

A Novel Fuzzy Optimized CNN-RNN Method for Facial Expression Recognition

Dong Zhang, Qichuan Tian*

College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture,
Beijing 100044, China
tianqichuan@bucea.edu.cn

Abstract—Facial expression is one of the important ways of transferring emotion in interpersonal communication, and it has been widely used in many interpersonal communication systems. The traditional facial expression recognition methods are not intelligent enough to manage the model uncertainty. The deep learning method has obvious ability to deal with model uncertainty in the image recognition. The deep learning method is able to complete the facial expression work, but the recognition rate can be further improved by a hybrid learning strategy. In this paper, a Fuzzy optimized convolutional neural network-recurrent neural network (CNN-RNN) method for facial expression recognition is proposed to solve the problems of direct image convolution without image enhancement and simple convolution stack ignoring feature layer-by-layer convolution resulting in information loss. Firstly, each face image is scaled by the bilinear interpolation and the affine transformation is adopted to expand the image data to avoid the shortage of the data set. Then the feature map of the facial expression is extracted by the CNN with small information loss. To deal with the uncertainty in the feature map, the Fuzzy logic is employed to reduce the uncertainty by recognizing the highly nonlinear relationship between the features. Then the output of the Fuzzy model is fed with the RNN to classify different facial expression images. The recognition results based on the open datasets CK, Jaffe, and FER2013 show that the proposed Fuzzy optimized CNN-RNN method has a certain improvement in the recognition effect of different facial expression data sets compared with current popular algorithms.

Index Terms—Facial expression; CNN; Fuzzy optimized CNN-RNN; Uncertainty reduction.

I. INTRODUCTION

Expression is one of the most important ways to convey emotions in interpersonal communication. It mainly refers to the formation of facial muscles and facial features. Common expressions include anger, disgust, fear, joy, surprise, etc. [1]. People can get their emotions directly through facial expressions. In addition, automatic facial expression analysis has been widely applied in social robots [2]–[5], medical treatment [6], driver fatigue monitoring [7]–[9], and many other human-computer interaction systems [10]–[13], and facial expression recognition has a very important research value.

In general, face expression recognition methods are mainly divided into three steps: face detection, feature extraction,

and classification recognition, and feature extraction is the key to ensure face detection [14]. The existing feature extraction methods mainly include traditional methods and deep learning methods [15]. Traditional methods for facial expression recognition usually use manual features or shallow learning, e.g., local binary mode [16], local binary patterns from three original planes, LBP-TOP [17]. However, the traditional machine learning method is difficult to meet the needs of the current era due to its long time of processing data, lack of intelligence, and weak performance [18]. Since 2013, the emotion recognition competitions, such as FER2013 and EmotiW, have collected relatively sufficient training data from challenging realistic scenes, which has virtually promoted the transformation of facial expression recognition from laboratory control to real environment [19].

With the rapid development of graphics processing unit (GPU) computing power of hardware, it is relatively easy to train large deep learning networks [20]. Therefore, the deep learning technology is increasingly applied to face expression recognition in the real world [21]. In terms of facial expression recognition, the deep learning algorithm does not rely on complex image preprocessing and does not need to design accurate manual features [22]. Compared with traditional machine learning algorithms, the deep learning algorithm performs better and has better robustness in the face of illumination, various postures, and occlusion problems. Hu *et al.* [23] used a convolutional network to complete facial expression recognition and detect facial key points through multi-task learning. In the multi-task training, the bottom layer of convolutional neural network (CNN) network is shared, and the layer that needs to be shared is determined by automatic learning, so the model cannot be affected by the distribution of data sets. To avoid complex feature extraction process and data manipulation problems involved in traditional facial expression recognition, Jiang and Learned-Miller [24] based on the Faster recurrent neural network (RNN) model in the field of target detection to recognize facial expressions. Firstly, the face expression images are normalized, and then the model convolutional network extracts the features and reduces dimension of feature maps. The candidate regions are generated by the region candidate network and detected by FERT-RNN. Finally, the expression recognition classification and border coordinates are given by Softmax classifier and regression

layer, respectively. Ghazouani [25] constructed the expression recognition method of principal component analysis (PCA) decomposed images and applied neural network to classify the features obtained from PCA decomposed images, and achieved good detection results on four data sets. Ly, Do, Kim, Yang, and Lee [26] transformed the images into 3D features, and then realized expression recognition with the CNN network constructed. Compared with the two-dimensional feature network, this method has better network generalization ability and can realize face expression recognition in multiple scenes. Li, Jin, Zhou, Kubota, and Ju [27] introduced the attention mechanism into the CNN network, fully and efficiently applied all feature graphs, so the detection accuracy was greatly improved. However, the introduction of the attention mechanism made the model more complex and poorer in timeliness. Yu, Zheng, Peng, Dong, and Du [28] constructed a detection network that separated local and global features, which improved the detection accuracy of similar expressions through local features and achieved good detection results on CK+ and Oull-CASIA databases with strong robustness.

Although deep learning method has achieved good results in expression recognition, different algorithms have their own limitations. For example, two problems of CNN are pointed out:

1. Most networks convolve directly on the expression image without enhancing the image, which reduces the detection accuracy of the network when the image distortion occurs;
2. Most of the current network stacks use simple convolution, ignoring the characteristics of the figure and the problem of information loss in the process of convolution step by step.

Recent literature shows that CNN can be combined with other algorithms to achieve better recognition performance

[29]. However, very limited work has been done to address this issue. To this end, this study proposed a novel combination of CNN-based facial recognition method, which increased the image preprocessing module in the CNN structure by the double linear quadratic interpolation to keep the detail characteristics of the image; in addition, the affine transformation increased the number of datasets to enhance the generalization ability of the RNN structure. Finally, the Fuzzy optimized CNN-RNN structure is used to train and recognize facial expressions in the public database to obtain better recognition results. The contributions of this work are summarized as follows.

- For the first time, a novel CNN combination model is proposed for facial recognition, which can improve the convolution process in the CNN operation;
- For the first time, the Fuzzy logic is employed to optimize the CNN-RNN model, which can improve the efficiency and effectiveness for facial recognition.

II. FUZZY OPTIMIZED CNN-RNN STRUCTURE

A. Overview of the Fuzzy-CNN-RNN Method

To solve the problem that the original CNN is difficult to make effective use of the global situation and information in the classification process, a new network structure is constructed in this paper. The network structure mainly includes four parts: a) The preprocessing part, which mainly makes the image zoom and affine transformation; b) In the feature extraction part of CNN, the pre-processed facial expression images are extracted by convolution; c) By adding Fuzzy-RNN module to realize the application of global information and improve the detection accuracy; d) Full connection layer, through which the prediction is realized. The network framework is shown in Fig. 1.

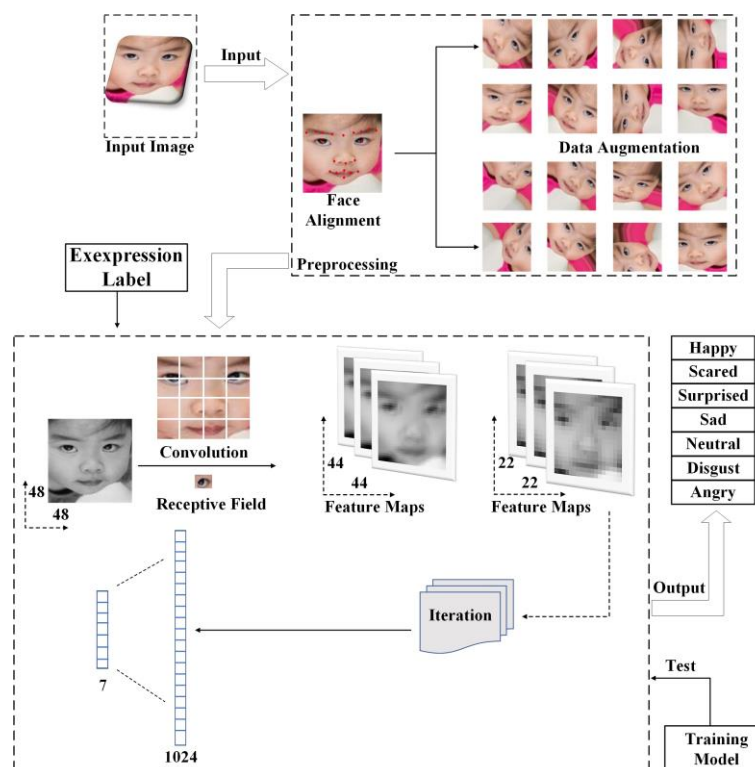


Fig. 1. An overview of the proposed model.

In Fig. 1, the original is mainly scaled by bilinear quadratic interpolation method. Image scaling cannot only reduce the image to shorten the training time, but also expand the facial expression data set by scaling the image at different multiples. Then the affine transformation of the scaled image is carried out to realize the data set expansion. Finally, the image is input to the improved convolutional neural network for classification. The main modules of the improved convolutional neural network in this paper are shown in Table I. The network mainly consists of five modules. The first module is C1, which is the convolution layer. The convolution features of the first layer are obtained through 5×5 convolution, and the obtained convolution features of the first layer are $32 \times 44 \times 44$, and the second layer is S2, and the feature data are reduced through the maximum pooling. The third layer is the convolution of 3×3 , and the size of the feature map is $64 \times 20 \times 20$. The feature map is reduced through the maximum pooling layer of the fourth layer, and the classification is finally achieved through Fuzzy-RNN.

TABLE I. DESCRIPTION OF EACH NETWORK MODULE.

Layer	Kernel Size	Output shape
Input		48×48
C1	$32 @ 5 \times 5$	$32 @ 44 \times 44$
S2	2×2	$32 @ 22 \times 22$
C3	$64 @ 3 \times 3$	$64 @ 20 \times 20$
S4	2×2	$64 @ 10 \times 10$
Fuzzy-RNN		7
Output		7

B. Dataset Preprocessing

In the task of facial expression recognition by computer vision, it is very necessary to preprocess the dataset. The preprocessing of this paper mainly includes the clipping of facial expression image and the expansion of facial expression image.

1. Image resizing

Image clipping is mainly to adjust the size of the image to fit the input of network Settings. It is a very common preprocessing method in the field of deep learning. Different from the linear interpolation in [30] and [31], in this work, the bilinear interpolation method is used to adjust the size of the image, which can keep the details of the image features to ensure the integrity of the face image and to avoid the identification accuracy reduction.

The principle of bilinear interpolation is shown in Fig. 2. The four adjacent grayscale values at (X, Y) are used to calculate the grayscale values at (X, Y) . Let four adjacent pixels be A, B, C, and D, and their coordinates are (i, j) , $(i + 1, j)$, $(i, j + 1)$, $(i + 1, j + 1)$, and their grayscale values are $g(A)$, $g(B)$, $g(C)$, and $g(D)$, respectively. First, calculate the grayscale values $g(E)$ and $g(F)$ at points E and F:

$$g(E) = (X - i)[g(B) - g(A)] + g(A) = (i + 1 - X)g(A) + (X - i)g(B), \quad (1)$$

$$g(F) = (X - i)[g(D) - g(C)] + g(C) = (i + 1 - X)g(C) + (X - i)g(D). \quad (1)$$

Then the gray value at (X, Y) is calculated by

$$g(X, Y) = (Y - j)[g(F) - g(E)] - g(E) = (j + 1 - Y)g(E) + (Y - j)g(F). \quad (2)$$

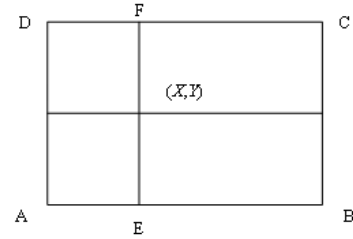


Fig. 2. Bilinear interpolation.

The above bilinear difference method is used to zoom the facial expression image. The zoomed expression image is shown in Fig. 3.

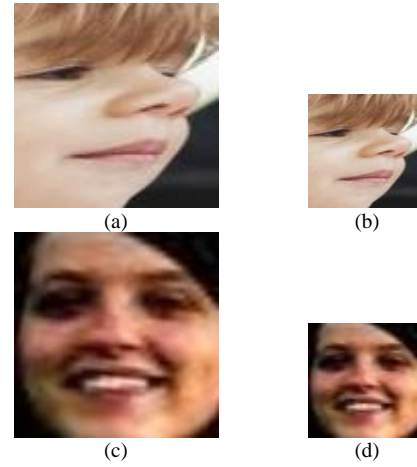


Fig. 3. Scaling of facial expression image: (a) Original emotional image; (b) Scale emotional image; (c) Original emotional image; (d) Scale emotional image.

2. Image expansion

The introduction of a large number of training data is to ensure the accuracy of network identification [32]. However, many studies directly use them for network training after scaling, which limits the generalization ability of the network to a certain extent, resulting in a low detection accuracy in the case of face data distortion. To solve the above problems, the affine transformation of the image is carried out in the process of preprocessing. It can not only increase the number of training data sets, but also improve the generalization ability of the network by training data sets from different angles.

The Affine Transformation [33], [34] is a linear transformation method of two-dimensional coordinates, which can maintain the straightness and parallelism of two-dimensional graphs at the same time. In other words, affine transformation allows the image to be inclined at will and can be contracted at will in two directions. It is applicable to the cases of horizontal shift, rotation, scaling, and inversion, and can be expressed by (4):

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (3)$$

In the formula, (t_x, t_y) represents the translation amount, and the parameter a_i reflects the size of the image rotation and scaling amount. (x', y') represents the image after transformation, and (x, y) represents the image before transformation. After affine transformation, the facial expression image is shown in Fig. 4, where $a_1, a_2, a_3,$ and a_4 are 1, 0, 0, and 1.5, respectively, and the translation is $(0, 0)$.

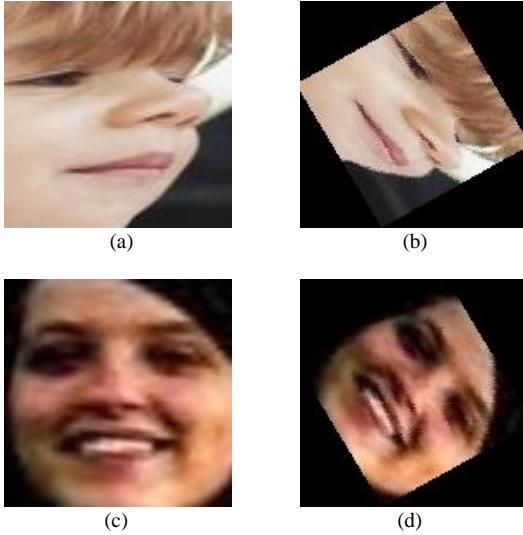


Fig. 4. Affine transformation of facial expression image: (a) Original emotional image; (b) Affine transform expression image; (c) Original emotional image; (d) Affine transform expression image.

3. Loop structure

At present, to improve the detection accuracy of neural network, people use the deep neural network model [35], but lack a deepening convolution neural network features to solve the figure low utilization ratio and disappearing problem. In this paper, the traditional CNN model is optimized by combining CNN with Fuzzy-RNN for facial expression recognition.

First, RNN has a powerful ability to capture contextual information in sequences [36]. It is more effective to use contextual clues for image-based sequence recognition than to process each symbol independently. Secondly, RNN can backpropagate errors to the convolutional layer so that the network can be trained end-to-end [37], [38]. By introducing the RNN module into the CNN model, the convolution feature can be recycled [39]. On this basis, not only the redundant features extracted from the network can be reduced, but also the disappearance of features can be avoided by effectively utilizing the convolution feature. Thirdly, the Fuzzy logic is powerful to manage the uncertainty in the feature map [40]; hence, the Fuzzy-RNN is able to significantly improve the recognition performance on the facial expression.

RNN was first used in natural language processing, where the entire sentence is defined as sequential data and each word is based on the understanding of the previous word [41], [42]. When an artificial neural network performs natural language processing, it needs a structure to reason about the next word based on the context of a sentence that combines the previous output as the input for the inference. RNN is a series of neural networks used to process sequential data. The RNN structure is shown in Fig. 5.

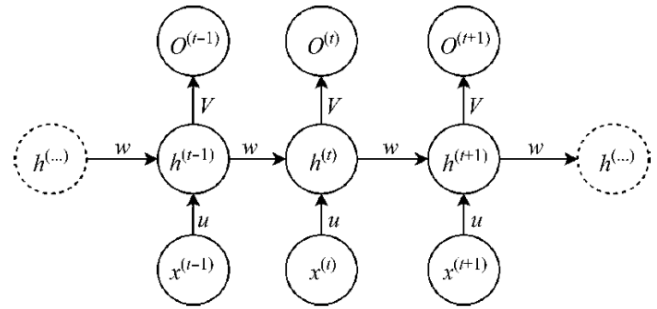


Fig. 5. Schematic diagram of RNN structure.

In Fig. 5, O^t is the output unit, X^t is the given input convolution feature sequence, and h^t is the hidden layer unit. A one-way flow of information from the input unit reaches the hidden unit, and another one-way flow of information from the hidden unit reaches the output unit. h^t is calculated based on the output of the current input layer and the state of the previous hidden layer h^{t-1} , as shown in (5) [37], [43]

$$h^{(t)} = f(Ux^{(t)} + Wh^{(t-1)}), \quad (4)$$

where f represents the nonlinear activation function, such as tan or Relu, and U and W are shared parameters. O^t is the output of step t , which depends on the activation function of the current neuron, as shown in (6) [38], [44]

$$O^{(t)} = \sigma(Vh^{(t)}), \quad (5)$$

where σ represents the activation function of the output layer.

Due to stochastic nature of the CNN deep learning process, the uncertainty often exists in the obtained feature map. To relieve the uncertainty effect, the Fuzzy logic is used to map the highly nonlinear relationship between different features. By doing so, the uncertainty of features can be reduced to allow the Fuzzy-RNN to start the circular convolution from the input sequence feature map so that the face expression feature map can be used more efficiently to realize the utilization of the global information in the context of the feature map. The details of the Fuzzy-RNN theory can refer to [44]–[46].

III. RESEARCH DESCRIPTION

A. Training Method

The experimental environment is carried out on the deep learning server with GPU Tesla P100, the operating system is Linux, and the version of Python2.7 is used. Firstly, the image size of the data set was processed into 48×48 pixels, and the batch size was set to 128 according to the GPU video memory in the training process. The training cycles of the three data sets were 60, 60, and 250, respectively. The momentum gradient descent method was adopted in the optimization algorithm. The initial learning rate is 0.01, the momentum is 0.9, and the weight attenuation is 5×10^{-4} . The overfitting can be prevented and the model generalization performance can be improved by weight attenuation.

During the training process, after each Epoch is completed, the model will be evaluated and saved through the test set. After all training rounds, the weight value with the best recognition effect will be saved into the final model file. The

training results of the proposed method are shown in Fig. 6. As can be seen, the training accuracy gradually increases and is close to 100 % after 3,000 epochs, while the entropy loss gradually decreases and is close to 0 after 3,000 epochs. There are no sudden fluctuations in the accuracy and entropy loss curves during the training process, indicating that the training of the proposed method is stable and reliable.

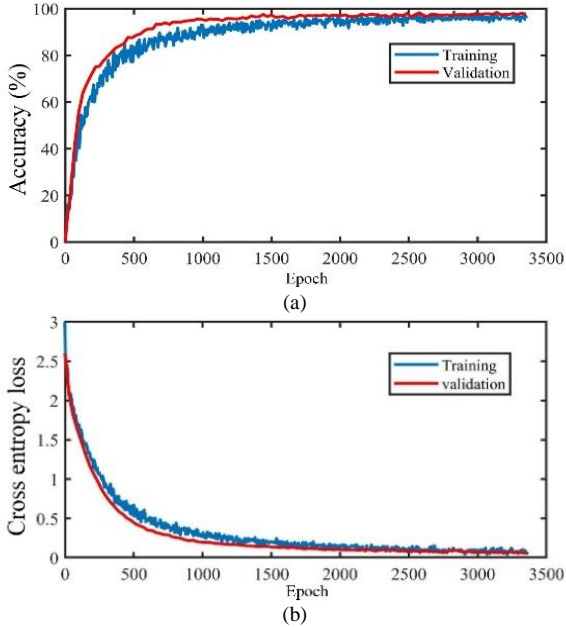


Fig. 6. Training performance curve: (a) Accuracy; (b) Entropy loss.

B. Evaluation Index

Cross-entropy loss function: In the classification task, the cross-entropy error is often used as a loss function [46]–[49]. The cross-entropy error is shown as follows

$$E = -\sum_k t_k \log y_k, \quad (6)$$

where \log represents the natural logarithm with E as the base, y_k is the output of the neural network, and t_k is the correct

solution label.

Accuracy rate: Accuracy rate is defined as the evaluation index of network performance, and the calculation method of the accuracy rate is

$$P = \frac{TP}{TP + FP}. \quad (7)$$

In the formula, P is the network accuracy rate, TP is the number of correctly classified images, and FP is the number of wrongly detected images. By calculating the above detection accuracy rate, the performance of different networks is compared and analysed.

C. Data Set Description

CK+, Jaffe, and FER2013 datasets from the open dataset were used for training and testing. All the datasets contained seven facial expressions, including anger, neutral, disgust, scared, happiness, sad, and surprised. CK+ dataset was extended from the Cohn-Kanade database and contained 327 tagged facial videos. The images used in the experiment were extracted from the last three frames of each sequence. A total of 981 pictures with seven facial expressions were selected. JAFFE dataset contains seven expression images of 213. In this study, the images of the human face will be processed by the bilinear interpolation, and then the affine transformation will be used to rotate the images' angles five times for data expansion. FER2013 facial expression dataset contains 35,886 facial expression images and is a common dataset for facial expression recognition competitions. The three kinds of facial expressions are shown in Fig. 7.

There are many wrong labels in the data samples of FER2013 data set, and some samples are not positive facial expression pictures and face occlusion. FER2013 data set is much higher than CK+ and Jaffe data sets in terms of scale and recognition difficulty. In the experiment, the recognition rate of FER2013 private test set was selected as the model recognition accuracy, and the final sample grouping of the three data sets was shown in Table II.

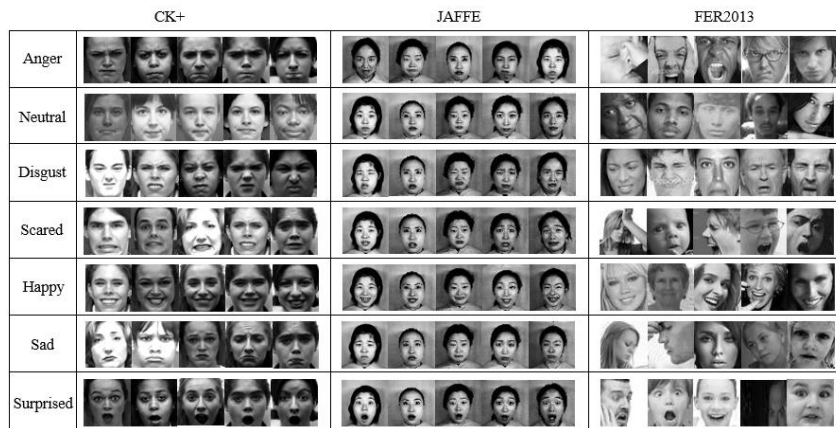


Fig. 7. Examples of various emoticons in three datasets.

TABLE II. COMPARISON OF TRAINING AND TEST SAMPLE SIZE.

Dataset	CK+	JAFFE	FER2013
Number of Expression Categories	7	7	7
The training set	880	860	28000
The test set	95	200	3500

IV. TEST RESULTS AND ANALYSIS

A. Comparison of Different Models

In the network model, Resnet18, Resnet50, Resnet50 + congenital pulmonary airway malformations (CPAM) and CNN-RNN were selected to compare the recognition effect,

and the test performance index was made through the recognition rate and the number of model parameters. The experimental results on the two data sets are shown in Table III.

TABLE III. COMPARISON OF DIFFERENT NETWORK MODELS.

Network model	CK+	JAFFE	FER2013	Params
ResNet18	90.91 %	88.67 %	69.10 %	11.69
ResNet50	92.93 %	89.66 %	70.10 %	25.56
ResNet50 + CBAM	96.97 %	93.59 %	71.00 %	27.55
CNN	98.48 %	94.37 %	50.32 %	26.62
RNN	99.01 %	96.06 %	55.56 %	25.13
Fuzzy optimized CNN-RNN	99.22 %	96.64 %	72.81 %	28.55

Compared with Resnet18 or its improved model and the separate deep learning model, the Fuzzy optimized CNN-RNN model has improved gender recognition, and the recognition rates on CK+, Jaffe, and FER2013 data sets are 99.22 %, 96.64 %, and 72.81 %, respectively. This is because by introducing the RNN module, the Fuzzy optimized CNN-RNN can make full use of the convolution feature to avoid the loss of context information. Combined with the expanded data set, the detection accuracy is improved to a certain extent.

By deepening the network depth, the RESNET model has improved the recognition performance of the training model to some extent, but the number of model parameters has increased too much, and the effectiveness is getting lower and lower. Resnet50 increased by 13.87M in the number of parameters compared with Resnet18, and the recognition rate was only improved by 1 %, 1 %, and 2 %, while the Fuzzy optimized CNN-RNN constructed in this paper improved by 6.29 %, 6.98 %, and 2.71 % compared with Resnet50 with

the increase of 2.99M in the number of fewer parameters. Compared with Resnet50 + CBAM, the increase of 1M fewer parameters improved the recognition rates by 2.25 %, 3.05 %, and 1.81 %, which proved the feasibility of the proposed method.

Compared with CNN and Fuzzy-RNN network models, the identification of CK+ and Jaffe has a certain improvement, but not significant, while the identification of FER2013 database has a significant improvement. At this time, the increase of the number of parameters is not significant, which also conforms to the statement in [24] and further proves the effectiveness of the method presented in this paper.

B. Comparison of CNN-RNN with other Methods

The various expression recognition rates of the Fuzzy optimized CNN-RNN on CK+ and FER2013 data sets are compared with those of other methods, as shown in Table IV and Table V.

As can be seen from Table IV and Table V, the Fuzzy optimized CNN-RNN method has some deficiencies in the judgment of fear expressions on CK+ and has achieved good recognition effect on the other 6 expressions. In FER2013 data set, the recognition effect of happy, surprise, and disgust are good, and there are some discrimination errors in anger, fear, and sadness, because there are similarities between these three expressions, while neutral and sad expressions have little change in facial features, which increases the difficulty of recognition. In general, compared with other literature in recent years, the Fuzzy optimized CNN-RNN method achieves a better recognition rate, which proves the effectiveness of the Fuzzy optimized CNN-RNN structure for facial expression recognition.

TABLE IV. COMPARISON BETWEEN CK+ AND OTHER METHODS (%).

Methods	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Recognition Rate
[51]	89.57	100	84.17	100	95.00	98.33	-	96.28
[52]	100	100	94.44	100	100	91.67	-	98.18
[53]	95	100	96	100	95	100	-	98.52
Improvement of RARNet	100	100	89	100	100	100	100	98.99
Fuzzy optimized CNN-RNN	100	100	92	100	100	100	100	99.22

TABLE V. COMPARISON WITH OTHER METHODS ON FER2013 (%).

Methods	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutral	Recognition Rate
[50]	67	60	64	75	80	83	67	70.59
[52]	65.38	69.09	53.41	90.67	61.45	81.73	71.09	72.11
Improvement of RARNet	65	76	55	89	64	84	72	72.78
Fuzzy optimized CNN-RNN	65	76	64	90	63	80	71	72.81

As can be seen from Table IV, the Fuzzy optimized CNN-RNN method on CK+ database is 2.71 % higher than that proposed by Gan, Chen, Yang, and Xu [51] for embedding single-pooled channel attention module in the convolutional layer. Compared with the feature fusion reclassification network model proposed by Zhang, Huang, and Tian [53], which extracted the original image and Local Binary Pattern (LBP) feature image through two VGG

network channels respectively, the reclassification network model achieved better recognition performance. The method of adding attention and the method presented in this paper still maintain the similar recognition results.

As can be seen from Table V, the total recognition rate of the Fuzzy optimized CNN-RNN method in the Fer2013 dataset is 1.22 % higher than that of the convolutional neural network enhanced in the preprocessing stage proposed by

Khemakhem and Ltifi, [50] and 0.7 % higher than that of adding the course learning strategy in the process of facial expression recognition training proposed by Liu and Zhou [52]. Both methods enhance feature learning at different stages of facial expression recognition. The bilinear interpolation and the attention mechanism were used to enhance the key feature of the images with more datasets than the traditional linear interpolation. As a result, the residual network integration in the present model can avoid the network depth degradation problem, which has the advantages over the RARNet model.

Lastly, we would like to mention possible applications of the proposed facial recognition method. The first application is for the intelligent cell phones; the second application is in the intelligent house equipment. Moreover, there are still many other applications of the facial recognition systems, including the social robots, medical treatment driver, fatigue monitoring, and human-computer interaction systems.

V. CONCLUSIONS

In this paper, a Fuzzy optimized CNN-RNN method is proposed for facial expression recognition. The recognition performance of this new method is compared with many existing popular methods using major testing datasets. The comparison demonstrates that the proposed method has improved the recognition rate of different facial expression datasets and performs better than existing CNN-based method. The max improvement can reach to 3.4 %, which is a good indicator to show the advantage of the proposed Fuzzy optimized CNN-RNN method. However, due to the influence of mislabeling, non-positive expression, occluded expression, and low-quality expression, there is still a lot of room for improvement in the facial expression recognition. The next step will be to further study the attention mechanism algorithm and Fuzzy-based uncertainty reduction.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local Gabor features and classifier synthesis", *Pattern Recognition*, vol. 45, no. 1, pp. 80–91, 2012. DOI: 10.1016/j.patcog.2011.05.006.
- [2] A. Bunde, "Design of a patient-therapist-social robot therapy system in neurorhabilitation therapies for engagement and motivation", in *Proc. of 26th Inter. Conf. Intel. User Interfaces*, 2021, pp. 93–95. DOI: 10.1145/3397482.3450712.
- [3] R. Søråa and M. E. Fostervold, "Social domestication of service robots: The secret lives of Automated Guided Vehicles (AGVs) at a Norwegian hospital", *Inter. J. Human Comput. Studies*, vol. 152, p. 102627, 2021. DOI: 10.1016/j.ijhcs.2021.102627.
- [4] Y. Song, L. A. Luximon, and Y. Luximon, "The effect of facial features on facial anthropomorphic trustworthiness in social robots", *Appl. Ergon.*, vol. 94, p. 103420, 2021. DOI: 10.1016/j.apergo.2021.103420.
- [5] H. Ru and S. H. Koo, "Big data analyses on key terms of wearable robots in social network services", *Inter. J. Cloth. Sci. Tech.*, to be published. DOI: 10.1108/IJCS-11-2020-0180.
- [6] M. H. Williams and S. A. Broadley, "SUNCT and SUNA: Clinical features and medical treatment", *J. Clinical Neurosci.*, vol. 15, no. 5, pp. 526–534, 2008. DOI: 10.1016/j.jocn.2006.09.006.
- [7] Y. Xu, W. Yang, X. Yu, H. Li, T. Cheng, X. Lu, and Z. L. Wang, "Real-time monitoring system of automobile driver status and intelligent fatigue warning based on triboelectric nanogenerator", *ACS Nano*, vol. 15, pp. 7271–7278, 2021. DOI: 10.1021/acsnano.1c00536.
- [8] Q. Zhao, J. Jiang, Z. Lei, and J. Yi, "Detection method of eyes opening and closing ratio for driver's fatigue monitoring", *IET Intel. Trans. Sys.*, vol. 15, no. 1, pp. 31–42, 2021. DOI: 10.1049/itr2.12002.
- [9] A. Laucka, D. Andriukaitis, "Research of the Defects in Anesthetic Masks", *Radioengineering*, vol. 24, no. 4, pp. 1033–1043, 2015. DOI: 10.13164/re.2015.1033.
- [10] C. Paton, A. W. Kushniruk, E. M. Borycki, M. English, and J. Warren, "Improving the usability and safety of digital health systems: The role of predictive human-computer interaction modeling", *J. Med. Intern. Res.*, vol. 23, no. 5, p. e25281, 2021. DOI: 10.2196/25281.
- [11] A. Glowacz, "Fault diagnosis of electric impact drills using thermal imaging", *Measurement*, vol. 171, art. 108815, 2021. DOI: 10.1016/j.measurement.2020.108815.
- [12] X. Chen, M. Cao, H. Wei, Z. Shang, and L. Zhang, "Patient emotion recognition in human computer interaction system based on machine learning method and interactive design theory", *J. Med. Imag. Health Infor.*, vol. 11, no. 2, pp. 307–312, 2021. DOI: 10.1166/jmihi.2021.3293.
- [13] T. Surgailis, A. Valinevicius, V. Markevicius, D. Navikas, D. Andriukaitis, "Avoiding forward car collision using stereo vision system", *Elektronika ir elektrotechnika*, vol. 18, no. 8, pp. 37–40, 2012. DOI: 10.5755/j01.eee.18.8.2609.
- [14] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study", *Image Vision Comput.*, vol. 27, no. 6, pp. 803–816, 2009. DOI: 10.1016/j.imavis.2008.08.005.
- [15] Y. Miao, B. Zhang, J. Lin, M. Zhao, H. Liu, Z. Liu, and H. Li, "A review on the application of blind deconvolution in machinery fault diagnosis", *Mech. Sys. Signal Process.*, vol. 163, art. 108202, 2022. DOI: 10.1016/j.ymsp.2021.108202.
- [16] K.-Y. Tsai, Y.-W. Tsai, Y.-C. Lee, J.-J. Ding, and R. Y. Chang, "Frontalization and adaptive exponential ensemble rule for deep-learning-based facial expression recognition system", *Signal Process. Image Comm.*, vol. 96, art. 116321, 2021. DOI: 10.1016/j.image.2021.116321.
- [17] B. Niu, Z. Gao, and B. Guo, "Facial expression recognition with LBP and ORB features", *Comput. Intel. Neuroscience*, vol. 2021, art. ID 8828245, 2021. DOI: 10.1155/2021/8828245.
- [18] Y. Li, S. Wang, Y. Yang, and Z. Deng, "Multiscale symbolic fuzzy entropy: An entropy denoising method for weak feature extraction of rotating machinery", *Mech. Sys. Signal Process.*, vol. 162, art. 108052, 2022. DOI: 10.1016/j.ymsp.2021.108052.
- [19] S. Li and W. Deng, "Deep facial expression recognition: A survey", *IEEE Trans. Affec. Comput.*, to be published. DOI: 10.1109/TAFCC.2020.2981446.
- [20] N.-M. Ho and W.-F. Wong, "Tensorox: Accelerating GPU applications via neural approximation on unused tensor cores", *IEEE Trans. Paral. Dist. Sys.*, vol. 33, no. 2, pp. 429–443, 2022. DOI: 10.1109/TPDS.2021.3093239.
- [21] F. C. Krause, E. Linardatos, D. M. Fresco, and M. T. Moore, "Facial emotion recognition in major depressive disorder: A meta-analytic review", *J. Affec. Disor.*, vol. 293, pp. 320–328, 2021. DOI: 10.1016/j.jad.2021.06.053.
- [22] Y. Lu, H. Zhang, L. Shi, F. Yang, and J. Li, "Expression-EEG bimodal fusion emotion recognition method based on deep learning", *Comput. Mathe. Method. Med.*, vol. 2021, art. ID 9940148, 2021. DOI: 10.1155/2021/9940148.
- [23] G. Hu *et al.*, "Deep multi-task learning to recognize subtle facial expressions of mental states", in *Computer Vision - ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*, vol. 11216. Springer, Cham, 2018. DOI: 10.1007/978-3-030-01258-8_7.
- [24] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN", in *Proc. of 12th IEEE Inter. Conf. Auto. Face Gesture Recog.*, 2017, pp. 650–657. DOI: 10.1109/FG.2017.82.
- [25] H. Ghazouani, "A genetic programming-based feature selection and fusion for facial expression recognition", *Appl. Soft Comput.*, vol. 103, art. 107173, 2021. DOI: 10.1016/j.asoc.2021.107173.
- [26] T. S. Ly, N.-T. Do, S.-H. Kim, H.-J. Yang, and G.-S. Lee, "A novel 2D and 3D multimodal approach for in-the-wild facial expression recognition", *Image Vision Comput.*, vol. 92, art. 103817, 2019. DOI: 10.1016/j.imavis.2019.10.003.
- [27] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, "Attention mechanism-based CNN for facial expression recognition", *Neurocomput.*, vol. 411, pp. 340–350, 2020. DOI: 10.1016/j.neucom.2020.06.014.
- [28] M. Yu, H. Zheng, Z. Peng, J. Dong, and H. Du, "Facial expression

- recognition based on a multi-task global-local network”, *Pattern Recognition Letters*, vol. 131, pp. 166–171, 2020. DOI: 10.1016/j.patrec.2020.01.016.
- [29] B. Fasel and J. Luettin, “Automatic facial expression analysis: A survey”, *Pattern Recog.*, vol. 36, no. 1, pp. 259–275, 2003. DOI: 10.1016/S0031-3203(02)00052-3.
- [30] M. S. S. Moustafa, S. A. Mohamed, S. Ahmed, and A. H. Nasr, “Hyperspectral change detection based on modification of UNet neural networks”, *J. Appl. Remote Sens.*, vol. 15, no. 2, ID 028505, 2021. DOI: 10.1117/1.JRS.15.028505.
- [31] J. Li, X. Lin, H. Che, H. Li, and X. Qian, “Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet”, *Phys. Med. Biology*, vol. 66, no. 11, art. ID 115010, 2021. DOI: 10.1088/1361-6560/abfce3.
- [32] H. Wang, Y. Wang, and Y. Cao, “Video-based face recognition: A survey”, *World Academy Sci. Eng. Tech.*, vol. 3, no. 12, pp. 2809–2818, 2009.
- [33] Q. Shu, X. He, C. Wang, Y. Yang, and Z. Cui, “Fast point cloud registration in multidirectional affine transformation”, *Optik*, vol. 229, art. 165884, 2021. DOI: 10.1016/j.ijleo.2020.165884.
- [34] S.-C. Chu, Z. Zhuang, J. Li, and J.-S. Pan, “A novel binary QUasi-affine transformation evolutionary (QUATRE) algorithm”, *Appl. Sci.*, vol. 11, no. 5, p. 2251, 2021. DOI: 10.3390/app11052251.
- [35] A. S. Bist, W. Febriani, C. Lukita, S. Kosasi, and U. Rahardja, “Design of face recognition attendX for recording student attendance data based on artificial intelligence technology”, *Solid State Tech.*, vol. 63, no. 2s, pp. 4505–4518, 2020.
- [36] J. Yu, J. Li, Z. Yu, and Q. Huang, “Multimodal transformer with multi-view visual representation for image captioning”, *IEEE Trans. Circuits Sys. Video Tech.*, vol. 30, no. 12, pp. 4467–4480, 2020. DOI: 10.1109/TCSVT.2019.2947482.
- [37] S. He and L. Schomaker, “GR-RNN: Global-context residual recurrent neural networks for writer identification”, *Pattern Recog.*, vol. 117, art. 107975, 2021. DOI: 10.1016/j.patcog.2021.107975.
- [38] S. Zheng, Y. Liu, and P. H. Siegel, “PR-NN: RNN-based detection for coded partial-response channels”, *IEEE J. Sel. Are. Comm.*, vol. 39, no. 7, pp. 1967–1982, 2021. DOI: 10.1109/JSAC.2020.3041377.
- [39] C. Singh, S. Wibowo, and S. Grandhi, “A deep learning approach for human face sentiment classification”, in *Proc. of 21st ACIS Inter. Winter Conf. Soft. Eng. Arti. Intel. Net. Paral. Dist. Comput.*, 2021, pp. 28–32. DOI: 10.1109/SNPDWinter52325.2021.00015.
- [40] W. Lv, J. Xiong, J. Shi, Y. Huang, and S. Qin, “A deep convolution generative adversarial networks based fuzzing framework for industry control protocols”, *J. Intel. Manu.*, vol. 32, pp. 441–457, 2021. DOI: 10.1007/s10845-020-01584-z.
- [41] Z. Zhou, L. Liu, S. Jere, J. Zhang, and Y. Yi, “RCNet: Incorporating structural information into deep RNN for online MIMO-OFDM symbol detection with limited training”, *IEEE Trans. Wirel. Comm.*, vol. 20, no. 6, pp. 3524–3537, 2021. DOI: 10.1109/TWC.2021.3051317.
- [42] W. Simayi, M. Ibrayim, and A. Hamdulla, “Study the preprocessing effect on RNN based online Uyghur handwritten word recognition”, *Wire. Net.*, to be published. DOI: 10.1007/s11276-021-02651-w.
- [43] J. Zhang and W. Huang, “Option hedging using LSTM-RNN: An empirical analysis”, *Quantitative Finance*, to be published. DOI: 10.1080/14697688.2021.1905171.
- [44] S. Sharma, R. K. Agrawal, and M. M. Tripathi, “Synergism of recurrent neural network and fuzzy logic for short term energy load forecasting”, in *Proc. of Fourth Inter. Conf. Comput. Method. Commu.*, 2020, pp. 165–169. DOI: 10.1109/ICCMC48092.2020.ICCMC-00033.
- [45] A. Laucka, V. Adaskeviciute, D. Andriukaitis, “Research of the Equipment Self-Calibration Methods for Different Shape Fertilizers Particles Distribution by Size Using Image Processing Measurement Method,” *Symmetry*, vol. 11, no. 7, p. 838, Jun. 2019. DOI: 10.3390/sym11070838.
- [46] M. G. Huddar, S. S. Sannakki, and V. S. Rajpurohit, “Attention-based multi-modal sentiment analysis and emotion detection in conversation using RNN”, *Int. J. Interact. Multimedia Artif. Intel.*, vol. 6, no. 6, pp. 112–121, 2021. DOI: 10.9781/ijimai.2020.07.004.
- [47] A. S. Bosman, A. Engelbrecht, and M. Helbig, “Visualising basins of attraction for the cross-entropy and the squared error neural network loss functions”, *Neurocomput.*, vol. 400, pp. 113–136, 2020. DOI: 10.1016/j.neucom.2020.02.113.
- [48] M.-P. Jarabo-Amores, D. de la Mata-Moya, R. Gil-Pita, and M. Rosa-Zurera, “Radar detection with the Neyman–Pearson criterion using supervised-learning-machines trained with the cross-entropy error”, *EURASIP J. Adv. Signal Process.*, vol. 2013, art. no. 44, 2013. DOI: 10.1186/1687-6180-2013-44.
- [49] D. M. Kline and V. L. Berardi, “Revisiting squared-error and cross-entropy functions for training neural network classifiers”, *Neural Comput. Appl.*, vol. 14, pp. 310–318, 2005. DOI: 10.1007/s00521-005-0467-y.
- [50] F. Khemakhem and H. Ltfi, “Facial expression recognition using convolution neural network enhancing with pre-processing stages”, in *Proc. of IEEE/ACS 16th Int. Conf. Comput. Sys. Appl.*, 2019, pp. 1–7. DOI: 10.1109/AICCSA47632.2019.9035249.
- [51] Y. Gan, J. Chen, Z. Yang, and L. Xu, “Multiple attention network for facial expression recognition”, *IEEE Access*, vol. 8, pp. 7383–7393, 2020. DOI: 10.1109/ACCESS.2020.2963913.
- [52] X. Liu and F. Zhou, “Improved curriculum learning using SSM for facial expression recognition”, *The Visual Computer*, vol. 36, pp. 1635–1649, 2020. DOI: 10.1007/s00371-019-01759-7.
- [53] H. Zhang, B. Huang, and G. Tian, “Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture”, *Pattern Recognition Letters*, vol. 131, pp. 128–134, 2020. DOI: 10.1016/j.patrec.2019.12.013.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (<http://creativecommons.org/licenses/by/4.0/>).