IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON INNOVATION AND APPLICATION OF INTELLIGENT PROCESSING OF DATA, INFORMATION AND KNOWLEDGE AS RESOURCES IN EDGE COMPUTING

Received February 27, 2020, accepted March 16, 2020, date of publication March 20, 2020, date of current version April 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982286

A Novel Facial Expression Intelligent Recognition Method Using Improved Convolutional Neural Network

MIN SHI[1], LIJUN XU[2], AND XIANG CHEN[3]

[1]School of Art and Design, Fuzhou University of International Studies and Trade, Fuzhou 350202, China
[2]Institute of Art and Design, Nanjing Institute of Technology, Nanjing 211167, China
[3]School of Design, Jiangnan University, Wuxi 214122, China

Corresponding author: Lijun Xu (xulijun_1122@aliyun.com)

This work was supported in part by the Major Project of Philosophy and Social Science Research in Colleges and Universities of 2019 Jiangsu Province, Research on Design of Human–Machine Interaction through the Application of AI Technology, under Grant 2019SJZDA118, in part by the Cultural Research Projects of 2019 Jiangsu Province, Research on Human–Machine Interaction of Cultural Tourism Industry through the Application of AI Technology, and in part by the Major Project of Philosophy and Social Science Research in Colleges and Universities of Jiangsu Province through the Innovation and Development of Industrial Design Driven by Big Data under Grant 2018SJZDA015.

ABSTRACT Human facial expression is the core carrier of feedback. Facial expression recognition(FER) has been introduced into mickle fields, such as auxiliary medical care, safe driving, marketing assistance, distance education. However, in the real production process, facial expression image samples collected in different scenarios have problems such as complex backgrounds, which causes the FER model to train very slowly, low recognition rate, and insufficient generalization, so it cannot meet the actual production requirements. As the originator of the clustering algorithm, fuzzy C-means clustering(FCM) algorithm has stable performance and good results. It is applied to the convolutional layer of a convolutional neural network(CNN) to obtain a convolution kernel with an initial value, so as to extract the expression image features in the training set and the test set. This can solve the problem of random initialization of the convolution kernel. Based on the CNN, this paper introduces FCM to optimize the feature extraction (FE) capability of the model, and proposes a novel FER algorithm using an improved CNN(F-CNN). Because traditional CNN has problems such as irrational layer settings and too many parameters. The proposed F-CNN first adjusts the CNN network structure to improve the nonlinear expression ability of CNN. Then, replace the Softmax classifier that comes with CNN with a support vector machine (SVM) to improve the model's classification ability. The comparison experiments with other models show that the improved model improve the FER rate. The introduced FCM algorithm can effectively improve the model's FE performance and shorten the time of F-CNN during training. On the whole, F-CNN has reference value.

INDEX TERMS Facial expression recognition, convolutional neural network, fuzzy C-means clustering, support vector machine, intelligent processing.

I. INTRODUCTION

Facial expressions, as the most intuitive reaction in the human heart, are an integral part of the smooth communication process. Humans can obtain the facial expressions of others through vision, and understand the inner states of others through the analysis of human brain, so as to achieve the purpose of communication between people. With the prosperity of scientific research, especially the gradual living of artificial intelligence [1]–[5] in recent years, people's demand for the intelligence of machines has also increased day by day. Humans hope that machines can recognize facial expressions relatively accurately, so as to complete the communication between humans and machines, not limited to human-to-human communication. With the gradual strengthening of network and computer hardware technology, massive video image data can be better stored, transmitted, and processed, which also provides convenience for the development

The associate editor coordinating the review of this manuscript and approving it for publication was Honghao Gao.

of FER. One of the benefits of rapid technology development is that it can better facilitate people's work and life, and FER technology can be better applied in the following fields.

A. SAFE DRIVING MONITORING

According to the survey, in recent years, the number of motor vehicles in China has increased year by year, and traffic accidents have remained high. The number of deaths is as high as 50,000 to 60,000 person-times per year. The main reason for traffic accidents is the behavior of non-compliance with traffic rules, such as: drunk Driving, fatigue driving, etc. By installing camera equipment in the vehicle, the driver's face is monitored in real time, and when a dangerous driving behavior is detected, a warning message is issued to remind the driver to ensure their personal and property safety [6].

B. MEDICAL SERVICES

FER can help patients with autism recover [7], [8]. In general, people with autism have difficulty understanding other people's thoughts and feelings through their facial expressions, and often show their reluctance to interact with others. Through real-time FER for autistic patients, understand the heart of autistic patients. Coupled with targeted conversation skills, it can effectively help autistic patients recover.

C. MARKETING ASSISTANCE

The most intuitive expression of people's likes and dislikes about something is the expression. If it is applied in a store, the owner analyzes the customer's expression through the FER system, and can judge the customer's degree of love for the product. Coupled with appropriate sales techniques, it will greatly increase the sales rate of the product. On the other hand, by capturing people's expressions when watching advertisements, it is possible to accurately detect the effect of advertising, and provide a reliable basis for accurate advertising marketing.

D. DISTANCE EDUCATION

In the context of the Internet, distance education [9] came into being. As a learning and teaching method that spans time and space, distance education is very popular among learners. Its advantage is that learners and educators can realize teaching anytime and anywhere through the Internet without being limited by space. However, this new type of online education method also has obvious shortcomings. For example, teachers and learners cannot achieve face-to-face communication, and teachers cannot grasp the knowledge of learners based on the expressions of student feedback. Based on this, FER technology has a wide range of applications in the distance education industry. Teachers obtained students' psychological state at this time in line with the expressions of students fed back by the expression recognition system, and adjust the rhythm of their classrooms according to the state of students, so as to obtain better teaching results.

E. GAMES

Video games [10] as a cultural event, are gradually changing the games' definition. Computer-based video games are difficult to interact with players, so they can't understand player feedback about the game. Apply FER technology to the interactive field of the game, analyze the player's expression in real time, and make corresponding adjustments to the game according to the expression. Therefore, the introduction of FER technology to video games industry can improve the interactive and entertaining aspects of the game.

In addition to the applications in these fields, FER has emerged in many other fields. It is a trend to introduce face recognition technology based on the Internet of Things [11]–[17]. The urgent needs of many fields have played a good role in promoting the improvement of its theoretical research. With the continuous development of this technology, we should take its essence and continue to innovate, make further research on FER, and better serve the society.

In the early 19th century, some scholars began to devote themselves to the FER field. The origin of this study was that British biologist Darwin pointed out in his research [18] that there is a correlation between human and animal expressions. In 1971, psychologist Ekman and his research partner Ekman and Friesen [19] made an in-depth study of the relationship between facial muscle movements and different expressions, and proposed the Facial Action Coding System. Prior to this, FER has always belonged to the psychology' field. Until 1978, Suwa [20] and others first applied FER to computer image processing. Its core purpose is to realize automatic FER. In 1991, Mase and Pentland [21] first introduced the optical flow method for FER. By extracting optical flow values as expression features, FER is realized. This method has laid a good foundation for the automatic FER technology's development.

The simultaneous advancement of science and technology and academic research, the current FER technology has become more mature. Wu *et al.* [22] introduced Gabor filters into FER systems, and compared and analyzed Gabor spatial energy filters and Gabor motion energy filters. The experimental results show that the Gabor motion energy filter is better at extracting expression features and can also achieve a higher expression recognition rate, but the disadvantage is that the implementation of the algorithm is too complicated. Luo *et al.* [23] proposed an FER algorithm using improved Principal Component Analysis. The local gray features extracted by the local binary mode are used to help the global gray FE of FER. Finally, SVM was used for classification. Simulation experiments demonstrate that the method is effective in classifying different expressions. Shin *et al.* [24] constructed a new FER algorithm, which trains 20 different CNN models to find a model structure with relatively strong recognition ability. The experimental results show that a simple convolutional layer with a histogram equalized image as input and a structural unit composed of a downsampling layer are relatively most effective, but

the recognition rate still needs to be improved. The above are some representative studies. In addition, many scholars have proposed different methods to contribute to the field of FER [25]–[33].

FER technology has many applications in real life. In 2000, the MIT Media Lab at MIT developed a robot called Kismet, which recognizes facial expressions and mimics human facial emotions. In 2009, Waseda University in Japan developed an expression robot called KOBIAN, which can interact with humans using 7 expressions. In 2012, during the 57th presidential election in the United States, Affectiva used Affdex technology to detect the facial expressions of more than 200 voters while watching the debate between Obama and Romney, and predicted the results of voters with 73% accuracy. In 2015, Italian designer Simone Rebaudengo and his companion Paul Adams designed a haze-proof mask called Unmask. This mask can convey people's facial expressions, such as smile, pain, surprise, etc. through an externally connected LED matrix screen. In 2017, Apple introduced its new phone, the iPhone X. One of the new features is that users can create their favorite 3D Animojis through Face ID and can send them via iMessage.

The above research is carried out from various aspects of FER. To improve the FER's rate and shorten the training time of recognition models, a new recognition method is given, which uses improved CNN combined with classic FCM algorithm for FER. The specific work is summarized as follows:

(1) This study improved CNNs. Replace the Softmax classifier that comes with the metamodel with the SVM to ameliorate the classification performance. Comparative Experiments verify that the F-CNN model can upgrade the expression recognition rate to a certain extent.

(2) Because the samples collected in the actual application scene have complex backgrounds, the training speed of the model is very slow. This paper introduces the classic FCM algorithm and proposes a FER algorithm with the combination of FCM clustering ideas and CNN. FCM acts on the convolutional layer of CNN, so as to obtain the convolution kernel with initial value to extract the expression image features in the training and test sets. Simulation experiments show that the proposed algorithm increases the FE capability and reduces the training time of the model.

II. RELATED KNOWLEDGE

A. FER FRAMEWORK

The FER flowchart is shown in Figure 1.

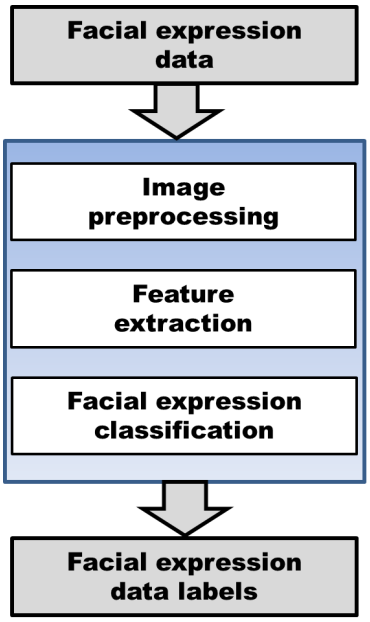

FIGURE 1. Identification framework.

1) FACE IMAGE DATA SET

The face image can be acquired in real time by the camera equipment. After illumination compensation and geometric normalization, face images are stored in the database in the form of still images or dynamic images. Facial expression database is the basic condition for research on FER. At present, the facial expression databases commonly used at home and abroad are: JAFFE dataset [34], Fer2013 dataset [35], Cohn-Kanade dataset and CK + dataset [36], [37], RAF2017 dataset [38] And other data sets [39]–[41].

2) DATA PREPROCESSING

For extracting high-quality features, it is best to discard unnecessary information in the original image when performing FE operations. Preprocessing of face images is necessary. Face expression image preprocessing mainly includes face cutting and image normalization. Because facial expression images may be collected with some other background information, which is not conducive to highlighting the expression features, it needs to be cut. The methods of removing background can be divided into two categories: coarsening and refinement. Rough background cutting uses the face recognition algorithm to identify the face area in the picture and remove the rest. The expression pictures obtained in this way often have a small amount of background information. Adaboost algorithm [42] can remove most facial features such as hair, neck, and background after processing facial images, and reduce some interference information. A more elaborate approach is to recognize the contour of the face through the active appearance model ASM algorithm [43] and intercept the face area by the contour. Because different people have different facial contour ranges and different ways of collecting images, the face image sizes obtained are different. In order to reduce the problems of uneven illumination and inconsistent image size, it is necessary to perform operations such as illumination normalization, grayscale normalization, and scale normalization on facial expression images. Black *et al.* gave a model of illumination changes, which solved the problem that Gabor wavelets could not effectively deal with local brightening of faces [44].

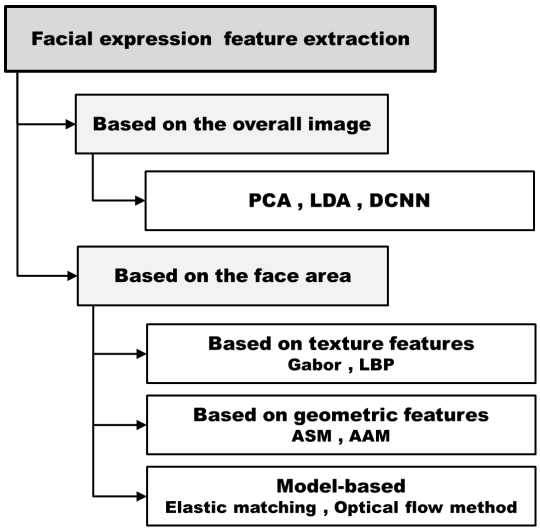

FIGURE 2. Basic classification of FE methods in expression recognition.

3) FEATURE EXTRACTION

The quality of FE of facial expression images determines the final effect of FER. Therefore, FE based on facial expressions has attracted many scholars' research interests, and many good FE algorithms have emerged. The current FE algorithms are roughly divided into two parts: FE based on the face area and FE based on the overall image. The classification framework and representative algorithms are shown in Figure 2.

The FE algorithm in the face area mainly considers the mapping relationship between the biological features of the face area and expression recognition. The method can be divided into 3 categories, namely texture-based, geometric-based and model-based FE methods. Typical texture FE methods include Gabor filters [45], Local Binary Patterns [46], Local Gabor Binary Patterns [47], Direction Gradient Histogram [48], and scale-independent Variable Feature Transform [49]. Typical geometric FE algorithms are Active Appearance Model [50] and Active Shape Model [43]. It turns out that the algorithm based on the face area is also flawed, and the accuracy of expression recognition cannot be guaranteed by using alone. The overall image-based algorithm considers the entire picture as a whole and analyzes the picture using a universal feature algorithm. The most typical are Deep CNNs (DCNN) [51]–[53].

4) CLASSIFIER

The expression classification problem is a classic multi-classification problem, and a variety of classifier models can be used for expression classification. In recent years, common expression feature classification algorithms include Bayesian classification, hidden Markov model algorithm, Softmax, random forest, nearest neighbor method, and SVM algorithm. The main feature classification methods are classified into distance-based methods [54]–[56] and Bayesian network-based methods [57]–[60] and neural network methods [61]–[64].

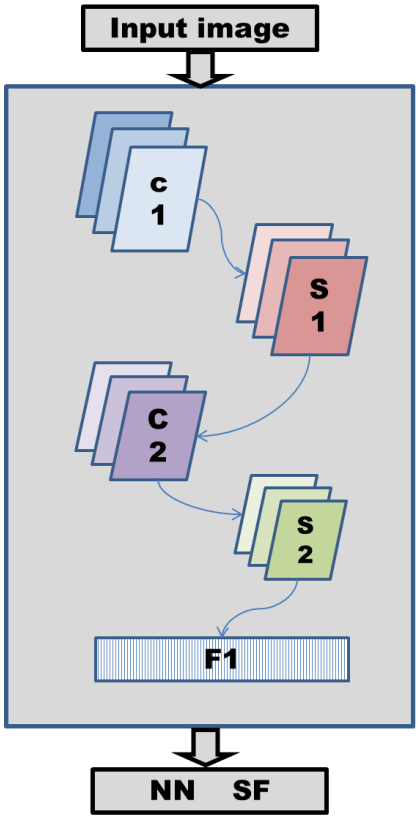

FIGURE 3. Schematic diagram of the CNN structure.

B. CONVOLUTIONAL NEURAL NETWORK

In 1958, Nobel Laureates in Physiology and Medicine Hubel and Wiesel [65] first proposed the concept of visual cortical neurons in cat visual cortex experiments, and made important contributions to the development of visual neural networks. In 1980, Japanese scientist Kunihiko Fukushima and Neocognitron [66] proposed a neural network cognitive model on the basis of his predecessors, which laid a solid foundation for the subsequent development of CNN. In 1995, LeCun and Bengio [67] proposed the concept of CNNs and successfully applied it to handwritten digit recognition. The CNN does not need artificial adjustment parameters, and can automatically realize the connection from input to output.

CNNs generally consist of an input layer, a convolutional layer, a downsampling layer, a fully connected layer, and an output layer. The input layer can input unprocessed raw image data, and the parameter adjustment of the entire network structure is mainly realized by the forward propagation algorithm and the backward propagation algorithm. The structure of the CNN is shown in Figure 3.

After the input image is processed by the convolutional layer, the downsampling layer, and the fully connected layer, the classification result is finally output by the softmax classification layer. As shown in Figure 3, the convolutional layer (C1, C2) and the downsampling layer (S1, S2) usually appear in pairs. Figure 3 shows only two sets of processing units for the convolutional layer and the downsampling layer.

And in the actual situation can be adjusted according to different graphic data. The input original image data and n convolutional layers are convolved. After the convolution processing, n feature maps will be obtained, and then the feature maps are input to the down-sampling layer for dimensionality reduction and other processing. All neurons in the fully connected layer, the downsampling layer and the output layer will be connected and the classification results will be output in the output layer.

As a kind of deep feed-forward neural network, the CNN has artificial neurons connected to each other as its basic structural unit. When the neuron is activated, it transmits signals to other neurons. The learning process of CNNs can be divided into two processes: feedforward neural networks and back propagation. The feedforward neural network algorithm calculates the raw data of the input layer layer by layer to obtain the output data, and compares the actual output data with the expected data to get the error. In the back-propagation algorithm, the parameters of the model are adjusted backwards based on the obtained error, and iteratively repeated until the objective function approaches a value within a range.

III. IMPROVED CONVOLUTIONAL NEURAL NETWORK

The selection of network model parameter initialization method is very important in a network training process. It not only affects the network's convergence ability, but also affects the network's convergence speed. Due to the lack of transparency in the middle layer of traditional CNNs, the convolution kernel is generally initialized by random methods during the FE stage. This makes it easy for CNN models to suffer from long model training time and insufficient nonlinear expression capabilities.

This study uses the FCM algorithm to train the data in the JAFFE database to obtain the convolution kernel. Extract the face patches of all the images and get a set of clustering centers through FCM. Then delete and subtract more convolution kernels through the convolution kernel, and finally obtain the volume needed for CNN Accumulate the initial value. The execution steps are as follows.

(1) Since the initial image is a complete facial image with a size of $256 * 256$, the Haar-like algorithm needs to be used to find the facial area in the image before FCM clustering;

(2) Crop and normalize the image to obtain a size of 48×48 pixels, and randomly extract a patch of $t * t$ size from it, with a step size of 1, as shown in Expression 1.

$$X_i = \left[x_{i,1}, x_{i,2}, \ldots, x_{i,m*n}\right] \in R^{k*k*mn} \quad (1)$$

(3) Among them, the value of t depends on the size of the convolution kernel of each layer. The parameters $m = 48\text{-}t + 1$ and $n = 48\text{-}t + 1$ can be obtained by calculation;

(4) Subtract the average of all the small patches on all faces to obtain Expression 2;

$$\bar{X}_i = \left[\bar{x}_{i,1}, \bar{x}_{i,2}, \ldots, \bar{x}_{i,m*n}\right] \in R^{k*k*mn} \quad (2)$$

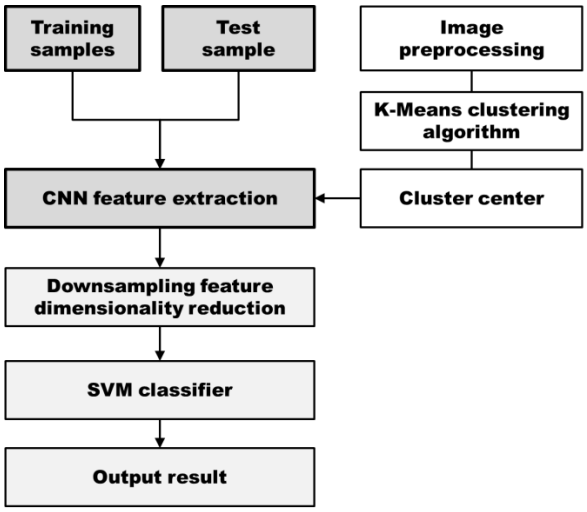

FIGURE 4. F-CNN algorithm framework.

(5) The FCM clustering method is used to cluster the face patches in the dataset into K clusters. Deduplication is performed to obtain a convolution kernel that can be input into the CNN.

This paper uses FCM to train a certain number of cluster centers as the convolution kernel's initial value in CNN model convolution layer. SVM is applied to make up for the shortcomings of the random initial convolution value of the CNN and the shortcomings of weak FE capability. The overall structure of the F-CNN algorithm is shown in Figure 4.

The steps of F-CNN to recognize facial expressions are as follows

Input	Using Haar-like algorithm to detect faces in facial expression images, and after trimming and normalizing, 48×48 pixel expression images are used as input data
Output	A one-to-many classification method is used. One sample is selected as one class, and the remaining samples are classified as one class. Take the category with the highest classification function value as the global result.
Training phase	1. Parameter initialization: In the first half, the learning rate set by the algorithm in this article is 0.0002, the middle and rear parts are set to 0.001, and it will be 0.00002 at the end of the experiment. The batch size is 80 or less, the epochs are 90 or less, and the momentum is 0.95. 2. Obtain the cluster center by performing FCM, and set this center to the initial value of the convolution kernel. 3. Train the CNN model using the initial CNN values obtained in steps 1 and 2. 4. The extracted features are input to the SVM to obtain recognition results. 5. Find the error between the recognition result obtained by the model and the

	known label. Use the method of minimizing the mean square error cost function to adjust the weight parameters until the iteration accuracy is satisfied, at this point the algorithm operation ends. The expression of the cost function is: $Loss = -\frac{1}{m}\sum_{i=1}^{m} y_i \log f(x_i) + \lambda \sum_{k=1}^{l} sum\left(\|W_k\|^2\right)$ where m represents the samples total number, and l represents the layers number in the network.
Testing Phase	1. The clustering center extracted by FCM during the training phase is input into CNN to extract features. 2. Execute SVM. 3. Repeat step 4 and step 5 to give the running result of the test image.

IV. EXPERIMENT DESIGN AND DISCUSSION

Two sets of experiments are performed in this section. The first group of experiments verified that the proposed algorithm compared with the traditional CNN algorithm's classification accuracy and training time, and the experimental data is Fer-2013; The purpose of another set of experiments is to evaluate the recognition rate of F-CNN under complex backgrounds. The experimental data is obtained by mixing a part of each of the two data sets Fer-2013 and LFW.

A. ANALYSIS EXPERIMENT OF RECOGNITION ACCURACY AND TRAINING TIME

To analyze whether the F-CNN algorithm has improved recognition accuracy and training time compared to CNN. The data set was selected from the Fer-2013 database. To select the most objective experimental data, 5 fold cross validation was used in the experiment. The 35,886 samples in the database were evenly divided into five. Four of them were selected as the training sample set, and the remaining one was used as the test sample set. The experiment was repeated 5 times and the average value was taken as the final experimental data. It should be noted that both the training set and test set each contain 7 expressions such as sadness, anger and happiness.

Figure 5 plots the relationship between iterations and Accuracy of the two models on the training set.

The experimental results in the figure are run on the training data set. The information in the figure shows that both models can complete convergence after a certain number of iterations. A total of 28,708 samples were selected in the training set, and 48 samples were processed each time, so it took 598 times to complete all the samples in the sample set. In the experiment, training 60 generations on the training set is equivalent to 30,000 iterations. The message conveyed in Figure 5 is:

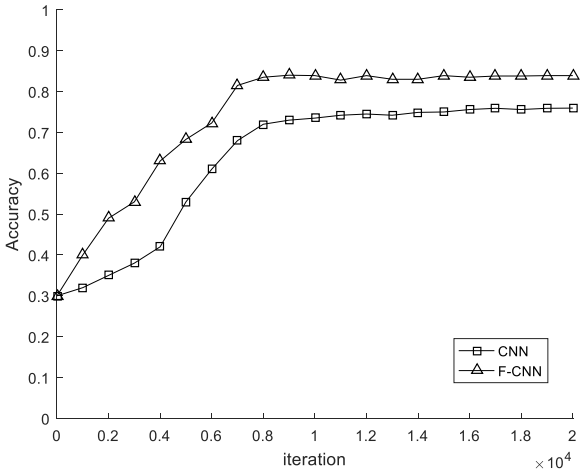

FIGURE 5. Comparison of the relationship between iterations and Accuracy of the two models on the training set.

(1) On the training set, the classification accuracy curves of the two models of F-CNN and CNN tend to be flat, which shows that both models can converge. When they started to converge, the recognition accuracy obtained by F-CNN was 83.86% and that of CNN was 76.95%. The difference between the recognition rates of the two models is 7%. So this can verify that the proposed F-CNN algorithm can indeed improve the FER rate.

(2) The F-CNN model started to converge after iterating to 12,000 times. After 139,00 iterations, the CNN started to converge. By comparing the number of iterations of the two models, it can be concluded that the F-CNN model can converge faster.

Table 1 is the comparison of F-CNN and CNN on the test set and training set. Note that the training time of the training set and test time of the test set described in Table 1 refer to the time-consuming process of processing a batch of images. Here, a batch of images refers to 48 images.

TABLE 1. F-CNN and CNN training time.

model	Training set training time (s)	Test set test time (s)	Layers	Recognition rate
F-CNN	195	27.43	13	83.86%
CNN	278	36.84	13	75.95%

Enlightenment from the data in Table 1:

The advantage of the F-CNN model over the CNN model is that the initialization of the convolution kernel in F-CNN uses the FCM algorithm to complete, and uses the SVM classifier for classification. Because SVM has no obvious advantage in algorithm complexity over Softmax classifier, the impact of SVM classifier on model training time can be ignored. As can be seen from the table 1, Compared with the CNN model, the F-CNN model takes less time in both cases in Table 1.

FIGURE 6. Part of the facial expression image of the LFW dataset.

Thence, the introduction of FCM on the basis of CNN has a certain effect on shortening the training time of the model.

B. ROBUSTNESS VERIFICATION EXPERIMENT

This group of experiments aims to evaluate the recognition effect of F-CNN on facial expression images with complex backgrounds. Since the complex background has been removed from the images in the Fer-2013 database, it is not suitable for the data of this experiment. This set of experimental data is a mixture of the Labeled Faces in the wild (LFW) dataset and the Fer-2013 dataset. Among them, the LFW dataset contains 14,000 face images. The LFW dataset is an image with a complex background collected from the network. Partial expression images of the database are shown in Figure 6. Since this set of experiments aims to verify the recognition performance of F-CNN for images with complex backgrounds, that is, whether the model is robust. This experiment did not perform the face image cropping process, but retained the background of the original face image to a certain extent.

In order to verify that F-CNN is more robust than CNN in FER in complex backgrounds, this experiment is designed as two links. In the first link, first 3000 images of each expression were randomly selected from the Fer-2013 database, for a total of 21,000 images. Then it is combined with

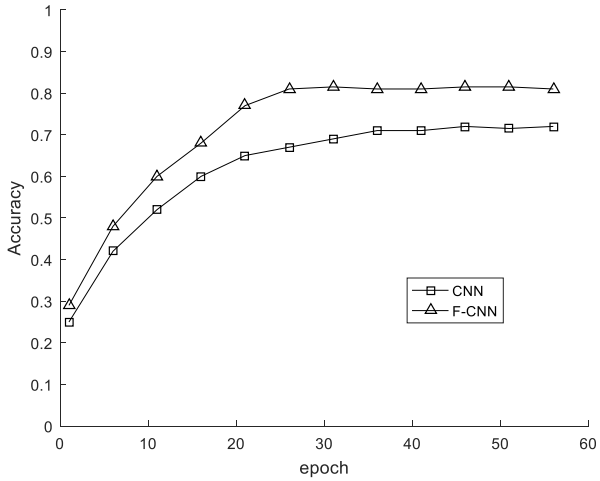

FIGURE 7. Recognition rate of each model under complex background.

14,000 images randomly selected in the LFW data set to construct a training set with a sample size of 35,000. Link 2 is to randomly select 27,000 pictures in the mixed data set as the training set and the remaining 8,000 pictures as the test set. The above algorithm obtains the FER rate under different test sets with different iterations. Figure 7 describes the recognition rate of each algorithm in a complex background.

From what is shown in Figure 7, it can be seen that both models reach a convergence state after iterating to a certain degree. The data set used in this experiment is the test set, the following conclusions can be obtained by analyzing Figure 7.

(1) Figure 7 shows that the proposed model begins to converge after 30 generations of training. After 36 generations, CNN began to converge. F-CNN has the same number of neural network layers as CNN, but F-CNN has a stronger FE capability. Moreover, F-CNN recognizes images with complex backgrounds and converges faster.

(2) The proposed F-CNN model has a lower recognition rate for data with complex backgrounds than standard datasets, such as the Fer-2013 dataset. However, compared with CNN, both iteration speed and recognition accuracy show that F-CNN still has certain advantages.

V. CONCLUSION

Although the CNN algorithm has successful application examples in the FER field, CNN has some obvious disadvantages, such as low model recognition rate in non-simple backgrounds, and model training takes more time. Based on this, this paper proposes a FER algorithm based on the combination of FCM clustering algorithm and CNN. The FCM algorithm is applied to the convolutional layer of the CNN to obtain a convolution kernel with an initial value to extract the expression image features in the training set and the test set. It is expected to increase the FE capability of the model and reduce model training time. From the results of simulation experiments, the F-CNN algorithm proposed in this paper can increase the recognition rate of facial expressions in complex

backgrounds to varying degrees, and can reduce the convergence time required to train CNN models. Although the algorithm in this paper has made good progress in processing FER in non-simple backgrounds, in reality, many collected images are multi-face images. So how to increase the recognition rate of facial expressions and ensure the robustness of the model in the case of non-simple backgrounds and multiple faces still needs further research.

REFERENCES

[1] H. Liu, J. Yin, X. Luo, and S. Zhang, "Foreword to the special issue on recent advances on pattern recognition and artificial intelligence," *Neural Comput. Appl.*, vol. 29, no. 1, pp. 1–2, Jan. 2018.

[2] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.

[3] Y. Jiang, K. Zhao, K. Xia, J. Xue, L. Zhou, Y. Ding, and P. Qian, "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *J. Med. Syst.*, vol. 43, no. 5, pp. 118:1–118:9, May 2019.

[4] P. Qian, J. Zhou, Y. Jiang, F. Liang, K. Zhao, S. Wang, K.-H. Su, and R. F. Muzic, "Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes," *IEEE Access*, vol. 6, pp. 28594–28610, 2018.

[5] P. Qian, C. Xi, M. Xu, Y. Jiang, K.-H. Su, S. Wang, and R. F. Muzic, "SSC-EKE: Semi-supervised classification with extensive knowledge exploitation," *Inf. Sci.*, vol. 422, pp. 51–76, Jan. 2018.

[6] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intell. Automat. Soft Comput.*, vol. 25, no. 3, pp. 547–559, 2019.

[7] T. Ahsan, T. Jabid, and U. P. Chong, "Facial expression recognition using local transitional pattern on Gabor filtered facial images," *IETE Tech. Rev.*, vol. 30, no. 1, pp. 47–52, 2013.

[8] N. Farajzadeh and M. Hashemzadeh, "Exemplar-based facial expression recognition," *Inf. Sci.*, vol. 460, pp. 318–330, Sep. 2018.

[9] J.-M. Sun, X.-S. Pei, and S.-S. Zhou, "Facial emotion recognition in modern distant education system using SVM," in *Proc. Int. Conf. Mach. Learn. Cybern.*, vol. 6, Jul. 2008, pp. 3545–3548.

[10] C. Zhan, W. Li, F. Safaei, and P. Ogunbona, "A real-time facial expression recognition system for online games," *Int. J. Comput. Games Technol.*, vol. 2008, pp. 1–10, Mar. 2008.

[11] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intell. Automat. Soft Comput.*, vol. 25, no. 3, pp. 547–559, 2019.

[12] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, "Research on cost-driven services composition in an uncertain environment," *J. Internet Technol.*, vol. 20, no. 3, pp. 755–769, 2019.

[13] H. Gao, Y. Duan, L. Shao, and X. Sun, "Transformation-based processing of typed resources for multimedia sources in the IoT environment," *Wireless Netw.*, pp. 1–17, Nov. 2019. [Online]. Available: https://doi.org/10.1007/s11276-019-02200-6

[14] H. Gao, Y. Xu, Y. Yin, W. Zhang, R. Li, and X. Wang, "Context-aware QoS prediction with neural collaborative filtering for Internet-of-Things services," *IEEE Internet Things J.*, early access, Dec. 2, 2019, doi: 10.1109/JIOT.2019.2956827.

[15] X. Ma, H. Gao, H. Xu, M. Bian, "An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, 2019, Art. no. 249. [Online]. Available: https://doi.org/10.1186/s13638-019-1557-3

[16] J. Yu, J. Li, Z. Yu, and Q. Huang, "Multimodal transformer with multi-view visual representation for image captioning," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Oct. 15, 2019, doi: 10.1109/TCSVT.2019.2947482.

[17] J. Yu, M. Tan, H. Zhang, D. Tao, and Y. Rui, "Hierarchical Deep Click Feature Prediction for Fine-grained Image Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jul. 30, 2019, doi: 10.1109/TPAMI.2019.2932058.

[18] C. Darwin, *The Expression of the Emotions in Man and Animals*. London, U.K.: John Murray, 1872.

[19] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Personality Social Psychol.*, vol. 17, no. 2, pp. 124–129, 1971.

[20] M. Suwa, "A preliminary note on pattern recognition of human emotional expression," in *Proc. 4th Int. Joint Conf. Pattern Recognit.*, 1978, pp. 408–410.

[21] K. Mase and A. Pentland, "Automatic lipreading by optical-flow analysis," *Syst. Comput. Jpn.*, vol. 22, no. 6, pp. 67–76, 1991.

[22] T. Wu, M. S. Bartlett, and J. R. Movellan, "Facial expression recognition using Gabor motion energy filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.-Workshops*, Jun. 2010, pp. 42–47.

[23] Y. Luo, C. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of PCA and LBP with SVM," *Optik-Int. J. Light Electron Opt.*, vol. 124, no. 17, pp. 2767–2770, 2013.

[24] M. Shin, M. Kim, and D. S. Kwon, "Baseline CNN structure analysis for facial expression recognition," in *Proc. 25th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 724–729.

[25] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, Jul. 1997.

[26] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[27] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 19–31, Feb. 1997.

[28] J. J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li, "Automated facial expression recognition based on FACS action units," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 390–395.

[29] P. Ekman, "An argument for basic emotions," *Cognition Emotion*, vol. 6, nos. 3–4, pp. 169–200, 1992.

[30] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[31] U. Mlakar, I. Fister, J. Brest, and B. Potočnik, "Multi-objective differential evolution for feature selection in facial expression recognition systems," *Expert Syst. Appl.*, vol. 89, pp. 129–137, Dec. 2017.

[32] A. M. Ashir and A. Eleyan, "Facial expression recognition based on image pyramid and single-branch decision tree," *Signal Image Video Process.*, vol. 11, no. 6, pp. 1017–1024, 2017.

[33] V. H. Duong, Y. S. Lee, J. J. Ding, B. T. Pham, Q. Bui, P. T. Bao, and J. C. Wang, "Projective complex matrix factorization for facial expression recognition," *Eurasip J. Adv. Signal Process.*, vol. 2018, no. 1, pp. 1–10, 2018.

[34] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1357–1362, Dec. 1999.

[35] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *Neural Information Processing*. Berlin, Germany: Springer, 2013, pp. 117–124.

[36] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Grenoble, France, Mar. 2000, pp. 46–53.

[37] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kande dataset (CK+): A complete facial expression dataset for action unit and emotionspecified expression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2010.

[38] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc. CVPR*, Jul. 2017, pp. 2852–2861.

[39] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proc. IEEE Int. Conf. Multimedia Expo*, Amsterdam, The Netherlands, Jul. 2005, p. 5.

[40] A. Dhall, R. Goecke, T. Gedeon, and S. Lucey, "Collecting large, richly annotated facial-expression databases from movies," *IEEE Multimedia*, vol. 19, no. 3, pp. 34–41, Jul./Sep. 2012.

[41] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Piscataway, NJ, USA: IEEE, Nov. 2011, pp. 2106–2112.

[42] E. Owusu, Y.-Z. Zhan, and Q.-R. Mao, "An SVM–AdaBoost-based face detection system," *J. Exp. Theor. Artif. Intell.*, vol. 26, no. 4, pp. 477–491, 2014.

[43] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 743–756, Jul. 1997.
[44] M. J. Black, D. J. Fleet, and Y. Yacoob, "A framework for modeling appearance change in image sequences," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 660–667.
[45] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Apr. 1998, pp. 454–459.
[46] A. Savran, H. Cao, A. Nenkova, and R. Verma, "Temporal Bayesian fusion for affect sensing: Combining video, audio, and lexical modalities," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1927–1941, Sep. 2015.
[47] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 786–791.
[48] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based HoG features," in *Proc. Face Gesture*, Mar. 2011, pp. 884–888.
[49] S. Berretti, B. Ben Amor, and M. Daoudi, and A. del Bimbo, "3D facial expression recognition using SIFT descriptors of automatically detected keypoints," *Vis. Comput.*, vol. 27, no. 11, p. 1021, 2011.
[50] H. Kang, T. F. Cootes, and C. J. Taylor, "Face expression detection and synthesis using statistical models of appearance," *Measuring Behav.*, pp. 126–128, Aug. 2002.
[51] M. Ranzato, J. Susskind, V. Mnih, and G. Hinton, "On deep generative models with applications to recognition," in *Proc. CVPR*, Jun. 2011, pp. 2857–2864.
[52] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[53] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2017, Jul. 2017, pp. 2261–2269.
[54] M. S. Bartlett, G. C. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan, "Automatic recognition of facial actions in spontaneous expressions," *J. Multimedia*, vol. 1, no. 6, pp. 22–35, 2006.
[55] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car-racing drivers: A biosignal processing approach," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 3, pp. 502–512, May 2008.
[56] C.-C. Hsieh, M.-H. Hsih, M.-K. Jiang, Y.-M. Cheng, and E.-H. Liang, "Effective semantic features for facial expressions recognition using SVM," *Multimedia Tools Appl.*, vol. 75, no. 11, pp. 6663–6682, Jun. 2016.
[57] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Comput. Vis. Image Understand.*, vol. 91, nos. 1–2, pp. 160–187, 2003.
[58] T.-H. Wang and J.-J. J. Lien, "Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estimation," *Pattren Recognit.*, vol. 42, pp. 962–977, May 2009
[59] K.-Y. Cheng, Y.-B. Chen, C.-J. Wen, Y.-Z. Zhan, "A new classifier for facial expression recognition: Fuzzy buried Markov model," *J. Comput. Sci. Technol.*, vol. 25, no. 3, pp. 641–650, 2010.
[60] 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.
[61] L. Ma and K. Khorasani, "Facial expression recognition using constructive feed forward neural networks," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 34, no. 3, pp. 1588–1595, Jun. 2004.
[62] S. V. Ioannou, A. T. Raouzaiou, T. P. Mailis, K. C. Karpouzis, S. D. Kollias, and V. A. Tzouvaras, "Emotion recognition through facial expression analysis based on a neurofuzzy network," *Neural Netw.*, vol. 18, no. 4, pp. 423–435, 2005.
[63] K. Takahashi, S. Takahashi, M. Hashimoto, and Y. Cui, "Remarks on computational facial expression recognition from HOG features using quaternion multi-layer neural network," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2014, pp. 15–24.
[64] J. S. Yoo, H. Ahn, and I. K. Choi, "Facial expression classification using deep convolutional neural network," *J. Elect. Eng. Technol.*, vol. 13, no. 1, pp. 485–492, 2018.
[65] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962.
[66] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.
[67] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10. Cambridge, MA, USA: MIT Press, 1995, p. 1995.

MIN SHI graduated from Great College, Daegu University, South Korea, in February 2011, and the Ph.D. degree in design. In September 2015, he left the Postdoctoral Mobile Station with the Southeast University, Nanjing, China. He is currently an Associate Professor with the Department of Product Design, Fuzhou University of International Studies and Trade. His main research interests include kansei engineering and interactive design.

LIJUN XU is currently an Associate Professor with the School of Art and Design, Nanjing Institute of Technology, and a Senior Visiting Professor with Northumbria University. The lectures include product sketch and expression, ergonomics, computer graphic design, and thematic design. In the past five years, she has completed one project of Humanities and Social Sciences of the Ministry of Education and one major project of Philosophy and Social Science Research in Jiangsu Province. She has mainly participated in the completion of two provincial- and municipal-level projects, successively in foreign SCI, EI journals, and CSSCI. She has more than ten academic articles, obtained seven invention patents and several utility model patents, and successfully transformed three invention patents.

XIANG CHEN received the M.D. degree from the School of Art Design, Dongseo University, Busan, South Korea, in 2011. She is currently an Associate Professor with the School of Design, Jiangnan University, Wuxi, China. Her research areas include design of human–machine interaction, computer science and technology, system innovation, and design strategy.