Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 5, No. 2, 2023

Occluded Face Recognition based on Deep Learning

Quanyi Guo *

Electronic information school, Wuhan University, Wuhan, Hubei, 130524, China
*** Corresponding author Email:** GUOQUANYI060515@outlook.com

Abstract: Compared to the traditional sparse representation and the dictionary processing method of occlusion, deep learning-based face recognition methods are being used more and more widely in the field of face recognition. However, in practice, face recognition results are greatly influenced by light intensity, shooting Angle, mask and sunglasses occlusion and other factors. Therefore, this paper will discuss the face recognition under the occlusion situation. In order to solve the problem of large pose change of human face and local occlusion respectively, an offset network and a weight network was introduced into the convolutional neural network. In the following paper, the facial recognition accuracy of the introduction of the offset network, the facial recognition accuracy of the weight network and the recognition accuracy of the unification of the two are compared with the traditional facial recognition model VGG16.

Keywords: Face Recognition; Occlusion; Deep Learning; Offset Network; Weight Network.

1. Introduction

Face recognition in the past decade has been active in the field of computer vision and biological statistics is an important research topic [1], face recognition technology is widely used in criminal identification, company attendance, community management and the station customs monitoring areas. Face recognition models such as linear discriminant analysis, Gaussian hybrid model, support vector machine face recognition has greatly promoted the positive face recognition technology progress [2], and the traditional face recognition technology has become increasingly mature, in normal conditions can achieve high accuracy.

However, the accuracy of face recognition in traditional algorithm is influenced by objective factors such as shooting Angle, the illumination intensity, mask shielding, eye shielding, which makes the recognition accuracy in the presence of face mask. In the background of disease transmission, face recognition of mask will increase the efficiency and at the same time, the face recognition maintains social security and the regulation of masked robbery and theft in the community. Therefore, improving the existing mask blocking face recognition algorithm can reduce the cost of epidemic prevention and public security, and can also have reference value for other related fields.

For occlusion of face recognition, academic has been studied for many years, in recent years, the emergence of deep learning technology gradually replaced the traditional, such as sparse representation and occlusion dictionary processing method, in 2011, Zhaohua Chen and Rui Min use the principal component analysis and improve the support vector machine method detection, and then use block based weighted local binary mode only handle the occlusion face area. [3] In 2014, A Morelli Andres proposed a face recognition algorithm for occlusion detection based on compression sensing, which extracts recognition information by excluding occlusion regions during the recognition process. [4] In 2018, Weitao Wan proposed a MaskNet model which can set higher weights for the hidden units of the non-occluded part of the face and lower weights for the hidden units of the occluded part of the face. The experimental results show that the MaskNet model can effectively improve the robustness of the convolutional neural network model in occluded face recognition. [5] In conclusion, the predecessors on improving recognition accuracy breakthrough generally to reduce the occlusion part of the weight value and through the mapping function correction face, this study will combine the two into a new recognition network optimization, design a new loss function detection model accuracy, in order to improve the algorithm under the condition of face with occlusion recognition accuracy.

2. This Paper Method

2.1. Multi-angle Face Recognition

When face Angle changes, it will not only cause some misses of facial features, but also the feature vector after face coding change. At present, there are the following ideas to solve the problem of face angle: 3D reconstruction of the face, and identification of facial information, which is costly, which is commonly used in deep learning. The face recognition method based on poses estimation estimates the face pose, so as to perform face alignment, which can solve the problem of face offset to a certain extent. The face recognition method based on facial key points is to detect the features of the face that has not changed due to rotation, and it is also a commonly used method in deep learning.

And this paper will adopt the attitude of positive, face Angle changes after the coded vector and face attitude after coding the vector difference between the vector, and this paper will use the offset vector to represent the vector difference, the vector difference is the connection between offset face and positive face, vector of different pose face plus offset vector can get the vector of positive face. There is a mapping relationship between the vector difference between the offset face and the front pose face. The frontal pose face can be regarded as an offset face with an offset angle of zero. Therefore, we designed the offset network function to use the fitting ability of the deep neural network to learn this mapping function [6]. By observing the face images of different poses in reality, we can find that even in the case of the posture face, we can still capture a part of the face information, but the information has geometric deformation relative to the positive image.

The generalization ability of the deep learning model is greatly affected by the distribution of data sets. It is difficult for the model to learn the accurate depth features when the face attitude changes greatly. Therefore, the model should not only be able to output the corresponding vector difference according to the face offset Angle, but also apply this method to the future attitude recognition task. So before training face recognition model need to establish a vector compensation mechanism, assuming that there is a complex function y=a (x), x represents the vector of posed face, and y represents the vector difference between offset face and positive face, the offset network mechanism is to learn a (x). The learning process of the offset network is the process of continuously assigning learning tasks to the network, and the network continuously generates corresponding mapping functions to complete the learning.

In this paper, the training method is based on supervised learning. The output vector of the offset network is the weighted sum of the output vector of the convolution layer and also the linear combination of the coding layer output, so the output of the ReLu function in the offset network can be applied to many nonlinear models, the ReLu function is a piecewise linear function that treats negative values as 0 while leaving positive values unchanged. At the same time, compared with other activation functions, so it does not have the problem of gradient disappearance.

When two images of the same person are encoded by the same coding layer, X and x are output, and then the two establish a vector compensation mechanism through the offset network, and the offset network outputs A and a to achieve the effect of X+A=a+x. During the learning process, the model will gradually learn the compensation function of the vector difference to realize face recognition in different poses.

The offset network will be directly stitched after the convolutional neural network and the batch normalization layer. In the case of supervised learning, the vectors difference of the front image and the side image can be learned without affecting any performance of the original model. And this learning result can be applied to multi-classification tasks of face data.

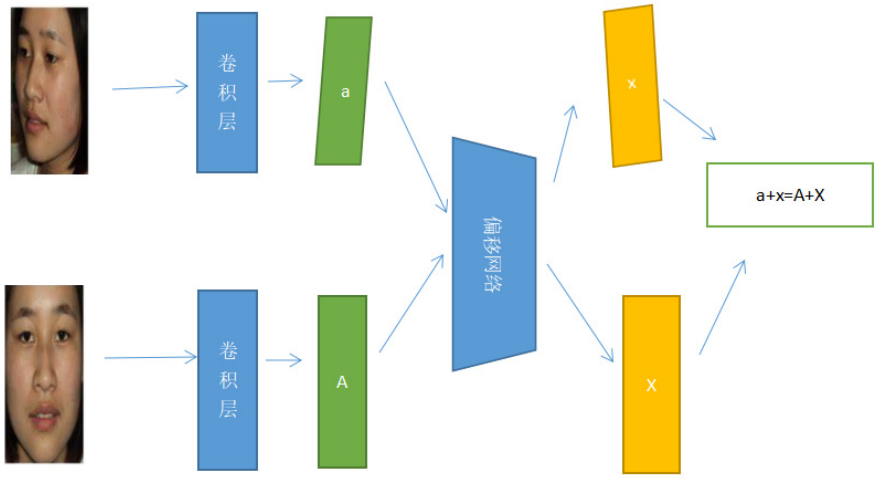

Figure 1. A Schematic diagram of the offset vector

2.2. Face Local Occlusion Recognition

For facial mask, glasses, shadow occlusion problem, this paper adopts the method of introducing a weight network, through the smaller weight value of occlusion area to reduce the influence on recognition results, this requires the establishment of a weight allocation model in face recognition to combine the coding vector after passing the coding layer with the weight vector after the allocation. Face recognition result comes after the coding layer coding vector and allocate the combination of the weight vector. In this paper, whether the learning of the compensation vector or the weight vector, the underlying idea of the network is based on the twin network.

When learning the weight vector, two face images are generally needed. At least some data of the triplet input data pair can meet the randomly selected front image as the reference sample. At the same time, the local occluded image of the same category is used as the positive sample. Each kind of other faces feature vector distance close and approximate overlap, at this time due to the reference samples and shade samples are different, the latter because shade will lose part of the facial information, and the feature vector coded by the coding layer is also useless. The role of the weight network is to emphasize the similarity in the shade part, weaken the influence of local occlusion on identification results.

Weight vector essence is the allocation of different parts of the feature vector, the reference sample and local occlusion of the positive samples for reference samples and samples after the convolution layer will be N * N feature matrix, the reference sample matrix is defined as X, shielding sample defined as x, the two matrices exist local different because of occlusion, and the weight vector of different parts of the two matrix characteristics gives different weight w. The smaller w is, the smaller the effect of the feature on the result. Thus, different weight vectors Y and y are output. For the weight vector output vector Y and y, the vector y has eliminated the invalid features of the occluded part, and assigned 0 to the value of some features in the vector, but the output vector Y does not assign 0 to any feature value. So, in order to ensure that the two images except for the same feature value, here defined vector z, each dimension in the vector z value, are the smallest vector of Y and y in the same dimension, the eigenvectors of the unconcluded part of the two images can be made the same, and the difference in the occluded part can be minimized. Then the vector z is matrix multiplied with the original encoded vector, making F (x) =x * z, f (x) = x * z, and making the two equal. In this learning process, the weight

vector can gradually learn the real weight, and the learning results of the model will be used for the classification task in future tests.

3. Model Training

In this paper, LFW [7] faces database and private data are used for result verification and analysis. The training set is divided into: (1) 1989 categories from LFW, with 1 to 10 images in each category. (2) 200 category labels come from the private data training set, in which 75 different categories are used as the test set, and the rest are used as the training set. There are about 10 images in each category in the private data set, and the pose faces data with different angles and facial occlusion data such as sunglasses are added to the private data set. Due to the diversity of LFW datasets in terms of partial occlusion and illumination, even in the same category of labels, the differences between different face images are very large, and there is only one image in some categories. In the private data, many pose images and partial occlusion images from different angles are used, such as wearing a hat, wearing sunglasses, and hairstyle occlusion. These data make the training set a face recognition data set that is currently difficult and challenging.

3.1. Data Set Selection

The training data used in this paper is extracted from the LFW and private data sets. The extraction method of the training data set is random sampling with replacement. Each set of training data retains the data of 125 different people in the private data set and six percent of the total number of categories is extracted from the LFW dataset. Then further extract the training set from it. The test data set used in this paper were generated from LFW and private data sets, and a total of 100 samples of different people was selected. The face data of these people comes from different data sets, 25 of them come from LFW, and a certain proportion of the data of these people is positive face data. The other 75 people come from the author's hand-made private data set. The data of these people are basically face data of different poses (side, slope) and face data under occlusion (sunglasses, hats, etc.). These data are combined into a set of sample pairs, and then 3000 samples are further extracted from several test samples composed of these data as a test set, of which the front face data accounts for about 25%, with poses of different deflection angles face and occluded face data account for about 75%, which shows that some of the samples are difficult data.

In the training process of the overall face model, the test data used in this paper is generated from LFW and private data, a total of 100 different samples was selected, 50 from LFW, 50 from private data, these data contain positive face with different posture or partial occlusion photos, they pair of samples, and then extract 3000 samples as the test set.

In the training process of the overall faces model, three important parts are included: the custom splicing convolutional network, the side face the vector difference of the offset network, and the adaptive weight for different local occlusions. The offset network is essentially the same as the weight network, both of which are composed of fully connected neural networks, and the ReLu function is added in the neurons to improve the applicability. Since the posture offset and facial occlusion may exist at the same time in the actual situation, the processing of the feature vector is carried out in no order at the same time. Assuming that the two exist at the same time, the algorithm in this paper will compensate the original vector of the face and then carry out the weight distribution. We define the original vector is a, the compensation vector is y, the weight vector is w, the above process can be expressed as H (x) =w (a + y). Then the network will perform Gaussian distance calculation on the output H(x), and the distance among the vectors of the same category will be shortened, and vice versa.

3.2. Performance Evaluation Method

According to the above data extraction method, images of about 250 different people are extracted as training data. They have a total of about 2000 images. Among these images, the images in the private data set are fixed data, and the images from the LFW data set are extracted. Obtaining data, even if a small proportion of data is extracted using the extraction method with replacement, it is difficult for these data to reappear.

When testing the network proposed in this chapter, in theory, the offset network and the weight network should be used together to deal with various complex situations in the data set. In order to verify the different networks in the process. Here, it is hoped that the two defined networks can be tested separately, so as to judge the effect of different networks on identifying difficult data. Finally, put the two together in the ordinary VGG [8] network, and use the test data extracted by the above method to test the effect, but in the effect comparison experiment, the test data obtained once should be saved. To ensure that different network models are tested using the same data.

In the test process of the images of sample input have trained model, get the face image encoding vector, then calculate the cosine similarity, both to measure the two feature vectors are similar, is similar to the feature vector to show the similarity between different people, where the similarity is greater than a certain threshold is the same person, similarity below the threshold as different people.

4. Analysis of Test Results

Adding the offset network proposed in this paper to the model, the performance of the original network is improved, and the performance of the weight network is still improved after adding to the model. The relationship between the number of training rounds and accuracy after adding the offset network is shown in Table 1.

Table 1. Model accuracy for the different number of training rounds

Epoch	1	2	3	4	5	6	7	8	9
accuracy	63%	66%	76%	84%	88%	90%	91%	89.5%	87%

The result of the addition of the weight network is shown in Table 2.

Table 2. Model accuracy under different training rounds

Epoch	1	2	3	4	5	6	7	8	9
accuracy	62%	70.5%	78%	82%	87%	88%	91%	88%	87%

After training all the two networks, the accuracy this paper and VGG network is shown in Table 3.

Table 3. Comparison of network training accuracy and VGG16

algorithm	VGG16	The offset network	Weight network	this text
accuracy	88.3%	91%	91%	93.5%

It can be seen that by introducing the offset network and the weight network into VGG 16 model, the accuracy of VGG network can be effectively improved, which can be increased by 5 percentage points in the complex background such as multiple poses and occlusion.

5. Summary

On the basis of VGG network, this paper proposes a method to improve the accuracy of face recognition in the case of multiple poses and occlusion, and significantly improves the accuracy of face recognition of VGG network in the case of occlusion, but only obtains a rough result for the identification of difficult data. At the same time, the subsequent research will explore the measurement scheme of different vectors, and train multiple models to improve the face recognition accuracy of the model in complex situations.

The face recognition model proposed in this paper also has some shortcomings. The offset network and the weight network are both fully connected networks. Trying more complex networks (such as adding convolution, residual and other structures) may bring further performance improvements. Promote. In addition, the premise that this method can be used normally is to capture RGB images in a natural environment, and in a dark environment, the images captured by the infrared camera may not be used directly, which will have a certain impact on normal face recognition. In future research work, we will consider designing a more compatible algorithm to improve the universality of the algorithm, and at the same time be able to deal with face images in different complex environments, which will be an important problem that this algorithm hopes to solve.

References

[1] Jiang, Y., Li, G., Ge, H., Wang, F., Li, L., Chen, X., ... & Zhang, Y. (2022). Machine learning and application in terahertz technology: A review on achievements and future challenges. IEEE Access, 10, 53761-53776.

[2] Guangcan, Y., & Huibin, L. (2021). Overview of face recognition methods based on deep learning. Journal of Engineering Mathematics, 38(04), 451-469.

[3] Chen, Z., Xu, T., & Han, Z. (2011). Occluded face recognition based on the improved SVM and block weighted LBP. 2011 International Conference on Image Analysis and Signal Processing. pp. 118-122.

[4] Andrés, A. M., Padovani, S., Tepper, M., & Jacobo-Berlles, J. (2014). Face recognition on partially occluded images using compressed sensing. Pattern Recognition Letters, 36, 235-242.

[5] Wan, W., Zhong, Y., Li, T., & Chen, J. (2018). Rethinking feature distribution for loss functions in image classification. Proceedings of the IEEE conference on computer vision and pattern recognition. 9117-9126.

[6] Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information theory, 39(3), 930-945.

[7] Lu, C., & Tang, X. (2015). Surpassing human-level face verification performance on LFW with GaussianFace. Proceedings of the AAAI conference on artificial intelligence, 29 (1).

[8] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR 2015). San Diego.1-14.