Survey on Deep Learning Based Face Expression Recognition Methods

Xiang Qi[1]

[1] School of computer and information Anhui Normal University, Anhui **241002**, China
22111303059@ahnu.edu.cn

Abstract. In recent years, the recognition of facial expressions has become a highly popular area of study, carrying significant importance for the advancement of scientific and technological progress. Face expression recognition has a very wide and practical application in many fields such as medical treatment, transportation, education and so on. This paper initially defines face recognition technology and examines its advancements. Following that, the operation and recognition efficiency of CNN and FaceNet, two deep learning-based face recognition techniques, are examined and analyzed. Subsequently, the two extraction techniques—conventional feature extraction and deep learning-based feature extraction—are presented together with their respective benefits and drawbacks. Finally, combined with the current development trend, the future development direction of face recognition technology based on image fusion is proposed, which is expected to combine the information from different image sources to improve the accuracy and robustness of recognition, and summarize and prospect the development of face recognition technology.

Keywords: Face Recognition, Deep Learning, Convolutional Neural Network, Feature Extraction.

1 Introduction

Face recognition technology is the use of computer programs to analyze and process the face, so as to realize the recognition and comparison of the face and identity information. Face recognition technology can be traced back to the 1970s, when it was mainly used to recognize faces by measuring the distance and angle of facial key points. As artificial intelligence technology advances and evolves, facial recognition has found extensive application across numerous areas of human activity. For example, in the medical field, face recognition can be used for the diagnosis and treatment of depression [1], In the realm of education, face recognition can be used in the classroom to monitor the state of students' learning [2]; In the realm of transportation, face recognition can monitor the fatigue state of drivers [3].

Techniques for recognising facial expressions can be split into two categories: those that rely on deep learning and those that use conventional techniques. The color, texture, and structure of a face image are mostly obtained using traditional approaches; nevertheless, facial expression recognition techniques based on old

© The Author(s) 2024
Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_48

methods suffer from issues like low recognition rates in complex contexts or with complex expressions. Better performance has been demonstrated by deep learning-based face expression recognition techniques, such as KRIZHEVSKY's AlexNet [4], which has made significant advances in image recognition.

For example, AlexNet, which was developed by KRIZHEVSKY,,made a breakthrough in image recognition, and building upon AlexNet's foundation, VGGNet [5], which was developed by the University of Oxford, also performed well. ResNet [6], proposed by Kaiming He, used jump connections for the first time to solve the problem of gradient vanishing that may occur in deep neural networks due to network depth. Among the many face expression methods, CNN and FaceNet are two very typical techniques [4,5].

Traditional technology feature extraction and deep learning-based feature extraction are the two main methods of feature extraction, both of which have been widely accepted and applied in multiple fields.

2 Deep learning based face expression recognition technology

2.1 Convolutional neural network (CNN)

Convolutional neural network CNN is a deep learning network model that is widely used in face recognition technology. As shown in Fig.1, the fundamental architecture of a CNN consists of: an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer.

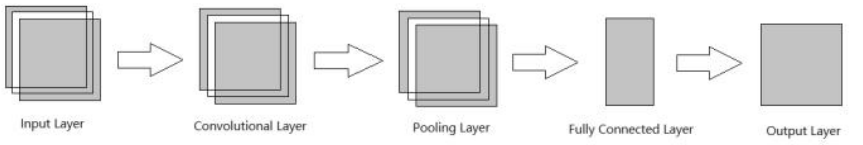

Fig.1 CNN basic structure(Photo/Picture credit : Original)

The two main processes performed by CNN are forward propagation and back propagation.

Forward propagation: The data is first received by the input layer and then forwarded to the first convolutional layer. The input data is then convolved using a series of convolutional kernels in each convolutional layer to carry out a convolution operation and produce a feature map. Subsequently, the feature map undergoes a nonlinear transformation via the activation function. To preserve the primary information, the feature map is then reduced through a pooling process. Through completely connected layers, the final classification result is acquired, and the output layer is then used to normalize the output.

Backpropagation: This method calculates the gradient of the loss function in relation to the parameters of the network. The network model is optimized when the gradient is used to update the network parameters, causing the loss function to progressively decline. The capacity of the network to adapt to the training data and to

generalize is increased when optimization algorithms such as gradient descent are utilized to adjust the network's parameters. By repeatedly changing the parameters until a predetermined stopping condition is met, the iterative training procedure optimizes the network model.

2.2 FaceNet

In order to determine whether or not the face image is of the same person, the conventional FaceNet [7] algorithm involves importing the face image, setting the optimal threshold, processing the image via a convolutional neural network to identify its features, mapping the face image features to the Euclidean space using L2 regularization, calculating the Euclidean distance between the face feature vectors in this space, and comparing the Euclidean distance with the optimal threshold.

FaceNet Framework is shown in Fig.2.

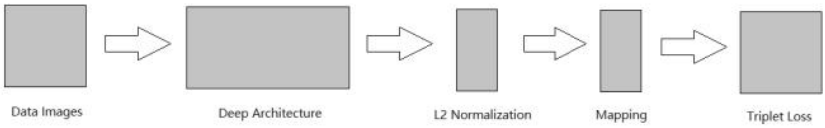

Fig.2 FaceNet Framework(Photo/Picture credit : Original)

Fig.3 illustrates the schematic diagram of one of the ternary losses:

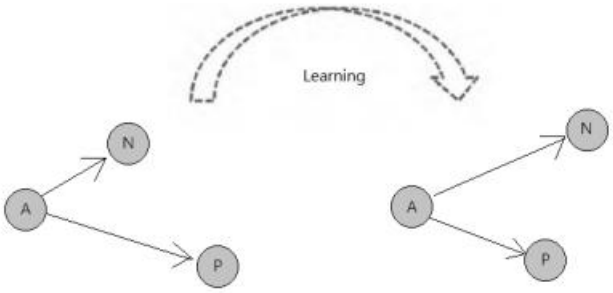

Fig.3 Triplet Loss(Photo/Picture credit : Original)

Anchor represents the reference sample, the Euclidean distance of Negative is greater than the threshold value, which means that the sample does not belong to the same category as Anchor, and the Euclidean distance of Positive and Anchor is less than the threshold value, which means that the two are the same category under the Anchor reference.

With the LFW dataset as its training set, the Face Net algorithm achieves 99.63% accuracy, 0.07% misrecognition rate, and 8.09% rejection rate [8]. Furthermore, the FaceNet and Retinanet-based dynamic face recognition system offers high recognition accuracy and real-time performance while resolving the issue of waiting for

collaboration in static face recognition [9]. Based on the aforementioned conclusions, it can be said that the FaceNet algorithm has a very remarkable impact on face recognition technology. It can enhance the precision and resilience of face recognition, encourage the practical use of face recognition technology, and potentially lead to even more extensive and far-reaching applications in the future.

3 Feature Extraction

3.1Feature extraction based on facial texture of face pictures

Since diverse texture blocks are frequently present in face photos, the facial texture-based feature extraction method will be utilized to accurately extract information from face images. The recognition of human facial emotions has made extensive use of the Gabor wavelet transform [10], which has a good sense of direction and frequency characteristics and can extract texture features and edge information of the image. The facial texture based feature extraction approach is a highly easy-to-use and efficient methodology that enhances face recognition performance by extracting important information.

3.2 Deep learning based feature extraction

A crucial stage in face recognition is feature extraction, which is based on deep learning and is extensively utilized. A face expression identification approach based on image fusion and deep learning is proposed [11], aiming to address the issue that deep learning cannot successfully extract local features of the picture and texture feature extraction method is single. One popular deep learning-based feature extraction technique that is frequently used to increase face recognition accuracy is multi-scale feature extraction. The feature pyramid structure is one type of multi-scale feature extraction that is more frequently utilized [12]. By increasing the variety of feature representations, this technique can enhance feature extraction accuracy.

4 Conclusion

Two face expression recognition algorithms based on deep learning are mostly covered in the review above, along with feature extraction techniques. Face recognition algorithms built on CNN and FaceNet are often utilized in society and have shown promise in terms of accuracy. However, in certain circumstances, such as variations in light and shade, facial occlusion, etc. the accuracy of face recognition may be compromised.

At present, one of the hottest research topics in computer vision is deep learning-based face recognition technology, which offers a plethora of socially useful applications. To increase face recognition accuracy in a range of scenarios, pertinent research has been done on a constant basis. In the future, face recognition technology

based on image fusion is expected to be a major development direction for face recognition, because the technology has many potentials to be developed in many fields, on the basis of which face recognition applications with higher accuracy and robustness and more adaptable to environmental changes can be developed.

References

1. Xu, L.: Research on Depression Identification Based on Facial Expression Behavior Patterns. Lanzhou University 1:84 (2021).
2. Zhang, L.J., Cao, J.T., Ji, X.F., Wang, T.H.: Design of Student Classroom Learning Status Monitoring System Based on Deep Learning. Journal of Electronic Measurement and Instrumentation 1-8 (2024).
3. Wang, W., Zhang, X.W., Wang, W.H.: Detection of Driver Fatigue State for Safety Alarm System. Transducer and Microsystem Technologies 41(06):82-85 (2022).
4. KRIZHEVSKY, A., SUTSKEVER, I., HINTON, G. E.: ImageNet classification with deep convolutional neural networks. Communications of the ACM 60(6): 84-90 (2017).
5. Simonyan, K., & Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv, https://arxiv.org/abs/1409.1556, (2015).
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. IEEE: 770-778 (2016).
7. William, I., Rachmawanto, E. H., Santoso, H. A., et al.: Face recognition using FaceNet (survey, performance test, and comparison). Long Beach:The Fourth International Conference on Informatics and Computing, IEEE 1-6 (2019).
8. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: A Unified Embedding for Face Recognition and Clustering. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 815-823 (2015).
9. Li, Y. P., Xi, Z. H.: Dynamic Face Recognition System Design Based on RetinaFace and FaceNet. Electronic Science and Technology 1-8 (2024).
10. Liu, S.X., Cheng, C.L.: An Improved Gabor Wavelet Transform Feature Extraction Algorithm. Application Research of Computers, 37(02): 606–610 (2018).
11. Jiao, Y.Y., Huang, R.C., Wan, W.T., Zhang, Y.: Face Expression Recognition Based on Image Fusion and Deep Learning. Transducer and Microsystem Technologies 43(03): 148–151 (2024).
12. Hou, Y. R., Di, L., Liang, J. J.: Low-Resolution Face Recognition by Fusing Gaussian Pyramid Features. Journal of Chinese Computer Systems 42(10): 2107–2115 (2021).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.