International Journal of Computational Intelligence Systems and Applications. Volume 11, Issue 1, (2024) 50-57
http://www.cayley-nielson.com/ijcisa/

Facial Expression Recognition Based on Mini-Xception Network

Cui Dongyan[1], Hou Shen[1]
[1] School of Artificial Intelligence, North China University of Science and Technology ,Tangshan, Hebei, China
E-mail: *cdy_xxz@163.com*

Abstract

Traditional facial expression recognition cannot obtain deeper high semantic features and their deep features from the original image, resulting in low accuracy. However, deep learning methods based on convolutional neural networks have become a feasible solution to the problem of facial expression recognition. This article is based on the Xception neural network and simplifies its network structure by removing the fully connected layers in traditional neural networks and replacing them with depthwise separable convolutions. Finally, a network model called Mini Xception is constructed for facial expression classification and recognition tasks. This model extracts features from the input image through convolution operations and trains and classifies the model through depthwise separable convolution. Its structure is simple, with good efficiency and accuracy.

Keywords: Facial Expression Recognition, Expression Recognition System, Deep Learning, Mini-Xception.

1. Introduction

In the 1970s, American scholar Ekman conducted extensive facial expression recognition experiments and for the first time classified facial expressions into six basic forms, including sadness, happiness, fear, disgust, surprise, and anger [1]. In 1978, researchers first proposed the recognition of facial expressions and took the first step in recognizing facial expressions. They drew design inspiration from animated videos, proposed relevant theories for expression recognition, and conducted further analysis, flexibly utilizing the functions of image encoding sequences [2].

The rapid development of facial expression recognition technology has attracted the attention of many experts and scholars [3-5]. Among them, some researchers have proposed different algorithms to solve the problem of facial expression recognition. For example, in 2008, S Bashyal et al. proposed using Gabor transform for facial expression feature extraction, and found that Gabor wavelet has good performance in facial expression feature extraction, and the extracted features also have good classification characteristics [6]. Almaev et al. combined the previously successful expression recognition based on LGBP with TOP extensions of other descriptors to propose a new dynamic appearance descriptor - Local Gabor Binary Patterns from Three Orthogonal Planes (LGBP-TOP). LGBP-TOP combines spatial and dynamic texture analysis with Gabor filtering, achieving unprecedented real-time recognition accuracy [7].

With the further development of research, the advantages of deep learning in the field of images are becoming increasingly prominent. The problems of cumbersome steps, low recognition rates, and weak robustness in traditional face recognition techniques are becoming increasingly difficult to handle,

especially the advantages of convolutional neural networks in the field of images, which make deep learning play an increasingly important role [8]. Reference [9] aims to address the issue of class imbalance in facial expression datasets by developing and implementing a deep learning-based classification model that uses synthetic images generated through Generative Adversarial Networks. The goal is to improve recognition accuracy for each expression.

Traditional facial expression recognition algorithms require a lot of manual design and tuning, and cannot automatically learn and summarize features, requiring a lot of time. This method requires a lot of manual design and tuning, and often cannot accurately recognize complex facial images, requiring more time and lower efficiency. Facial recognition algorithms using deep learning can recognize changing factors such as angles, lighting, facial expressions, age, and skin color, and have adaptability and learning ability. Therefore, this article is based on Mini Perception for facial expression recognition.

2. Facial expression recognition based on Mini-Xception Network

2.1. Dataset

This article uses the Fer2013 dataset, which is divided into three parts: training set, validation set, and testing set. Composed of 35887 facial images, 28709 are used for training, 3859 are used for validation, and another 3859 are used for testing. Fer2013 dataset is composed of images published on the Internet, which has higher complexity and diversity, and is more consistent with face images in practical application scenarios. The Fer2013 dataset has more expression categories, which can enable algorithms to perform emotion recognition more comprehensively and accurately.

2.2. Data Preprocessing

Before establishing a neural network model, the collected image data needs to be preprocessed first. Preprocess the detection, normalization, and data enhancement of facial positions to ensure the usability and trainability of the face, in response to uneven lighting and angular deviation in the original image.

Firstly, it is necessary to perform image normalization on the dataset. By normalizing the images, the contrast of the images can be increased, making the targets more prominent. Image normalization is a commonly used preprocessing step that has a certain impact on the performance and effectiveness of various algorithms. The linear normalization formula is shown in (1).

$$X_{\text{norm}} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Image enhancement technology is a method of generating more similar but different images by performing a series of processing on the original image, thereby expanding the size of the training dataset. Through these operations, the dataset size can be expanded, the training samples of the model can be increased, the problems of overfitting and underfitting can be avoided, and the generalization performance of the model can be improved. The dataset size is relatively small, and using image enhancement techniques can expand the dataset size without adding new data, improving the training ability and accuracy of the model. For example, by flipping or rotating images, images from different angles can be generated, increasing the diversity of the dataset. By cropping images, the size of the dataset can be increased without losing target information. By adjusting parameters such as brightness and contrast, the dynamic range of the image can be increased and the adaptability of the model can be improved.

The data augmentation effect is shown in Fig. 1.

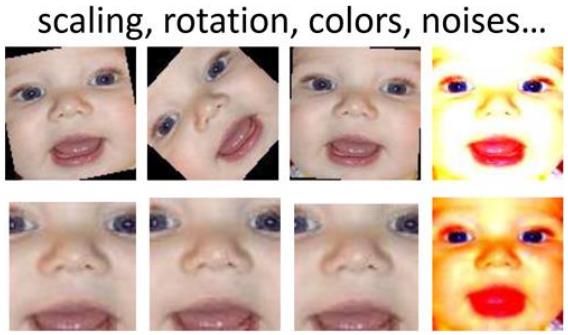

Fig.1. Data augmentation effect

2.3. Building a Mini-Xception Network Model

Facial expression recognition is a multi-classification problem, and increasing the number of hidden layers in the network can enable the model to learn and process abstract features of data more deeply. However, this can also lead to problems such as vanishing and exploding gradients, and long training times. This article uses the Mini Perception model, which is modified based on the Xception model and utilizes techniques such as depthwise separable convolution and global average pooling. This can further reduce the number of model parameters and computational complexity while maintaining high accuracy, allowing the model to run in real-time on embedded or mobile devices.

As shown in Fig. 2, first establish a Mini Perception convolutional neural network model and perform data augmentation processing such as scaling and rotation on the data to increase the data volume. By incorporating data into the constructed network model, feature extraction and classification of facial expressions can be performed. Train the network using category labels derived from the dataset, obtain the optimal training model after training, and finally evaluate and test it. During the testing period, the neural network will maximize the predicted confidence value of a certain dimension in the output feature vector to determine the final classification of the input expression image. By using the Softmax activation function, 7-dimensional feature vectors of 7 different expression types were obtained.

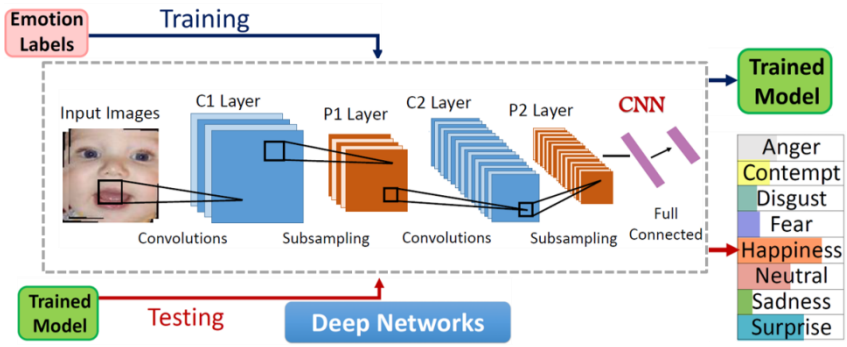

Fig.2. Process of building models

3. Experimental Results and Analysis

3.1. Model Training

Build a model on the Keras framework and train it using the preprocessed dataset. The training parameters include batch size, number of training rounds, number of output categories, etc. The parameter settings are shown in Table 1.

Table 1. Training parameter settings

Parameter	Value
Batch Size	32
Epoch	10000
Patience	50
Num Classes	7

The Fer2013 dataset was expanded through data augmentation during model training, and the parameter settings for data augmentation are shown in Table 2.

Table 2. Data augmentation parameter settings

Expansion type	Random rotation of images / °	Horizontal Offset	Vertical offset	Flip horizontal	Scale
Parameter	[-10, 10]	0.1	0.1	Yes	0.1

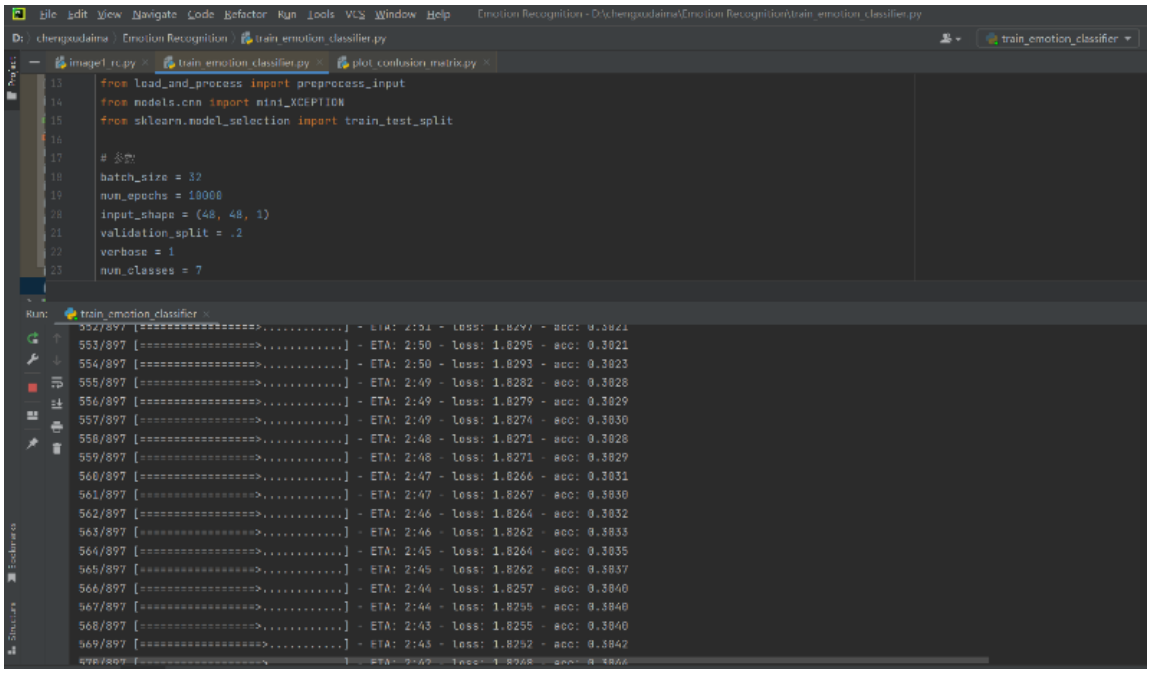

Fig.3. Output Results of Training Process

When compiling the model, Adam optimizer (adaptive moment estimation) was used to optimize the loss function of the model, categorical cross entropy was used as the loss function, and accuracy was used as the evaluation metric. The Adam algorithm is a commonly used optimization algorithm that can be applied to the training process of almost all deep learning models. It combines gradient descent and momentum methods, while also using adaptive learning rates. Compared to traditional gradient descent methods, Adam is more convenient and efficient, can find the global optimal solution faster, and is more robust to hyperparameter selection. Adam's parameter update formula is:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\widehat{v_t}}}\widehat{m_t} \tag{2}$$

Among them, θ is the parameter to be updated, η is the learning rate, $\widehat{m_t}$ is the mean of the gradient at the first time, $\sqrt{\widehat{v_t}}$ is the variance at the second time, and t is the time step.

The model needs to minimize the loss function during the training process to achieve the best fitting effect on the samples and improve the accuracy of predictions. The formula is as follows:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^i \log\left(h_\theta(x^i)\right) + (1-y^i)\log\left(1-h_\theta(x^i)\right)\right] \tag{3}$$

After setting the parameters and models for training and data augmentation, the program code can be run to train the model, as shown in Fig. 3, which shows the training process and changes in accuracy.

Accuracy is one of the indicators to measure the performance of a model. It can help us understand the performance of the model and also assist us in selecting model parameters, adjusting network structure, and other tasks. As the training progresses, the accuracy of the training and validation sets gradually improves, and both curves eventually reach a steady state. The resulting training curve is shown in Fig. 4.

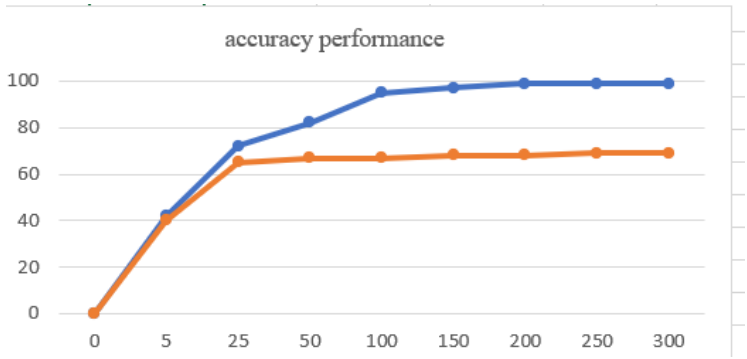

Fig.4. Training Curve

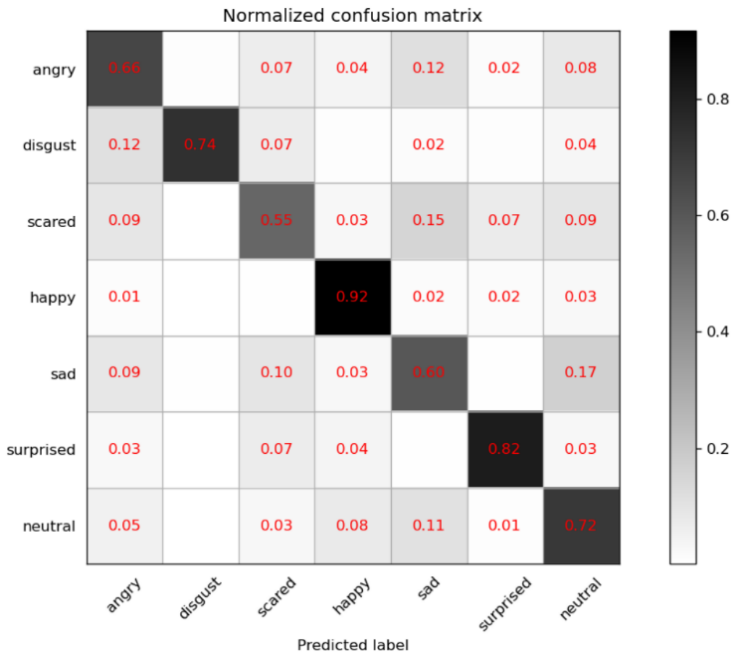

Fig.5. Test set confusion matrix results

3.2. Result Analysis

In facial expression recognition tasks, confusion matrix is one of the commonly used evaluation methods. Usually, facial expression recognition tasks are multi class tasks, with the number of categories classified equal to the number of different facial expressions to be recognized. The rows and columns of the confusion matrix represent seven different types of expressions, such as anger, disgust, fear, happiness, sadness, surprise, and neutrality. By analyzing the confusion matrix, we can better understand the classification performance of the model. The confusion matrix is an N × N matrix, where N represents the number of categories classified. In this design, the category of classification is facial expressions, so N equals the number of expressions. The result of the confusion matrix is shown in Figure 5.

Among them, the Mini-Xception model has a recognition rate of 66% for angry expressions, 74% for disgusted expressions, 55% for fearful expressions, 92% for happy expressions, 60% for sad expressions, 82% for surprised expressions, and 72% for neutral expressions. The average recognition accuracy calculated is 71.57%, and the overall recognition result is relatively accurate. The recognition probability of different expressions varies, which may be due to the uneven number of different expressions.

In order to verify the effectiveness of the algorithm and eliminate the randomness of the results, for static images: in this design, a total of 140 expression images were collected and divided into 7 types for expression recognition testing. The accurate number of recognition results for each type of expression was plotted in a bar chart, as shown in Figure 6, with 20 samples for each type of expression. Based on the recognition results, it can be seen that this method has the highest accuracy in recognizing happy and normal expressions, with 18 and 18 being recognized respectively; Sadness and anger can accurately identify 17 and 17; 16 expressions of surprise can be accurately identified; The recognition rates of fear and disgust are the lowest, with 15 and 15 being recognized respectively. According to calculations, the average recognition rate reaches 80.7%, indicating excellent recognition performance. Sometimes, fear can be mistaken for surprise, which may be due to differences in facial expressions among different people, as well as environmental factors such as image quality and lighting conditions. Overall, the facial expression recognition system can accurately recognize various expressions.

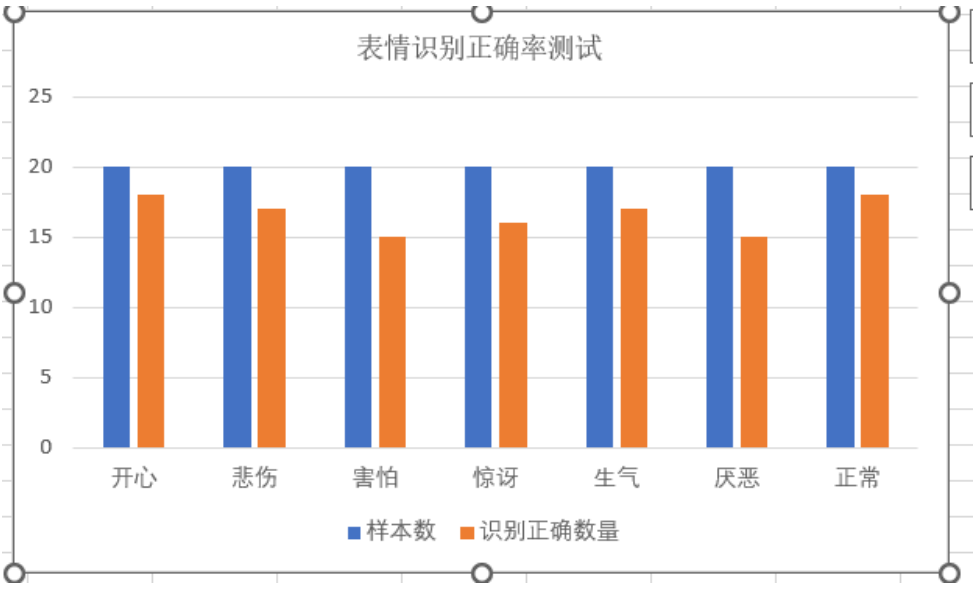

Fig.6. Test results of static expression recognition accuracy

For dynamic video expressions: In order to evaluate the stability and generalization performance of the model under different backgrounds, lighting, personnel, poses, angles, and other conditions, comparative experiments are needed. Therefore, different people were selected to record videos with different poses and angles in different backgrounds. Screenshots were taken from the videos, and 20 facial expressions were selected for each category. 19 happy expressions, 18 normal expressions, 17 angry and sad expressions, 15 scared and surprised expressions, and 14 disgusted expressions were correctly recognized. As shown in Figure 7. Through statistics, it can be concluded that the average recognition accuracy of the entire real-time facial expression recognition system is 82.1%, indicating that the system has high recognition accuracy in practical applications.

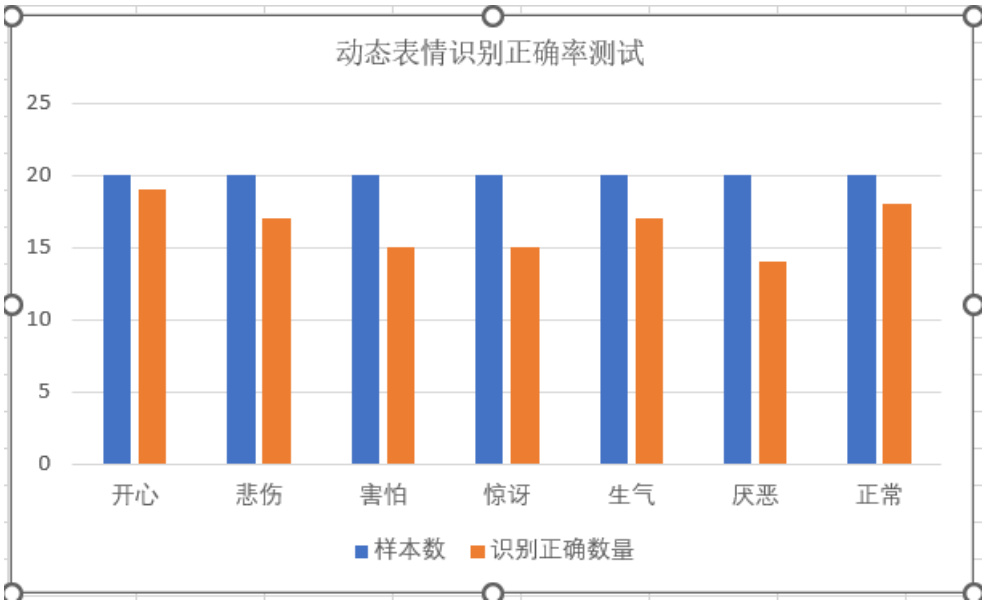

Fig.7. Test results of dynamic expression recognition accuracy

According to the experimental results, it can be concluded that the model can effectively extract emotional information that humans can understand from facial expressions and make predictions on the captured emotions.

4. Conclusion

This design introduces the facial detection methods, feature extraction principles, and classification algorithms used. Based on the above theory, a Mini-Xception architecture was constructed, which uses Convolutional Neural Networks (CNN) to extract features from facial expression images and has strong learning ability and computational efficiency. It has the following advantages:

(1) High accuracy: The Mini-Xception model has extremely high accuracy for facial expression classification, reaching around 72%. It achieves very high accuracy through lightweight design, data augmentation, etc., and can be widely applied in practical life.

(2) Efficiency: The Mini-Xception model has a smaller model size and parameter count, allowing for training and testing to be completed in less time, resulting in higher efficiency.

(3) Scalability: The facial expression recognition method based on Mini-Xception can be expanded and optimized by increasing the dataset and adjusting model parameters, and has good scalability.

References

[1] Prkachin Kenneth M.The consistency of facial expressions of pain: a comparison across modalities. *No longer published by Elsevier*,1992,51(3):382-391.

[2] Anna Tcherkassof, Damien Dupré. The emotion–facial expression link: evidence from human and automatic expression recognition. *Psychological Research*,2020.

[3] Tanja S.H., Wingenbach, Chris, Ashwin, et al. Diminished sensitivity and specificity at recognising facial emotional expressions of varying intensity underlie emotion-specific recognition deficits in autism spectrum disorders. *Research in Autism Spectrum Disorders*,2017,34(3): 12-18.

[4] Aihua Li, Lei An, Zihui Che. A Facial Expression Recognition Model Based on Texture and Shape Features. *Traitement du Signal*,2020,37(4): 4-15.

[5] Jenni Kommineni, Mohd Shahrizal Sunar, Parvathaneni Midhu, et al. Accurate computing of facial expression recognition using a hybrid feature extraction technique. *The Journal of Supercomputing*,2020,77(5): 26-29.

[6] ShishirBashyal,Ganesh K. Venayagamoorthy. Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence*,2007,21(7):1056-1064.

[7] Almaev T R , Valstar M F .Local Gabor Binary Patterns from Three Orthogonal Planes for Automatic Facial Expression Recognition, *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on. IEEE Computer Society*, 2013.

[8] Pablo Barros, Nikhil Churamani, Alessandra Sciutti. The FaceChannel: A Fast and Furious Deep Neural Network for Facial Expression Recognition. *SN computer science*,2020,1(6): 1-3.

[9] Rani R , Arora S , Verma S S R .Enhancing facial expression recognition through generative adversarial networks-based augmentation. *International journal of systems assurance engineering and management*, 2024, 15(3):1037-1056.