International Conference on Humanities, Social and Management Sciences (HSMS 2024)

Facial Expression Recognition Algorithm Integrating Semantic Features and Attention Mechanisms

Xue Tian[1], Shuili Zhang[1,2,*], Rui Huo[1]
[1]*College of Physics and Electronic Information, Yanan University, Yan'an, Shaanxi, China*
[2]*Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data, Yan'an, Shaanxi, China*
**Corresponding Author.*

Abstract：Due to the insufficient feature values extracted by traditional CNN networks, existing facial expression recognition algorithms struggle to effectively cope with the diversity and complexity of human emotional expressions, as well as the influence of lighting conditions and environmental factors, this paper proposes a facial expression recognition algorithm that integrates semantic features and attention mechanisms. By embedding attention mechanisms into the U-Net network, more prominent facial expression features are extracted, and the improved U-Net structure network is integrated in each layer of Resnet34 to extract richer feature values. Utilizing each layer of ResNet34 to process features, combined with U-Net and ECA -Net to generate weights, and outputting the weights and feature values through a residual network, more significant facial expression features can be extracted, enhancing the robustness and generalization of the model. Experimental evidence shows that the algorithm achieves a recognition rate of 78.1% on the FER2013 facial expression dataset, indicating superior accuracy.

Keyword: U-Net; Resnet34; Attention Mechanism; FER2013 Dataset

1. Introduction

In the realm of computer vision, the research direction of facial expression recognition has been concerned, it plays a crucial role in various fields such as emotion analysis, mental health, and also plays a role in the field of healthcare for automatically assessing the emotional state of patients, so facial expression recognition still faces a series of challenging problems [1]. Facial expression recognition needs to go through four steps: image acquisition, image preprocessing, feature extraction, refining facial expression features and classification, among them, feature extraction is the most important and has a decisive impact on the expression recognition results [2]. Traditional facial expression feature extraction methods include global methods such as independent component analysis (ICA) [3] and local methods such as LBP algorithm [4]. Shan et al [5] proposed a new facial expression recognition method, which extracts features from face images using Gabor filters across five scales and eight directions, and then use the LBP and LPQ of the face image to encode the Gabor image, to achieve the classification of the expression. Zhu et al [6] used linear regression classification as an expression classifier, which reduces the dimensionality of the data while preserving important information from the original image. Shi et al [7] proposed a feature point constraint algorithm that utilizes the SIFT descriptor as a feature parameter to determine the optimal position of feature points in the region, thereby accurately capturing expression changes, and extract regional gradient information. These traditional methods have contributed greatly to expression recognition, but manually designed feature extraction algorithms are cumbersome and time-consuming processes , usually demands a substantial quantity of labeled data for training the classifier, and might fail to capture complex image features, and the generalization ability is insufficient.

With the great breakthroughs made by deep learning in image recognition and other related fields, researchers have also changed facial expression recognition from traditional methods to methods based on deep learning, which to a certain extent solves the problem of the diversity and imbalance of training data

caused by different individuals showing a variety of facial features and expression intensity when expressing emotions, as well as lighting conditions. The change of face posture and expression intensity leads to the inconspicuous features of face extraction, which reduces the recognition performance. For example, Wang et al [8]proposed a novel region attention network (RAN) for adaptively capturing facial regions; Nie [9] proposed an improved VGG16 network and implemented expression classification on the FER2013 dataset, which can capture channel dimension features, thereby improving accuracy and convergence speed; The Facial Motion Priority Network (FMPN) was proposed by Chen et al [10], which introduced an additional branch to generate facial masks, thereby focusing on facial muscle movement areas, but the detection accuracy of these methods is not particularly high. Wang et al [11] introduced an ECA-Net, which involves only a minimal number of parameters, yet significantly performance, this module employs a local cross-channel interaction strategy without reducing dimensionality, it overcomes the problem of SE-Net proposed by Hu et al [12] which increases the model complexity due to dimensionality reduction. For automatic facial emotion detection, Li et al [13] unveiled a novel end-to-end network that combines LBP feature extraction with an enhanced attention mechanism, resulting in enhanced performance. Wang et al [14] used multiple convolutional blocks to extract high-level semantic features and adopted attention branches similar to the U-Net architecture to obtain local highlighting information, thereby achieving high-precision facial expression recognition. Zhang and Zhao[15] proposed a facial expression recognition method that integrating attention mechanism in the improved ResNet, reconstructed the feature map in the middle of the network, emphasized important features, suppressed common features, and replaced the original ReLU in ResNet with activation function PReLU to overcome the problem that the model could not perform backpropagation during training, and adding a Dropout to suppress the overfitting between the avgpool and the fc, in an effort to further enhance the robustness and generalizability of the network model.

To focus the extracted facial expression features more on the areas of interest and to enhance the accuracy of the detected expression categories, this paper proposes an expression recognition algorithm that integrates semantic features with an attention mechanism. It first introduces the attention mechanism ECA-Net, which reconstructs the feature maps, meanwhile, the attention mechanism network is embedded into the down sampling part of U-Net to extract more significant facial expression features; subsequently, the enhanced U-Net is integrated into ResNet34, enriching the feature extraction capabilities of the network. Importantly, the depth of each U-Net and ECA-Net integrated network varies, the depth decreases sequentially based on the input and output features of each layer in ResNet34, ensuring that the features of each network layer match, this enhances the model's robustness. By implementing these improvements, training and testing on the dataset have demonstrated that this model significantly enhances the accuracy of expression recognition.

2. Relevant Basic Theories

2.1 Residual Network

In this paper, image features are extracted using residual networks [16], which can construct very deep networks, so it can learn more complex and abstract feature representations. The structure of a residual block is illustrated in Figure 1.

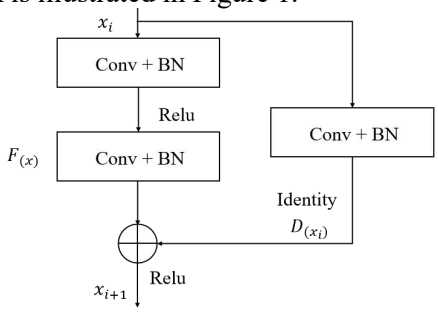

Figure 1. Residual Block Structure Diagram

A residual Network consists of many residual blocks, which are divided into two parts: identity mapping and the residual. Each residual block performs a series of feature transformations and adds the input directly to the output by bypassing the intermediate layers. This architecture enables the model to learn residual mappings instead of complete mappings, which reduces the difficulty of training.

Table 1. Resnet34 Network Structure

Layer Number	Layer name	Output size	configuration	Residual unit type
	Conv1	112×112	7×7 conv, 64, stride2 3×3 max pool, stride2	
layer1	Conv2_x	56×56	$\begin{bmatrix} 3\times3, & 64 \\ 3\times3, & 64 \end{bmatrix}\times3$	basicblock
layer2	Conv3_x	28×28	$\begin{bmatrix} 3\times3, & 128 \\ 3\times3, & 128 \end{bmatrix}\times4$	basicblock
layer3	Conv4_x	14×14	$\begin{bmatrix} 3\times3, & 256 \\ 3\times3, & 256 \end{bmatrix}\times6$	basicblock
layer4	Conv5_x	7×7	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix}\times3$	basicblock
		1×1	*Average pool, 1000-d fc, softmax*	

As shown in Figure 1, $D_{(x_i)}$ is the direct mapping part, $F_{(x)}$ for the residual structure part, which consists of two to three convolution operations, in this paper, we use Resnet34 [17], which has a cell type of basicblock, so it consists of two convolution operations, 1×1 convolution operation is used to upscale or downscale the dimensions, in an effort to match the number of feature maps with those added in the output. The expression for the residual block is as shown in Equation (1).

$$x_{i+1} = D_{(x_i)} + F_{(x)} \quad (1)$$

Table 1 shows the basic structure of Resnet34. The network contains five convolutional groups, each comprising one or more basic convolutions, with each layer processing through Conv, BN, and Relu. Initially, a single convolution operation using a 7×7 kernel is employed, followed by 3×3 max pooling, the entire network model consists of four layers, each containing multiple basic blocks, where Conv2_x, Conv3_x, Conv4_x, Conv5_x correspond to layer1, layer2, layer3, and layer4 layers, respectively, and finally the image features are obtained by average pooling.

2.2 Attention Mechanism Network ECA-Net

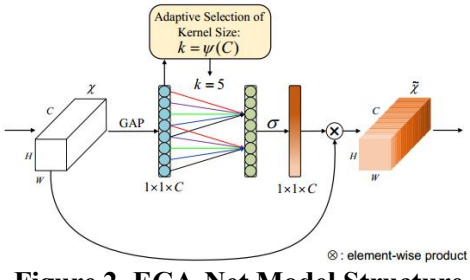

Figure 2. ECA-Net Model Structure Diagram

ECA-Net (Efficient Channel Attention Network) [18] is an efficient neural network for image categorization that enhances performance and efficiency through Efficient Channel Attention, improving inter-channel information representation and cross-channel interaction without dimensionality reduction.

Figure 2 illustrates the ECA-Net architecture. The ECA module aggregates data from each channel in the input feature map and uses the k to control the generated channel weights, determining which channels' information is more important. The larger the value of k, the more channels are considered and the module relies more on the overall channel information to generate weights. This is illustrated in the model structure diagram of ECA-Net, the input feature maps, after global average pooling of channels without dimensionality reduction, followed by an efficient implementation through a fast 1D convolution with the kernel size of k, the k denotes the extent of local cross-channel interactions.

The primary objective of the ECA module in ECA-Net is to efficiently capture local cross-channel interactions within a convolutional neural network. To circumvent the computational demands associated with manual tuning via cross-validation, ECA-Net employs an adaptive approach to select the size of the one-dimensional convolution kernel. The channel dimension is used in this selection to automatically calculate the magnitude of local cross-channel interactions. Equation (2) illustrates the proportional relationship between the kernel size k and the channel dimension C, which is commonly set to a power of two, in the context of 1D convolution.

$$C = \varphi(k) = 2^{(\gamma * k - b)} \quad (2)$$

Equation (3) illustrates how the C determines the k. The nearest odd integer to t is shown by the symbol |t|odd. In this paper, γ and b have been set to 2 and 1, respectively.

$$\mathbf{k = \varphi(C) = |\frac{\log_2(C)}{\gamma} + \frac{b}{\gamma}|_{odd}} \quad \mathbf{(3)}$$

3. Improved Programs

3.1 Embedding ECA-Net into the U-Net Architecture

Due to inter-individual and intra-individual differences, the faces of individuals vary based on gender, age, or ethnicity, and the primary cues for facial expression recognition primarily come from crucial areas of the face, such as the eyes, mouth, and eyebrows, so as to enable the model to learn information from key regions of expressions, an attention mechanism is embedded within the U-Net structure [19],the primary goal of U-Net, a deep learning framework for image segmentation tasks, is to split the input image into pixel-level masks of the same size. The encoder, located on the left side of the structure, consists of a sequence of convolutional and pooling layers that lower the image's resolution. The middle feature layer extracts global contextual information from the input image. Finally, the decoder, located on the right side, consists of a sequence of deconvolution layers and skip connections that enable the network to combine feature information. Either Sigmoid or Softmax are used in the output layer.

The attention mechanism network is the ECA-Net introduced in the previous section. This method uses the U-Net input semantic features to refine the feature maps output from the Resnet section, embedding ECA-Net is intended to focus on more prominent facial expression features, cover up the unimportant parts of the face, reduce the loss of important information and enhance detection accuracy, so that the final expression classification results are more accurate.

In the U-Net encoding stage, after each extraction layer extracts effective features, ECA-Net is embedded to refine the features. It generates weights to learn the relevance of the initial feature channels. Key facial expression features receive higher weights, while irrelevant features get lower weights, prioritizing useful information and enhancing the network's focus and sensitivity to primary features. Figure 3 illustrates the network structure diagram where ECA-Net is embedded into the U-Net architecture the each layer of Resnet34 extracts common features. In the encoding part of the U-Net, each convolutional layer is followed by the facial expression weights processed through ECA-Net, which filters out less prominent features and extracts the most evident expression weights, and it is observable that the number of channels on the left side has been increasing, while the number of channels on the right side returns to the original count, maintaining consistency with the downsampling process, the encoder feature maps are concatenated with the decoder feature maps along the channel dimension using the concat operation, finally, the result is outputted using a 1×1 Softmax function, this method enhances the prominence of the extracted expression features. Importantly, to match the output channels of each layer in ResNet34, the entire model has different numbers of convolutional blocks, pooling, upsampling, and downsampling layers, as well as attention layers when passing through this network layer, this depends on the spatial feature size of the input to the residual network.

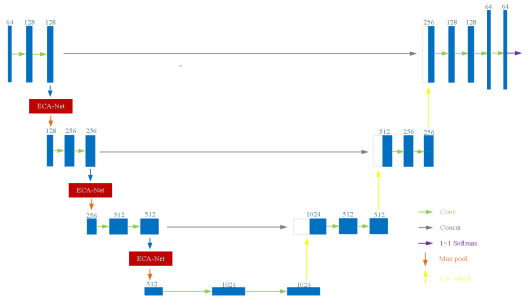

Figure 3. Network Structure of ECA-Net Embedded in U-Net Network

3.2 Facial Expression Recognition Algorithm Integrating Semantic Features and Attention Mechanisms

To extract richer facial expression information from multiple viewpoints, enhance image quality and feature extraction accuracy, and accelerate the network's convergence speed, and to make the facial cues more focused on the parts with obvious expression feature information such as the eyes, the mouth, the eyebrows, etc. The U-Net and the ECA-Net

are embedded into each layer of Resnet34, and Resnet34 has a total of 4 layers, so the model can be divided into 4 blocks, each block goes through each layer of Resnet34 to extract the image features, and then use U-Net and ECA-Net to generate the expression weights to extract richer semantic features, making the feature information more complete and making the expression feature information more prominent, due to the different output channels of each layer in ResNet34, the depth of each part of the U-Net network varies to align the feature information, in Figure 3, the input channels in the first layer are 64, then the depth of the U-Net and ECA-Net fusion network is 4, this means that the encoder goes through 4 convolution blocks to increase channel quantity from 64 to 1024, and the decoder returns the number of channels from 1024 back to 64 through 4 inverse convolutions, which allows it to be fitted to each layer of the residual network, the second block of Resnet34 , which has an input channel count of 128 and a U-Net network depth minus 1.

The primary steps of this model can be divided into the following categories, according to the overall algorithmic model flowchart displayed in Figure 4:

(1) Firstly, the input image size is 224×224, it will go through a step of 2, 7×7 convolutional layer and a step of 2, 3×3 maximum pooling layer to reduce its spatial size to 112×112, and then transferred to block1, in block1, the features generated by layer1 of Resnet34 are multiplied with the weights generated by a U-Net and Attention Mechanism ECA-Net fusion network with a depth of 4 and then added to the features generated by layer1 to output the expression features of block1, which is equivalent to a residual structure, and the size of the block1 generated feature map space is 56×56.

(2) The input of block 2 is the block 1 output features, following layer 2 of Resnet34. The channel quantity of the input is 128 in size, and the U-shaped fusion network's depth is 3, the U-shaped network begins at the convolutional layer with 128 channels, and block 2 creates a 28×28 feature space.

(3) The output of the upper layer is passed to block3, which passes through layer3 of Resnet34, at which point the depth of the U-shaped fusion network is reduced to 2, block 3 produces a 14x14 feature space size with 256 input channels.

(4) Block4 has an input channel count of 512 and a U-shaped fusion network depth of 1. Starting from a convolutional layer with a channel count of 512, a feature space size of 7×7 is generated. The end of the network generates outputs corresponding to the 7 expression states (happy, angry, fearful, disgusted, sad, surprised, and neutral) using a 7-way fully connected layer with Softmax and an average pooling layer.

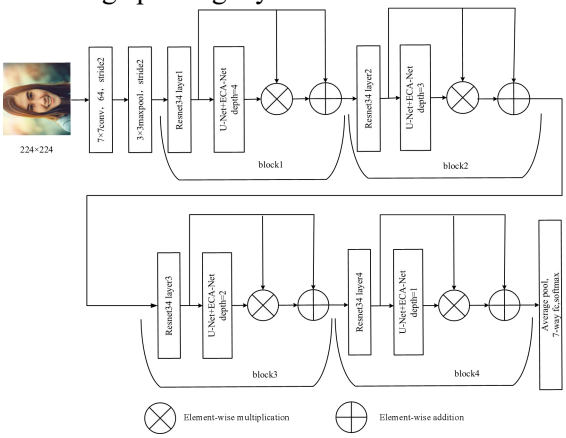

Figure 4. Flowchart of the Overall Algorithm Model

4. Experimental Results and Analysis

4.1 Training Environment

The Python programming language and a Tesla T4 GPU with 16GB of VRAM are used in this experiment's server hardware setup. Model training, testing, and validation are carried out using an Ubuntu 18.04 operating system with CUDA 10.2 and CuDNN 7.6.5.

4.2 Dataset

The FER2013 facial expression dataset, which was first assembled via the internet and is freely accessible, is used in this work. The dataset is widely used for training and assessing facial expression recognition models since it contains the expressions of people of various ages and ethnicities. There are 35,887 grayscale pictures in it. Resizing the photos to 224x224 is part of the data preparation procedure because of the deeper network layers. 28,709 training samples, 3,589 public test samples, and 3,589 private test samples make up the dataset. The seven categories of expressions in the dataset are: fear, rage, disgust, surprise, sadness, and neutral. The

dataset's CSV file, which links the image filenames to their associated expression labels, is used in this work. Typically, the CSV file format has two columns: the image pixel data column and the expression label column. Figure 5 shows some examples of images from the FER2013 dataset.

Figure 5. FER2013 Dataset

4.3 Training Strategies

This paper trains and tests the designed model on the FER2013 dataset. Initially, the images are reshaped to 224×224 and converted to RGB format, each experiment has a 50 period limit, training stops if there is no change in validation accuracy for more than 8 epochs, the starting learning rate is 0.0001, the learning rate will drop by an order of magnitude if the validation accuracy remains unchanged for two epochs. There is a 0.9 starting momentum and a 0.001 weight decay. Table 2 displays the configuration used for the experiment.

4.4 Analysis and Comparison of Experimental Results

This experiment trains and tests on the FER2013 public dataset, utilizing Resnet34, U-Net, and the attention mechanism network ECA-Net, to incorporate the attention mechanism to enrich the extracted facial expression features, and the focus is placed on distinct facial components, such as eyes and mouth, enhancing the model's performance and detection accuracy. As shown in the model comparison results in Table 3, this model has higher detection accuracy and better robustness compared to the previous face expression recognition models on the FER2013 dataset.

Table 2. Experimental Setup

parameter category	parameter name	parameter value
software environment	Ubuntu	18.04
	Python	3.6
hardware environment	GPU	Tesla T4
model training	Image_size	224×224
	Max_epoch_num	50
	Batch_size	48
	initial value of momentum	0.9
	learning rate	0.0001

As can be seen from the Table 3, integrating the attention mechanism network ECA-Net into each convolution layer of the downsampling layers in U-Net within each layer of Resnet34, the model achieves a facial expression recognition accuracy of 78.1%, this is 4.82% higher compared to the VGGNet network, 4.71% higher than the method that embedding channel attention and spatial attention mechanisms in Resnet50, and 3.96% higher than ResmaskingNet. Through a series of experiments and data comparisons, it is demonstrated that the algorithm designed in this paper, based on semantic features and attention mechanisms, possesses superior recognition performance.

Table 3. Comparison of Experimental Data

Algorithm name	Recognition rate/%
ResmaskingNet [20]	74.14
VGGNet [21]	73.28
Cbam_resnet50 [22]	73.39
VGGSpinalNet [23]	74.45
LHC-Net [24]	74.42
EmoNeXt [25]	76.12
Ours	78.1

4.5 Confusion Matrix Evaluation Model

The confusion matrix of the FER2013 dataset for this algorithm is displayed in Figure 6. The expression recognition accuracy is represented by the values along the diagonal, the higher the value, the better the model performs. Read the values on the diagonal of the confusion matrix from the statistical results, the recognition effect of the two expressions fear and sad is relatively poor, which may be caused by the greater similarity between them, however, for expressions like happiness and surprise, the model achieves high recognition rates of 95% and 87%, respectively; the off-diagonal values represent incorrect classifications, and the smaller the value, the better, from the matrix, it's evident that these values are relatively small, even the accuracy of identifying disgust as angry, happy, surprise, and neutral is 0%, which shows that the model can distinguish these expressions very well. Overall, the algorithm demonstrates a low rate of confusion on the FER2013 dataset, suggesting it performs well in minimizing misclassifications.

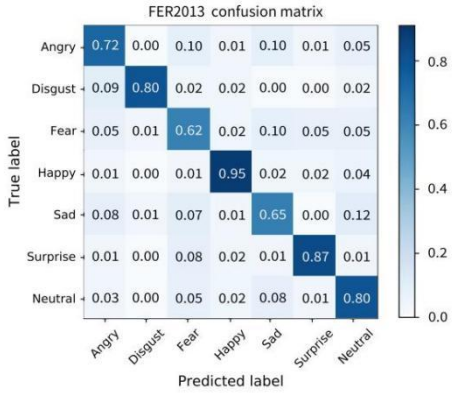

Figure 6. FER2013 Confusion Matrix

5. Conclusion

This paper proposes a facial expression recognition algorithm that integrates semantic features and attention mechanisms, fully leveraging the advantages of the U-Net network and attention mechanism fusion network, extracting higher-level semantic features, and focusing the model on more interesting parts. Firstly, the attention mechanism ECA-Net is embedded into U-Net to generate important expression weights, filtering out less distinct features. Secondly, to increase the resilience and generalization capacity of the ResNet34 four-layer network, the upgraded U-Net is fused into each layer. The depth of the fused network in each layer varies depending on the input and output sizes of ResNet34. Utilizing each layer of ResNet34 to obtain image features and employing the U-Net network structure embedded with the attention mechanism ECA-Net to generate important expression weights, richer and more effective features are extracted. By training and testing this model on the FER2013 dataset, it is found that this network can extract the most prominent parts of facial expression features and focus attention on the interested features, thus demonstrating better recognition performance, proving the effectiveness and superiority of this method.

Acknowledgements

This study was approved by the Foundation of the Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data (Grant No.IPBED17), Graduate Innovation Program Project of Yan'an University, China (Grant No. YCX2023026), Teaching Reform Project of Yan'an University, China (Grant No. YDJG23-34).

References

[1] He J. "Research on the Different Meanings of Facial Expressions under Various Cultural Backgrounds". Proceedings of the International Conference on Interdisciplinary Humanities and Communication Studies (ICIHCS 2022) (part5). Ed. University California Davis; 2022, 300-304.

[2] Luo Sishi, Li Maojun, Chen Man. Facial expression recognition network with multi-scale fusion attention mechanism. Computer Engineering and Applications, 2023, 59 (01):199-206.

[3] Buciu I,Pitas I.ICA and Gabor representation for facial expression recognition//Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429). IEEE, 2003, 2: II-855.

[4] Zhang B, Liu G, Xie G. Facial expression recognition using LBP and LPQ based on Gabor wavelet transform //2016 2nd IEEE international conference on computer and communications (ICCC). IEEE, 2016: 365-369.

[5] Shan C, Gong S, Mcowan P W. Facial Expression Recognition Based on Local Binarypatterns: A Comprehensive Study. Image and Vision Computing, 2009, 27(6):803.

[6] Zhu Y, Li X, Wu G. Face expression recognition based on equable principal component analysis and linear regression classification//2016 3rd International Conference on Systems and Informatics (ICSAI). IEEE, 2016: 876-880.

[7] Shi Y, Lv Z, Bi N and Zhang C. An improved SIFT algorithm for robust emotion recognition under various face poses and illuminations. Neural Computing and Applications, 2020, 32: 9267-9281.

[8] Wang K, Peng X, Yang J, Meng D and Qiao Y. Region attention networks for pose and occlusion robust facial expression recognition. IEEE Transactions on Image Processing, 2020, 29: 4057-4069.

[9] Nie H. Face expression classification using squeeze-excitation based VGG16 network//2022 2nd International Conference on Consumer Electronics and

Computer Engineering (ICCECE). IEEE, 2022: 482-485.
[10]Chen Y, Wang J, Chen S, Shi Z and Cai J. Facial motion prior networks for facial expression recognition//2019 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2019: 1-4.
[11]Wang Q, Wu B, Zhu P, Li P, Zuo W and Hu Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 11531-11539.
[12]Hu J, Shen L, Sun G. Squeeze-and-excitation networks//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
[13]Li J,Jin K,Zhou D, Kubota N and Ju Z. "Attention mechanism-based CNN for facial expression recognition." Neurocomputing 411 (2020): 340-350.
[14]Wang Z, Zeng F, Liu S, Zeng B."OAENet: Oriented attention ensemble for accurate facial expression recognition." Pattern Recognition 112 (2021): 107694.
[15]Zhang Dongyu, Zhao Lei. Facial Expression Recognition with Improved ResNet Integrating Attention Mechanism. Computer Technology and Development, 2023, 33(05):130-137.
[16]He K, Zhang X, Ren S and Sun J. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[17]Miao K, Zhao N, Lv Q, He X, Xu M, et al. Prediction of benign and malignant ovarian tumors using Resnet34 on ultrasound images. Journal of Obstetrics and Gynaecology Research, 2023, 49(12): 2910-2917.
[18]Zhang P, Jiang M, Li Y, Ling X,Wang Z,et al. An efficient ECG denoising method by fusing ECA-Net and CycleGAN. Mathematical Biosciences and Engineering, 2023, 20(7): 13415-13433.
[19]Ronneberger O, Fischer P, Brox T .U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR, 2015, abs/1505.04597.
[20]Pham L, Vu T H, Tran T A. Facial expression recognition using residual masking network//2020 25Th international conference on pattern recognition (ICPR). IEEE, 2021: 4513-4519.
[21]Khaireddin Y, Chen Z. Facial emotion recognition: State of the art performance on FER2013. arxiv preprint arxiv:2105.03588, 2021.
[22]Woo S, Park J, Lee J Y, Kweon I S. Cbam: Convolutional block attention module//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
[23]Santoso B E, Kusuma G P. Facial emotion recognition on FER2013 using VGGSPINALNET. Journal of Theoretical and Applied Information Technology, 2022, 100(7): 2088-2102.
[24]Pecoraro R, Basile V, Bono V. Local multi-head channel self-attention for facial expression recognition. Information, 2022, 13(9): 419.
[25]Boudouri Y, Bohi A. EmoNeXt: an Adapted ConvNeXt for Facial Emotion Recognition//2023 IEEE 25th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 2023: 1-6.