Facial Expression Recognition Based on Deep Learning: A Survey

Ting Zhang[(✉)]

School of Information Engineering, Minzu University
of China, Beijing 100081, China
`tozhangting@126.com`

Abstract. Facial expression recognition (FER) enables computers to under-stand human emotions and is the basis and prerequisite for quantitative analysis of human emotions. As a challenging interdisciplinary in biometrics and emotional computing, FER has become a research hotspot in the field of pattern recognition, computer vision and artificial intelligence both at home and abroad. As a new machine learning theory, deep learning not only emphasizes the depth of learning model, but also highlights the importance of feature learning for network model, and has made some research achievements in facial expression recognition. In this paper, the current research states are analyzed mostly from the latest facial expression extraction algorithm and the FER algorithm based on deep learning a comparison is made of these methods. Finally, the research challenges are generally concluded, and the possible trends are outlined.

Keywords: Facial expression · Deep learning · Feature extraction · Computer vision

1 Introduction

As the psychologist Mehrabian said "the facial expression sends the 55% information during the process of human communications while language sends only 7%". Therefore, the expression in the daily communication of mankind occupies an important position in the daily communication of mankind, is an important way to exchange emotional information among people. In order to have the ability of human thinking, it is necessary for the computer to understand the human emotions, that is the reason why after several decades of development the researches on facial expression recognition have been becoming the focus.

Usually facial expression recognition is divided into four processes: image pre-processing, face detection, feature extraction and facial expression classification, as shown in Fig. 1. The basic task of feature extraction is to produce some basic features according to the object to be identified, and then find the most effective representation from these characteristics, so that the difference is big in the different categories of expressions, and is small in the same type of expression difference is small. Feature extraction is a key problem in pattern recognition and one of the most difficult tasks in facial expression recognition. How to extract these features effectively is the key to realize accurate and feasible expression recognition. Due to the different research

background, the researchers have different definitions of basic emotions. Among them, Ekman's emotional theory has the greatest impact, where there are six basic emotions: anger, disgust, fear, pleasure, sadness and surprise. The expression classifier divides the extracted features into the corresponding categories according to the corresponding classification mechanism.

Fig. 1. Facial expression classification block diagram

2 Expression Feature Extraction Algorithm

At present, there are many kinds of feature extraction methods, which can be divided into geometric shapes, texture features and the combination of the two.

2.1 Methods Based Geometric Features Based

In the expression recognition algorithm based on geometric deformation, ASM (Active Shape Model) [1] and FAUs (Facial Action Units) [2] are the two most typical methods. ASM constructs geometric features by extracting facial feature points. The geometric features were extracted by the improving ASM and the use of the triangle features in the model, and achieved good results [3]. Combined with the texture and shape, AAM (Active Appearance Model) as a statistical model is used to model the face, to identify and analyze the expression. An expression recognition algorithm is proposed, combining AAM and ASM to improve the problem of feature point convergence in AAM [4].

According to FACS, as another important geometric feature, by analyzing the constant facial features and instantaneous characteristics. In the image sequence, a variety of features were tracked and modeled in order to recognize facial expression [5]. Combined with multi-scale feature extraction and multi-core learning algorithm, the contour of the face shape was outlined [6]. The face geometry and the activity appearance of the activity in the Candide grid frame are feature vector of expression [7]. Through the self-organizing mapping, the feature points that describes the shape of the eye, lips, and eyebrows in the face constitutes a low-dimensional eigenvector for expression recognition [8].

2.2 Method Based on Texture Feature

LBP (Local Binary Patterns, LBP) as a local operator can be used for image texture analysis, face recognition and facial expression recognition. Based on the idea of LBP, an LDP (Local Directional Patterns, LDP) operator is proposed, which is used to extract facial texture features [9]. The LDP feature is weighted by LDP variance, and then the most significant direction of LDP is calculated by PCA (Principle Component

Analysis) to reduce the feature dimension [11]. The proposed Gradient Directional Pattern calculates the texture information of the local area by quantifying the gradient direction to form the facial representation [12].

Lyons et al. applied the Gabor feature to calculate the expression characteristics, combining with elasticity graph matching and linear analysis [13]. But the high-dimensional Gabor feature is an obstacle to quickly identify facial expressions. The image is divided into local sub-regions to extract multi-scale Gabor features in order to describe the topology structure of human visual [14]. The algorithm combined Gauss-Laguerre wavelet and face reference to describe the expression feature, the former in which extracted the texture feature. The latter computed geometric features [15].

2.3 Multi-feature Fusion Method

Multi-feature fusion is an effective way of the facial expressions representation, and the different features in many algorithms are combined to describe facial expressions. The combination of Gabor and LBP is used in [16], and the minimized error rate and feature dimension is selected by genetic algorithm (GA). Li et al. Used SIFT and PHOG (Pyramid Histogram of Oriented Gradient) to calculate local texture and shape information [17]. TC (Topographic Context) is used as a feature operator to describe the static facial expression feature [18]. The haar-like feature was used to characterize facial expression changes over time, and was encoded to form a dynamic feature according to the binary pattern [19]. According to the main direction extracted in the video sequence and the average optical flow characteristics in the face block, MDMO (Main directional mean optical flow feature, MDMO) was proposed in [20]. Table 1 lists the main feature extraction methods in facial expression recognition.

Table 1. Main feature extraction methods for expression recognition

Literature	Feature	Category	Dataset	Rate of recognition (%)
[3]	ASM	7	JAFFE	86.96
[5]	FACS	7	Cohn-Kanade	96.7
[7]	Candide	4	Cohn-Kanade	85
[9]	LDP	7	Cohn-Kanade	93.4
		7	JAFFE	85.4
[10]	LDN	7	Cohn-Kanade	95.1
		7	JAFFE	91.1
[14]	Gabor	7	JAFFE	89.67
		7	Cohn-Kanade	91.51
[15]	Gauss-Laguerre	6	Cohn-Kanade	90.37
		7	MMI	85.97
[16]	LBP + Gabor	7	Cohn-Kanade	99.2
		6	MMI	94.1
[17]	PHOG + SIFT	6	Cohn-Kanade	96.33
		7	JAFFE	96.2

3 Facial Expression Recognition Algorithm Based on Deep Learning

By constructing a nonlinear neural network with multiple hidden layers, the deep learning simulates the human brain for analysis and learning. By transforming the input data by layer, the feature representation of the sample in the original space is transformed into a more abstract feature space, so as to study the essential characteristics of the data set and further to imitate the human brain to interpret the data such as image, sound and text. At present, the theory research for facial expression recognition based on the deep learning mainly focused on the two methods: the method based on DBN (deep belief network) and the method based on CNN (convolution neural network). Block diagram of FER based on deep learning as shown in Fig. 2.
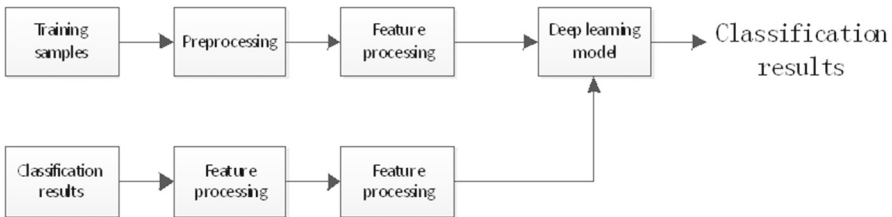

Fig. 2. Block diagram of FER based on deep learning

3.1 Method Based on DBN

As a classic method deep learning, DBN has aroused widespread concern, which can contain more hidden layer, and can better learn a variety of complex data structure and distribution. As shown in Fig. 3, the structure of DDBN (discriminative deep belief networks), includes an input layer h^0, N hidden layers h^1, h^2, ..., h^N and a label layer with C units at the top, equivalent to the number of classes in the tag data y. Look for the mapping function $X \rightarrow Y$, where it can be converted into a deep architecture to find the parameter space. The problem of finding the mapping function $X \rightarrow Y$ can be transformed into a deep architecture to find the parameter space W.

Combined with other methods, DBN has achieved significant results for the facial expression recognition. In [23], first the face was cut at different scales, and then the HOG (Histogram of Oriented Gradient) feature was extracted from these local cut, finally DBN were detected through the DBN. The Gabor wavelet features were extracted from the face cuts detected by DBN, and the fused feature fusion is input to the automatic coding machine for expression recognition [24].The strong classifier combined DBN and the adaboost method got a higher expression recognition accuracy rate for FER [25, 28]. A set of deep confidence network model was proposed, which can improve the performance of expression classification. MLDP-GDA (Modified Local Directional Patterns, Generalized Discriminant Analysis) was proposed, which was a novel method to extract salient features from depth faces was applied to DBN for training different facial expressions [31].
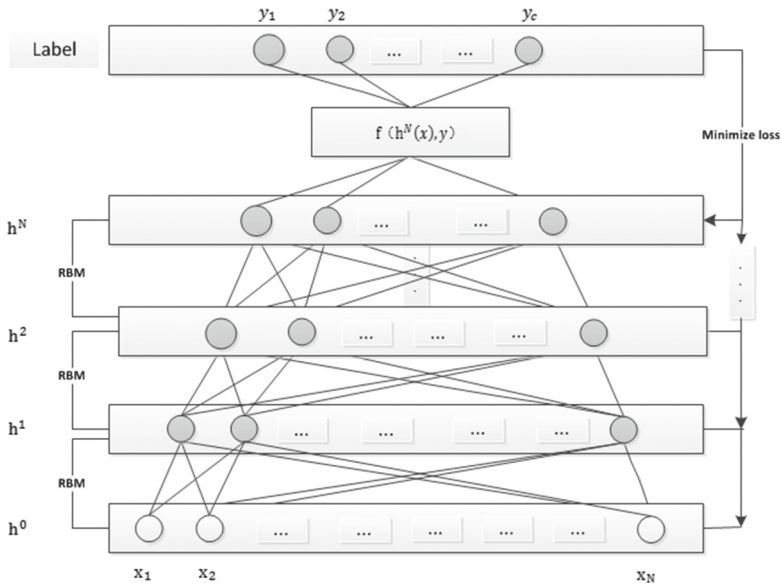
Fig. 3. Structure of DDBN

3.2 Method Based on CNN

As early as 2002, in order to solve the often encountered problems of the face gestures and the uneven illumination, the convolution neural network was applied to identify the facial expression. A set of FER system based on CNN was developed, which could predict the location of facial feature points recognize the facial expression by training a multi-task deep neural network. A novel approach based on based on the combination of optical flow and a deep neural network-stacked sparse autoencoder could analyze video image sequences effectively and reduce the influence of personal appearance difference on facial expression recognition [29]. Combined region of interest (ROI) and K-nearest neighbors (KNN), a fast and simple improved method called ROI-KNN for facial expression classification was proposed, which relieves the poor generalization of deep neural networks due to lacking of data and decreases the testing error rate apparently and generally [30]. A new architecture had the two hard-coded feature extractors: a Convolutional Autoencoder (CAE) and a standard CNN, which could can significantly boost accuracy and reduce the overall training time [32]. A 3D Convolutional Neural Network architecture consists of 3D Inception-ResNet layers followed by an LSTM unit that together extracts the spatial relations within facial images as well as the temporal relations between different frames in the video [33]. The attention model, which consisted of a deep architecture which implements convolutional neural networks to learn the location of emotional expressions in a cluttered scene, greatly improved the expression recognition [34].

4 Challenge and Future Work

Facial expression recognition is a very challenging subject, and is still in the elementary exploratory stage. With the non-rigid of the face and the complexity of FER, there are still many challenges:

(1) Improve the algorithm robustness of the expression recognition. At present, most research is the expression dataset build in the lab, but the facial expressions in real life are more complex and constantly changing.
(2) The difference differences in the race, the geography, and the culture.
(3) Gender differences in expression.
(4) Difficult to real-time recognition.

With the study of computer graphics, machine learning, artificial intelligence and other disciplines, the ability to make computers and robots able to understand and express emotions like humans will fundamentally change the relationship between people and computers, Facial expression recognition, as a long-term challenging subject, as a long-term challenging issues, may be the following trends:

(1) 3-d facial expression recognition.
(2) Robust facial expression representation algorithm and generalization recognition classification algorithm.
(3) Improve the deep learning model.

The deep learning model is widely used because of its strong feature learning ability. However, most of the models are time-consuming in training. It is necessary to train the model in a shorter time and improve the real-time ability.

References

1. Alemy, R., Shiri, M.E., Didehvar, F., Hajimohammadi, Z.: New facial feature localization algorithm using adaptive active shape model. Int. J. Pattern Recogn. Artif. Intell. **20**(1), 1–18 (2012)
2. Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans. Image Proc. **16**(1), 172–187 (2007)
3. Huang, K.C., Kuo, Y.H., Horng, M.F.: Emotion recognition by a novel triangular facial feature extraction method. Int. J. Innovative Comput. Inf. Control **8**(11), 7729–7746 (2012)
4. Sung, J.W., Kaneda, T., Kim, D.: A unified gradient-based approach for combining ASM into AAM. Int. J. Comput. Vis. **75**(2), 297–310 (2007)
5. Tian, Y.I., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Trans. Pattern Anal. Mach. Intell. **23**(2), 97–115 (2001)
6. Rapp, V., Bailly, K., Senechal, T., Prevost, L.: Multi-kernel appearance model. Image Vis. Comput. **31**(8), 542–554 (2013)

7. Patil, R.A., Sahula, V., Mandal, A.S.: Features classification using geometrical deformation feature vector of support vector machine and active appearance algorithm for automatic facial expression recognition. Mach. Vis. Appl. **25**(3), 747–761 (2014)

8. Majumder, A., Behera, L., Subramanian, V.K.: Emotion recognition from geometric facial features using self-organizing map. Pattern Recogn. **47**(3), 1282–1293 (2014)

9. Jabid, T., Kabir, M.H., Chae, O.: Robust facial expression recognition based on local directional pattern. ETRI J. **32**(5), 784–794 (2010)

10. Rivera, A.R., Castillo, J.R., Chae, O.: Local directional number pattern for face analysis: face and expression recognition. IEEE Trans. Image Process. **22**(5), 1740–1752 (2013)

11. Kabir, M.H., Jabid, T., Chae, O.: Local Directional Pattern Variance (LDPV): a robust feature descriptor for facial expression recognition. Int. Arab J. Inf. Technol. **9**(4), 382–391 (2012)

12. Ahmed, F.: Gradient directional pattern: a robust feature descriptor for facial expression recognition. Electron. Lett. **48**(19), 1203–1204 (2012)

13. Lyons, M.J., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. IEEE Trans. Pattern Anal. Mach. Intell. **21**(12), 1357–1362 (1999)

14. Gu, W.F., Xiang, C., Venkatesh, Y.V., Huang, D., Lin, H.: Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recogn. **45**(1), 80–91 (2012)

15. Poursaberi, A., Noubari, H.A., Gavrilova, M., Yanushkevich, S.N.: Gauss-Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. EURASIP J. Image Video Process. **17**, 1–13 (2012)

16. Zavaschi, T.H.H., Britto, A.S., Oliveira, L.E.S., Koerich, A.L.: Fusion of feature sets and classifiers for facial expression recognition. Expert Syst. Appl. **40**(2), 646–655 (2013)

17. Li, Z., Imai, J., Kaneko, M.: Facial-component-based bag of words and PHOG descriptor for facial expression recognition. In: Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, vol. 64(2), pp. 1353–1358. San Antonio, TX, USA (2009)

18. Wang, J., Yin, L.J.: Static topographic modeling for facial expression recognition and analysis. Comput. Vis. Image Underst. **108**(1–2), 19–34 (2007)

19. Yang, P., Liu, Q.S., Metaxas, D.N.: Boosting encoded dynamic features for facial expression recognition. Pattern Recogn. Lett. **30**(2), 132–139 (2009)

20. Liu, Y.J., Zhang, J.K., Yan, W.J., Wang, S.J., Zhao, G.Y., Fu, X.L.: A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Trans. Affect. Comput. **7**(4), 299–310 (2016)

21. Fasel, B.: Mutliscale facial expression recognition using convolutional neural networks. In: Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 02), pp. 1–9. Ahmedabad, India (2002)

22. Liu, Y., Hou, X., Chen, J., et al.: Facial expression recognition and generation using sparse autoencoder. In: International Conference on Smart Computing, pp. 125–130. Hong Kong, China (2014)

23. Lv, Y., Feng, Z., Xu, C.: Facial expression recognition via deep learning. Int. Conf. Smart Comput. **32**(5), 347–355 (2015)

24. Lv, Y., Feng, Z., Xu, C.: Facial expression recognition via deep learning. In: International Conference on Smart Computing, pp. 303–308. Hong Kong, China (2014)

25. Jung, H., Lee, S., Park, S., et al.: Development of deep learning-based facial expression recognition system. In: 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp. 1–4. Mokpo, South Korea (2015)

26. Liu, P., Han, S., Meng, Z., et al.: Facial expression recognition via a boosted deep belief Network. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1805–1812. Columbus, USA (2014)

27. Devries, T., Biswaranjan, K., Taylor, G.W.: Multi-task learning of facial landmarks and expression. In: Proceedings of IEEE International Conference on Computer and Robot Vision (CRV), pp. 98–103. Montréal, Canada (2014)
28. Kim, Y., Lee, H., Provost, E.M.: Deep learning for robust feature generation in audiovisual emotion recognition. In: Proceedings of IEEE International Conference on Speech and Signal Processing (ICASSP), pp. 3687–3691. Vancouver, Canada (2013)
29. Liu, Y., Hou, X., Chen, J., et al.: Facial expression recognition and generation using sparse autoencoder. In: Proceedings of IEEE International Conference on Smart Computing (SMARTCOMP), pp. 125–130. Hong Kong, China (2014)
30. Xiao, S., Ting, P., Fu-Ji, Rn: Facial expression recognition using ROI-KNN deep convolutional neural networks. Acta Automatica Sinica **42**(6), 883–891 (2016)
31. Uddin, M.Z. et al.: A facial expression recognition system using robust face features from depth videos and deep learning, Computers and Electrical Engineering. Available online 29 Apr 2017
32. Hamester, D., Barros, P., Wermter, S.: Face expression recognition with a 2-channel convolutional neural network. In: Proceedings of International Joint Conference on Neural Networks (IJCNN), pp. 1787–1794 (2015)
33. Hasani, B., Mahoor, M.H.: Facial expression recognition using enhanced deep 3D convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition. https://arxiv.org/pdf/1705.07871.pdf (2017)
34. Barros, Pablo, et al.: Emotion-modulated attention improves expression recognition: a deep learning model. Neurocomputing **30**(253), 104–114 (2017)