International Journal of Innovative
Computing, Information and Control
Volume **19**, Number **4**, August **2023**

ICIC International ©2023 ISSN 1349-4198
pp. **1239–1250**

FACIAL EXPRESSION RECOGNITION ALGORITHM BASED ON MULTI-ATTENTION MECHANISM

Huixin Wu, Zehuan Huang, Wei Jiang* and Xin Zhao

School of Information Engineering
North China University of Water Resources and Electric Power
No. 136, Jinshui East Road, Zhengzhou 450046, P. R. China
wuhuixin@ncwu.edu.cn; { huangzehuan; zhaoxin }@stu.ncwu.edu.cn
*Corresponding author: jiangwei@ncwu.edu.cn

Received October 2022; revised February 2023

Abstract. *This paper suggests a facial expression recognition network based on a multi-attention mechanism called SCSNet to address factors like pose change and occlusion in facial expression recognition while also taking account of the issue of inadequate facial information extraction using a single attention mechanism (Spatial-attention, Channel-attention, and Self-attention Network). The network is put together by the multi-attention mechanism using channel, spatial, and Self-attention. In comparison to using a single attention mechanism, the network concentrates on more useful expression traits when channel attention and spatial attention are combined. The Self-attention technique can reduce ambiguous annotation-related elements including illumination, occlusion, human position changes, and others, which enhances the model's capacity for discriminative expression. This study used the RAF-DB public dataset for the experimental evaluation, which had an accuracy of 86.89%. The outcomes of the experiments on different datasets further demonstrate the efficacy of this approach.*

Keywords: Facial expression recognition, Multi-attention mechanism, Channel-attention mechanism, Spatial-attention mechanism, Self-attention

1. **Introduction.** In the realm of computer vision, facial expression recognition is an area of active research. Facial expressions can transmit emotion from the standpoint of emotional comprehension because high-resolution facial expression photos contain a lot of information [1]. As a result, it is frequently utilized in a variety of industries, including transportation [2], service industry [3], lie detection [4], auxiliary medical diagnosis [5], Human-Computer Interface (HCI) [6], etc.

The development of deep learning has significantly improved facial expression identification. Facial expression recognition uses a variety of network frameworks, including CNN (Convolutional Neural Network) [7], VGGNet (Visual Geometry Group Network) [8], ResNet (Residual Network) [9], and others. In testing AffectNet [11], Li et al. [10] used an enhanced VGG-16 facial expression recognition model based on CNN and achieved notable accuracy. Nevertheless, the feature learning capacity of the network model is influenced by a number of non-facial expression-related parameters (such as illumination, posture, and occlusion), making the expression identification process susceptible to overfitting and limiting the model's ability to generalize. Li et al. [12] created a facial expression identification convolutional neural network with an attention mechanism in 2020 to improve the neural network's capacity for feature learning. To enhance network performance, this technique combines attention mechanisms, convolutional features, and LBP (Local Binary Patterns) features.

DOI: 10.24507/ijicic.19.04.1239

The aforementioned techniques have significantly improved facial expression recognition in constrained settings. Facial expression recognition accuracy in natural scenarios still has to be improved, despite facial expression recognition's growing use in the wild. This research suggests a paradigm for recognizing facial expressions that incorporates many attentional mechanisms based on this. Based on ResNet, this paradigm suggests a novel schema that integrates Channel, Spatial, and Self-attention methods. The article includes seven main kinds of expression labeling: anger, disgust, fear, happy, sad, surprise, and normal.

In summary, our contributions are the following.

- We propose a facial expression recognition network with multi-attention mechanism called SCSNet (Spatial-attention, Channel-attention, and Self-attention Network), which achieves high performance compared with single attention mechanism.
- We reveal a discovery that the feature extraction network can better extract facial expression features by combining channel and spatial attention mechanism.

The rest of the paper is organized as follows. Section 2 discusses the principle and implementation of this algorithm, Section 3 describes the results and analysis of the experiment, and Section 4 concludes the paper.

2. **Methods.**

A facial expression recognition network called SCSNet is suggested in this paper. The SCSNet model is first introduced in this section, followed by an explanation of its feature extraction network and three crucial parts.

2.1. **SCSNet model.**

The SCSNet model put forth in this paper has four essential parts: feature extraction network, Channel-attention mechanism [13], Spatial-attention mechanism [14], and Self-attention mechanism [15].

The ResNet-34 feature extraction network along with the Channel-attention and Spatial-attention mechanisms is used to extract facial expression features for a collection of facial expression images with unclear factors (illumination, occlusion, etc.). The weighing, regularization, and relabeling processes are all a part of the Self-attention mechanism. Each image is given an essential weight during the weighting process utilizing the Full Connection (FC) layer and sigmoid function. Attention weights are regularized during the regularization process to lessen the significance of questionable samples. The learnt attention weights are divided into groups of high and low priority during the regularization process. To further enhance the network, a relabeling method is used, which alters certain ambiguous samples of low relevance groups. Finding more dependable labels with the intention of enhancing the model is the goal of relabeling. The SCSNet model procedure is shown in Figure 1.

2.2. **Feature extraction network.**

The residual network ResNet-34, which integrates Channel-attention and Spatial-attention mechanisms, is chosen as the fundamental structure in this article taking account of the performance and computational cost of the feature extraction. The ResNet-34 network, which is organized into six modules in Figure 2, consists of 33 convolution layers and 3 complete connection layers. Initially, the fifth module adds the Channel-attention and Spatial-attention mechanisms to suppress background noise and uninteresting feature information from the channel and space, respectively. The network focuses primarily on the successful extraction of facial expression feature information by minimizing the influence of other elements at the low level. Second, to improve the connection between the pertinent channel properties, the average pooling layer is utilized in the final module.

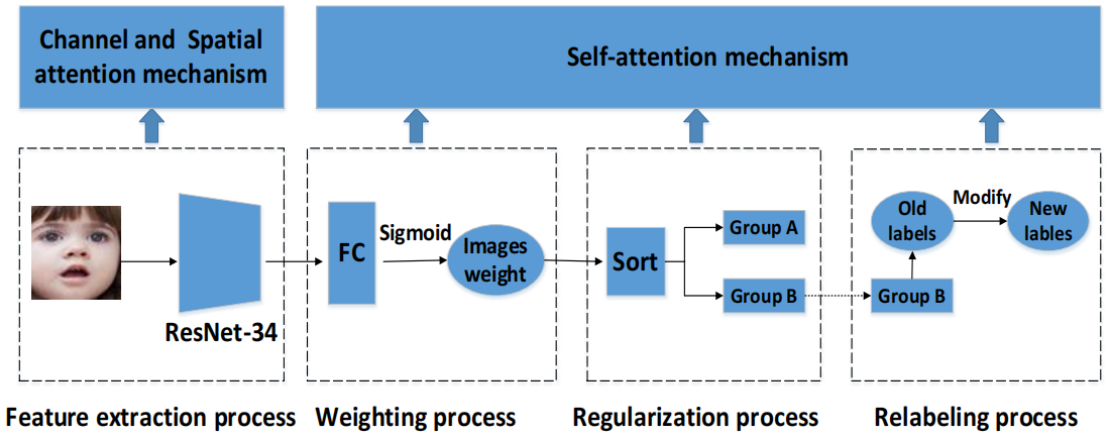


FIGURE 1. SCSNet structure

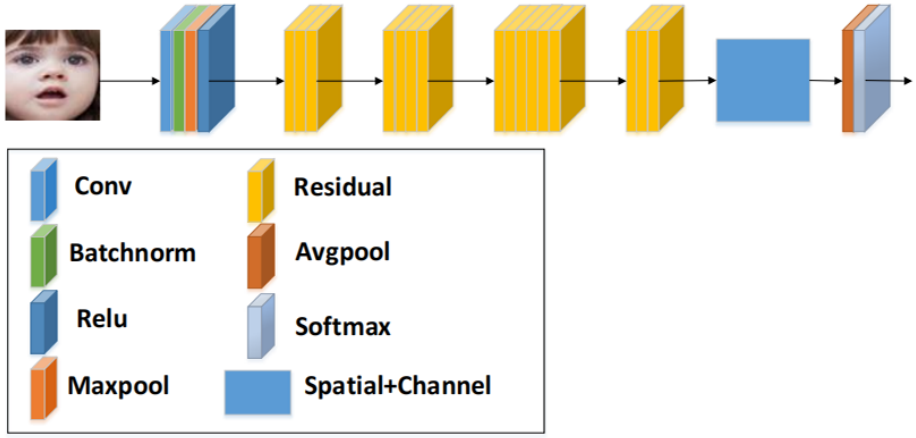


FIGURE 2. Spatial and Channel-attention based ResNet-34 for feature extraction

2.3. **Channel-attention mechanism.** The Channel-attention mechanism's job is to give each channel a varied weight so that the network can concentrate on key features and inhibit uninteresting ones. In other words, the neural network decides which channel is crucial and gives it the right weight automatically. Two $1 \times 1 \times C$ channel descriptions are created by global maximum pooling and average pooling of one channel dimension, respectively, from the input feature map of the Channel-attention mechanism, which is $H \times W \times C$ (H and W indicate height and weight of feature maps, respectively). Next, as illustrated in Figure 3, a weight coefficient is created via feature superposition. The input features and weight coefficient can be multiplied to produce the new features.

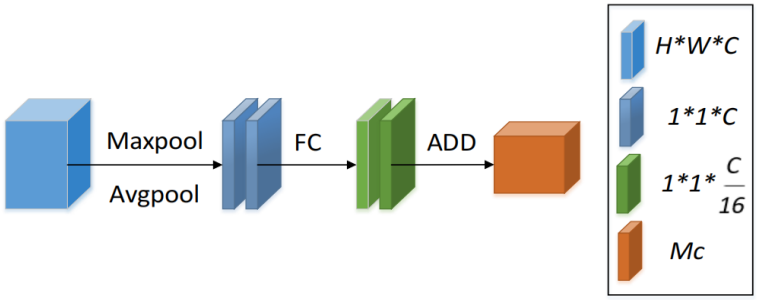


FIGURE 3. Channel-attention

The operation process is as follows. Firstly, the maximum global pooling and average pooling of the input feature map are carried out according to the channel. The two one-dimensional vectors after pooling are added to the full connection layer, and the one-dimensional Channel-attention weight coefficient $M_c \in R^{1\times1\times C}$ is generated. Then, M_c multiplied by the input element to obtain the feature map after channel attention adjustment.

The process can be described as

$$F' = M_c(F) \otimes F \tag{1}$$

where $\otimes$ represents the multiplication of elements, F is the input feature map, and F' is the feature map after the multiplication.

2.4. **Spatial-attention mechanism.** Not all regions in the image are equally important for the task, and only the task-related regions need to be concerned. Therefore, spatial attention is utilized to find the most important part of the network processing in the classification task.

The size of the input feature map for spatial attention is $H \times W \times C$, just like it is for Channel-attention. First, two feature maps with the size of $H \times W \times 1$ are created by performing the maximum and average pooling of a spatial dimension, respectively. Second, a 7×7 convolution layer is used to combine these two feature maps along the channel dimension into $H \times W \times 1$. As seen in Figure 4, the Sigmoid function is then used to get the spatial weight coefficient. Finally, it can be multiplied by the input feature map to obtain the feature map after Spatial-attention adjustment.

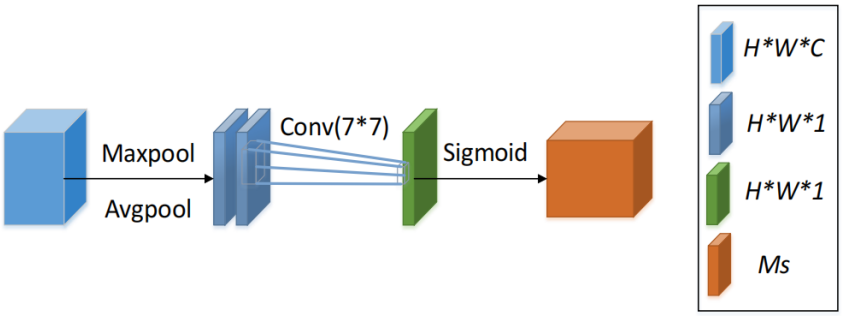


FIGURE 4. Spatial-attention

The process can be described as

$$F'' = M_s(F') \otimes F' \tag{2}$$

where $\otimes$ represents the multiplication of elements, F' is the input feature map, and F'' is the feature map after operation.

The channel attention and spatial attention are combined in the feature extraction network shown in the following Figure 5.

2.5. **Self-attention mechanism.** The SCSNet model also includes the Self-attention mechanism [15]. To determine the relevance of each image, weight must first be learned. Uncertainty in some images caused by lighting, occlusion, and other factors lowers the network's performance. In order to solve the issue, samples with uncertainty are given a low weight, whereas ones with certainty are given a high weight. The samples are then separated into two groups and arranged in descending order according to importance weights (high and low weights). Finally, some uncertain samples of low importance weight groups are modified to enhance the final model, as shown in Figure 6.

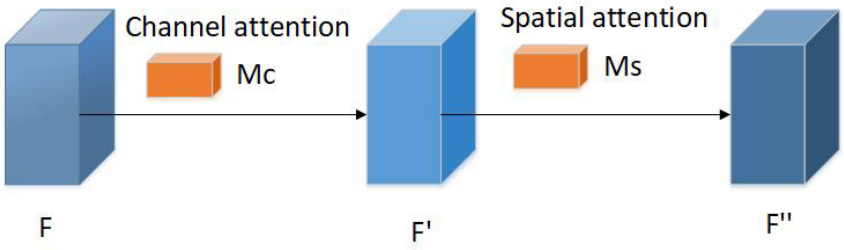


FIGURE 5. Attention mechanism combination

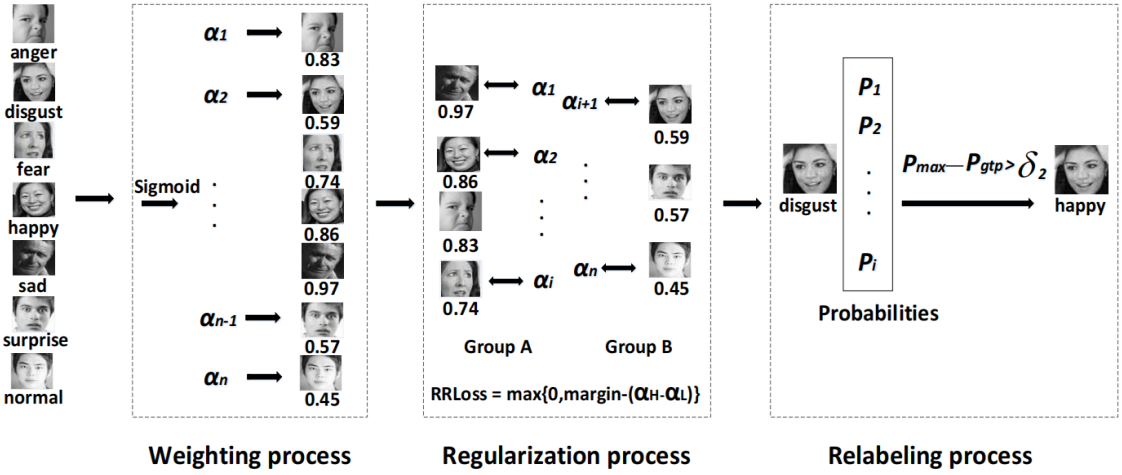


FIGURE 6. Self-attention mechanism

2.5.1. *Importance weighting process.* This paper introduces the Self-attention mechanism. To suppress the uncertainty, the expectation is that the certain samples have higher importance weights, and the uncertain samples have lower importance weights. $F = [X_1, X_2, \ldots, X_N] \in R^{(D\times N)}$ is denoted as the facial features of N images. The input is F in the importance weighting process, as the output is the importance weighting of each feature.

This process is described as the following:

$$\alpha_i = \sigma\left(W_a^T X_i\right) \tag{3}$$

It is made up of a Fully Connected (FC) layer and a Sigmoid activation function, where α_i is the important weight of the i sample. W_a is the coefficient for attention, and σ is the Sigmoid function.

With attention weight, the Weighted Cross-Entropy loss (WCE-loss) is chosen to compute the multi-class cross-entropy loss.

The formula is

$$WCE = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\alpha_i W_{y_i}^T X_i}}{\sum_{j=1}^{C} e^{\alpha_i W_j^T X_i}} \tag{4}$$

where W_j is the j classifier, and WCE is positively correlated with α.

2.5.2. *Regularization process.* The range of the aforementioned Self-attention weights is $(0, 1)$. This phase regularizes the attentional weights to limit the significance of unclear samples. The learnt attention weights are separated into two groups using a ratio of after being sorted in descending order. The procedure of regularization makes sure that the

average attentional weight of the high-importance group is, by a margin value, greater than the average attentional weight of the low-importance group.

The Rank Regularization Loss (RR-Loss) is defined as

$$L_{RR} = \max\{0, \delta_1 - (\alpha_H - \alpha_L)\} \tag{5}$$

with

$$\alpha_H = \frac{1}{M}\sum_{i=0}^{M}\alpha_i, \quad \alpha_L = \frac{1}{N-M}\sum_{i=M}^{N}\alpha_i \tag{6}$$

where δ_1 is a margin value, which can be an empirical value, or a learning parameter, the α_H and α_L are respectively the mean values of the high importance group with $\beta \times N = m$ samples and the low importance group with an $N - M$ sample (a mini-batch sample number is N).

2.5.3. *Relabeling process.* After the regularization process, each batch of data is divided into high-importance and low-importance groups. Since the importance of uncertain samples is usually low, the strategy is designed to relabel the samples. In the relabeling process, only the samples in the low-importance group are considered, and softmax function is performed. For each sample in the low-importance group, the maximum prediction probability is compared with the probability of a given label through the softmax function. Suppose the maximum prediction probability is higher than the probability of the originally given label with a threshold. In that case, the sample is assigned a new pseudo label and corrected to a certain sample. Eventually, the final model is enhanced through the relabeling.

This process is defined as

$$y' = \begin{cases} l_{\max} & \text{if } P_{\max} - P_{gtp} > \delta_2 \\ l_{prm} & \text{others} \end{cases} \tag{7}$$

where y' denotes the new label, δ_2 is the threshold, $P_{\max}$ is the maximum prediction probability, and P_{gtp} is the prediction probability of the given label. l_{prm} and $l_{\max}$ represent the indexes corresponding to the original given label and the maximum prediction probability, respectively.

3. Experiment and Result Analysis.

3.1. **Datasets.** Several experiments are carried out on RAF-DB [16], FER2013 [17], and CK+ [18] datasets. The RAF-DB dataset includes about 30000 facial images, the FER2013 dataset includes about 30000 facial images, and the CK+ dataset contains 593 image sequences of 123 individuals. Sample images of three datasets are shown in Figure 7.

3.2. **Experimental parameters.**

Facial feature preprocessing. SCSNet is implemented by Pytorch 1.7.1 as the basic framework, and facial features are extracted by the ResNet-34 residual network, which fuses channel attention and spatial attention.

Training. The system used in this experiment is Linux, and the GPU is NVIDIA GeForce GTX 1080 Ti. The cross-entropy loss function is used in training. The optimizer is AdaBound [19], and the learning rate is set to be exponential decay (Formula (8)), with the initial set of 0.01. The parameter between the average values of high and low importance groups can be set as 0.15 by default or a learning parameter, and the parameter of relabeling can be set as 0.2 by default.

$$Learning\_rate = Decline\_rate \times \frac{Global\_step}{Decay\_steps} \tag{8}$$

Namely learning rate = initial learning-rate decline rate (current step/number of rounds updated once).

(a) FER2013 image example

(b) RAF-DB image example

(c) CK+ image example

FIGURE 7. Examples of images on datasets

3.3. **Results and analysis.**

3.3.1. *Algorithm evaluation.* The algorithm is assessed by making use of the confusion matrix, which is frequently employed in the current facial expression recognition techniques, in order to ensure the validity of the experimental results.

The numerical value of each parameter can be seen intuitively from the confusion matrix, including the accuracy and error rate indicators.

Accuracy: the proportion or number of samples correctly classified. The formula is shown in (9):

$$Acc = \frac{TP + TN}{Total} \tag{9}$$

True Positive (TP): the real class, that is, the real class of the sample is the positive class, and the result of model identification is also a positive class.

True Negative (TN): the true category of the sample is negative, and the model predicts it as negative.

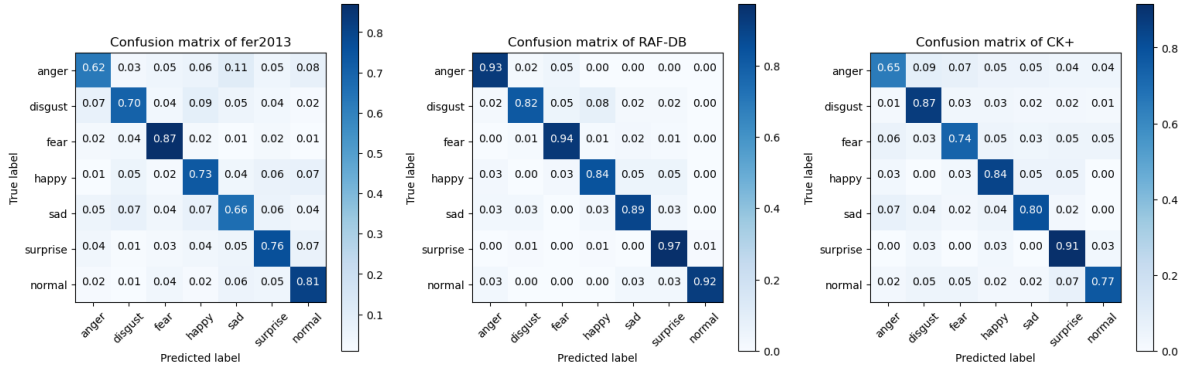
Error-Rate (ER): the proportion or number of samples correctly classified. The formula is shown in (10):

$$ER = \frac{FP + FN}{Total} \tag{10}$$

False Positive (FP): the true category of the sample is negative, but the model predicts it as a positive case.

False Negative (FN): the true category of the sample is positive, but the model predicts it as negative.

Figure 8 shows the confusion matrix of SCSNet on FER2013, RAF-DB, and CK+ datasets.



(a) FER2013 matrix (b) RAF-DB matrix (c) CK+ matrix

FIGURE 8. Confusion matrix of SCSNet model on datasets

3.3.2. *Algorithm comparison and analysis.*

1) CK+ dataset comparison

For the CK+ dataset, the suggested model is contrasted with other well used techniques. The findings are displayed in Table 1. Our method's accuracy, as shown in Table 1, is 72.29%, which is greater than Ensemble DCNNs, SCNN, HOG-TOP, and SVM, respectively, by 5.29%, 11.29%, 7.29%, and 17.82%. It is clear that this method's accuracy is higher than that of certain popular algorithms already in use.

TABLE 1. Accuracy comparison on CK+ dataset

Algorithm	Precision
DCNNs [20]	67%
SCNN [21]	61%
HOG-TOP [22]	65%
SVM [23]	54.47%
Ours	72.29%

2) FER2013 dataset comparison

Using the FER2013 dataset, the suggested model is contrasted with a few popular approaches currently in use. Table 2 presents the outcomes. Table 2 shows that the accuracy of our technique is 69.03% which is 1.63%, 3.03%, 3.83%, and 2.72% greater than the accuracy of Bag of Words, FER on SoC, GoogleNet, and VGG+SVM, respectively. It is clear that the suggested method outperforms some popular current methods.

TABLE 2. Accuracy comparison on FER2013 dataset

Algorithm	Precision
Bag of Words [24]	67.4%
FER on SoC [25]	66%
GoogleNet [26]	65.2%
VGG+SVM [27]	66.31%
Ours	69.03%

3) RAF-DB dataset comparison

For the RAF-DB dataset, the suggested model is contrasted with a few popular approaches currently in use. Table 3 presents the outcomes. Table 3 shows that the accuracy

Table 3. Accuracy comparison on RAF-DB dataset

Algorithm	Precision
LDL-ALSG [28]	85.53%
APM [29]	85.17%
DLP-CNN [30]	84.13%
gACNN [31]	85.07%
Ours	86.89%

of the approach used in this study is 86.89%, which is successively 1.36%, 1.72%, 2.76%, and 1.82% greater than that of LDL-ALSG, APM, DLP-CNN, and gACNN. The fact that the suggested method outperforms some widely used existing algorithms demonstrates the development of the suggested model.

3.3.3. *Ablation experiment.* In order to further verify the effectiveness of the multi-attention mechanism, three key components of this paper are compared:

- The Channel-attention mechanism module
- The Spatial-attention mechanism module
- The Self-attention mechanism module

On the datasets CK+, FER2013, and RAF-DB, ablation experiments are run. Each dataset is run five times to reach a higher level of accuracy, and Table 4 shows the average value of the outcomes. The table indicates that, in comparison to single Channel-attention, single Spatial-attention, and single Self-attention mechanisms, the multi-attention mechanism can increase accuracy. The accuracy of the multi-attention method is, respectively, 3.75%, 2.71%, and 3.12% greater than others on the CK+ dataset. It is 4.26%, 4.5%, and 2.64% higher on the FER2013 dataset, respectively. It is 0.91%, 0.71%, and 1.23% higher on the RAF-DB dataset, respectively. Also, the accuracy of the CK+, FER2013, and RAF-DB datasets increased by 1.68%, 1.19%, and 0.52%, respectively, when compared with the fusion channel attention and spatial attention. The outcomes of the experiments demonstrate that a multi-attention mechanism can significantly enhance network performance.

Table 4. Ablation experiment

Module	Precision		
	CK+	FER2013	RAF-DB
Channel-attention	68.54%	64.77%	85.98%
Spatial-attention	69.58%	64.53%	86.18%
Self-attention	69.17%	66.39%	85.66%
Channel + Spatial	70.61%	67.84%	86.37%
Multi-attention	72.29%	69.03%	86.89%

4. **Conclusion.** The paper suggests a method for recognizing facial expressions that integrates the Channel-attention, Spatial-attention, and Self-attention mechanisms. Firstly, to more correctly extract face expression features, the ResNet-34 feature extraction network combining Channel-attention and Spatial-attention mechanisms is utilized. Second, the Self-attention process is employed to improve the final model by altering the uncertain samples. Finally, the RAF-DB, CK+, and FER2013 datasets are used to verify and assess the model. The experimental findings demonstrate that the model outperforms the

various face expression recognition techniques currently in use. In conclusion, the face expression recognition model is significant and performs well.

However, there are still some shortcomings in this study that need to be further improved. In the future work, the following two aspects will be studied.

1) In the research process, this paper uses several large-scale facial expression datasets, and this method requires large-scale data samples to support. However, human beings only need one or several samples to establish the cognition of objective things, so in the case of using a small number of data samples, the model can analyze the nature of things.

2) This paper studies static expression recognition, and the application scenarios have certain limitations. The next step is to expand the scene of expression recognition to the field of dynamic expression recognition, and apply facial expression recognition to real-time application scenarios such as safe driving and medical monitoring.

REFERENCES

[1] P. Ekman, Facial expression and emotion, *American Psychologist*, vol.48, no.4, 384, 1993.

[2] W. Li, Y. Cui, Y. Ma et al., A spontaneous driver emotion facial expression (DEFE) dataset for intelligent vehicles: Emotions triggered by video-audio clips in driving scenarios, *IEEE Transactions on Affective Computing*, 2021.

[3] R. Li, F. Yang and X. Zhu, The Janus face of grandiose narcissism in the service industry: Self-enhancement and self-protection, *Journal of Business Ethics*, pp.1-19, 2022.

[4] N. Mehendale, Facial emotion recognition using convolutional neural networks (FERC), *SN Applied Sciences*, vol.2, no.3, pp.1-8, 2020.

[5] B. Zhang et al., Many faces of the hidden souls: Medical and neurological complications and comorbidities in disorders of consciousness, *Brain Sciences*, vol.11, no.5, 608, 2021.

[6] Y. Shi, Z. Zhang, K. Huang et al., Human-computer interaction based on face feature localization, *Journal of Visual Communication and Image Representation*, vol.70, 102740, 2020.

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Communications of the ACM*, vol.60, no.6, pp.84-90, 2017.

[8] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv Preprint*, arXiv: 1409.1556, 2014.

[9] K. He, X. Zhang, S. Ren et al., Identity mappings in deep residual networks, *European Conference on Computer Vision*, pp.630-645, 2016.

[10] S. Li, W. Zheng, Y. Zong et al., Bi-modality fusion for emotion recognition in the wild, *2019 International Conference on Multimodal Interaction*, pp.589-594, 2019.

[11] A. Mollahosseini, B. Hasani and M. H. Mahoor, AffectNet: A database for facial expression, valence, and arousal computing in the wild, *IEEE Transactions on Affective Computing*, vol.10, no.1, pp.18-31, 2017.

[12] J. Li, K. Jin, D. Zhou, N. Kubota and Z. Ju, Attention mechanism-based CNN for facial expression recognition, *Neurocomputing*, vol.411, pp.340-350, 2020.

[13] J. Hu, S. Li and G. Sun, Squeeze-and-excitation networks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132-7141, 2018.

[14] S. Woo et al., CBAM: Convolutional block attention module, *Proc. of the European Conference on Computer Vision (ECCV)*, 2018.

[15] K. Wang et al., Suppressing uncertainties for large-scale facial expression recognition, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[16] Z. Lian, Y. Li, J. H. Tao et al., Expression analysis based on face regions in real-world conditions, *International Journal of Automation and Computing*, vol.17, pp.96-107, 2020.

[17] E. Barsoum, C. Zhang, C. C. Ferrer and Z. Zhang, Training deep networks for facial expression recognition with crowd-sourced label distribution, *Proc. of the 18th ACM International Conference on Multimodal Interaction*, pp.279-283, 2016.

[18] P. Lucey et al., The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp.94-101, 2010.

[19] L. Luo, Y. Xiong, Y. Liu et al., Adaptive gradient methods with dynamic bound of learning rate, *arXiv Preprint*, arXiv: 1902.09843, 2019.

[20] G. Pons and D. Masip, Supervised committee of convolutional neural networks in automated facial expression analysis, *IEEE Transactions on Affective Computing*, vol.9, no.3, pp.343-350, 2018.
[21] X. Pan, J. Shi, P. Luo et al., Spatial as deep: Spatial CNN for traffic scene understanding, *Proc. of the AAAI Conference on Artificial Intelligence*, vol.32, no.1, 2018.
[22] J. Chen, Z. Chen, Z. Chi et al., Facial expression recognition in video with multiple feature fusion, *IEEE Transactions on Affective Computing*, vol.9, no.1, pp.38-50, 2018.
[23] M. Sert and N. Aksoy, Recognizing facial expressions of emotion using action unit specific decision thresholds, *Proc. of the 2nd Workshop on Advancements in Social Signal Processing for Multimodal Interaction*, pp.16-21, 2016.
[24] I. R. Tudor, M. Popescu and C. Grozea, Local learning to improve bag of visual words model for facial expression recognition, *ICML*, 2013.
[25] P. T. Vinh and T. Q. Vinh, Facial expression recognition system on SoC FPGA, *2019 International Symposium on Electrical and Electronics Engineering (ISEE)*, pp.1-4, 2019.
[26] P. Giannopoulos, I. Perikos and I. Hatzilygeroudis, Deep learning approaches for facial emotion recognition: A case study on FER-2013, *Advances in Hybridization of Intelligent Methods*, pp.1-16, 2018.
[27] M. I. Georgescu, R. T. Ionescu and M. Popescu, Local learning with deep and handcrafted features for facial expression recognition, *IEEE Access*, vol.7, pp.64827-64836, 2019.
[28] N. Perveen, D. Roy and K. M. Chalavadi, Facial expression recognition in videos using dynamic kernels, *IEEE Transactions on Image Processing*, vol.29, pp.8316-8325, 2020.
[29] Z. Li, S. Han, A. S. Khan et al., Pooling map adaptation in convolutional neural network for facial expression recognition, *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp.1108-1113, 2019.
[30] S. Li, W. Deng and J. Du, Reliable crowdsourcing and deep locality preserving learning for unconstrained facial expression recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2852-2861, 2017.
[31] Y. Li, J. Zeng, S. Shan et al., Occlusion aware facial expression recognition using CNN with attention mechanism, *IEEE Transactions on Image Processing*, vol.28, no.5, pp.2439-2450, 2018.

Author Biography

Huixin Wu received Ph.D. degree in Systems Engineering from Northwestern Polytecnical University, Xi'an, China. He is currently Professor and Dean in School of Information Engneering, North China University of Water Resources and Electric Power, Zhengzhou, China. He is doing graphics and image processing, 3D modeling and simulation, virtual reality, artificial intelligence and big data analysis research.

Zehuan Huang is currently a graduate student in School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, China. Prior to this, he received a Bachelor of Engineering degree in School of Information Engineering of Shandong Youth University of Political Science in 2019. His research interests include graphics and pattern recognition.

Wei Jiang received the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China. He is currently as an Associate Professor with the School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, China. He is doing computer vision and remote sensing research.

Xin Zhao is currently a graduate student in School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou, China. Prior to this, he received his Bachelor of Engineering degree from School of Software, Zhongyuan University of Technology in 2019. His research directions include image processing and deep learning technology, focusing on computer vision.