Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 9, No. 2, 2024

Research and Analysis of Facial Expression Recognition Based on Deep Neural Networks

Yali Yu

Guangdong University of Science and Technology, Dongguan Guangdong, 523668, China

Abstract: Since the traditional feature extraction algorithm cannot extract a large number of effective high-dimensional expression features, and the traditional convolutional network model has a large number of parameters in expression recognition and weak generalization ability, this paper selects the deep convolutional neural network Xception architecture as the basis for improvement. The core operations in the model are residual module and depthwise separable convolution, and ReLU6 activation function is used. The improved model is trained and tested using the public dataset CK+. Through multiple training experiments, the results show that the improved model has achieved a certain level of facial expression recognition performance.

Keywords: Expression Recognition; Neural Network; Depthwise Separable Convolution.

1. Introduction

With the development of Internet technology and intelligent machines, it is more and more important for machine intelligence to recognize inner feelings and needs through human expressions. Facial expressions are a non-verbal form of communication, but the amount of information conveyed through facial expressions is as high as 55%, indicating the importance of facial expression recognition in the process of communication. Facial expression recognition is the foundation of emotional understanding, which refers to separating specific expression states from facial images or video sequences with facial features, and then determining the psychological state and inner emotions of the recognition object, achieving computer understanding and recognition of facial expressions. Facial expressions express different emotions through the different activity states of facial muscles, such as when happy, the corners of the mouth retract and lift upwards, the cheeks lift upwards, and crow's feet increase; When feeling sad, the corners of the mouth droop, the lips are tightly closed, and the eyebrows tighten and even wrinkle into the shape of "inverted eight". In 1971, psychologists Ekman et al. [1] made groundbreaking work on facial expression recognition, first proposing that humans have six basic emotions that reflect their unique psychological activities, and dividing facial expressions into six basic expressions: happy, sad, surprised, fearful, angry, and disgusted. In recent years, the application fields of facial expression recognition have gradually expanded, such as remote classroom education, fatigue driving, medical care, etc., greatly promoting social progress and improving living standards.

There are two common methods for facial expression recognition. The first is traditional feature extraction methods, such as Local Binary Patterns (LBP), Active Appearance Models (AAM), and other algorithms. The second method is to extract facial expression features through deep neural networks, by constructing a neural network model with fewer parameters and smaller size, and training and predicting the model for facial expression recognition. Dang Xin et al. proposed a network model for facial expression recognition based on the driving status of drivers, which introduced the YOLOv5 module to enhance the network's perception ability [2]. Li Jing et al. developed a multi-scale network model for extracting global facial expression features to address the issue of facial occlusion in natural environments, which improved the performance of facial expression recognition in natural environments [3]. Li Chunhong et al. constructed a segmentation network to extract important facial features for expression recognition, and then constructed a base classifier to extract different levels of expression features. Experimental verification showed that it can effectively improve the expression recognition rate [4]. Liu Jin et al. proposed a deep convolutional residual network module to extract features, which can effectively extract subtle facial features by fusing with global features, improving the discriminative ability of facial expression changes [5].

In order to solve the problem of the large number of network parameters that cannot extract a large number of effective high-dimensional facial expression features, this paper improves the deep convolutional neural network Xception architecture. The core of the network module is the residual module and depthwise separable convolution. At the same time, the ReLU6 activation function is used to input images into the improved network model for multiple training experiments. The experimental results show that the improved model has achieved certain facial expression recognition effects.

2. Depthwise Separable Convolution

Due to the fact that the conventional convolution in general network structures processes all channels with a single convolution kernel, which not only increases the number of parameters but also has poor performance in extracting facial features, the depth separable convolution used in this paper processes one channel with a single convolution kernel. The idea of this article is to first perform convolution operations on the channel dimension (1x1 convolution), and then perform convolution operations on each dimension in the spatial dimension (3x3 convolution) [6,7], using the activation function ReLU6. In order to ensure that the data is not corrupted, no activation function layer is added, and the activation function ReLU6 is used for other layers [8]. This network module achieves decoupling by separating the channel dimension and spatial dimension, making the entire process more efficient. When building the model, all

separable convolutional layers are followed by batch normalization, and a depth multiplier of 1 is used. The SeparableConv2D layer in Keras can establish corresponding functionality. As shown in Fig. 1.

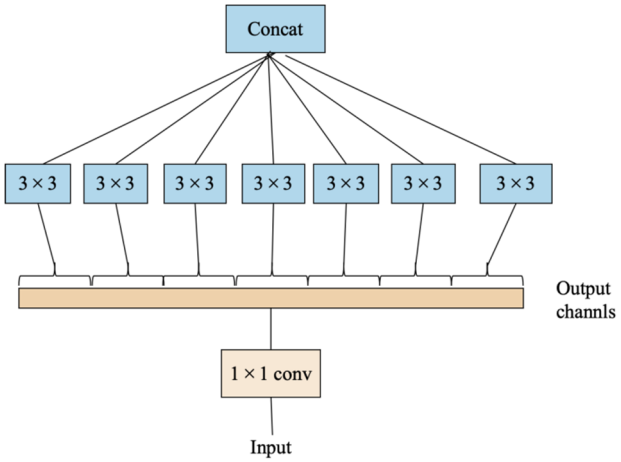

Fig 1. Depthwise separable convolution structure diagram

3. Deep Residual Network

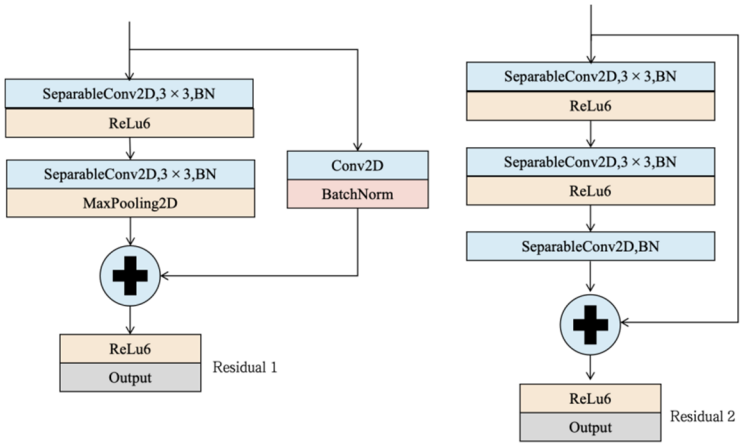

Fig 2. Residual 1 and Residual 2

In order to overcome the problems of gradient vanishing and decreased learning efficiency caused by the deepening of the network, this paper uses residual modules to construct the network structure. Residual network is a method of directly skipping multiple layers and introducing the data output from one of the previous layers into the input part of the subsequent layers. The difference between residual networks and ordinary networks is the addition of a shortcut branch, which allows the loss of the network during backpropagation training to be directly transmitted to the earlier network through this shortcut, thereby slowing down the problem of network degradation [9]. As shown in Fig. 2, it is a linear stack of depthwise separable convolutional layers with residual connections. Adding residual structures to separable convolutions can make the network converge faster and more accurately.

4. Network Structure and Experimental Analysis

Table 1. Recognition results of CK+dataset on different network models

Numble	Methods	CK+ Accuracy
1	Xception	87.24%
2	InceptionV4	89.92%
3	Our	95.1%

The Xception algorithm was proposed by a research team from Google in 2016. This article improves the model structure based on this model and conducts training and prediction experiments on facial expression images. The overall structure of the network is first passed through two conv2D convolutional layers, with 32 and 64 channels respectively, a convolution kernel size of 3 × 3, and a stride of 1; Then, the output is passed through Residual 1, Residual 2, Residual 1, Residual 2, Residual 2, Residual 1, and SeparableConv2D, with channel numbers of 128, 256, 728, 728, 728, 728, 728. The stride sizes are 2, 1, 2, 1, 1, 2, and 1, respectively; Finally, the output image is fed into GlobalAveragePooling2D and a conv2D operation with channel 7. The convolution kernel size is 1 × 1, resulting in a feature vector of 1 × 1 × 728. The dataset used in this article is CK+, and the selected expression types are angry, neutral, disgust, scared, happy, sad, and surprised. Through multiple experiments, the accuracy of facial expressions has reached 95.1%. This model reduces the number of parameters and calculations, achieving a certain level of facial expression recognition rate. As shown in Table 1.

5. Summary

With the continuous advancement of feature extraction in

neural network models, network models with small size and few parameters are receiving increasing attention. Due to the fact that traditional feature extraction algorithms spend a lot of time and cannot extract a large number of effective high-dimensional facial expression features, and traditional convolutional network models have a large number of parameters and volume in facial expression recognition, this paper conducts experiments by constructing a network model. The core operations of the model are residual module and depthwise separable convolution, using ReLU6 activation function, and training and testing the improved model using the publicly available dataset CK+. Through multiple training experiments, the results show that the improved model has achieved a certain level of facial expression recognition performance. Due to the limited image data in the CK+dataset and its proximity to natural facial expression images in daily life, the next step of the experiment will use facial expression images from daily life or videos to train the model. In addition, channel attention mechanism will be used to construct a network and data augmentation methods to expand the limited number of images in the dataset. This approach can more fully train the model and prevent overfitting of the network.

Acknowledgments

Focus analysis of cloud classroom based on expression recognition+GKY-2023KYQNK-2.

References

[1] Ekman P, Friesen W V. Constants across cultures in the face and emotion[J]. Journal of Personality and Social Psychology, 1971, 17(2): 124.

[2] Dang Xin, Xu Hua. Driving state Analysis based on Facial Expression Recognition [J]. Journal of Information Recording Materials, 2019,25(3):108-111,114.

[3] Li Jing, Li Jian, Chen Haifeng, et al. Facial expression recognition based on occlusion and reconstruction of key areas [J]. Computer Engineering, 2019,50(5):241-249.

[4] Li Chunhong, Lu Yu. Facial expression recognition based on depth-separable Convolution [J]. Computer Engineering and Design, 2019,42(5):1448-1454.

[5] Liu Jin, Luo Xiaoshu, Xu Zhaoxing. Lightweight Facial expression recognition using spatial grouping to enhance attention [J]. Computer Engineering and Applications, 2019, 59 (22):233-241.

[6] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1251-1258.

[7] Chen L, Peng L, Yao G, et al. A modified inception-ResNet network with discriminant weighting loss for handwritten chinese character recognition[C]. 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019: 1220-1225.

[8] Opschoor J A A, Petersen P C, Schwab C. Deep ReLU networks and high-order finite element methods[J]. Analysis and Applications, 2020, 18(05): 715-770.

[9] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.