Advancements in Human Eye Micro-Expression Detection Using Convolutional Neural Networks: Database Development and Real-Time Application in Security and Education

Arjun Singh*, Ziyu Wang

Loyola Marymount University*,Loyola Marymount University
Arjuns7@gmail.com*,wangziy88@gmail.com

Abstract:

A convolutional neural network (CNN) represents a type of artificial neural network distinguished by its efficiency in handling complex artificial intelligence tasks. It presents new opportunities for the pattern recognition of micro-expressions characterized by short duration and small movement amplitude. In the realm of micro-expression recognition research, studies focusing exclusively on eye micro-expressions are limited, yet their significance should not be underestimated. This paper's primary contributions include the creation of a database dedicated to human eye expressions. By segmenting existing facial expression databases and isolating only the eye region, a specialized micro-expression database for the human eye was established. Utilizing TensorFlow, CNN is employed to train, predict, classify, and identify facial expressions within the human eye. This approach successfully achieves micro-expression recognition in the eye area. Additionally, the integration of CUDA significantly reduces the training time of the system model.

Keywords:

Convolutional neural network, eye micro-expression recognition, deep learning.

1. Introduction

Human motion recognition is a research hotspot in the field of computer vision in recent years. It is widely used in human-computer intelligent interaction, virtual reality and video surveillance. Micro-expression is also a kind of human movement [1].

Human eye micro-expression recognition has great significance and value in driver fatigue driving supervision, police patrol screening, criminal suspect behavior prediction, marriage relationship prediction, communication negotiation, teaching evaluation and so on.

In the field of security, the polygraph has been around for a long time as an auxiliary tool. It can realize the function of polygraph detection by checking the fluctuation of a person's emotional signal. Although this equipment can greatly help the security personnel, due to the complexity of the inspection equipment, it has great limitations in practical applications. It is like in some places where the mobility of the crowd is relatively large, such as railway stations, airports, and subways. Usually, in order to ensure the safety of the people, some security checkpoints will be set up. Some experienced security personnel can identify some high-risk groups in time, but when the flow of people in these places is particularly large, it is relatively weak to rely on the observation of security personnel. At this time, micro-expressions are a way of detecting lies. It is able to quickly identify and judge lies without being noticed by the other party, which is undoubtedly a great help to security personnel.

In addition, in the field of education, teachers can improve the classroom environment by observing the micro-expressions of the students, improve the teaching methods, and enable students to enhance their interest in learning. In the field of criminal investigation, the police can ask the criminal suspects many problems while observing each other's micro-expressions, which will undoubtedly help the case to detect the progress. Due to the short duration of the micro-expressions, the small movements, etc., this requires a lot of manpower and time to manually identify the micro-expressions, and the reliability cannot be completely guaranteed. Therefore, the development of a micro-expression automatic recognition system is potential. It's a must, and if more and more researchers can join the research team of micro-expression technology, it will undoubtedly speed up this progress.The recognition of subtle expressions and subtle movements is still in a weak stage in the field of behavior recognition, and it is one of the key research directions. Eye micro-expression recognition can not only play a huge role in the fields of security, political psychology, education, etc., but also a small branch of facial expression recognition, which has an important role in promoting facial expression recognition.

2. Eye micro-expression recognition technology

2.1 Basic steps.

Micro-expression recognition is the process of first detecting and determining the position of a face from dynamic scene and complex background by algorithm, detecting and dividing the sub-image of the face and detecting the micro-expression in the image sequence, then extracting the micro-emoticon sequence, classifying and recognizing the micro-emoticon sequence, and finally determining the category of micro-emoticon. The following are the basic steps for micro-emoticon recognition, as shown in Figure 1. It includes face image acquisition and pre-processing, micro-emoticons detection, feature extraction and classification recognition. The essence of micro-expression feature extraction is pattern classification, extraction features are used for later classification recognition, the main process is to define categories, design classification mechanism, identify categories, according to different categories to judge different hidden real psychological emotions.

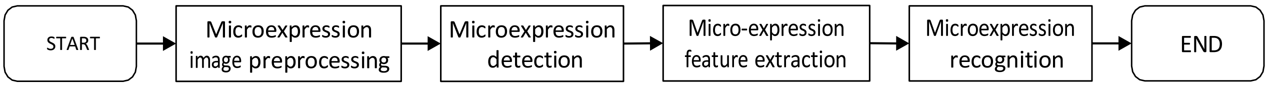
START
Microexpression image preprocessing
Microexpression detection
Micro-expression feature extraction
Microexpression recognition
END

Figure 1 The basic process of micro-emoticon recognition

2.2 Micro-expression Database.

As the existing micro-expression database is human facial expression data, it is difficult to show the advantages of only micro-expression for the human eye, there is currently no database containing only the micro-expression of the human eye. And an emoticon database contains tens of thousands of facial features, and the effort to build a new database of micro-expressions about the human eye is enormous. Therefore, this paper for the human eye micro-expression research needs from the existing facial micro-expression database to extract data processing, the generation of only contains the human eye expression database.

Common micro-expressions include USF-HD, Polikovsky's database, YorkDDT, SMIC, AVEC, CASME, and CASMEII. The emoticon database processed in this article is the fer2013 [2] emoticon recognition database on the kaggle website, and the fer2013 database emoticons contain seven types, namely: anger, disgust, fear, happy, normal, sad, surprised. Each expression has 56 to 895 images of the emoticons, and is divided into three model categories, Private Test, Public Test, and Training, for different stages of model building training, testing, etc. The original file is stored in csv format. Figure 2,4,6 shows some of the different emoticons in the fer2013 database, and Figure 3,5,7 is a processed eye micro-emoticon image of fer2013.

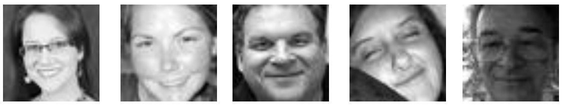

Fig. 2 Original Emoticon Data -Happy

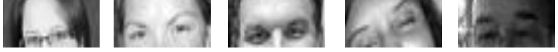

Fig. 3 Eye Micro-Emoticons -Happy

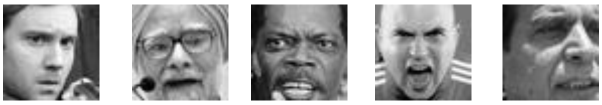

Fig. 4 Original Emoticon Data - anger

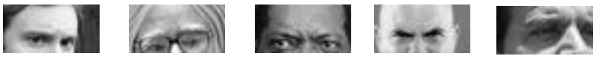

Fig. 5 Eye Micro-Emoticons – anger

Fig. 6 Original Emoticon Data -surprised

Fig. 7 Eye Micro-Emoticons -surprised

2.3 Micro-expression feature extraction method.

A variety of effective methods for micro-expression recognition have been proposed, such as optical flow method, SVM, BP, CNN, and so on. BP algorithm is a local search optimization algorithm; Vapnik et al. put forward a new design criterion for linear classifier based on years of research on statistical learning theory, support vector machine, which is SVM. The optical flow method is applied to micro-expression recognition, which can successfully capture small expression changes, detect and distinguish macro expressions and micro expressions.

2.4 Micro-expression classification and recognition method.

Like ordinary expressions, micro-expressions also contain changes in human mood. In a video clip with micro-expressions, identify the emotions it contains, called micro-expression symes.

After Fu Xiaolan's team extracted the micro-expression features [3], the micro-expression classification was identified using the Gent-leSVM classifier, and the average recognition rate on the CASME micro-expression database reached 85. 42%。

RandomForest (RF) is a machine learning method first proposed by Leo Breiman in 2001 as a randomly established classifier with multiple decision trees. Therefore, random forests can directly implement multi-classification function. In addition, random forests have several advantages: the ability to process high-dimensional data, relatively simple implementation, high efficiency of algorithm execution, and good noise resistance. But random forests also have some drawbacks: the classification of small and low-dimensional datasets may not work very well. Zhang Xuange combined the light flow characteristics with the LBP-TOP operator characteristics, classified by RF classifier, and achieved a recognition rate of 64.46% on the CASMEII database.

Extreme Learning Machine (ELM) is a machine learning method, which is the first algorithm proposed by Huang Guangbin and others to solve single hidden neural networks. Compared with traditional BP neural networks or SVM, ELM algorithms can not only guarantee learning accuracy, but also guarantee learning speed. Fu Xiaolan's team used extreme learning machines and differential quantum space analysis to identify micro-expressions.

The Hidden Markov Model (HMM) [4] is a probability statistical model and a double random process, one of which is the Markov chain and the other is a random process. Wu Xue constructed a double-layer hidden Markov model, improved the model training method, made full use of all the information of the training sample, and achieved an overall recognition rate of 86. 72%。

Deep Belief Network (DBN) is a deep learning model widely used in the field of pattern recognition. DBN combines unsupervised learning with supervised learning, not only for feature recognition, data classification, but also for generating data. Liu Yuzhen [6] designed a DBN network including a two-layer restricted Boltzmann machine network and a top-level back propagation network, and put the extracted micro-expression features into the DBN network. After pre-training and fine-tuning, A 40.24% recognition rate was obtained on the CASMEII database.

3. Methods and implementations

3.1 Convolutional neural networks.

The traditional neural network is to transfer the input floating-point data array to each layer, and the layers are multiplied by the full connection weights, and then the corresponding nonlinear function is used to obtain the output result. Convolutional neural networks are based on traditional neural networks. It differs from traditional neural networks in that each layer of CNN is only connected to a small area and is divided into a set of feature maps, all of which correspond to the same input. Since the use of the network does not require complex preprocessing of the original image, it has been widely used in the fields of character recognition, image recognition, and expression classification. It makes the characteristics learned by the network more robust in three ways: local receptive field, weight sharing and down sampling. The CNN has four layers [5], a feature extraction layer (C-layer), a feature mapping layer (S-layer), a pooling, and a fully-connected layer. The network structure of the convolutional neural network is shown in Figure 8.

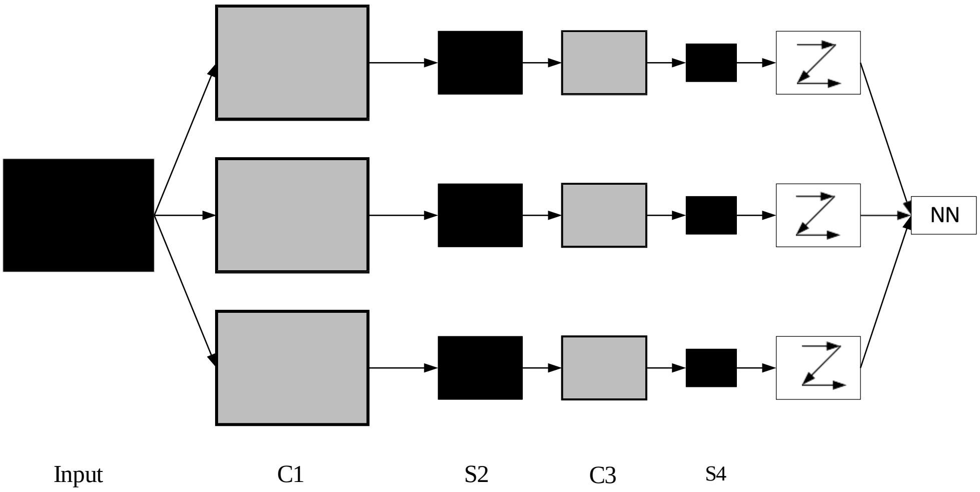

Fig. 8 Structural schematic of convolutional neural networks

The network structure of a typical convolutional neural network includes:

(1) Input layer):32*32, convolution core:5*5*6, then the size of C1(convolution result layer 1) is28*28*6. There are six differentC1layers, each within the C1 layer with the same weight.

(2) The C1 layer generates three feature maps. Each of these feature maps is summed, weighted, and offset. The feature maps of the three S2 layers are obtained by a sigmoid function. S2 is the lower sample layer and the size is 14*14*6.

(3) C3 convolution core: 5*5*16, then the size of C3 layer (convolution result layer 2) is 10 x 01 x 16. If the S2 layer has only one plane. Then the Input-C1 layer and the S2-C3 layer are the same.

(4) The size of the S4 layer is 5*5*16.

(5) The number of nodes in each layer after the S4 layer is very small, that is, the fully connected layer. The fully connected layer aggregates the features learned through the convolutional layer to form global features, and then uses these features for classification detection and recognition.

3.2 BP and SGD.

Before the reverse propagation algorithm was proposed, people should think of using SGD learning model, and some ways to solve the bias derivative of the network model, but these algorithms are relatively low efficiency, so the reverse propagation algorithm is proposed to calculate the bias derivative more efficiently. Reverse propagation utilizes chain law, which greatly accelerates the learning speed.

Gradient descent (GD) [6] is a common method for minimizing loss functions. Random Gradient Drop (SGD) is updated iteratively per sample. In cases where the sample size is large (e.g. hundreds of thousands), compared to the volume gradient drop, the iteration requires hundreds of thousands of training samples at a time, one iteration is not optimal, and if it is outs 10 times you need to traverse the training sample 10 times.

3.3 Analysis of each layer's characteristics.

The following is a characteristic analysis and comparative analysis of the output image obtained by a micro-emoticon image through different quantitative pooled layers [7]. After many convolutions and pooling, the advanced feature points are obtained. This shows that the convolutional pooled structure is effective for micro-expression feature extraction, and the efficiency of the convolution neural network applied to micro-expression feature extraction has been successfully tested.

3.4 CNN training process.

Since CNN itself is an input-to-output mapping, the process of training the CNN model is equivalent to training a function map. CNN's sample set consists of an input vector and an ideal output vector pair, as convolutional neural networks perform supervised training. To ensure that the network does not reach saturation due to excessive weights, the weight initialization is set to a random value between -1 and 1 before starting training. CNN's training process is as follows:

(1) Initialize the ownership value to a smaller random value between -1 and 1.

(2) Take a sample X as input from the micro-expression database, and its target vector is D.

(3) Forward communication stage:

$$x_j^l = f(\sum_{i=M_j} x_i^{l-1} \times \mathrm{ker}\,nel_{ij}^l + B^l) \tag{1}$$

For the calculation formula for the sampling layer:

$$x_j^l = f(\alpha^l \sum_{i=M_j} x_i^{l-1} + B^l) \tag{2}$$

For the full connection layer:

$$x_j^l = f(\sum_{i\in(i-1)layer} x_i^{j-1} + B^l) \tag{3}$$

Where $f(x)$ is the Relu activation function:

$$f(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{4}$$

(4) The error term of each layer is calculated in reverse: the error of the output layer. If the output layer has m nodes, the error term for the output layer node k is:

$$\delta_k = (d_k - y_k)y_k(1-y_k) \tag{5}$$

Where d_k is the target output of node k , y_k is the predicted output of node k .

The error of the middle full connection layer: If the current layer is the first layer, the total L node, the first layer has M nodes. Then the error term for node j of the first layer is:

$$\delta_j = h_j(1-h_j)\sum \delta_k W_{jk} \tag{6}$$

Where: h_j is the output of the j-node, W_{jk} is the node of the lth layer, and the weight of the node k of the j to $l+1$layer.

(5) The adjustment amount of each weight is calculated step by step from the back to the front, and the change amount of the weight vector of the k th input of the node j in the n th iteration is:

$$\Delta W_{jk}(n) = \frac{\eta}{(\Delta W\;1+N}{}^{jk}(n-1)+1)\delta_k h_j \tag{7}$$

The amount of change in the threshold is:

$$\Delta B_k(n) = \frac{\eta}{1+N}(\Delta B_k(n-1)+1)\delta_k \tag{8}$$

(6) Adjust the weights, and the updated weights are:

$$\Delta W_{jk}(n+1) = W_{jk}(n) + \Delta W_{jk}(n) \tag{9}$$

The updated threshold is:

$$B_k(n+1) = B_k(n) + \Delta B_k(n) \tag{10}$$

(7) Repeat (2) to (6) the process until the error function is less than the set threshold. The error function is:

$$E = \frac{1}{2}\sum_{k=1}^{M}(d_k - y_k)^2 \tag{11}$$

The advantages of convolutional neural networks in image processing:(1) input images and network topology coincide; Weight sharing can reduce network training parameters and make the network simpler and more adaptable.

4. Conclusion

Building on human facial expression recognition, this thesis conducts an in-depth study of human eye micro-expressions and optimizes the model using the CUDA platform. The model, developed on the TensorFlow platform, employs a convolutional neural network for the extraction, classification, and recognition of micro-expression features. Due to the absence of a dedicated database for eye micro-expressions, the current model remains rudimentary. Nonetheless, this research represents a crucial branch in the field of facial expression recognition. It is anticipated that further research and analysis will be facilitated with increased time and funding.

References

[1] Ekman P, Friesen W V. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding[J]. Semiotica, 1969, 1(1):49-98.

[2] Information on https://www.kaggle.com

[3] Wang S J, Chen H L, Yan W J, et al. Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine [J]. Neural Processing Letters, 2014:39(1)：25-43.

[4] Sabrina Hoppe, Andreas Bulling. End-to-End Eye Movement Detection Using Convolutional Neural Networks. arXiv:1609.02452 [cs.CV].2016,09(08).

[5] Wu Q, Shen X, Fu X. The machine knows what you are hiding: an automatic micro-expression recognition system[M]. Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2011: 152-162.

[6] Shreve M, Godavarthy S, Goldgof D, et al. Macro and micro-expression spotting in long videos using spatiotemporal strain[C]. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops. IEEE, 2011:51-56.

[7] Wang S J, Yan W J, Li X B, Zhao G Y, Zhou C G, Fu XL, Yang M H, Tao J H. Micro-expression recognition usingcolor spaces. IEEE Transactions on Image Processing, 2015,24(12): 6034−6047.