International Journal of Computer Applications Technology and Research
Volume 12–Issue 04, 55 - 60, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1204.1012

Facial Expression Recognition Based on Improved Densenet Network

Tao JunChen
Chengdu University of
Information Technology
Chengdu, China

Abstract: As a branch of image recognition, facial expression recognition helps to carry out medical, educational, security and other work more efficiently. This article combines deep learning knowledge and conducts research on expression recognition based on DenseNet121, a dense convolutional neural network that integrates attention mechanisms for multi-scale feature extraction. Firstly, in response to the insufficient ability of DenseNet121 to extract complex facial expression features, multi-scale feature extraction dense blocks were introduced to replace DenseBlocks used to extract features of different sizes; Secondly, using multi-scale feature extraction convolutional blocks to replace the large convolutional kernel at the head of DenseNet121 further enriches feature extraction; Finally, in order to extract more important features from the channel dimension, we consider combining ECA channel attention mechanism to help improve model performance. The experiment proves that the model proposed in this chapter has improved recognition accuracy by 2.034% and 3.031% compared to DenseNet121 on the FER2013 and CK+datasets, respectively. It also has certain advantages compared to other commonly used classification models.

Keywords: Facial expression recognition, Deep Learning, Convolutional Neural Network,DenseNet

International Journal of Computer Applications Technology and Research
Volume 12–Issue 04, 55 - 60, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1204.1012

1. INTRODUCTION

There are various ways for humans to convey emotions, such as language, voice, facial expressions, and body movements. Research has shown that among various emotional expression methods, facial expressions carry the most abundant emotional expression and can comprehensively reflect the emotions and psychological status of the speaker at that time. Facial expressions, as an important way of conveying information between people in our daily lives, also convey a rich and efficient amount of information. Therefore, studying facial expression recognition has significant social significance and practical value for our human emotional analysis and psychological assessment. The field of facial expression recognition, as an important component of image recognition technology, is showing an unprecedented development trend. Currently, large-scale software and hardware foundations have been fully popularized. With the support of hardware foundations, the application market of facial expression recognition is very broad, and the demand in various fields is also very large. This technology can be applied to multiple fields, such as medical treatment, education, driving safety, etc. Traditional facial expression recognition algorithms mainly use manually designed feature extraction algorithms, such as LBP, HOG, and other feature operators, which have average extraction results and low efficiency. After the emergence of deep learning technology, it can perform end-to-end feature extraction and classification work, opening up a new direction for the field of facial expression recognition. This article takes into account the shortcomings of traditional feature extraction algorithms, and uses the DenseNet convolutional neural network model in deep learning to conduct research. In order to enrich its feature level extraction capabilities, multi-scale feature extraction convolutional blocks, multi-scale feature extraction dense blocks, ECA channel attention mechanism are introduced, and an improved DenseNet model is constructed and obtained, The improved DenseNet model has better feature extraction capabilities in the field of facial expression recognition.

2. RELATED WORK

The research on facial expression recognition was first proposed by British biologist Darwin in the 19th century. Early facial expression recognition work belongs to the field of psychology. By studying the relationship between facial expressions and psychological states, it helps to discover the relationship between facial expressions and psychological states and explore some observation and treatment methods. In 1971, psychologist Ekman[1] and his research partner Friesen proposed a facial action encoding system by dividing facial muscle action units and analyzing the corresponding relationships between different expression categories and facial muscle units. The system marked the facial muscle action units and provided corresponding formulas for the expression types obtained from different combinations of markers, which had a certain impact on subsequent facial expression recognition work. Until 1978, Suwa et al. [2] were the first to use computer technology for facial expression recognition, with the aim of achieving efficient automatic recognition of facial expressions. In 1991, Mase et al. [3] proposed an optical flow method for facial expression recognition using optical flow as the expression feature of expressions. The development of computer technology has driven the maturity of facial expression recognition technology. Traditional methods commonly use handmade features, and Gabor filters [4] have shown certain advantages in feature extraction for facial expressions, improving the recognition rate of facial expressions, but the algorithm complexity is relatively high. Other traditional manual feature algorithms include extracting local grayscale features using local binary patterns [5] to assist in recognition, and finally using SVM [6] for classification, LBP-TOP [7] on three orthogonal planes, etc. Luo et al. [8] proposed an improved principal component analysis facial expression recognition algorithm for static expression recognition. The grayscale features extracted from local binary patterns were used to assist in the global grayscale processing of facial expression recognition, and SVM was used for classification. Simulation experiments showed that this method can effectively classify different expressions.

In recent years, with the development of modern information technology and the continuous improvement of computer computing power, deep learning technology has rapidly developed and widely applied in the field of image processing. The focus of facial expression recognition has gradually shifted from traditional manual feature operators to generalization and strong classification capabilities in neural networks. By using computer vision algorithms and deep learning algorithms, visual observations similar to those of the "human eye" can be achieved in some scenes in daily life, Recognition functions, such as facial recognition for entering and exiting large venues, facial expression recognition for analyzing emotions, and scene recognition for the environment, can be applied. The rapid development of deep learning has led foreign scholars to use this technology to carry out facial expression recognition work on large datasets, and have made certain progress. For example, Lopes et al. [9] preprocessed facial expression data, constructed multi-layer convolutional neural networks for feature features, and inputted the extracted features into a classifier for facial expression classification. Experiments have shown that good recognition rates have been achieved on the CK+dataset. Later, the development direction of deep learning in classification tasks was mainly to improve recognition rate by deepening the depth and width of the network. A series of models began to emerge, such as AlexNet, VGGNet, ResNet, MobileNet, DenseNet, etc. The models had better performance in classification tasks.

3. PROPOSED METHOD

3.1 Multi scale feature extraction of dense blocks

In response to the lack of complex feature extraction ability in the DenseNet121 model, in order to make the features extracted by DenseNet121 richer and more significant, this paper proposes a multi-scale feature extraction dense block, which improves the structure of the DenseLayer single branch in the original DenseNet121 dense block Denseblock. As shown in Figure 1,

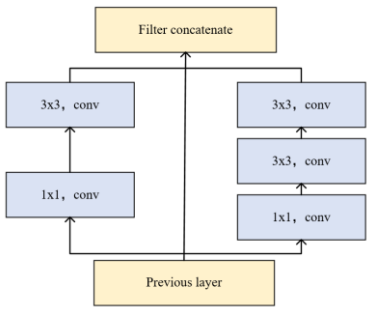

Figure. 1 Multi scale feature extraction of dense blocks

the number of convolutional kernels in the original Denselayer is reduced by half, At the same time, another convolution branch is developed in the feature extraction process. The new branch is composed of two convolutions of size 3 and one convolution of size 1. The purpose of this design is to provide

International Journal of Computer Applications Technology and Research
Volume 12–Issue 04, 55 - 60, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1204.1012

different scales of Receptive field for feature extraction, compared with the 3 provided by the original single trunk branch 3 × 3 Receptive field, multi-scale feature extraction dense block can provide 3 × 3 and 5 × 5 With two sizes of Receptive field, this structural design is convenient to better capture targets of different sizes and improve the effectiveness of feature extraction.

3.2 Multi-scale feature extraction convolutional blocks

The first layer of DenseNet121 is a convolutional layer with a size of 7 and a step size of 2. This large convolutional kernel is used to extract the boundary information of the image, which consumes a lot of computation. In order to further improve the feature expression ability, this paper proposes a multi-scale feature extraction convolution block to convolution with a size of 7. Compared to methods such as enhancing channels, the effect is better. The specific structure of the multi branch feature extraction convolution block is shown in Figure 2,The specific structure of the multi branch feature extraction convolution block first uses a convolution with size 3 for dimensionality reduction, followed by two branch structures for processing. The first branch uses a convolution with size 1 and a convolution with size 3 for feature extraction, and the second branch uses a maximum pooling with size 2. This strategy similar to combination pooling can effectively enrich the feature layer.

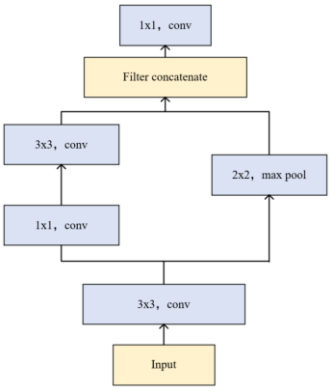

Figure. 2 Multi scale feature extraction of convolutional blocks

3.3 Lightweight Channel Attention Mechanism ECA-Net

In response to the problem of weak correlation between channels caused by the deepening of Dense121 layers, distinguishing different channels in the channel dimension will be helpful for the overall feature extraction work in facial expression recognition. Research has shown that the channel attention mechanism can help improve the feature extraction ability of convolutional neural networks. However, some current attention modules are designed more complex to achieve better performance indicators, making the overall model more complex. Here, an effective and lightweight channel attention module ECA-Net[10] is used, which has fewer overall parameters and excellent performance, and has better recognition performance gain for the overall model, ECA-Net is improved based on SENet. The core of ECA-Net is to use a 1D convolution and set an adaptive Kernel size to replace the channel dimensionality reduction of the fully connected layer, as shown in Figure 3.

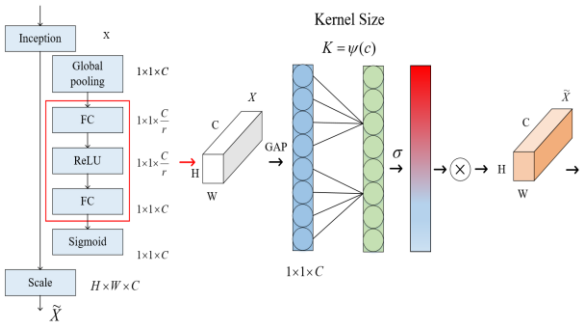

Figure. 3 Lightweight Channel Attention Mechanism ECA-Net

3.4 Improved DenseNet Network Architecture

On the basis of DenseNet121 model of dense convolutional neural network, in order to further improve its performance for facial expression recognition, the following work has been done: 1. A multi-scale feature extraction dense block is proposed to replace Denseblock in DenseNet121. The multi-scale feature extraction dense block has a multi branch feature extraction architecture, providing Receptive field of different sizes, It can extract richer and more refined features. 2. A multi-scale feature extraction convolutional block has been proposed to replace the top level convolution with a size of 7 in DenseNet121. Multiscale convolutional blocks can reduce computational consumption and better enhance feature expression capabilities, with better performance compared to methods such as channel enhancement. 3. On the basis of the improved model, the ECA channel attention mechanism was introduced, which increased the correlation between channels and improved the overall performance of the model in recognition work. Finally, the overall structure of a multi-scale dense convolutional neural network model integrating attention mechanism was proposed, as shown in Figure 4.

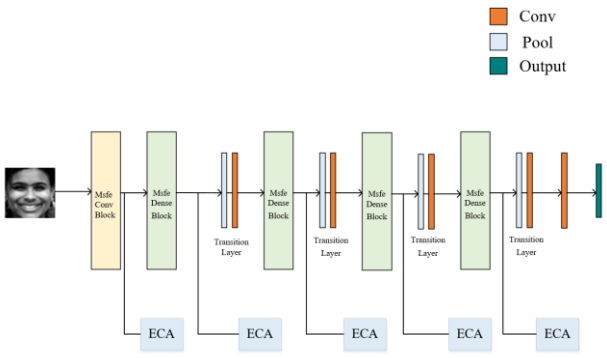

Figure. 4 Improved DenseNet Network Architecture

4. ANALYSIS OF EXPERIMENTAL RESULTS

4.1 Analysis Dataset

4.1.1 FER2013 Dataset

We proposed a facial expression recognition and classification technology based on convolutional neural network.ln order to

International Journal of Computer Applications Technology and Research
Volume 12–Issue 04, 55 - 60, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1204.1012

improve the robustness of fer task, we used multiple Data sets for training to improve the practicability of the network in expression recognition task. The first Data set is FER2013[11], a large facial image database randomlycollected by Google.After rejecting incorrectly marked frames and adjusting the clipping area, all images are registeredand adjusted to 48 pixels.FER2013 Dataset includes 35886 facial images, including 28709 training images, 3589verification images and 3589 test images. All images may correspond to one of the seven expression classification labels:0 anger, 1 Disgust, 2 fear;, 3 happy, 4 sad; 5 surprised; 6 Normal. The specific example is shown in Figure 5.

Figure. 5 FER2013 Dataset

4.1.2 CK+ Dataset

The second Data set is CK+ Data set[12].CK+ Data set recorded the facial movements of 210 adults with two hardware synchronized cameras in the laboratory environment. The participants are aged between 18 and 50 years old.CK+ Data set records 593 facial movement image data,which are based on the labels of the subjects' seven basic emotion categories: fear, happiness, disgust, sadness, anger, contempt and surprise.

4.2 Experimental pretreatment

The experiment compared the results on the FER2013 dataset and the CK+dataset. In terms of data preprocessing, considering the large number of image datasets, in order to accelerate the convergence speed of model training, improve training efficiency, and shorten training cycles, the processing of the dataset is divided into the following steps: image grayscale processing, image size normalization processing, facial expression localization and cropping processing, The processed dataset images are uniformly presented as 48 × The pre processed grayscale image of 48 is shown in Figure 6.

Figure. 6 Preprocessed images

4.3 Experimental configuration

The model proposed in this paper is based on pytorch deep learning architecture and runs on windows10 operating system. The GPU used is NVIDIA TiTan Xp. In the experiments of FER2013, the number of iterations is set to 250, the initial leaming rate is 0.01, and the learming rate is attenuated by 10 times every 20 iterations.In the CK + dataset experiment, the number of iterations is set to 60, the initial leaming rate is 0.01, and the learning rate is attenuated by 5 times every 5 iterations. Multiple datasets are set to 128 batches.

4.4 Experimental results and analysis

4.4.1 Experimental results on FER2013 Dataset

The confusion matrix obtained from the experiment of the improved model and DenseNet121 model on the Fer2013 dataset is shown in Figure 7 and Figure 8.

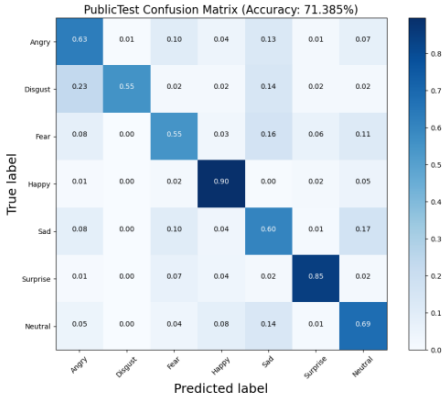

Figure. 7 The confusion matrix of this improved model on FER2013 dataset

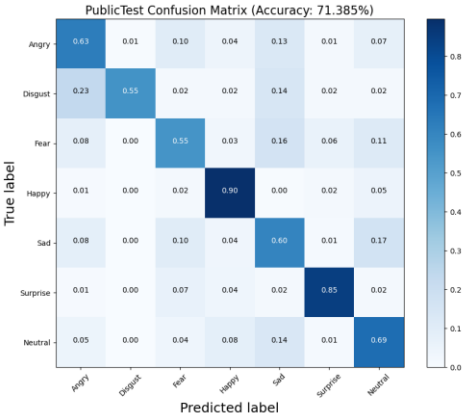

Figure. 8 The confusion matrix of DenseNet121 on FER2013 dataset

In order to further verify the progressiveness of the improved DenseNet model in the computer vision field of static expression recognition, we compared the accuracy with DenseNet121 and the public work of some common models of static expression recognition work scenes. The comparison results are shown in Table 1.

Table 1. The comparison results of the improved model and common network models on the FER2013 dataset in this article

Network mode	Recognition rate（%）
VGG19-BN	70.102
DenseNet121	71.385
ResNet32+CBAM	72.232
ResNet50+SVM	72.552
This paper	73.419

4.4.2 Experimental results on CK+ Dataset

The confusion matrix obtained from the experiment of the improved model and DenseNet121 model on the Fer2013 dataset is shown in Figure 7 and Figure 8.

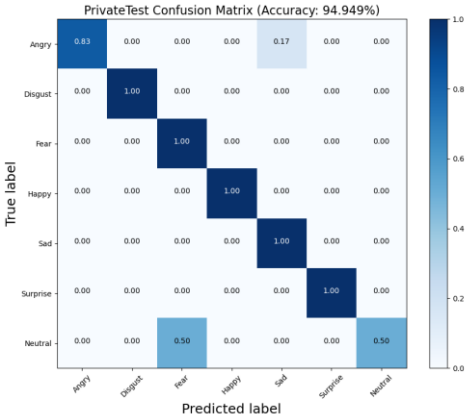

Figure. 8 The confusion matrix of this improved model on CK+ dataset

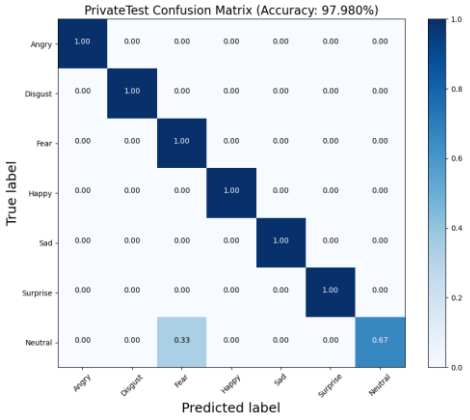

Figure. 9 The confusion matrix of this improved model on CK+ dataset

The comparison results between the improved model and common network models in this article on the CK+dataset are shown in Table 2.

Table 2. The comparison results of the improved model and common network models on the FER2013 dataset in this article

Network mode	Recognition rate（%）
VGG19-BN	93.816
DenseNet121	94.949
ResNet32+CBAM	94.512
ResNet50+SVM	95.183
This paper	97.980

4.4.3 Analysis of experimental results

By analyzing the experimental results of the improved DenseNet network and DenseNet121 on the FER2013 and CK+datasets, we found that due to the limitations of the network structure, DenseNet has average ability to extract complex facial expression features. After improvement, the DenseNet network improved its feature extraction ability, achieving the highest accuracy of 73.419% on the FER2013 dataset, and an improvement of 2.034% compared to DenseNet121; The accuracy rate of 97.980% was achieved on the CK+dataset, and 3.031% was increased compared with DenseNet121. Observing the confusion matrix of the improved DenseNet network on the above dataset, it can be found that the model has a very high recognition rate for happy and surprised expressions, almost reaching 90% - 100%, but for some more complex expressions, such as fear and neutral expressions, the recognition rate is relatively general, On the FER2013 dataset, the recognition rate is only 55% -69%, indicating that the network has good recognition performance for facial expressions with obvious features and high discrimination. However, there is still room for improvement in recognition performance for facial expressions with insignificant category features, such as neutral facial expressions.

Analyzing the recognition performance of the improved DenseNet model and commonly used network models for facial expression recognition on the public datasets FER2013 and CK+, it was found that the model proposed in this chapter has a certain leading recognition rate compared to the improved VGG19-BN, ResNet32+CBAM, ResNet50+SVM models on the two facial expression datasets FER2013 and CK+, reflecting the advantages of the model proposed in this chapter in facial expression recognition tasks.

5. CONCLUSION

This article selects DenseNet121, a dense convolutional neural network, for expression recognition work. It is found that DenseNet121 has a relatively average ability to extract complex facial features. After analyzing the structural characteristics of DenseNet, we consider improving the model structure of DenseNet121. Firstly, we improve the Denseblock in DenseNet by replacing the original Denseblock with dense blocks extracted from multi-scale features, Provide multi-scale Receptive field to facilitate the extraction of features of different sizes; Using multi-scale feature extraction convolutional blocks instead of the top level convolution with size 7 in DenseNet121, and introducing the ECA-Net channel attention mechanism to enhance the model's feature extraction ability from the channel dimension, an improved DenseNet model was finally proposed. The effectiveness of the proposed network improvement on DenseNet121 was verified through experiments on FER2013 and CK+datasets, And progressiveness compared with other network models in facial expression recognition.

6. ACKNOWLEDGMENTS

Thank you to the classmates and teachers who have contributed and helped with the work of this article. Thank you to the reviewers for their work.

7. REFERENCES

[1] Ekman P, Friesen w V, Tomkins s S. Facial Affect Scoring Technique: A First Validity Study[U]. Semiot ica,1971,3(1)37-58.

[2] Suwa M, Sugie N, Fujimora K.A preliminary note on pattern recognition of human emotional expression[U]. 1978.

[3] Mase K.Recognition of facial expression from optical flow[J]. IEICE Transactions on Information and Systems,1991,74 (10): 3474-3483.

[4] Gabor, D. Theory of communication. Part l: The analysis of information[J]. Electrical Engineers Part llI Radio & Communication Engineering Journal of the Institution of,1 946,93(26): 429-441.

[5] Prabhakar S, Sharma J,Gupta S. Facial expression recognition in video using adaboost and SVM[U]. International Journal of Computer Applications,2014,104(2):1-4.

[6] Huang D, Shan C, Ardabilian M, et al. Local binary patterns and its application to facial image analysis: a survey. IEEE Transactions on Systems, Man, and Cybernetics. 2011,41(6);765~781.

[7] Shan C, Gong S, Mcowan P.W. Facial expression recognition based on Local Binary Patterns: A comprehensive study[J]. Image and Vision Computing, 2009,27(6):803-816.

[8] Sun W, Ruan Q .Two-Dimension PCA for Facial Expression Recognition[C]l/ International Conference on Signal Processing.IEEE, 2007.

[9] Lopes A T, De Aguiar E, De Souza A F, et al.Facial expression recognition with convolutional neural networks: coping with few data and the training sample order[J].Pattern Recognition,2017,61:610-628.

[10] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[11] Khaireddin, Y. , and Z. Chen . "Facial Emotion Recognition: State of the Art Performance on FER2013." (2021).

[12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.