

Deep Learning-Based Micro-Expression Recognition Algorithm Research

Huanjing Yang¹, Shibao Sun^{1, *}, Jingyu Chen²

¹College of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China;

²College of Political Science and Public Administration, Henan Normal University, Xinxiang 453000, China.

ABSTRACT

In order to improve the accuracy and speed of micro-expressions, a modified model based on densenet and eca is proposed. Microfacial expression is a brief, weak facial change, its characteristics are similar, dense, difficult to extract and identify, and the improved model can be adapted to the characteristics and location of the interest. In particular, the eca attention module was added after the densenet model, using the densenet network to extract the rich characteristics of micro-expressions, and the eca attention module to recalibrate the feature channel and focus on the more subtle expression changes. In order to verify the validity of this method, the experiment was conducted in the micro-emotive data set, and compared with the resnet network and the densenet network, the experimental results showed that the improved model significantly improved the performance of micro-expression recognition, and had strong generalized ability and robustness.

KEYWORDS

Micro-Expression Recognition; Deep Learning; Residual Network; Densely Connected Convolutional Networks; Efficient Channel Attention.

1. INTRODUCTION

The difference between a microexpression and a normal facial expression is that the duration of a microexpression is very short, only 1/25 to 1/5 of a second. Normal facial expressions can last longer and can even be controlled voluntarily. Microexpressions are extremely short and uncontrollable facial expressions that appear when people try to suppress or cover up their true emotions, which can just show our inner feelings and emotions. Microexpression is an important window to understand the true emotions of human beings, which makes it widely used in many application fields, such as medical diagnosis, national defense and security, education, criminal trial, criminal investigation, advertising and consumption[1-3].

With the deepening of deep learning, micro-expression recognition algorithms based on deep learning have been greatly developed. Li Sicheng[4] proposed an improved ECANet34-DA network, which added DA module and non-dimensionally reducing local cross-channel interaction strategy to the backbone of the residual network, efficient channel attention[5], which can pay attention to more subtle expression changes. Using the sequence near the peak frame to compose the intermediate frame sequence as the input image, the problem of efficient data volume is effectively solved. Ren Yu[6] proposed to use the improved residual network SE-RESNEXT-50, which was added to the SE module, to extract the features of the peak frames of microexpressions, which solved the problems of difficult recognition and low recognition accuracy due to the small number of data sets and uneven distribution

of different categories. Although the above algorithms have achieved good results for expression recognition, micro-expressions are short and hidden, resulting in difficulty in recognition and low recognition accuracy.

In this paper, a deep learning-based microexpression recognition algorithm is proposed, which combines the DenseNet network and ECA attention mechanism to effectively extract and screen the feature information in microexpressions. The DenseNet network can capture subtle changes in microexpressions and build dense feature maps. ECA attentional mechanisms can adaptively adjust the weight of feature channels, highlight features related to microexpressions, and suppress irrelevant interference. Experimental results show that the proposed algorithm is superior to other classification network models in both accuracy and robustness of microexpression recognition.

2. BASIC THEORY

2.1. Overview of deep learning background

Deep learning is a machine learning method that uses artificial neural networks to simulate the human brain, which can automatically learn and extract valuable features and rules from massive data, so as to complete a variety of difficult intelligent tasks. Microexpression recognition is a method that combines computer vision and deep learning techniques to analyze and recognize momentary and subtle emotional movements on a person's face.

Convolutional neural networks are an important milestone in deep learning, which developed from neural networks and made major breakthroughs in the field of image recognition. Convolutional neural network can automatically learn and extract image features. Its core is to capture local information of image by convolutional operation and reduce the size and complexity of image by pooling operation. Figure 1 below is a diagram of a classic convolutional neural network, which consists of an input layer, a convolutional layer, a pooling layer, and finally a fully connected layer.

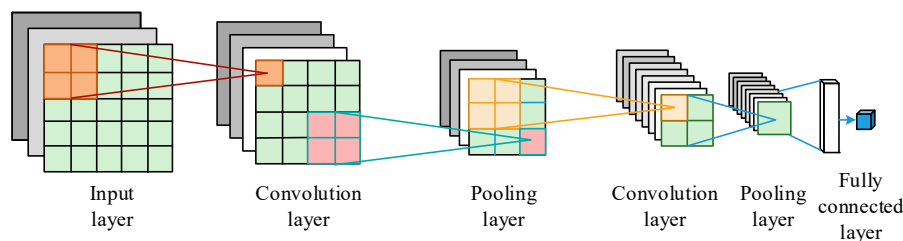


Fig.1 Convolutional neural network structure diagram

Convolutional neural network is a kind of deep learning model for processing image, speech, text and other data. Its structure is as follows: First, the original data is input and pre-processed, such as de-mean, normalization, PCA, etc., so that the data can reduce noise and redundancy, conform to certain distribution and specifications, and improve the quality and availability of data. Then, multiple convolution check data are used for locally correlated linear transformations to extract edge, texture, shape and other features in the data[7], enhancing the feature extraction capability of the model. Parameters such as the size, number, step size, and padding of the convolution kernel can be adjusted according to different tasks and data. This step can capture the local information and structure of the data and enhance the feature extraction capability of the model. Then, the output of the convolutional layer is nonlinear activated to enhance the expressibility of the model, such as using ReLU functions. Nonlinear activation can make the model fit more complex functional relationships, increase the nonlinearity and flexibility of the model, and improve the adaptability and generalization ability of the model. Then, the feature map is pooled, such as using maximum pooling, average pooling and other methods, to reduce the dimension and complexity of data, and enhance the invariance and robustness of features. Pooling operation can reduce model parameters and computation, prevent overfitting, and improve model efficiency and stability. Finally, the feature map

is flattened into a vector and input into the fully connected layer to perform tasks such as classification or regression, such as using softmax, sigmoid and other functions to output the final prediction results. The fully connected layer can map the feature vector to the target space, complete the output and evaluation of the model, and improve the accuracy and interpretability of the model. The function of convolutional neural network is to extract the deep semantic information of the original data at the input layer layer by layer and finally feed it into the classifier.

2.2. Residual network

Convolutional neural network (CNN) is a neural network architecture inspired by cerebral cortex visual research, which has excellent performance in image recognition, object detection and other fields[8]. However, with the deepening of the layers of CNN, the problems of gradient disappearance or explosion and network degradation will be encountered in the process of model training. In order to solve these problems, He Keming[9] and other scholars proposed deep residual network (ResNet). ResNet is a deep convolutional neural network architecture composed of multiple residual modules, each of which contains multiple convolutional layers and batch normalization layers. The core idea of ResNet is to add a skip connection to each residual module, so that the input can bypass the convolutional layer and directly add to the subsequent layer, so that the network learns the residual function rather than the original function, achieving the accumulation of features and the optimization of information flow. The purpose of ResNet is to solve the degradation problem in the training of deep neural networks, that is, to improve the performance of the network by increasing the depth of the network without causing the degradation of the network. The residual calculation formula is as follows:

$$H(x) = F(x) + x \quad (1)$$

Where: x is input; $F(x)$ is the residual function; $H(x)$ is the output. This formula shows how the input x is transformed by the residual function $F(x)$ to get the output $H(x)$, which is then added to the input x to form the final output. The characteristic of the residual network is that it can let the output of one layer skip several layers and directly serve as the input of a later layer, rather than the output of each layer can only serve as the input of the next layer, as in traditional neural networks. The residual module is shown in Figure 2.

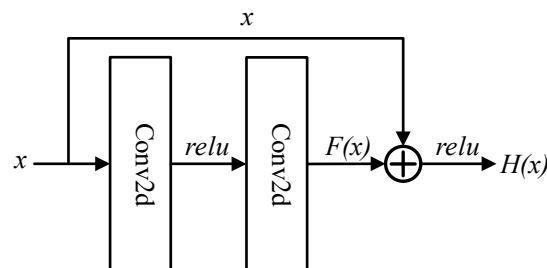


Fig.2 ResNet residual block

Specifically, the core of ResNet is the residual block, which consists of two or more convolutional layers, each of which is followed by a batch normalization layer and an activation function (usually a ReLU). Instead of simply passing the output of the convolution layer to the next layer, the output of the residual block is added to the input x and the residual function $F(x)$, and the final output is obtained by activating the function.

2.3. Dense connected network

DenseNet is a deep convolutional neural network proposed by Huang et al in 2017. It borrows from the design ideas of ResNet, using skip connections to avoid gradient dissipation, allowing the network to deepen the number of layers. However, unlike ResNet, which only creates skip connections

between adjacent layers, DenseNet connects the output of each layer with the input of all the layers that follow, forming a dense network of connections. In this way, each layer can directly obtain the feature maps of all previous layers, thus realizing feature reuse and information flow, and improving the expression ability of the network[10].

DenseNet differs from normal neural networks in that each of its layers is connected to all the others, forming $L(L+1)/2$ connections instead of just L connections. In this way, each layer can obtain the output of all previous layers. Through this connection, the feature maps of different layers are combined to realize the effect of feature reuse and improve the efficiency of the network[11]. The dense connection structure of the DenseNet network is shown in Figure 3.

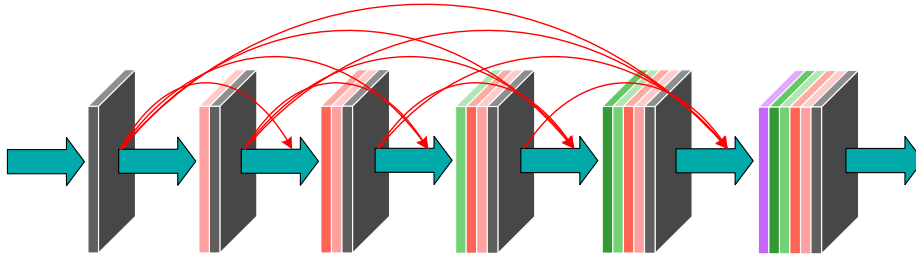


Fig.3 Dense connection structure

In DenseNet, each layer receives the feature output of all previous layers, and then extracts the feature of the corresponding layer, which is expressed as follows:

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (2)$$

$H_l (*)$ represents a function composed of multiple operations, which is a common method in convolutional neural networks, including Conv, Batch Normalization, activation function and Poolingsteps. $[X_0, X_1, \dots, X_{l-1}]$ represents the concatenation of output feature maps from layers 0 to $l - 1$ in the channel dimension, similar to the Inception model. In the previous ResNet, the output feature maps of different layers were numerically added, and the number of channels remained unchanged.

The DenseNet network model is mainly composed of several Dense blocks and a Transition Layer to connect the two dense blocks. The common DenseNet network structure is shown in Figure 4.

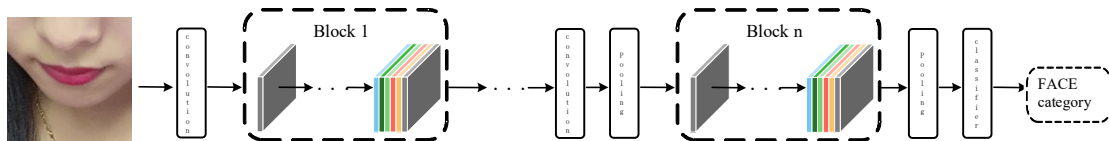


Fig.4 DenseNet network structure

The input image passes through a convolutional layer and then into several Dense Block layers and transition layers. Each Dense Block layer consists of multiple convolutional layers, and the output of each convolutional layer is connected to all the inputs of that Dense Block layer. Each transition layer consists of a convolution layer and a pooling layer, which are used to reduce the number and size of feature maps. Finally, through a fully connected layer, the softmax function is used to predict the classification results. The input data of DenseNet is two-dimensional image, which can use two-dimensional convolution kernel to carry out effective convolution operation[12].

2.4. Efficient channel attention network

The Efficient Channel Attention(ECA) module is a channel attention module that improves on the SE module[13]. The SE module is a method for learning the channel weights of convolutional blocks by applying channel compression to the input feature map and then using the full connection layer to

calculate the importance of each channel. However, this compression and dimensionality reduction results in information loss, and the fully connected layer does not capture local dependencies between different channels well. The ECA module avoids compression and dimensionality reduction, and it uses one-dimensional convolution to achieve local cross-channel interactions, thereby extracting the dependencies between channels. The advantage of this method is that it can reduce the parameters and calculation, while ensuring the effect of the model. The specific structure of ECA module is shown in Figure 5.

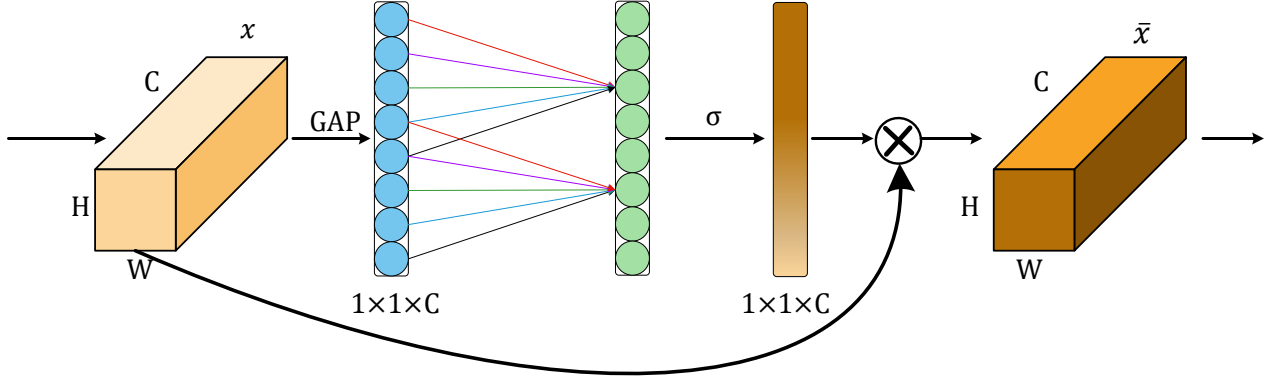


Fig.5 ECA module structure diagram

Firstly, the features are aggregated through global average pooling to obtain global channel information. The global average pooling formula is as follows:

$$y = \frac{1}{H \times W} \sum_a^H \sum_b^W x_i(a, b) \quad (3)$$

Where: x_i represents the i th feature map with input size $H \times W$, and y represents the global feature.

Secondly, the number of cross channels K is calculated adaptively using channel dimension C . The formula of the adaptive function is as follows:

$$k = \Psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{odd} \quad (4)$$

Where: $\lfloor t \rfloor_{odd}$ represents the nearest odd number of t ; C is the channel dimension; Both b and γ are constants, where $b=1$ and $\gamma=2$.

Then, the channel weights are calculated using 1D convolution with the convolution kernel size of K , and the interdependencies between channels are obtained. 1D convolution formula is as follows:

$$\omega = \sigma(C_1 D_k(y)) \quad (5)$$

Where: ω is the channel weight, σ is the Sigmoid function, and $C_1 D$ is the 1D convolution; y is the result of global average pooling; k is the size of the convolution kernel. Finally, the dot product of the original input features and channel weights is performed to obtain features with channel attention[14].

It can be seen that ECA attention mechanism is a simple and effective attention mechanism that can enhance the performance of convolutional neural networks. The core of it is to introduce a learnable one-dimensional convolution to carry out adaptive weighted features on channel dimensions, highlight important features, and improve the attention of the network to important features. Compared with the traditional attention mechanism, ECA attention mechanism has the advantages of high computational efficiency, fewer parameters, and less influence on the network processing speed. The function of ECA attention mechanism is to effectively filter the noise of non-expression region and enhance the feature expression of expression region.

3. MODEL PROPOSAL

In this paper, we propose a micro-expression recognition algorithm based on ECA-Net and DenseNet, which combines the advantages of convolutional neural networks and attention mechanisms in image classification and feature extraction. Convolutional neural networks extract deeper features by deepening the number of layers, but when the number of layers is too large, it will encounter problems such as gradient disappearance, explosion and network degradation, which will affect the accuracy of the model. DenseNet can deepen the depth of the network by increasing the number of dense blocks or the number of convolutional layers in each dense block, thereby improving the performance of the network. Microexpressions are short, weak facial changes that involve only localized muscle movements, increasing the complexity of the data and reducing the model's ability to extract key features. ECA-Net is a method to improve network performance by using channel attention mechanism, which can adaptively weight features on channel dimensions through a learnable one-dimensional convolution to highlight important features. ECA-Net has good embeddedness, which can establish the connection between feature channels, realize the control of global information, enhance the sensitivity of the model to important features, and improve the accuracy of the model. In this paper, ECA-Net is embedded in the dense connection network module, and the improved dense connection module is shown in the figure.

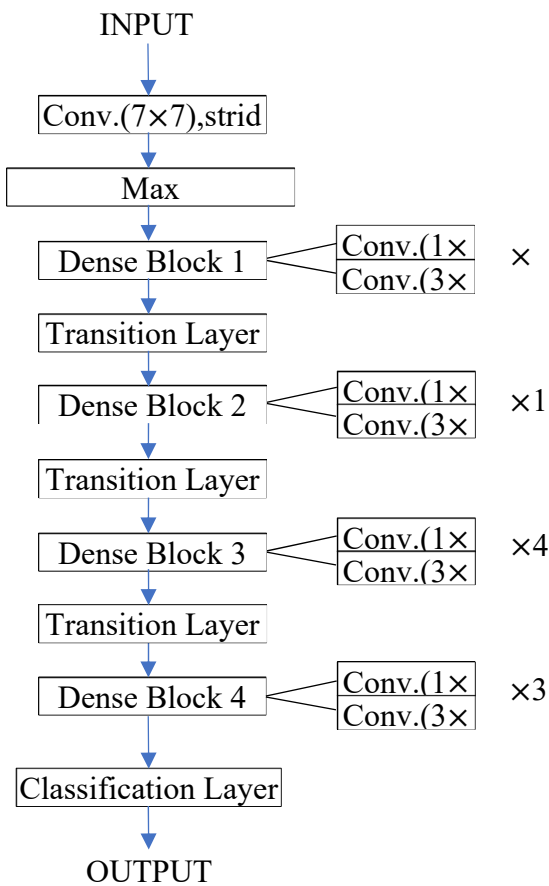


Fig.6 Layered architecture of DenseNet201 Diagram

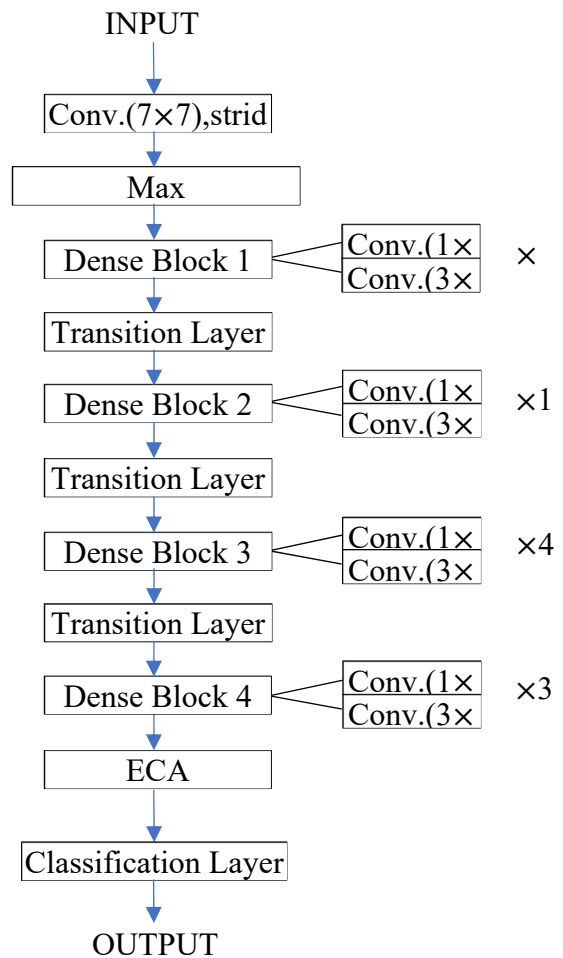


Fig.7 Improved DenseNet201 Layered Architecture Diagram

In order to improve the accuracy of microexpression recognition, an optimized network structure based on DenseNet was designed. DenseNet is a deep convolutional neural network that consists of

four densely connected modules, each connected by a transition layer. This structure can effectively integrate the features of different layers and improve the performance of the network. However, this structure does not make full use of channel information, which is very important for microexpression recognition. In order to solve this problem, this study added ECA module in front of each dense connection module to enhance the channel attention to the input features, thus enhancing the recognition ability of the model.

4. MODEL ESTABLISHMENT

This paper presents the following parts: first, a convolution layer is extracted for the input image, and then a pool layer is used to sample the feature diagram, and then several improved dense blocks are used to enhance the characteristics of the feature diagram, and finally, a full connecting layer is classified as the eigenvector, and the network structure is shown in the diagram.

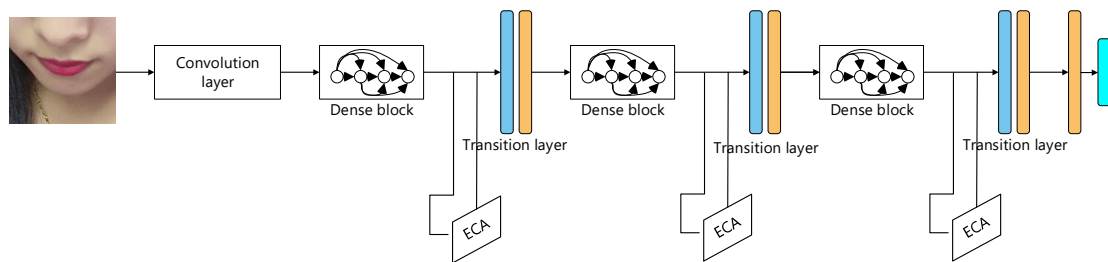


Fig.8 Effective Attention Channel Module DenseNet Structure

The proposed micro-expression recognition model based on DenseNet and ECA-Net has the following processing flow: First, the faces in the data set are preprocessed, including cropping, alignment, interpolation and amplification, so as to unify the position and size of the faces and enhance the visibility of the micro-expression. Then, the processed images are input into the convolutional neural network, and the edge features of the images are extracted through convolution and maximum pooling operations, and the redundant information is removed to improve the prediction speed of the model. Then, through several improved dense connection modules, the depth of the network is deepened and the deep features of the image are extracted. In the back of each dense connection module, the ECA-Net module is embedded to realize the channel attention mechanism, automatically select useful feature channels, suppress useless feature channels, and enhance the recognition ability of the model.

5. EXPERIMENTAL ANALYSIS

5.1. Data description

There are a total of 4 classified expression data in the data set, and the total data set is 15106. The format is 128*128.jpg images. The requirements of the training model are as follows: ① the input size is 96*96; ② Using the simplified MobileNet model, from the input to conv5_6, the last layer feature output size is 3*3 and the classification class is 4.

Tab.1 Data set of experiment

category	Training set	Test set	Total data set
Expressionless	4287	476	4763
pout	2839	315	3154
smile	4357	484	44841
laugh	2114	234	2348

5.2. Experimental environment

This experiment is based on Windows10 operating system, uses tensorflow deep learning framework, uses python version 3.7, and uses Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59GHz. The graphics is NVIDIA GeForce RTX 2060.

5.3. Identification process

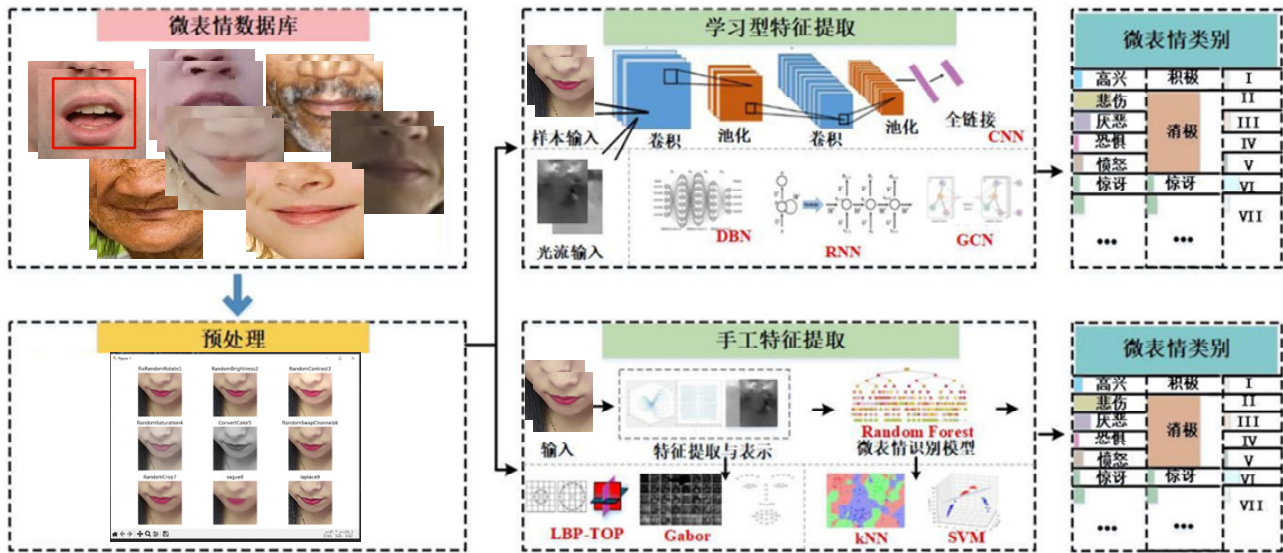


Fig.9 Schematic diagram of micro expression recognition process

5.3.1. Data processing

The original data needs to undergo some necessary processing in order to facilitate subsequent feature extraction and classification. The purpose of data preprocessing is to eliminate noise and interference in the data, strengthen the microexpression signal in the data, align the face and frame number in the data, and improve the availability and consistency of the data. There are many methods for data preprocessing, such as face detection, face registration, time domain image interpolation, action amplification, etc. Face detection refers to the recognition of the position and size of the face from the image, face registration refers to the alignment of the feature points of the face to a standard position, time-domain image interpolation refers to the insertion of new frames between the video frames, so that the frame rate of the video is consistent, action amplification refers to the method of signal processing to enlarge the small changes in the video, so that the micro-expression is more obvious.

5.3.2. Feature extraction

Microexpression recognition is a method to analyze and identify the small and imperceptible changes in facial expression produced in a short period of time. In this paper, we propose a microexpression recognition method that combines DenseNet network and ECA attention mechanism, which can extract effective features from facial microexpressions. Specifically, this approach consists of the following steps:

Firstly, the facial feature point localization algorithm is used to extract the lip Angle, lips and other key points of the face, and the image of the mouth area is cut out according to these key points. Then, the DenseNet network is used to extract convolutional features from the images of the mouth region. DenseNet network is a densely connected convolutional network, which connects the output of each layer with the input of all subsequent layers, realizes feature reuse and propagation, and enhances the depth and expressiveness of the network. Then, ECA attention mechanism is used to adjust the channel attention of convolutional features. ECA attention mechanism is an efficient channel

attention module, which avoids dimension reduction, realizes local cross-channel interaction through one-dimensional convolution, adaptively determines the size of convolutional kernel, and realizes the refinement and adaptive learning of channel weights. Finally, the convolutional features adjusted by ECA attention mechanism are spliced into a feature vector, which is used as the feature representation of mouth microexpressions and input into a classifier for the recognition of mouth microexpressions. The advantage of this method is that it can make full use of the depth and density of the DenseNet network, as well as the light weight and efficiency of the ECA attention mechanism, to extract features that can reflect the movements and emotions of the mouth microexpression, thereby improving the performance of mouth microexpression recognition.

5.4. Experimental analysis and comparison

Microexpression recognition is a method of analyzing and identifying small and imperceptible changes in facial expression produced in a short period of time through computer vision and machine learning technology. The performance of microexpression recognition can be measured by accuracy, that is, the proportion of microexpression categories that the classifier correctly identifies. In this paper, we use three different network structures to carry out experiments on microexpression recognition, they are ResNet network, DenseNet network and DenseNet network combined with ECA attention mechanism. The results of our experiments on these three networks are shown in the table below.

Tab.2 experimental result

method	Accuracy rate
ResNet	93.241
DenseNet	95.494
DenseNet+ECA	95.759

ResNet network is a kind of network that uses residual connection to solve the problem of deep network degradation, which can make a direct channel between the input and output of the network, thus accelerating the propagation and optimization of features. In this paper, ResNet network is used to carry out micro-expression recognition experiment, and the accuracy rate is 93.241, which indicates that ResNet network has a certain effect on micro-expression recognition, but there is still room for improvement.

DenseNet network is a kind of network that enhances feature extraction and propagation ability through dense connection. It can connect the output of each layer with the input of all subsequent layers to realize feature reuse and propagation, and enhance the depth and expression ability of the network. In this paper, DenseNet network was used to carry out the experiment of micro-expression recognition, and the accuracy rate was 95.494, which was significantly improved compared with 93.241 of ResNet network, indicating that DenseNet network could better extract features in micro-expression, thus improving the accuracy rate of micro-expression recognition.

ECA attention mechanism is a simple and efficient channel attention module, which can realize the fine and adaptive learning of channel weights by using a learnable one-dimensional convolution. In this paper, ECA attention mechanism is combined with DenseNet network to adjust the channel attention of the convolutional features of DenseNet network to improve the feature differentiation and selectivity. In this paper, the ECA-Densenet network is used to carry out the experiment of micro-expression recognition, and the accuracy rate is 95.759, which is slightly higher than the 95.494 of DenseNet network, indicating that ECA attention mechanism can optimize the feature extraction of DenseNet network to a certain extent. Thus, the accuracy of microexpression recognition can be further improved.

To sum up, this paper uses a network structure that combines the DenseNet network and ECA attention mechanism to carry out the experiment of microexpression recognition, and its accuracy

rate reaches 95.759, which is the highest among the three network structures. This shows that this network structure can effectively extract and recognize the emotion categories in microexpressions, and has a strong ability to recognize microexpressions.

In order to verify the superiority of the network combining the DenseNet network and ECA attention mechanism in micro-expression recognition, the accuracy of different network structures is compared on the experimental data set. It can be seen that the network with the combination of DenseNet network and ECA attention mechanism has the highest accuracy, which is obviously better than other network structures. The specific comparison results are shown in the figure.

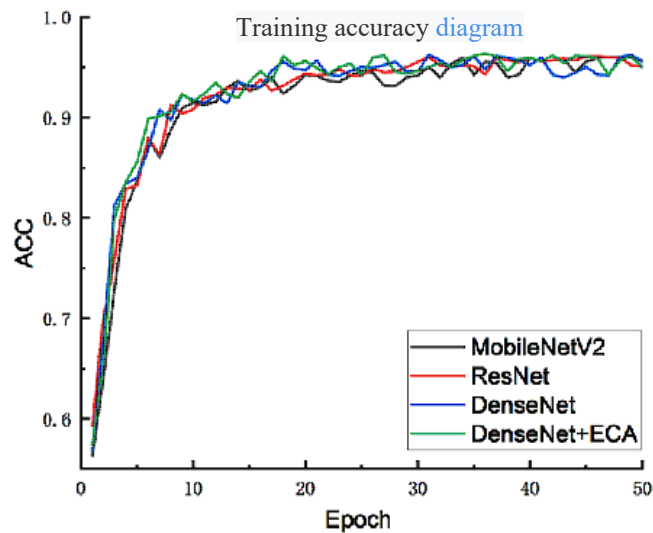


Fig.10 Training accuracy chart

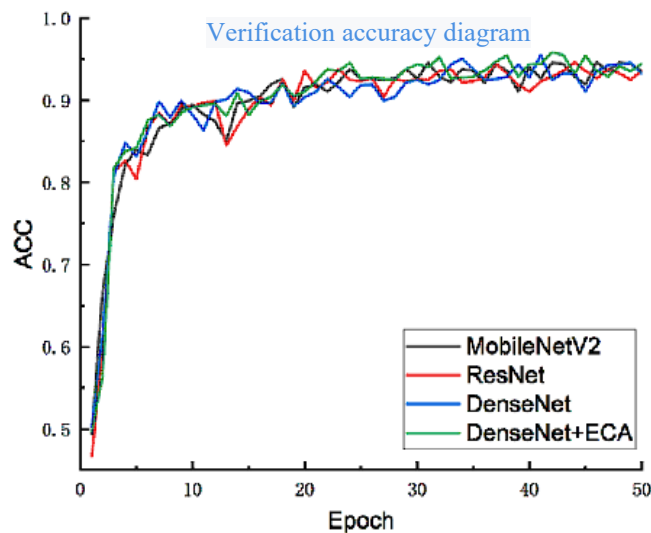


Fig.11 Verification accuracy diagram of the model

6. CLOSING REMARKS

Aiming at the challenging task of computer vision, this paper proposes a deep learning-based microexpression recognition algorithm, which combines the DenseNet network and ECA attention mechanism to effectively extract and screen the feature information in microexpressions. The results show that the proposed algorithm is superior to other classification network models in the accuracy and robustness of microexpression recognition. The main contributions and innovations of this paper are as follows:

①In this paper, DenseNet network is applied to micro-expression recognition for the first time, and its powerful feature extraction and efficient self-learning ability are used to capture the subtle changes of micro-expression and construct intensive feature maps.

②In this paper, ECA attention mechanism is introduced into microexpression recognition for the first time. It adaptively adjusts the weight of feature channels to highlight features related to microexpression and suppress irrelevant interference.

③In this paper, the effectiveness and superiority of the algorithm are verified by experiments on two microexpression data sets, and the influence of different parameters and modules on the performance of the algorithm is also analyzed.

The work of this article provides a new way of thinking and method for microexpression recognition, but there are still some shortcomings and improvement Spaces, such as:

①The algorithm in this paper only considers the recalibration of the feature channel, and does not consider the recalibration of the feature space, which may ignore some important spatial information.

②The algorithm in this paper is only tested on two microexpression data sets, and other types of microexpression data, such as dynamic microexpression and compound microexpression, are not considered, which may affect the generalization ability of the algorithm.

③The algorithm in this paper only uses a single network model and does not consider the fusion of multiple models, which may limit the performance improvement of the algorithm.

Therefore, in future work, this paper will further explore the re-calibration of feature space, expand the scope of experimental data, and try to integrate multiple models, in order to improve the effect and application value of microexpression recognition.

REFERENCE

- [1] Xie, T.,Sun, G., Sun, H., Lin, Q., & Ben, X.(2022).Decoupling facial motion features and identity features for micro-expression recognition. PeerJ Computer Science, 8, e1140.
- [2] Zeng X, Zhao X, Zhong X, et al. A Survey of Micro-expression Recognition Methods Based on LBP, Optical Flow and Deep Learning[J]. Neural Processing Letters, 2023: 1-32.
- [3] Tang M, Ling M, Tang J, et al. A micro-expression recognition algorithm based on feature enhancement and attention mechanisms[J]. Virtual Reality, 2023: 1-12.
- [4] Li Sicheng, Zhou Shunyong, Zhu Hao, et al. Microexpression recognition combining multi-attention mechanism and intermediate frame sequence [J]. Radio Engineering,2023,53(3):636-643. DOI:10.3969/j.issn.1003-3106.2023.03.017.
- [5] Li Wentao, Peng Li. Small Target Detection Algorithm for Multi-scale Channel Attention Fusion Networks [J]. Computer Science and Exploration,15(12):2390-2400. DOI:10.3778/j.issn.1673-9418.2011028.
- [6] Ren Yu, Chen Xinquan, Wang Dalong et al. Improved Residual Network and Peak Frame Microexpression recognition [J]. Journal of Chongqing Technology and Business University (Natural Science Edition).
- [7] Liu Fang, Li Ge, Hu Xing, et al. Research progress of program Understanding based on Deep learning [J]. Computer Research and Development, 2019,56(8):1604-1620. DOI:10.7544/issn1000-1239.2019.20190185.
- [8] Bi Mei. Research on Transfer learning and Multi-Object Recognition in Facial Expression Domain [D]. Jilin University,2023.DOI:10.27162/d.cnki.gjlin.2023.006925.
- [9] HE K M, ZHANG X Y REN S Q, et al. Deep residual learning for image recognition [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway , NJ, USA IEEE, 2016.
- [10] Lv Peng. Research on Facial Expression Recognition Algorithm Based on Deep Learning [D]. Nanjing University of Posts and Telecommunications,2022.DOI:10.27251/d.cnki.gnjdc.2022.001195.
- [11] Liu Jin. Research on Facial Expression Recognition based on Deep Convolutional Neural networks [D]. Guangxi Normal University,2023.DOI:10.27036/d.cnki.ggxsu.2023.000846.

- [12] Yan Chunman, Wang Cheng. Development and application of Convolutional Neural network model [J]. Journal of Frontiers of Computer Science & Technology, 2021, 15(1).
- [13] Luo Sishi, LI Maojun, Chen Man. Facial expression Recognition Networks with multi-scale integration of attention mechanisms [J]. Journal of Computer Engineering & Applications, 2023, 59(1).
- [14] Chen Jiamin, Xu Yang. Expression recognition in attention pyramid Convolutional residual networks [J]. Journal of Computer Engineering & Applications, 2022, 58(22).