a large number of faces with masks in daily and frontal states for training and testing the model.

Acknowledgement

Xiaopeng Feng and Boyun Zhou contributed equally to this work and should be considered co-first authors.

References

[1] Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A. (2003) Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4), 399-458.
[2] Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W. C., Wang, C. B., Bernardini, S. (2020) The COVID-19 pandemic. Critical reviews in clinical laboratory sciences, 57(6), 365-388.
[3] Damer, N., Grebe, J. H., Chen, C., Boutros, F., Kirchbuchner, F., Kuijper, A. (2020, September) The effect of wearing a mask on face recognition performance: an exploratory study. In 2020 International Conference of the Biometrics Special Interest Group (BIOSIG) (pp. 1- 6). IEEE.
[4] Petrescu, R. V. (2019) Face recognition as a biometric application. Journal of Mechatronics and Robotics, 3, 237-257.
[5] Turk, M., & Pentland, A. (1991) Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1), 71-86.
[6] Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W., Li, S. Z., Hospedales,
[7] T. (2015) When face recognition meets with deep learning: an evalua- tion of convolutional neural networks for face recognition. In Proceedings of the IEEE international conference on computer vision workshops (pp. 142-150).
[8] djabi, I., Ouahabi, A., Benzaoui, A., Taleb-Ahmed, A. (2020) Past, present, and future of face recognition: A review. Electronics, 9(8), 1188.
[9] Schroff, F., Kalenichenko, D., Philbin, J. (2015) Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815- 823).
[10] Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018) Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).
[11] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008) Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.
[12] Prasoon Kottarathil. (2020) Face Mask Lite Dataset. https://www.kaggle.com/prasoonkottarathil/face-mask-lite-dataset>,2020.