A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing

03CS6901 Seminar I

00/MCS/2020 CHN20CSIP002 Vishnu Vinod
M. Tech. Computer Science & Engineering (Image Processing)

Department of Computer Engineering
College of Engineering Chengannur
Alappuzha 689121
Phone: +91.479.2165706
http://www.ceconline.edu
hodcs@ceconline.edu

College of Engineering Chengannur
Dept. of Computer Engineering

C E R T I F I C A T E

This is to certify that, this report titled ***A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing*** is a bonafide record of the **03CS6901 Seminar I** presented on March 17, 2021 by

00/MCS/2020 CHN20CSIP002 Vishnu Vinod

First Semester M. Tech. Computer Science & Engineering (Image Processing)

scholar, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M. Tech. Computer Science & Engineering (Image Processing)** of **APJ Abdul Kalam Technological University**.

Guide	Coordinator
Radhu Krishna R Asst. Professor in Computer Engineering	Ahammed Siraj K K Associate Professor in Computer Engineering

Head of the Department

March 17, 2021

Dr. Smithadharan
Professor
in Computer Engineering

Acknowledgments

Primarily, I thank Lord Almighty for his eternal support through out my Seminar work.

I express my sincere thanks to **Dr. Jacob Thomas V**, Principal, College of Engineering Chengannur for extending all the facilities required for doing my seminar. My heartfelt words of gratitude to **Dr. Smitha Dharan**, Professor and Head of Department of Computer Engineering, for providing constant support.

Now I express my gratitude to my seminar co-ordinator **Mr. Ahammed Siraj K K**, Associate Professor in Computer Engineering and my seminar guide **Ms. Radhu Krishna R**, Assistant Professor in Computer Engineering who played a great role for valuable suggestions and expert guidance.

Vishnu Vinod

Abstract

With the rapid development of intelligent information technology and the wide use of computers, people can turn the complicated work to computers, which not only changes the traditional way of life, but also provides great convenience for human beings. We hope that computers can understand their intentions more intelligently, so as to better serve us. Facial expressions are an essential way to express emotions in nonverbal communication. With the development of computers, facial expression recognition plays an important role in many applications, such as human-computer interaction and medical escort. However, facial expression recognition is a complex task for computers. To avoid the complex process of explicit feature extraction in traditional facial expression recognition, a face expression recognition method based on a convolutional neural network (CNN) and an image edge detection is proposed. Firstly, the facial expression image is normalized, and the edgeof each layer of the image is extracted in the convolution process. The extracted edge information is superimposed on each feature image to preserve the edge structure information of the texture image. Then, the dimensionality reduction of the extracted implicit features is processed by the maximum pooling method. Finally, the expression of the test sample image is classified and recognized by using a Softmax classifier. To verify the robustness of this method for facial expression recognition under a complex background, a simulation experiment is designed by scientifically mixing the Fer-2013 facial expression database with the LFW data set. The experimental results show that the proposed algorithm can achieve an average recognition rate of 88.56% with fewer iterations, and the training speed on the training set is about 1.5 times faster than that on the contrast algorithm.

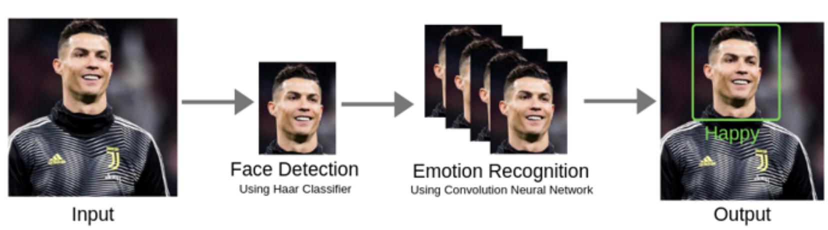
Contents

Chapter 1

Introduction

Facial expressions are part of human language and are often used to convey emotions. With the development of human-computer interaction technology, people pay more and more attention to facial expression recognition (FER) technology. Besides, in the domain of FER, human beings have made some progress. The facial expression recognition has a broad application background. It has been applied in the fields of assistant medicine, distance education, interactive games and public security.

Human emotions can be classified as: fear,contempt,disgust,anger,surprise, sad, happy, and neutral. These emotions are very subtle. Facial muscle contortions are very minimal and detecting these differences can be very challenging as even a small difference results in different expressions. Also, expressions of different or even the same people might vary for the same emotion, as emotions are hugely context dependent. While we can focus on only those areas of the face which display a maximum of emotions like around the mouth and eyes, how we extract these gestures and categorize them is still an important question. Neural networks and machine learning have been used for these tasks and have obtained good results. Here a face emotion recognition system based on image processing and convolutional neural network is proposed. The facial expression recognition extracts the information representing the facial expression features from the original input facial expression images through computer image processing technology, and classifies the facial emotion features according to human emotional expression, such as happiness, surprise, aversion and neutrality using a convolutional neural network.



Chapter 2

Literature Survey

1. **Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors** [2], IEEE Access, vol. 5, pp. 14797–14806, 2017

   In this method, they proposed an EEG feature extraction and selection method for emotion (happy, calm, sad, and scared) recognition. They employed Hjorth parameters in the frequency domain. The Hjorth parameters are useful for differentiating the power spectra that are used as a feature vector. To select the optimal feature set, they analyzed the extracted feature set using a balanced one-way ANOVA (p-value ¡ 0.05) method. Furthermore, top ranked classifiers the were used for the emotion classification using the optimal feature sets for each subject separately. Comparatively, the proposed method performs better than existing emotion recognition methods. The proposed feature selection method OF obtained the best emotion recognition rates of 76.6Based on the results, the conclusion is optimal feature selection is a good choice for enhancing the performance of EEG-based emotion recognition.

2. **Facial expression recognition based on local region specific features and support vector machines** [7], Multimed. Tools Appl., vol. 76, no. 6, pp. 7803–7821, Mar. 2017

   They propose a new method for the recognition of facial expressions from single image frame that uses combination of appearance and geometric features with support vector machines classification. The proposed face representation in provides better face registration than mainstream face representation, i.e. holistic representation. Performance improvement as well as dimensionality reduction is obtained with searching local face regions carrying the most discriminating information for facial expression classification.

3. **Facial expression recognition with faster R-CNN** [6], Procedia Comput. Sci., vol. 107, pp. 135–140, Jan. 2017

   In order to avoid the complex explicit feature extraction process and the problem of low-level data operation involved in traditional facial expression recognition, they proposed a method of Faster R-CNN (Faster Regions with Convolutional Neural Network Features) for facial expression recognition in this paper. Firstly, the facial expression image is normalized and the implicit features are extracted by using the trainable convolution kernel. Then, the maximum pooling is used to reduce the dimensions of the extracted implicit features. After that, RPNs (Region Proposal Networks) is used to generate high-quality region proposals, which are used by Faster R-CNN for detection. Finally, the Softmax classifier and regression layer is used to classify the facial expressions and predict boundary box of the test sample, respectively. The dataset is provided by Chinese Linguistic Data Consortium (CLDC), which is composed of multimodal emotional audio and video data. Experimental results show the performance and the generalization ability of the Faster R-CNN for facial expression recognition. The value of the mAP is around 0.82.

4. **Contrast limited fuzzy adaptive histogram equalization for enhancement of brain images** [5], Int. J. Imag. Syst. Technol., vol. 27, no. 1, pp. 98–103, 2017

   Contrast limited fuzzy adaptive histogram equalization (CLFAHE) is proposed to improve the contrast of MRI Brain images. The proposed method consists of three stages. First, the gray level intensities are transformed into membership plane and membership plane is modified with Contrast intensification operator. In the second stage, the contrast limited adaptive histogram equalization is applied to the modified membership plane to prevent excessive enhancement in contrast by preserving the original brightness. Finally, membership plane is mapped back to the gray level intensities. The performance of proposed method is evaluated and compared with the existing methods in terms of qualitative measures such as entropy, PSNR, AMBE, and FSIM. The proposed method provides enhanced results by giving better contrast enhancement and preserving the local information of the original image.

5. **A recursive framework for expression recognition: From Web images to deep models to game dataset** [4], Mach. Vis. Appl., vol. 29, no. 3, pp. 489–502, 2018

   In this paper, they propose a recursive framework to recognize facial expressions from images in real scenes. Unlike traditional approaches that typically focus on developing and refining algorithms for improving recognition performance on an existing dataset, they integrate three important components in a recursive manner: facial dataset generation, facial expression recognition model building, and interactive interfaces for testing and new data collection. To start with, they first create a candid-images-for-facial-expression (CIFE) dataset. they then apply a convolutional neural network (CNN) to CIFE and build a CNN model for web image expression classification. In order to increase the expression recognition accuracy, they also fine-tune the CNN model and thus obtain a better CNN facial expression recognition model. Based on the fine-tuned CNN model, they design a facial expression game engine and collect a new and more balanced dataset, GaMo. The images of this dataset are collected from the different expressions our game users make when playing the game. Finally, they evaluate the GaMo and CIFE datasets and show that our recursive framework can help build a better facial expression model for dealing with real scene facial expression tasks.

6. **Semantic-emotion neural network for emotion recognition from text** [3], IEEE Access, vol. 7, pp. 111866–111878, 2019

   In this paper, they propose a recursive framework to recognize facial expressions from images in real scenes. Unlike traditional approaches that typically focus on developing and refining algorithms for improving recognition performance on an existing dataset, they integrate three important components in a recursive manner: facial dataset generation, facial expression recognition model building, and interactive interfaces for testing and new data collection. To start with, they first create a candid-images-for-facial-expression (CIFE) dataset. they then apply a convolutional neural network (CNN) to CIFE and build a CNN model for web image expression classification. In order to increase the expression recognition accuracy, they also fine-tune the CNN model and thus obtain a better CNN facial expression recognition model. Based on the fine-tuned CNN model, they design a facial expression game engine and collect a new and more balanced dataset, GaMo. The images of this dataset are collected from the different expressions our game users make when playing the game. Finally, they evaluate the GaMo and CIFE datasets and show that our recursive framework can help build a better facial expression model for dealing with real scene facial expression tasks.

Chapter 3

Method

The proposed system has two stages facial emotion data pre-processing and facial emotion recognition using convolutional neural network. In the first stage the input facial image will be processed by using some image processing techniques to extract emotion features and in the second stage the facial image will be classified to the specific emotion based on the features using convolutional neural network developed.

3.1 Facial Emotion Data Preprocessing

Because the original pictures of facial expressions have complex background, different sizes, different shades and other factors, a series of image pre-processing processes have to be completed before facial expressions are input into the network for training. Firstly, we locate the face in the image and cut out the face image. Then, we normalize the face image to a specific size. Next, we equalize the histogram of the image to reduce the influence of illumination and other factors. Finally, we extract the edge of each layer of the image in the convolution process. The extracted edge information is superimposed on each feature image to preserve the edge structure information of texture image.

3.1.1 Face Detection And Location

This method uses the Haar-like to extract facial features, and uses an integral graph to realize fast calculation of Haar-like features, and screens out important features from a large number of Haar-like features. Then, we use the Adaboost algorithm to train and integrate the weak classifier into a strong classifier

Haar-like feature

Haar-like features are digital image features used in object recognition. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums.

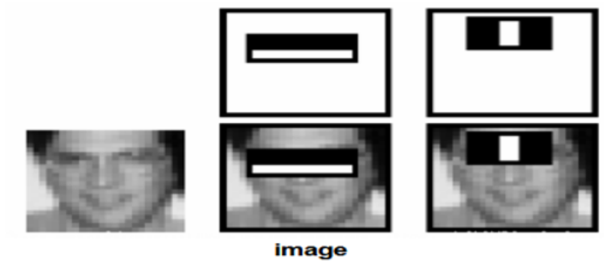


Figure 3.1: Haar Feature

Integral Images

An Integral Image is an intermediate representation of an image . Value for location (x, y) on the integral image equals the sum of the pixels above and to the left (inclusive) of the (x, y) location on the original image . This intermediate representation is essential because it allows for fast calculation of rectangular region.

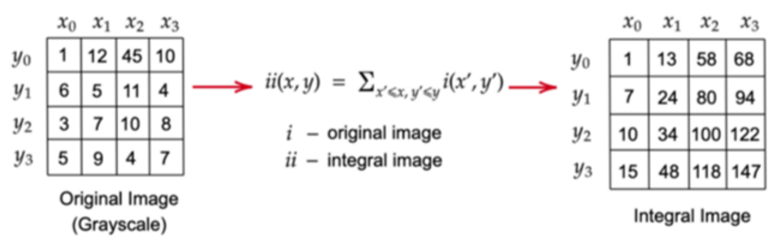


Figure 3.2: Integral Image

Haar Cascade Classifier

This is basically a machine learning based approach where a cascade function is trained from a lot of images both positive and negative. Based on the training it is then used to detect the objects in the other images. Based on the training it is then used to detect the face from the images. And using certain program codes(like OpenCV) we can extract the face from the input image as an image.

Adaboost Algorithm

AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique that is used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights to incorrectly classified instances. Boosting is used to reduce bias as well as the variance for supervised learning. It works on the principle

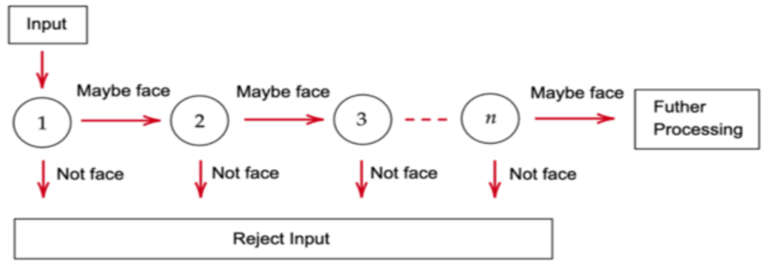


Figure 3.3: Haar Classifier

where learners are grown sequentially. Except for the first, each subsequent learner is grown from previously grown learners. In simple words, weak learners are converted into strong ones. Adaboost algorithm also works on the same principle as boosting, but there is a slight difference in working. Let's discuss the difference in detail.

3.1.2 Scale Normalization

Through normalization, the input image is scaled to 128*128 size. Let point (x, y) in the original picture be normalized and mapped to point x 0 , y 0 . The mapping is as follows

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Where sx represents the scaling ratio of the image in the direction of x axis and sy represents the scaling ratio of the image in the direction of y axis. In the process of image scaling, bilinear interpolation algorithm is also needed to fill the image.

Bilinear interpolation algorithm

Bilinear Interpolation is a resampling method that uses the distance weighted average of the four nearest pixel values to estimate a new pixel value. The four cell centers from the input raster are closest to the cell center for the output processing cell will be weighted and based on distance and then averaged.
A, B, C and D are the four points around the pixel (x, y). The corresponding gray values are g (A), g (B), g (C), g (D). To get the gray value of point (x, y) and calculate the gray value of points E and F, the formula is as follows:
g (E) = (x - xD) (g (C) - g (D)) + g (D)
g (F) = (x - xA) (g (B) - g (A)) + g (A)

xA and xD are the abscissa of point A and point D, respectively. The gray scale formula of (x, y) is as follows: g (x, y) = (y - yD) (g (F) - g (E)) + g(E) where yD represents the ordinates of CD points.

3.1.3 Gray Level Equalization

In the actual image acquisition process, it is easy to be affected by illumination, shadows and other factors. It is necessary to average the gray level of the image to enhance the contrast of the image. The Histogram Equalization (HE) method is used to process images. Histogram equalization is a technique for adjusting image intensities to enhance contrast.

Histogram Equalization

Histogram Equalization is a computer image processing technique used to improve contrast in images. It accomplishes this by effectively spreading out the most frequent intensity values, i.e. stretching out the intensity range of the image. This method usually increases the global contrast of images when its usable data is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast.The basic idea is to transform the histogram of the original graph into a uniform distribution form. If the gray

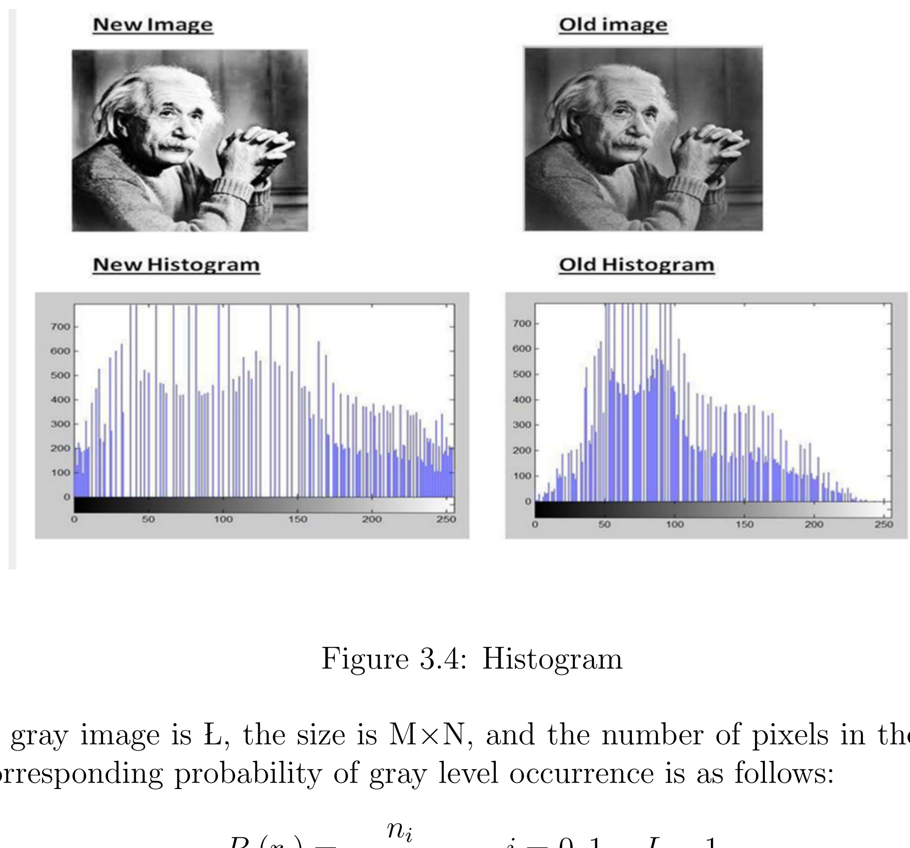


Figure 3.4: Histogram

level of the gray image is Ł, the size is M×N, and the number of pixels in the r_i gray level is E, the corresponding probability of gray level occurrence is as follows:

$$P_r(r_i) = \frac{n_i}{M \times N}, \quad i = 0, 1..., L-1 \tag{3.1}$$

Subsequently the cumilative distribution function is calculates using the following equation.

$$T(r_i) = \sum P_r(r_j), \quad i = 0, 1..., L-1 \tag{3.2}$$

Finally the image histogram is averaged using the following mapping relations:

$$e_j = INT[(e_{max} - e_{min})T(r) + e_{min} + 0.5], \quad j = 0, 1..., L-1 \tag{3.3}$$

3.1.4 Image Edge Detection

Kirsch edge operator is used to extract image edge information. The Kirsch operator or Kirsch compass kernel is a non-linear edge detector that finds the maximum edge strength in a few predetermined directions. The operator takes a single kernel mask and rotates it in

$$N = \begin{vmatrix} +5 & +5 & +5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{vmatrix} \text{NW}= \begin{vmatrix} +5 & +5 & -3 \\ +5 & 0 & -3 \\ -3 & -3 & -3 \end{vmatrix} \text{W}= \begin{vmatrix} +5 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & -3 & -3 \end{vmatrix}$$

$$\text{SW}= \begin{vmatrix} -3 & -3 & -3 \\ +5 & 0 & -3 \\ +5 & +5 & -3 \end{vmatrix} \text{E}= \begin{vmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & +5 \end{vmatrix} \text{NE} = \begin{vmatrix} -3 & +5 & +5 \\ -3 & 0 & +5 \\ -3 & -3 & -3 \end{vmatrix}$$

$$\text{S}= \begin{vmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ +5 & +5 & +5 \end{vmatrix} \text{SE} = \begin{vmatrix} -3 & -3 & +5 \\ -3 & 0 & +5 \\ -3 & +5 & +5 \end{vmatrix}$$

Figure 3.5: Kernels Used

45 degree increments through all 8 compass directions: N, NW, W, SW, S, SE, E, and NE. Each pixel in the image is operated with these 8 kernels (convolution) . The edge magnitude of the Kirsch operator is calculated as the maximum magnitude across all directions

Convolving mask over image

Place the center of the mask at each element of an image. Multiply the corresponding elements and then add them , and paste the result onto the element of the image on which you place the center of mask.

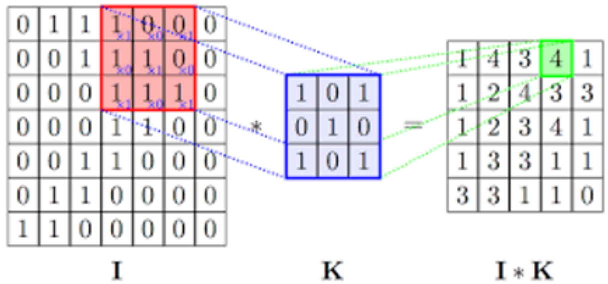


Figure 3.6: Convolution of Mask

3.2 Face Emotion Recognition Model Based On CNN

Image classification is the process of labeling images according to predefined categories. The process of image classification is based on supervised learning. An image classification model is fed a set of images within a specific category. Based on this set, the algorithm learns which class the test images belong to, and can then predict the correct class of future image inputs, and can even measure how accurate the predictions are. Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, a rough imitation of their biological counterparts, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value. The behavior of each neuron is defined by its weights. When fed with the pixel values, the artificial neurons of a CNN pick out various visual features.

When you input an image into a ConvNet, each of its layers generates several activation maps. Activation maps highlight the relevant features of the image. Each of the neurons takes a patch of pixels as input, multiplies their color values by its weights, sums them up, and runs them through the activation function.

3.2.1 Structure of CNN

The convolution layer C1 convolutes 96 × 96 pixels of input image with a 5 × 5 convolution core, i.e. each neuron specifies a 5×5 local receptive field, so the size of the feature map obtained by convolution operation is (96 5 + 1) × (965+1) = 92×92. Through convolution operations of 32 different convolution kernels, 32 feature maps are obtained, that is, 32 different local expression features are extracted. Convolution layer C2 uses 64 5 × 5 convolution kernels and then convolutes the characteristic graphs of convolution layer C1 output. 64 feature graphs are obtained. The size of each feature graph is (92 5 + 1) × (92 5 + 1) = 88 × 88. In convolution layer C3, 128 5 × 5 convolution kernels are used to convolute the characteristic maps of pool layer S1 output, and 128 feature maps are obtained. The size of each feature map is (44 5 + 1) × (44 5 + 1) = 40 × 40. The main purpose of the pooling operation is to reduce the dimension. A pooling window of 2 × 2 step size can reduce the dimension of the next feature map by half. Although there is no direct reduc-

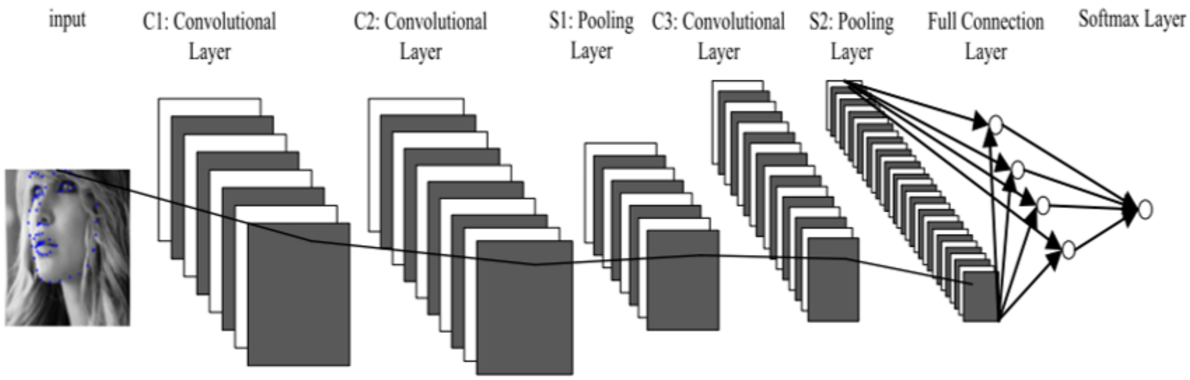


Figure 3.7: Structure of CNN

tion in the number of training parameters, halving the dimension of feature graph means that the computational complexity of convolution operation will be greatly reduced, which greatly improves the training speed. The input of the full connection layer must be a one-dimensional array, whereas the output of the previous pooling layer S2 is a two-dimensional array. First, the twodimensional array corresponding to each feature graph is converted into a one-dimensional array, and then 128 onedimensional arrays are connected in series to a feature vector of 51200 dimensions ($20 \times 20 \times 128 = 51200$) as the full connection.

The last layer of the CNN uses a Softmax classifier. The Softmax classifier is a multi-output competitive classifier. When a given sample is input, each neuron outputs a value between 0 and 1, which represents the probability that the input sample belongs to that class. Therefore, the category corresponding to the neuron with the largest output value is selected as the classification result.

Chapter 4

Conclusions

4.1 Performance Analysis of CNN

The proposed CNN is compared with the classical convolutional neural network FRR-CNN model and R-CNN algorithm. The experimental data come from the images of Fer-2013 expression database. The Fer2013 facial database contains 28,709 training pictures and 7,178 test pictures, each of which is a 48 $\times$ 48 gray scale image. Table 4.1 shows that

Method	Training Time(s)	Test Time(s)	Accuracy
Proposed Algorithm	178	24.89	88.56
R-CNN	256	33.97	79.34
FRR-CNN	148	17.92	70.63

Table 4.1: Comparison Table

the proposed algorithm is smaller than R-CNN algorithm and FRR-CNN algorithm in both training time and testing time. It can be concluded that the maximum pooling method is used to reduce the dimension of the extracted implicit features, which can shorten the training time of the convolutional neural network model. Moreover, the proposed algorithm achieves the highest average recognition rate of 88.56

4.2 Comparative Analysis

The purpose of this experiment is to verify the robustness of the proposed algorithm for facial expression recognition under complex background. The proposed algorithm is compared with the classical convolutional neural network FRR-CNN model and R-CNN algorithm. The experimental data is a mixture of LFW data set and Fer-2013 facial expression database. The LFW data set contains 13,000 face images, each of which is named after the person being photographed. Used 28,341 pictures in the mixed data set as training set and 7,542 pictures as test set to get the expression recognition rate of the above algorithm under different iteration times in different test sets. The experimental results are shown in the figure.

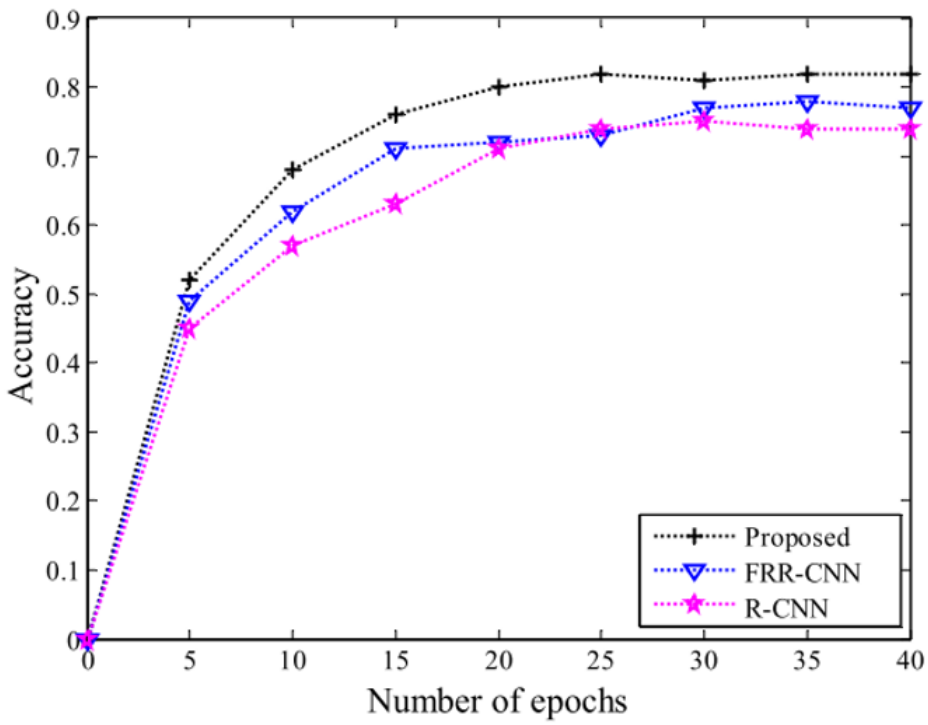


Figure 4.1: Comparison Graph

4.3 Future Scope

Facial expressions captured in reality may have various noises, such as face posture, occlusion, and blurring. To address this concern, as a future work, proper investigation will be performed. And the reduction of complexity of the CNN model developed is another future concern.

4.4 Suggestion

The dataset FER-2013 used for training the convolutional neural network has problem class imbalanace. Imbalanced data typically refers to a problem with classification problems where the classes are not represented equally. Here the images belongs to each of the emotion are differ in their quantity, and this will affect the perfomance of convolutional neural network.

Emotion	Number of Images
Disgust	547
Anger	4953
Surprise	4002
Sadness	6077
Calm	6198
Happiness	8989
Fear	5121

Table 4.2: Dataset Table

To solve this I prefer data augmentation method. By using this method we can generate new images from available images. And we can create a new balanced dataset to improve the performance of convolutional neural network.

4.5 Conclusion

Compared to traditional methods, the proposed method can automatically learn pattern features and reduce the incompleteness caused by artificial design features. The proposed method directly inputs the image pixel value through training sample image data. Autonomous learning can implicitly acquire more abstract feature expression of the image. The training process of the proposed method uses appropriate initialization of weights which has a great impact on the updating of weights.

References

[1] H. Zhang, A. Jolfaei, and M. Alazab. A face emotion recognition method using convolutional neural network and image edge computing. IEEE Access, 7: 159081–159089, 2019

[2] R. M. Mehmood, R. Du, and H. J. Lee, “Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors,” IEEE Access, vol. 5, pp. 14797–14806, 2017

[3] E. Batbaatar, M. Li, and K. H. Ryu, “Semantic-emotion neural network for emotion recognition from text,” IEEE Access, vol. 7, pp. 111866–111878, 2019

[4] L. Wei, C. Tsangouri, F. Abtahi, and Z. Zhu, “A recursive framework for expression recognition: From Web images to deep models to game dataset,” Mach. Vis. Appl., vol. 29, no. 3, pp. 489–502, 2018

[5] V. Magudeeswaran and J. F. Singh, “Contrast limited fuzzy adaptive histogram equalization for enhancement of brain images,” Int. J. Imag. Syst. Technol., vol. 27, no. 1, pp. 98–103, 2017

[6] J. Li, D. Zhang, J. Zhang, J. Zhang, T. Li, Y. Xia, Q. Yan, and L. Xun, “Facial expression recognition with faster R-CNN,” Procedia Comput. Sci., vol. 107, pp. 135–140, Jan. 2017

[7] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, “Facial expression recognition based on local region specific features and support vector machines,” Multimed. Tools Appl., vol. 76, no. 6, pp. 7803–7821, Mar. 2017.

[8] S. K. A. Kamarol, M. H. Jaward, H. Kälviäinen, J. Parkkinen, and R. Parthiban, “Joint facial expression recognition and intensity estimation based on weighted votes of image sequences,” Pattern Recognit. Lett., vol. 92, pp. 25–32, Jun. 2017.

[9] X. Yu, S. Zhang, Z. Yan, F. Yang, J. Huang, N. E. Dunbar, M. L. Jensen, J. K. Burgoon, and D. N. Metaxas, “Is interactional dissynchrony a clue to deception? Insights from automated analysis of nonverbal visual cues,” IEEE Trans. Cybern., vol. 45, no. 3, pp. 492–506, Mar. 2015

[10] J. Cai, Q. Chang, X.-L. Tang, C. Xue, and C. Wei, "Facial expression recognition method based on sparse batch normalization CNN," in Proc. 37th Chin. Control Conf. (CCC), Jul. 2018, pp. 9608–9613

[11] S. Escalera, X. Baró, I. Guyon, H. J. Escalante, G. Tzimiropoulos, M. Valstar, M. Pantic, J. Cohn, and T. Kanade, "Guest editorial: The computational face," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 11, pp. 2541–2545, Nov. 2018