Research on Facial Expression Recognition Based on Convolutional Neural Network

Wang Xuyang[1], Chen Jinheng[2]

[1]College of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, 730050, China

[2]College of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, 730050, China

Keywords: Expression recognition, Convolutional neural network, Image processing

Abstract: The main research methods of face recognition technology include geometric features and algebraic features. With the advent of artificial intelligence era, the rapid development of machine learning, the BP neural network and CNN (Convolutional Neural Network) model under deep learning have greatly improved the speed and accuracy of face recognition, which has greatly improved the stability, accuracy and rapidity of face recognition system. For a long time, as an important part of computer vision research, the traditional method of selecting features manually by machine learning and then training shallow classifiers for recognition is not satisfactory in the field of facial expression recognition. Facial expression involves machine learning contents such as image processing, computer vision, pattern recognition, etc. It is a multidisciplinary field. This paper attempts to apply the method of deep learning to facial expression recognition, construct a deep convolution neural network suitable for facial expression recognition, and put forward some effective improvement schemes in training algorithm and structure.

1. Introduction

People's facial expression conveys a person's happiness, anger, sadness and happiness, which is an important means for people to interact with the outside world, and also an important basis for the outside world to capture a person's emotional changes [1]. In life, people's expression often represents the emotion he expresses and his real feelings, so expression is another way of communication beyond language [2]. Class conveys the true reflection of the inner emotional world through the subtle changes of facial expressions. By quickly capturing and accurately recognizing human expressions, the machine can make different responses and effectively improve the friendliness and intelligence of human-computer interaction [3]. Facial expression recognition has a very wide range of application prospects, such as the interaction between human and robot, the application of facial expression recognition to monitor the details in the process of criminal confession, and the diagnosis of facial paralysis through facial expression recognition in medicine [4]. Facial expression recognition is an indispensable part of machine learning research, which has a very broad application value in today's society with the popularity of human-computer interaction. The breakthrough of facial expression recognition method also has great reference significance in the field of Intelligent Computing and brain like research [5].

With the increasing popularity of deep learning research, the application of deep learning in facial expression recognition has been seen, but compared with other fields of research, the application of deep learning in facial expression recognition is only a preliminary combination [6]. Facial expression recognition system generally includes face detection, image preprocessing, feature extraction and expression classification [7]. Among them, facial expression feature extraction and classification is the focus of facial expression technology research, which is related to the final result of facial expression discrimination [8]. The collision of facial expression recognition and artificial intelligence is bound to burst out more brilliant fireworks. Expression recognition has laid a solid foundation for human-computer interaction in the era of artificial

  DOI: 10.25236/mmmce.2021.045

intelligence, and deep learning has also guided a new way for the research of facial expression recognition [9]. Feature based recognition method is the key to facial expression recognition by classifier. Traditional classification methods need to extract features artificially for classification. The quality of feature selection directly determines the recognition accuracy, while feature selection requires certain professional knowledge, low recognition rate and time-consuming [10, 11]. In this paper, we try to apply deep learning method to facial expression recognition, construct deep convolution neural network for facial expression recognition, and put forward some effective improvement schemes on training algorithm and structure.

2. Basic Theory of Deep Convolution Network

One of the characteristics of the structure of artificial neural network is that it has distinct layers, which are composed of several neurons and then networks. The signal is input by one layer, processed by this layer, and then transmitted to the next layer, processed and output again. In this way, the signal is transmitted backward layer by layer until the final processing result is output. Convolutional neural network is a kind of feedforward neural network, its basic structure includes data input layer, convolution layer, pooling layer, full connection layer and output layer [12]. Input layer: fill the image into the input layer in the form of matrix. Convolution layer: convolution layer is the core of convolution neural network, convolution is the operation of two functions, called convolution operation. Pooling layer: the function is to retain the part with high strength and filter out the part with low strength, so as to reduce the size of the matrix after convolution operation. Full connection layer: put a series of data after convolution and pooling into the full connection layer, and output the class score vector. Output layer: output image.

The evolution process of convolution neural network is shown in Fig. 1. The development of convolution neural network follows a main line, from BP neural network to shallow convolution neural network, and then gradually increase the number of network layers to form deep convolution network. Although each evolution has a huge innovation in structure, it is essentially the same.

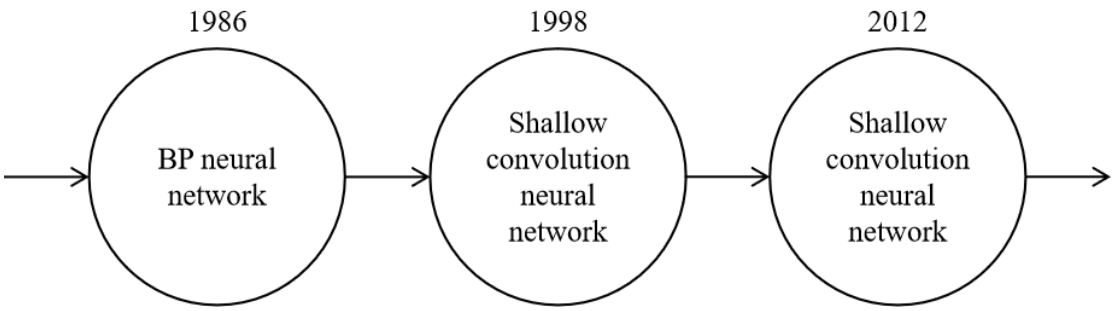

Fig.1 Evolution Process of Convolutional Neural Network

In the actual operation process, a receptive field window will also be used, which moves on the original image in line priority order, that is, each line moves from left to right first. When the window moves to the end of the line, it moves down one line, and then continues to move from left to right, and this is repeated until the end of the image. In this way, a neuron arranged in a two-dimensional matrix can be mapped to a two-dimensional matrix neuron with a slightly smaller dimension, and its operation mode is shown in Fig. 2.

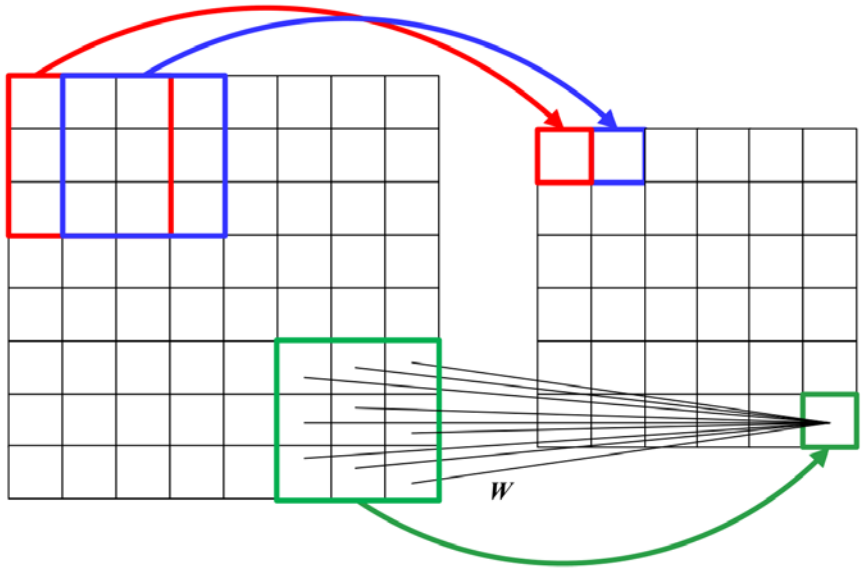

Fig.2 Schematic Diagram of Operation Mode of Partial Receptive Field Window

After applying the local receptive field and weight sharing technology, the mapping from neurons in the upper layer to neurons in the next layer can be expressed by the following formula:

$$y_{j,k} = \sigma\left(b + \sum_{l=1}^{n}\sum_{m=1}^{n} w_{l,m} x_{j+l,k+m}\right) \quad (1)$$

Among them, $y_{j,k}$ represents the output of the kth neuron in the jth row of the hidden layer.

3. The Structure of Convolutional Neural Network

In each convolution layer, the upper expression features are convoluted by a learnable convolution kernel to output higher dimensional expression features to the lower convolution layer. The calculation of facial expression feature map is as follows

$$G_i = f\left(G_{i-1} * W_i + b_i\right) \quad (2)$$

Among them, G_i is the neuron of the i-th layer, and G_{i-1} is the output of the neuron of the i-1th layer as the input of the i-th layer. W_i is the weight vector of the convolution kernel of the i-th neuron, b_i is the offset vector, and f is the activation function.

The pooling layer is located after the convolutional layer, which can maintain the feature scale to a certain extent while reducing the dimension of the feature map. The pooling formula is:

$$G_i = Pooling(G_{i-1}) \quad (3)$$

Among them, G_i is the downsampling layer, and Pooling is the pooling function.

The fully connected layer connects the neurons in the upper layer with each neuron in this layer to achieve the purpose of synthesizing the extracted features, so it is also called a multilayer perceptron, and its calculation formula is:

$$F(x) = f\left(x * W + b\right) \quad (4)$$

Among them, F(x) is the fully connected layer, f is the activation function, W is the weight vector, and b is the offset vector. Use SoftMax to map the output of multiple neurons to a value from 0 to 1, indicating that the probability that the expression image x belongs to category j is:

$$p\left(y^{(i)} = j \middle| x^{(i)}, \theta\right) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^{k} e^{\theta_j^T x^{(i)}}} \quad (5)$$

In the formula, $p\left(y^{(i)} = j \middle| x^{(i)}, \theta\right)$ is the probability that the expression image x corresponds to each label category, and θ is the parameter to be fitted.

4. Facial Expression Recognition Based on Convolution Neural Network

The research of facial expression recognition has wide application value. Fast facial expression recognition is helpful to analyze the emotion of the recognized object, and can realize the emotional communication between machines and people in the field of intelligent machines. The changes of people's own emotions lead to significant changes in their physiology and behavior, and facial expression is the most intuitive aspect of emotional manifestation. Facial expression recognition technology refers to extracting facial expression features from a given facial expression image and attributing them to a specific facial expression [13]. In facial expression recognition system, the object is aimed at facial expression. Therefore, in order to improve the accuracy and discard the hair, ears, neck and other parts that have nothing to do with expression, we can use the corresponding software to extract face features from the pictures. The steps of facial expression recognition generally include obtaining facial expression images, clipping and normalizing the original facial expression images, extracting expression features, training models and classifying expressions [14]. The key step is facial expression feature extraction, and the effectiveness of extracted features determines the performance of facial expression recognition. In order to solve the diversity of pictures and extract as many facial expression features as possible, a series of expansion processes are carried out on the facial expression data set. When training convolutional neural networks, a large number of training data are needed, which can improve the accuracy of facial expression recognition on the one hand, and prevent over-fitting caused by excessive training on the other hand.

At the beginning of the process of facial expression recognition, it is necessary to detect, locate and segment faces from the facial expression database. Then, face images segmented in the background are studied or transformed in a certain form to extract features. Finally, according to the features we extracted, we use a suitable classifier for recognition. The process is shown in Fig. 3.

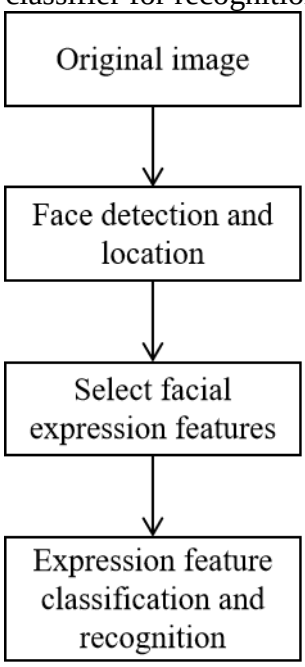

Fig.3 Facial Expression Recognition System Process

For images, there is usually a local spatial connection, so that each neuron only needs to deal with the input of a local area, and does not need to deal with the whole image globally. If facial expressions containing a frame of pictures are recognized, a simple convolution network structure can be designed, as shown in Fig. 4.

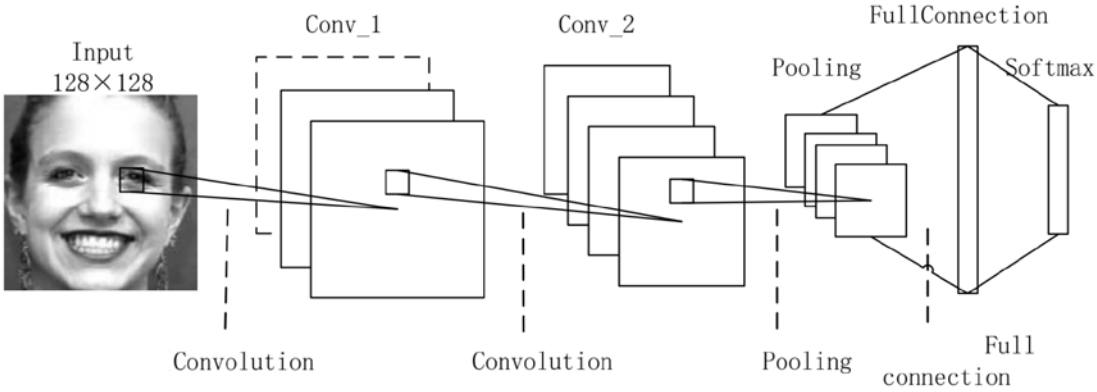

Fig.4 Convolutional Neural Network Structure for Processing a Single Frame

Because the unsupervised pre-training data is a face image, and the correlation between adjacent pixels in the image is very strong, it can be considered that the training input is redundant, and whitening processing is needed to reduce the redundancy of the input data. In order to avoid a lot of redundant information of non-expression features in the training and learning of expression features in the deep convolutional neural network, the facial expression image is segmented by local expression feature regions, and the segmented expression feature regions are used as the training and testing samples of the deep network, thus avoiding the redundant influence of other regions of the face in the learning of expression features [15]. Specifically, from the perspective of network structure design, on the premise of meeting the expression ability required by the network, try to use a small network scale and reduce the learning time and complexity of the network. From the scope of action, the receptive field obtained by using multiple small-scale convolution kernels can be the same as that obtained by one large-scale convolution kernel. Wavepooling not only preserves the information integrity of the output feature map of the upper convolution layer in the low-frequency information, but also adds the high-frequency texture information as the low-level feature expression to the final feature vector to realize the combination of local and abstract.

5. Conclusions

The research of facial expression recognition has wide application value. Fast facial expression recognition is helpful to analyze the emotion of the recognized object, and can realize the emotional communication between machines and people in the field of intelligent machines. When training convolutional neural networks, a large number of training data are needed, which can improve the accuracy of facial expression recognition on the one hand, and prevent over-fitting caused by excessive training on the other hand. With the increase of the number of sequence frames, it is necessary to add more levels of convolutional networks, which undoubtedly increases the complexity of network structure, and the training difficulty and training time will also increase. In order to improve the recognition rate of self-portrait images, in addition to building a richer expression database, it is necessary to test different convolutional neural networks to find a convolutional neural network structure that is more suitable for expression recognition.

In order to solve the diversity of pictures and extract as many facial expression features as possible, a series of expansion processes are carried out on the facial expression data set. When training convolutional neural networks, a large number of training data are needed. From the point of view of network structure design, on the premise of meeting the expression ability required by the network, try to use the small network scale and reduce the learning time and complexity of the network. From the scope of action, the receptive field obtained by using multiple small-scale convolution kernels can be the same as that obtained by one large-scale convolution kernel. In the future work, we can focus on the performance comparison between multi-scale transformation methods combined with deep network, so as to provide performance reference for the future use of multi-scale transformation combined with deep network.

Acknowledgments

The national key research and development plan “Research and development of network collaborative manufacturing platform supporting mass customization production” (project number: 2018YFB1702900)

References

[1] Li Siquan, Zhang Xuanxiong. Research on facial expression recognition based on convolutional neural network[J]. Software Guide, 2018, 17(001):28-31.

[2] Liu Shangwang, Liu Chengwei, Zhang Aili. Real-time facial expression and gender classification based on deep separable convolutional neural network[J]. Journal of Computer Applications, 2020, 040(004):990-995.

[3] Yu Ming, An Mengtao, Liu Yi. Facial expression recognition based on multiple features and convolutional neural networks[J]. Science Technology and Engineering, 2018, 018(013):104-110.

[4] He Yongqiang, Qin Qin, Wang Junpeng. Improved deep learning block convolutional neural network for facial expression recognition[J]. Computer Engineering and Design, 2019, 40(003):850-855.

[5] Du Yun, Zhang Lulu, Pan Tao. Miner facial expression recognition method based on convolutional neural network[J]. Industry and Mine Automation, 2018, 266(05):99-104.

[6] Zhang Jing. Research on facial expression recognition based on convolutional neural network [J]. Computer Knowledge and Technology, 2019, 015(016):212-213+215.

[7] Kang Jie, Li Jiawei, Yang Sili. Facial expression recognition based on domain-adapted convolutional neural network[J]. Computer Engineering, 2019, 045(012):201-206.

[8] Xia Chengjing. Research and implementation of facial expression recognition based on data augmentation and convolutional neural network[J]. Computer Knowledge and Technology, 2020, 16(03):219-221.

[9] Pan Xianzhang, Zhang Shiqing, Guo Wenping. Application of multi-mode deep convolutional neural network to video expression recognition[J]. Optics and Precision Engineering, 2019, 27(04):230-237.

[10] Xue Jiao, Zheng Jinjin. Facial expression recognition based on convolutional neural network[J]. Industrial Control Computer, 2020, 33(04):52-53+57.

[11] Fu Qianqian, Li Ang. An Improved Convolutional Neural Network Expression Recognition Algorithm[J]. Computer Technology and Development, 2020, 283(11):86-89.

[12] He Zhichao, Zhao Longzhang, Chen Chuang. Multi-resolution feature fusion convolutional neural network for facial expression recognition[J]. Progress in Laser and Optoelectronics, 2018, 630(07):364-369.

[13] Liu Quanming, Xin Yangyang. Facial expression recognition based on convolutional neural network feature graph clustering[J]. Computing Technology and Automation, 2020, 039(001):106-111.

[14] Qian Yongsheng, Shao Jie, Ji Xinxin, et al. Multi-view facial expression recognition based on improved convolutional neural network[J]. Computer Engineering and Applications, 2018, 54(24): 12-19.

[15] Yao Lisha, Xu Guoming, Zhao Feng. Facial expression recognition based on local feature fusion of convolutional neural network[J]. Progress in Laser and Optoelectronics, 2020, 663(04):338-345.