Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1707–1717

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

LIGHTWEIGHT EXPRESSION RECOGNITION NETWORK IN PSYCHOLOGICAL WARFARE APPLICATIONS

TIANCHENG DONG*, SHUPEI XIE†, YONG DU, AND QIMING SHI

Abstract. Psychological warfare has become an important aspect and form of modern warfare, with the essence of psychological confrontation being the ability to discern changes in the psychological state of enemies rapidly. Rapid and efficient recognition of expressions, a crucial component in a modern psychological confrontation, can support combat commanders' strategic decisions. The current classic facial expression recognition network is unsuitable for psychological confrontation due to its multiple levels and large number of parameters. Therefore, a lightweight convolutional neural network expression recognition algorithm is proposed by introducing depthwise separable residual and interleaved group convolutional structures. This algorithm effectively improves recognition accuracy while reducing the network's required processing data. For the Jaffe dataset, the ShuffleNet V2, MobileNet V2, and the proposed network accuracies are 86.35, 90.90, and 95.47%, reaching 95.00, 96.67, and 98.30% for the CK+ dataset. This proves the proposed network's superiority and feasibility in psychological warfare applications.

1. Introduction

As a new form of combat, psychological warfare is becoming another distinct field of confrontation following physical, firepower, and information warfare. The application of facial expression recognition technology in military operations holds multiple significance. Firstly, by accurately identifying the expressions of enemy commanders or key individuals, one can gain insight into their psychological state and emotional changes, thereby analyzing potential tactical intentions and directions of action, providing a strong basis for strategic decisions [4]. Secondly, in battlefield surveillance and intelligence analysis, using lightweight convolutional networks for facial expression classification can automatically filter out valuable information, enhancing the efficiency and accuracy of information processing. Additionally, in human-machine interaction and intelligent equipment control, facial expression recognition technology also contributes to elevating the level of intelligence in combat systems. The facial expression recognition network in psychological confrontation should have the characteristic of significantly reducing computational complexity and storage requirements while ensuring recognition accuracy. Therefore, building a lightweight convolutional network that can achieve low data processing and high accuracy in a big data environment has broad application value in psychological confrontation.

2020 *Mathematics Subject Classification.* 68T07, 92B20.

Key words and phrases. Machine vision, convolutional neural network, expression recognition, psychological warfare.

*The first author is Logistics university of people's armed police force, Tianjin 300300.

†Corresponding author.

To achieve the application of facial expression recognition networks in psychological confrontation, this paper adopts an interleaved group convolution structure to optimize channel connections and achieve information flow between group convolutions. The residual structure of depthwise separable convolutions is introduced to optimize feature extraction further. The performance was evaluated using data analysis combined with visualization technology. Based on the comparative analysis of various aspects, the network model proposed in this article performs better in psychological confrontation.

2. Geometric background

The network's feature extraction layer is the key to efficiently recognizing expressions in the application environment of psychological warfare. The feature extraction layer is a key step in extracting effective facial features for facial expression recognition. The architecture of the network's feature extraction layer is divided into three stages. The first and second stages focus on the extraction of effective features. Among them, the first stage mainly enhances the information communication between networks and increases the information fault tolerance rate of the networks. The second stage reduces network parameters to optimize the feature extraction performance. The third stage mainly involves classifying and counting the detection results [14].

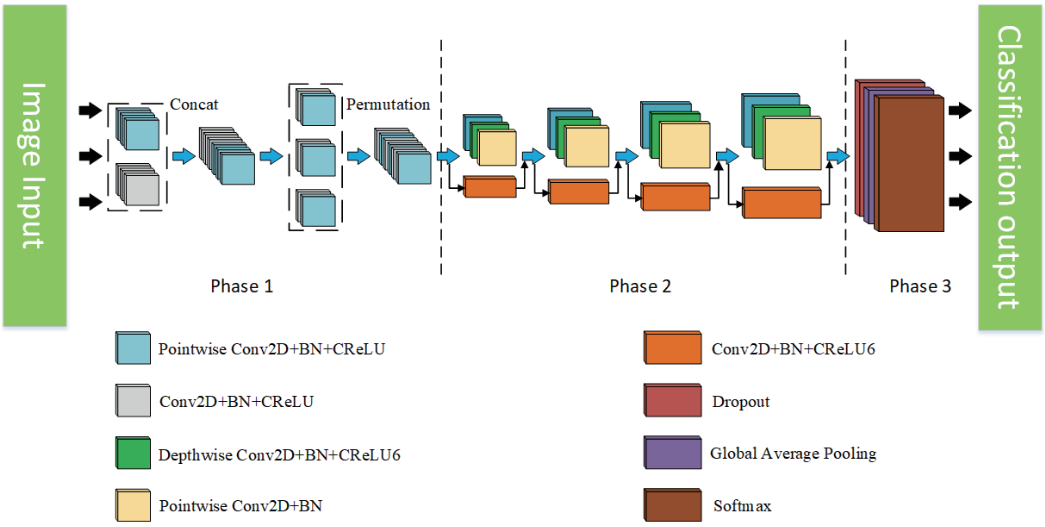

Figure 1. The network construction phases

The first stage is at the front of the network, with the overall framework of the feature extraction layer using an interleaved group convolution structure. Group convolution, as the basic structure of this part, divides the input and output layers into corresponding partitions with no relationship between them. The input layer only convolves with the corresponding partition of the output layer, aiming to reduce the number of channels between layers. The simple group convolution network has a problem of insufficient communication between network channels [7, 13].

An interleaved structure is added based on the group convolution to optimize the problems in the network. The interleaved group convolution consists of two group convolutions, performing two different partitions. The first group convolution divides into M partitions, each containing N channels. After the first group convolution, the results of the M partitions are concatenated to obtain M × N channels. The second group convolution is divided into N partitions, each containing M channels. These M channels contain information from different partitions in the first group convolution, effectively communicating channel information between different partitions.

The first group convolution performs spatial domain convolution (3 × 3 convolution kernel), dividing into M partitions, each containing N channels to obtain the spatial position of group convolution. The first group convolution is as follows:

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{\mathrm{M}} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{\mathbf{1}}^{\mathbf{P}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ & \mathbf{Y}_{22}^{\mathbf{P}} & \mathbf{0} & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{Y}_{\mathrm{MM}}^{\mathbf{P}} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_{\mathrm{M}} \end{bmatrix} \tag{2.1}$$

where$[z_1, z_2, \ldots, z_M]^T$serves as the input of the first group convolution, S represents the convolution kernel of the spatial convolution (3 × 3), z_M represents the dimension vector of (NS), a matrix of size N × NS, corresponding to the convolution kernel of the Mth partition.

The second group convolution permutes the number of groups and channels, dividing into N partitions, each containing M channels so that the channels of the second partition come from different partitions in the first group convolution. The N channels from the first partition constitute the nth partition of the second group convolution.

$$\bar{x}_n = [x_{1n}, x_{2n}, \ldots x_{M,n}]^{\boldsymbol{T}} = P_n^{\boldsymbol{T}} \boldsymbol{x}. \tag{2.2}$$

$\bar{x}_n$represents the nth partition of the second group convolution, y_{mn} is the mth element of y_n , P is the transpose matrix.

$$P = [P_1, P_2, \ldots P_N],$$
$$\bar{x} = [\bar{x}_1, \bar{x}_2, \ldots \bar{x}_M]^T,$$
$$x = \left[x_1^T, x_2^T, \ldots x_N^T\right]^T.$$

The second group convolution is executed on the N partitions:

$$\bar{z}_m = Y_m^d \bar{x}_n \tag{2.3}$$

where Y_{nn}^d is a matrix of size M × M, corresponding to the 1 × 1 convolution kernel on the nth partition of the second group convolution.

Since this stage tends to capture both positive and negative phase information simultaneously, the CReLU activation function is adopted as the frontend network's activation function to reduce the redundancy of convolutional kernels caused by the ReLU function, eliminating the negative phase response. The output dimension of CReLU will double automatically [5]. Based on this principle, CReLU reduces half of the output channels and concatenates the remaining channels with the negative output in a simple connection manner. Therefore, CReLU enables the frontend network to achieve a 2x speed increase without loss of precision. Changing the slope and activation threshold within the opposite channels increases the adaptability of the activation function.

The mathematical model of CReLU is:

$$\mathrm{CRe}\,LU(x) = [\mathrm{Re}\,LU(x), \mathrm{Re}\,LU(-x)]. \tag{2.4}$$

The optimized CReLU structure is shown in Figure 2:

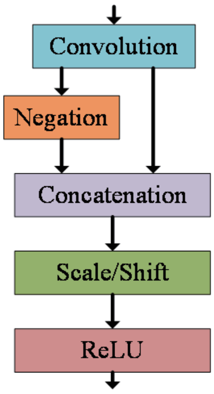

Figure 2. The optimized CReLU structure

The second stage is a network entirely composed of convolutions. This stage combines Depthwise Separable Convolutions with Residuals Networks, using pointwise convolutions first to increase and then decrease the number of feature channels, allowing the Depthwise Convolutional layer to operate in high-dimensional features. The output of the first stage's interleaved group convolution is the second stage's input. To reduce the loss of compressed input feature volume, a linear bottleneck structure is used, which can fully extract the compressed feature vectors from the first stage, thereby enhancing the precision of expression recognition. The network comprises four units. The first unit features 16 convolution kernels of size 11 Pointwise Conv, followed by a 33 Depthwise conv layer, culminating in an 11 Pointwise Conv layer. To further reduce network computational costs and enhance network lightweight, a residual network conv2d with 16 convolution kernels of size 11 is incorporated into the network. The output of the first unit is fed into the second unit, with the network comprising 32 convolution kernels, 64 in the third unit, and 128 in the fourth unit. Consequently, augmenting the number of convolution kernels in each unit ensures more thorough image feature extraction. The network structure is shown in Figure 3.

The third stage consists of a convolutional layer, a random connection layer, and a global average pooling layer. This stage reduces the number of features and performs regularization processing to prevent overfitting. The output vector of the

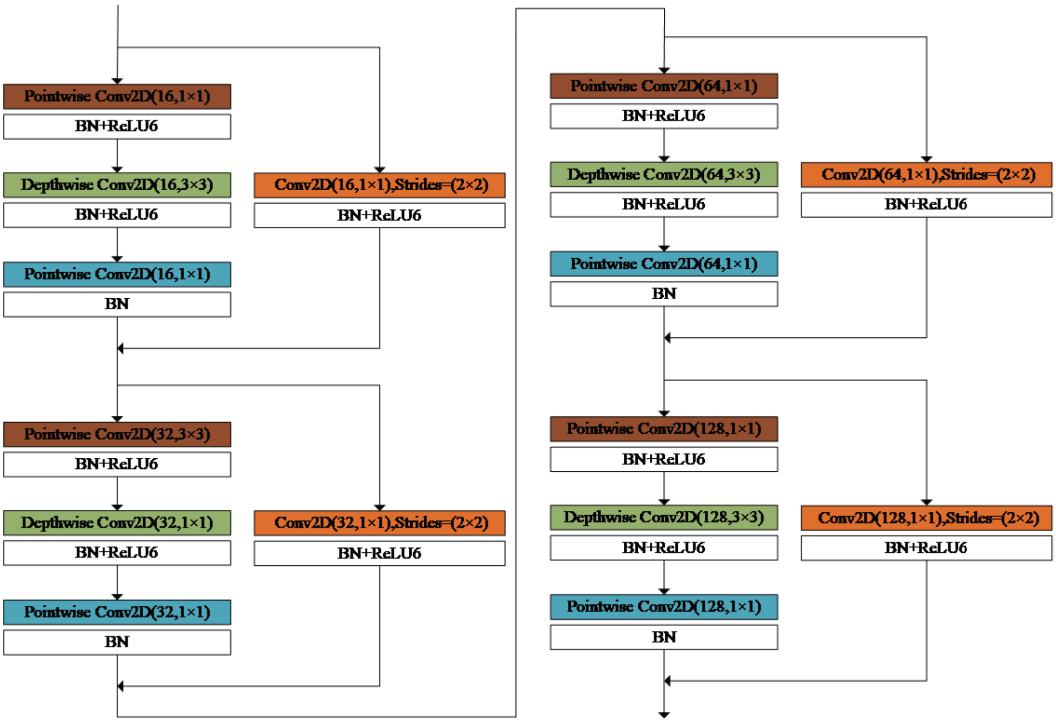

FIGURE 3. The parameters of the network in the second stage

third stage enters the Softmax layer for the final expression classification.

$$\text{Softmax } = S_{\mathrm{j}} = \frac{1}{\sum_{j=1}^{T} e_j x^{(i)}} \begin{bmatrix} e^{\theta^{(n)}_{\mathrm{g}(n)}} \\ e^{\theta_2(\theta^n} \\ \vdots \\ e^{\theta_{x(n}(0)} \end{bmatrix} \tag{2.5}$$

where T represents the total number of classifications, $\theta_1, \theta_2, \ldots, \theta_T$ represents the model's parameters, and $\mathbf{x}^{(\mathrm{i})}$ represents the classification layer $\frac{1}{\sum_{j=1}^{T} e_j x^{(i)}}$'s input features and the normalized probability distribution. Its loss function is as follows:

$$Loss = -\sum_{j=1}^{T} y_i \log\left(S_j\right) \tag{2.6}$$

where y^i is a 1 × T vector. Some structures retain the initial parameters of the original network.

3. Experiments and applications

The Fer2013 dataset was constructed and added to European and African facial expression information. The expression dataset created by FER two seven carrier and Courville in 2013 for the ICML2013 facial expression recognition competition.

The FER2013 dataset contains 35886 facial expressions, including those from different age groups in daily life. The pixel values of the images are all 48*48, divided into seven types of expressions: anger, disgust, fear, happiness, sadness, surprise, and neutrality. It is divided into three parts: training, validation, and testing. The training section includes 28709 images, the validation section includes 3589 images, and the testing section includes 3589 images. Each image is labeled with the correct facial expression category. Due to collection and labeling errors, the highest human recognition accuracy of the FER2013 dataset is only 65% to 70%. CK Company's Twenty-eight database is an extension of the Cohn Kanade dataset, which consists of seven expressions: anger, temptation, disgust, fear, joy, sadness, and surprise, with 123 participants and 593 image sequences. The CK company dataset is a universal facial expression dataset suitable for research on facial expression recognition. Then, the Jaffe and CK+ datasets were selected for application practice, targeting the facial expression characteristics of North American and East Asian personnel. Figure 4 shows some sample expression datasets.

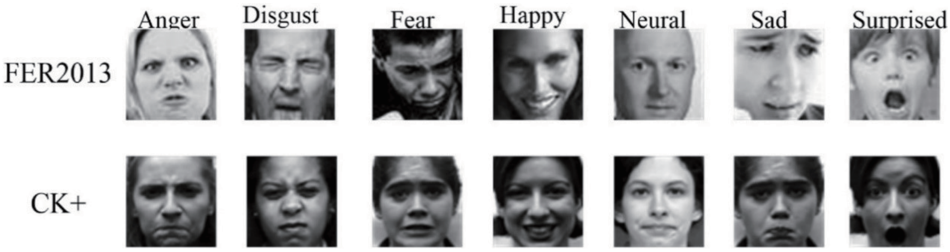

Figure 4. Dataset samples

The optimal settings were selected by analyzing the decline of the loss function and the reduction of error rates. The impact of different step size factors on the optimizer was also analyzed. In the table, Beta1 and Beta2 represent the exponential decay rates of the first-order and second-order matrix estimates, respectively. The parameters for different datasets are listed in Table 1.

Table 1. Some network parameters

Model	Parameters	Values		
		Jaffe	CK+	Fer2013
Our network	Optimizer	Adam	Adam	Adam
	Alpha	1E − 4	1E − 3	1E − 4
	Betal	0.95	0.95	0.95
	Beta2	0.999	0.999	0.999
	Image size	96 × 96	96 × 96	96 × 96

Experiments were conducted on the Jaffe, CK+, and expanded Fer2013 datasets. Figure 5 shows the change in learning rates during the training process.

Figure 5 suggests that the learning rates of the three datasets change dynamically with the number of training rounds without significant jumps, eventually converging to 0. This verifies that the network's fit is suitable and will not result in overfitting.

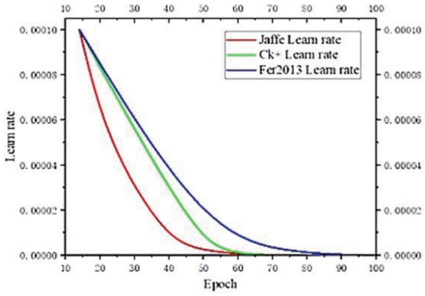

Figure 5. Learning rate variation

Figure 6 shows the change in accuracy during the training process of the three different networks under the same experimental conditions, using the Jaffe, CK+, and expanded Fer2013 datasets. Due to its small size, the Jaffe dataset experienced significant fluctuations at the beginning of training but gradually stabilized with the increase in rounds, eventually reaching a stable value. This preliminary judgment indicates that the network structure is robust for small datasets.

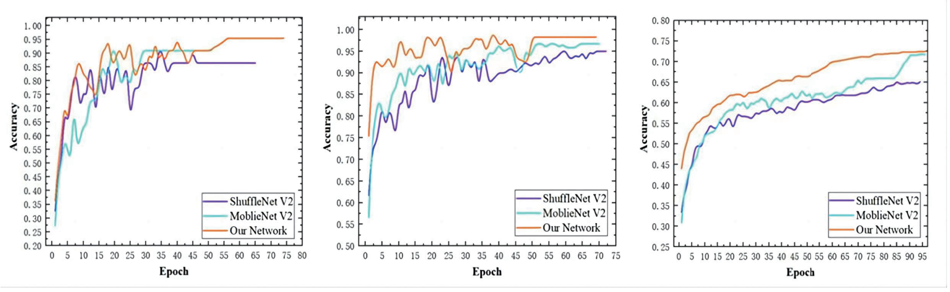

Figure 6. Training rounds

Figure 7 shows the change in training data in the lightweight convolutional network for each dataset. Analyzing the training process of various datasets shows that the prediction accuracy of emotion classification and the overall accuracy rate have reached the expected goal. Through extensive training and learning of a large amount of data, it is shown that the network is convergent, indicating that the network can accurately identify various complex facial expressions, including subtle emotional fluctuations. This high accuracy ensures the reliability of expression recognition in psychological warfare, fully demonstrating the superiority of the proposed method in this paper.

Table 2 lists the comparison results of the proposed method with other networks, with recognition accuracy reaching optimal results. Compared with the recognition accuracy of the MobileNet V2 and ShuffleNet V2 networks, the improvements are 4.57%, 1.63%, and 2.46%, respectively, with a more significant increase in accuracy compared to other networks. Due to the reference to various network structures, the running time is longer than the optimal networks, with an average of about 0.2 seconds less. Running time is an intuitive reflection of operational efficiency.

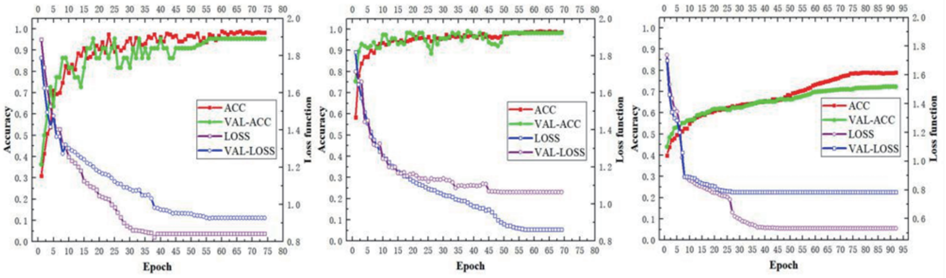

Figure 7. Training rounds

Compared with the improvement ratio of accuracy, insufficient running efficiency does not hurt practical psychological warfare applications.

Table 2. Experimental results

	Darams (M)	Jaffe		Ck+		Fer2013	
		Acc (%)	Run time (ms/step)	Acc (%)	Run time (ms/step)	Acc (%)	Run time (ms/step)
ShuffleNet V2	3.2	86.35	**264 ± 5**	95.00	306 ± 2	63.06	329 ± 6
MobleNet V2	3.4	90.90	353 ± 4	96.67	413 ± 3	69.87	453 ± 5
I-MobileNetV2 [15]	3.26	1	1	95.96	**179 ± 2**	68.62	**186 ± 5**
Our network	4.6	**95.47**	501 ± 6	98.30	511 ± 5	**72.33**	588 ± 5

To further evaluate the actual application performance of the lightweight convolutional network in psychological warfare, the efficiency parameters of expression recognition were compared between the lightweight convolutional network and ten classic expression recognition algorithm networks on the CK+ dataset. Table 3 lists the accuracy data of ten networks and the lightweight network's expression recognition algorithm. The second column of the table shows the number of layers and methods of network expression.

As shown in Table 3, compared with ten representative expression recognition networks, the proposed network's recognition rate reaches 98.30%, superior to the existing representative algorithm networks, achieving precise recognition, indicating that this network meets the accuracy requirements for applications in psychological warfare. To avoid the specificity of the experiment, the efficiency parameters of expression recognition were compared between nine expression recognition algorithm networks and the lightweight convolutional network on the newly constructed expanded Fer2013 dataset. Since the Fer2013 dataset consists of unrestricted facial images collected in the wild, including uneven lighting, occlusion, and low resolution, and the dataset sample was expanded, many interference factors were artificially added, resulting in an overall lower accuracy than the former.

Table 4 compares the accuracy between typical expression recognition algorithms and the lightweight convolutional network on the Fer2013 dataset.

Table 3. Comparison of results (CK+)

Methods	# of Expression	Accuracy (%)
LBP	6 +neutral	87.20
HOG [6]	6 +neutral	89.70
Gabor fifilter [4]	7	84.80
Poursaberi et al. [11]	6	92.02
Deepak Ghimire [12]	6	94.10
AU-DNN [8]	6 neutral	92.05
JFDNN [5]	6	97.30
CNN [14]	6	93.20
C-CNN [3]	6	91.64
Pre-trained CNN [2]	7	95.29
I-MobileNetV2 [15]	6	95.96
Our network	**7**	**98.30**

Table 4. Comparison of results (Fer2013)

Method	Accuracy (%)
RBM	70.56
Kim et al.	69.18
Jeon et al.	69.58
Devries et al.	66.11
CNN	65.01
Liu et al.	64.09
Shen et al.	60.06
Ergen et al.	56.40
Pre-trained CNN	70.03
I-MobileNetV2 [l5]	68.62
Our network	**72.33**

According to Table 4, the proposed lightweight network's recognition efficiency reaches 72.33%, outperforming all other expression recognition algorithms when applied to the most challenging Fer2013 dataset, with its numerous changes. This recognition efficiency can meet the application requirements in psychological warfare. Due to the numerous changes in this dataset, the key variable affecting the recognition efficiency of the expression recognition network, which is not high, needs further research.

4. Conclusions

This paper proposes a lightweight convolutional expression recognition network with lightweight characteristics. The design of the lightweight convolutional network enables it to process a large deal of facial expression data and achieve rapid recognition. Through experimental verification, the lightweight convolutional network has certain advantages in expression recognition. Compared with other methods, it shows better performance in terms of accuracy and robustness. The application of the lightweight convolutional network in expression recognition provides strong technical support for psychological warfare, proving the further application of the lightweight convolutional network in psychological warfare.

The next step should focus on adding small classification layers to the neural network for specific expression recognition and consider using transfer learning methods to improve the network model's generalization ability. Future research is also needed on key factors of recognition efficiency, such as the low quality of image input during the expression recognition process.

References

[1] M. S. Bartlett, B.G . Littlewort, I. Fasel and J. R. Movellan, *Real-time face detection and facial expression recognition: Development and applications to human computer interaction*, in: Proceedings of the Computer Vision and Pattern Recognition Workshop, 2003. CVPRW ' 03. Conference, IEEE, 2003, pp. 53–53.

[2] D. Ghimire, S. Jeong, S. Yoon, J. Choi and J. Lee, *Facial expression recognition based on regionspeciffc appearance an geometric features*, in: Proceedings of the Tenth International Conference on Digital Information Management, IEEE, 2016, pp. 142–147.

[3] W. Gong, Z. La, Y. Qian and W. Zhou, *Hybrid attention-aware learning network for facial expression recognition in the wild*, Arab J. Sci. Eng. **49** (2024), 12203–12217.

[4] S. Hossain, S. Umer, R. K. Rout and H. A. Marzouqi, *A Deep quantum convolutional neural network based facial expression recognition for mental health analysis*, in: IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 32, IEEE, 2024, pp. 1556–1565.

[5] H. Jung, S. Lee, J. Yim, S. Park and J. Kim, *Joint ffne-tuning in deep neural networks for facial expression recognition*, in Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2015, pp. 2983–2991

[6] P. Kumar, S. L. Happy and A. Routray, *A real-time robust facial expression recognition system using hog features*, in: Proceedings of the International Conference on Computing, Analytics and Security Trends, IEEE, 2017, pp. 289–293.

[7] C. J. Liu, X. Q. Liu, C. Chen and K. Zhou, *Deep global multiple-scale and local patches attention dual-branch network for pose-invariant facial expression recognition*, CMES - Computer Modeling in Engineering and Sciences **139** (2023), 405–440.

[8] M. Liu, S. Li, S. Shan and X. Chen, *Au-aware deep networks for facial expression recognition*, in: Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, IEEE, 2013, pp. 1–6.

[9] A. T. Lopes, E. D. Aguiar, A. F. D. Souza and T. Oliveira-Santos, *Facial expression recognition with convolutional neural networks: coping with few data and the training sample order*, Pattern Recognition **61** (2017), 610–628.

[10] A. Mollahosseini, D. Chan and M. H. Mahoor, *Goingdeeper in facial expression recognition using deep neural networks*, in: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016, pp. 1–10.

[11] J. A. Obaid and K. H. Alrammahi, *An intelligent facial expression recognition system using a hybrid deep convolutional neural network for multimedia applications*, Appl. Sci. **13** (2023): 12049.

[12] X. Wang, J. K. Yan, J. Y. Cai, J. H. Deng, Q. Qin, Y. Cheng, *Super-resolution reconstruction of single image for latentfeatures*, Comp. Visual Media **10** (2024), 1219–1239.

[13] M. Yu, J. Shi and C. Xue, *A review of single image super-resolution reconstruction based on deep learning*, Multimed Tools Appl. **83** (2024), 55921–55962.

[14] M. Zhu and M. Wen, *A multi-channel convolutional neural network based on attention mechanism fusion for facial expression recognition*, Appl. Math. Nonlinear Sci. **9** (2024), 1–14.

[15] Q. Zhu, H. Zhuang and M. Zhao, *A study on expression recognition based on improved mobilenetV2 network*, Sci. Reports **14** (2024): Article number 8121.

Manuscript received March 29, 2024
revised September 12, 2024

T. C. Dong
Logistics university of people's armed police force, Tianjin, 300300, No.1 Huizhi street, Dongli District, Tianjin
E-mail address: `WJUdtc@163.com`

S. P. Xie
Master of Science, Director of Textbook Construction Center of University of International Business and Economics, Beijing, 10036, No.10, Huixin East Street, Chaoyang District, Beijing
E-mail address: `xieshupei1987@163.com`

Y. Du
Logistics university of people's armed police force, Tianjin, 300300, No.1 Huizhi street, Dongli District, Tianjin
E-mail address: `duyong367@163.com`

Q. M. Shi
Logistics university of people's armed police force, Tianjin, 300300, No.1 Huizhi street, Dongli District, Tianjin
E-mail address: `1909744485@qq.com`