IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 April 2024, accepted 20 May 2024, date of publication 27 May 2024, date of current version 5 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3406037

RESEARCH ARTICLE

Graph Convolutional Neural Networks for Micro-Expression Recognition–Fusion of Facial Action Units for Optical Flow Extraction

XULIANG YANG[1], YONG FANG[2], AND C. RAGA RODOLFO JR.[3]
[1]Institute for the Synergistic Development of Urban-Academic Dynamics, Dongguan City University, Dongguan 523419, China
[2]School of Artificial Intelligence, Dongguan City University, Dongguan 523419, China
[3]National University CCIT, Manila 999005, Philippines

Corresponding author: Yong Fang (fangyong_2023@163.com)

This work was supported in part by the Application and Research in the field of Emotion Recognition Based on Multi-Modal Pre-Training Model under Grant 2023YZD001Z; in part by the Characteristic Innovation Project of General Universities in Guangdong Province through the Project Name: Exploration and Practice Based on Multimodal Pre-Training Digital Human under Grant 2023KTSCX217; in part by the 2022 Higher Education Teaching Reform through the Project Name: Curriculum Clustering Construction in Discipline Construction-Taking OBE-CDIO-Oriented Object-Oriented Curriculum under Grant 2022yjjg001; in part by the Young Innovative Talents Project of Guangdong Province under Grant 2018KQNCX359; in part by the Social Development Science and Technology Project of Dongguan City under Grant 20211800900282, Grant 2022180090952, and Grant 20231800936732; and in part by the Research on Key Technologies of 3D Object Acquisition and Reconstruction Based on Artificial Intelligence under Grant 20221800905202.

ABSTRACT Micro-expression recognition is an important problem in the field of computer vision and affective computing. To improve the accuracy of micro-expression recognition, the study proposes a novel graph convolutional neural network model oriented to micro-expression recognition, which divides facial features into regions and uses the optical flow method for feature extraction of facial action units. The model utilizes graph structure to encode facial features intuitively, while obtaining dynamic changes of facial features through the change information of optical flow, thus obtaining richer micro-expression feature information. The results show that the model performance can be maximized when the model is set to 5-way 5-shot and is taken as 1.4, at which time the model's accuracy on the dataset CAMSE II is 79.168%. The proposed algorithm performs well when compared with other algorithms in terms of accuracy and F1 score, the proposed algorithm realizes an accuracy of 0.795 on the CAMSE II dataset compared to the other algorithms which is up to 0.785. The accuracy of the proposed algorithm on the SAMM dataset is 0.738, which is only lower than that of the spatio-temporal recurrent convolutional neural network. The algorithm proposed in the study shows good performance in micro-expression recognition and promotes the development of the field of computer vision and affective computing.

INDEX TERMS Facial action, feature extraction, graph convolutional neural network, micro-expression recognition.

I. INTRODUCTION

In interpersonal communication, there are two core modes of communication, one is through verbal expressions such as direct conversations or text messaging; the other relies on non-verbal expressions such as facial expressions, body movements, and tone of voice [1], [2]. Micro-expressions are intense, but brief, changes in facial expressions that people experience in their daily lives and are often associated with real, hidden emotional states [3], [4]. However, they are difficult to capture with the naked eye due to their extremely short duration, usually only 1/25 to 1/5 of a second [5], [6]. In micro expression recognition, the shadows caused by lighting on the face directly affect the facial expression of

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

the character, resulting in inconsistency between the facial expression and the expression information of the character. In computer vision processing technology, when recognizing facial expressions of characters, it is necessary to first capture animation frames of the character's expressions. The various color features in the captured image will be more obvious, and different colors correspond to different emotions, which can also cause errors in facial expression feature extraction. Facial expression recognition is a technology for feature recognition of faces. When extracting basic facial features, it is necessary to first determine the facial recognition area in the image [7], [8]. The shape similar to the face in the image can interfere with the judgment of the recognition area, leading to facial expression recognition errors. The division of facial regions is a prerequisite for determining the extraction area of special zones. The division of facial regions is based on shape edges, and misjudgment of facial edges can directly lead to errors in the recognition area. To address these issues, the study proposes a graphical convolutional neural network (GCN) based on Facial Action Unit (FAU) for facial MER. The algorithm employs FAU to depict the detailed changes of facial expressions and extracts the key facial emotion signals to get the dynamic features of micro-expressions. Finally, GCN is then utilized to process these features for MER.

The innovation of the research lies in the use of multi block partitioning to extract FAU features, dividing FAU into different regions and performing separate feature extraction on FAU in different regions. Use the area with the greatest change in feature amplitude as the representative feature. The main contribution of the research is to enhance the feature extraction ability of FAU, improve the accuracy of facial micro expression recognition in machine vision, and achieve home interaction in the field of micro expressions.

The study will investigate facial MER from four aspects, firstly, a review of the current state of research on FAU and GCN, secondly, a study on micro-expression feature extraction and recognition based on FAU-GCN, thirdly, an experimental validation of the method raised in the study, and finally, a summary of the study.

II. RELATED WORKS

Facial expressions are a display of people's current mood state, Cowen A. S. et al. In order to explore the connection between human facial expressions and specific cultural and social contexts, they proposed the use of machine learning methods to analyze dynamic behaviors in the real world. The results show that each facial expression is associated with a specific context, these associations are preserved by 70% in different cultural regions, and the frequency of generating different facial expressions varies from region to region [9]. Li et al. to solve the issue of deep CNN, which has too large parameter sizes in the facial expression recognition task, they propose a dedicated lightweight facial expression recognition network with differentiable neural structure search model for automatic search. The outcomes denoted that the authors' proposed algorithm exhibits high accuracy on the facial expression dataset, which verifies the robustness of the system [10]. Chen et al. to solve the issue of data inconsistency between facial expression recognition datasets and to improve the model's ability to generalize across different datasets. proposed to construct a unified cross-domain FER evaluation benchmark and reconstructed cross-domain facial expression recognition and generalized domain adaptation algorithms. The results show that the proposed framework outperforms previous state-of-the-art methods on a unified evaluation benchmark and facilitates the development of cross-domain expression recognition [11]. Ben et al. put forward a video-based micro-expression analysis method in order to reveal the real emotions that people try to hide. Due to the transient nature and low intensity of micro-expressions, their detection and recognition are difficult and require expert experience. The authors published a new dataset and performed a unified comparison of representative methods. The results show the potential of micro-expression analysis in applications such as lie detection and criminal investigation [12].

GCN is a commonly applied algorithm for recognition classification processing, Levie et al. to study the transferability of spectral GCNs, they proposed to analyze the importance of graph-to-graph transformation. The outcomes indicated that if two graphs discretize the same continuous space, then the spectral filter or GCN has approximately the same effect on them [13]. Zhao in order to recognize new drug-target interaction (DTI) more efficiently, proposed a new framework based on GCN "GCN-DTI". The outcomes showcased that the framework outperforms the current state-of-the-art methods [14]. Tong et al. proposed a GCN-based fault detection method in order to detect and categorize transient faults in transmission lines more efficiently. The method utilizes spatial information as a priori knowledge and establishes a framework for transient fault detection and classification. Experimental findings indicated that the method can achieve the objectives more effectively [15]. Choi et al. proposed a method for large-scale GCNN training and data management using HydraGNN and ADIOS in order to efficiently utilize high-performance computational resources to train GCNs to predict material properties. The outcomes indicated that the method reduces the data loading time and achieves linear scaling performance on multiple GPUs [16]. Nguyen et al. to predict the drug response more accurately, they proposed a new method based on graph convolutional networks with graphical representation of drugs and cell lines. The findings indicated that the method proposed by the authors outperforms traditional methods in all performance metrics and finds the contribution of genomic aberrations to the response [17].

In summary, the recognition of facial micro-expressions, is the current popular research direction of facial expression recognition, but all kinds of deep learning networks can show good performance only in the feature extraction and recognition of facial macro-expressions, when facing micro-expressions, there is always insufficient feature extraction,

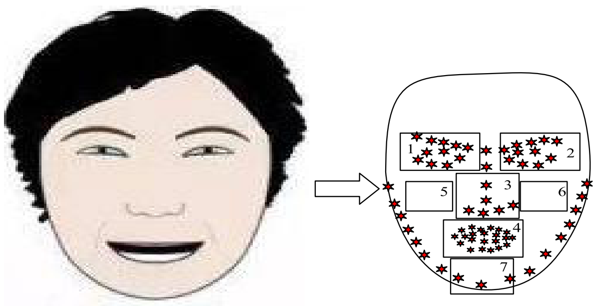

FIGURE 1. Division of facial feature regions.

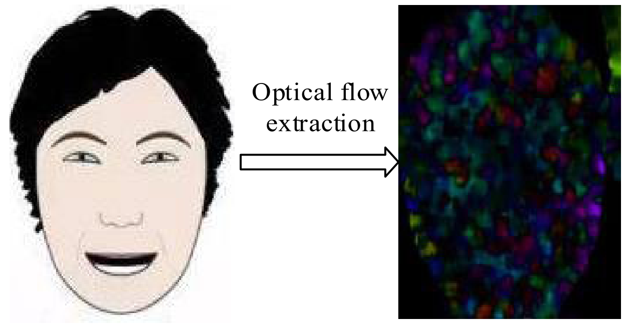

FIGURE 2. Facial optical flow field.

resulting in the failure of MER. the GCN performs well in all kinds of subtle domains, with excellent feature extraction i.e., recognition, so the study proposes GCN algorithm with FAU as input for feature extraction and recognition of people's facial micro-expressions.

III. MICRO-EXPRESSION FEATURE EXTRACTION AND RECOGNITION BASED ON FAU-GCNN

The MER can help to understand the emotional changes of others, the chapter 3's for the micro-expression feature extraction and MER is investigated, which is divided into two parts, the first part is based on FAU-GCNN micro-expression feature extraction, and the second part is based on prototype CNN MER.

A. RESEARCH ON MICRO-EXPRESSION FEATURE EXTRACTION BASED ON FAU-GCNN

Compared with macro-expressions, micro-expressions are characterized by short performance time and small facial changes, and it is difficult for the human eye to recognize changes in facial micro-expressions. The visual information capture accuracy of computers is much higher than that of the human eye, so it is inevitable to use computers in MER for information capture to assist recognition. When generating micro-expressions, people's facial feature changes are small, if feature extraction is performed from the whole face, or due to the small feature changes, it leads to insufficient feature extraction, resulting in the failure of MER [18]. Therefore, it is necessary to divide the facial features of the task into several regions and perform separate feature extraction for each region. When using MER as the basis of facial feature region division, it is necessary to determine the key points of facial regions related to expressions with the help of the response degree fitting method, and then according to the facial key points corresponding to micro-expressions, the FAU location of micro-expressions can be determined and the facial regions can be divided. The key points of facial features in micro-expressions and the feature extraction region division unit are denoted in Figure 1.

As can be seen in Figure 1, most of the key points of the character's facial expression are located in the character's five senses, therefore, the character's five senses are divided separately when dividing the feature extraction region. When the facial expression changes, it will involve the cheek nerves and cause the variable ratio of cheek FAUs, therefore, the study includes both cheeks in the feature extraction region of FAUs in addition to the five senses. The study delineates a total of seven micro-expression feature extraction units, region 1 for the left eye, region 2 for the right eye, region 3 for the nose, region 4 for the lips, region 5 for the left cheek, region 6 for the right cheek, and region 7 for the chin, and the FAUs corresponding to micro-expressions when they are generated are all in the above seven regions. After delineating the feature extraction regions of micro-expressions, the study extracts the facial features of the task with the help of optical flow method (OFM). OFM is a computational method for estimating pixel motion in an image sequence. In facial feature extraction, the OFM can be used to track the dynamic changes of the face in the video. Since micro-expressions are characterized by small feature changes and short expression maintenance time, it is sufficient to select the FAU position of the first frame in the task's expression change as the FAU position of the sample when extracting the features using the OFM, and the optical flow field of the facial features is shown in Figure 2.

According to the optical flow field of the sample, calculate the sum of the optical flow amplitude of FAUs in each region, sort the regions, and the region with the largest change in the amplitude of the optical flow field can be taken as the representative feature. The feature adjacency matrix is used as the expression of the regional features, and the GCN is constructed based on it. When constructing the adjacency matrix, it is necessary to find the FAUs that co-occur in the dataset as the correlation matrix, and the elements in the adjacency matrix are shown in Equation (1).

$$A_{ij} = P\left(U_i \middle| U_j\right) \tag{1}$$

In Equation (1), A_{ij} denotes the element in the adjacency matrix; U_i denotes the probability of the occurrence of the i th FAU, and U_j denotes the probability of the occurrence of the j th FAU. When constructing the GCN for facial feature recognition, the introduction of the attention mechanism can effectively improve the network's allocation of resources, and

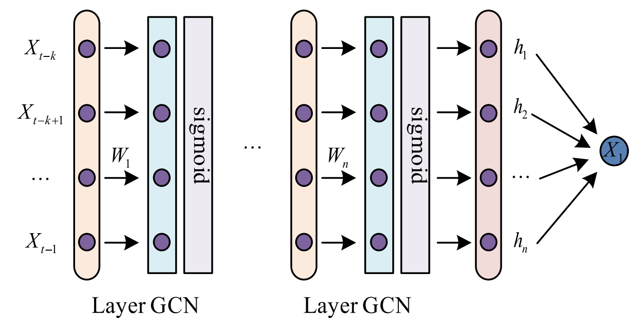

FIGURE 3. Self-attention mechanism structure.

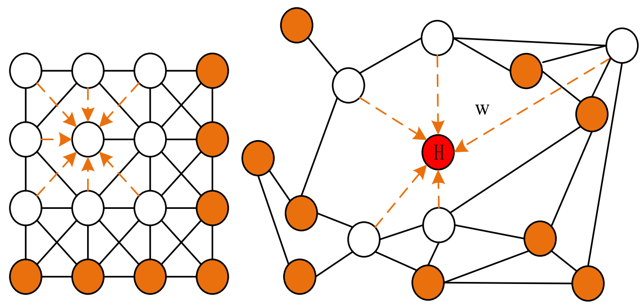

FIGURE 4. Feature extraction process of GCN.

the attention mechanism is shown in Equation (2).

$$Attention\,(Query, Source) = \sum_{i=1}^{L_x} similarity\,(Query, Key_i) \times Value \quad (2)$$

In Equation (2), L_x refers to the length of *Source*; *Key* and *Value* represent the attributes of *Source* after obtaining the weight coefficients of each key value, a weighted summation is performed to obtain the *Attention* value. Aiming at the characteristics of micro-expressions themselves, the studied attention mechanism adopts the self-attention mechanism, which can reduce the interference of external information, as shown in Equation (3).

$$Attention\,(Query, Key, Value) = soft\max\left(\frac{Query \bullet Key^T}{\sqrt{d_{key}}}\right) \times Value \quad (3)$$

In Equation (3), d_{key} denotes the dimension size of the *Key* value;$Query \times Key^T$ denotes the desired output. The structure of the self-attention mechanism is shown in Figure 3.

Assuming k d dimensional feature vectors h_i, the d dimensional feature vectors can be obtained by fusing the k features, which are computed in Equation (4).

$$\hat{h} = \sum_{i=1}^{k} a_i h_i \quad (4)$$

In Equation (4), the weights of the feature vectors are indicated. After determining the d dimensional feature vectors, it is also necessary to calculate the score of each feature vector, which is used to judge the influence of h_i on $\hat{h}$, and the score calculation is shown in Equation (5).

$$s_i = F\,(h_i) = \tanh\left(w^T h_i + b_i\right) \quad (5)$$

In Equation (5), s_i denotes the score; w^T denotes the matrix transposition. Finally, the weight of each region in the micro-expression feature change can be obtained by using softmax function to normalize, as shown in Equation (6).

$$\alpha_i = soft\max\,(s_i) = \frac{e^{s_i}}{\sum_{j=1}^{N} e^{s_i}} \quad (6)$$

In Equation (6), N denotes the feature extraction region. GCN is a deep learning model specialized in processing graph data. Different from the traditional CNN that performs convolution operation by sliding the window on the matrix data, GCN directly performs convolution operation on the structure of the graph, which can extract the complex associated features in the graph data. The core of the principle of GCN feature extraction lies in the propagation and fusion of the features through the multiplication operation of the neighboring matrices of the graph and the feature matrices. In this process, the new features of a node will be obtained by the weighted sum of the features of its neighboring nodes, reflecting the correlation between the nodes in the graph. The feature extraction process of GCN is shown in Figure 4.

Feature transfer between layers in a GCN network is in the form of Equation (7).

$$H^{(I+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(I)}W^{(I)}\right) \quad (7)$$

In Equation (7), H denotes the features of each layer; W denotes the weight feature matrix of the features; σ denotes the nonlinear activation function; $\tilde{D}$ is the degree matrix of $\tilde{A}$; I is the unit matrix; $\tilde{A} = A + I$. The micro-expression feature extraction model constructed in the study consists of three layers of GCN with self-attention mechanism, and the weight matrix is trained together with the optical flow field during the training, and the calculation of the cross-entropy loss during the training is shown in Equation (8).

$$L = -\sum_{c=1}^{n} y_{o,c} \log P_{o,c} \quad (8)$$

In Equation (8), $y_{o,c}$ denotes the binary indicator that the label is the same as the observation; $P_{o,c}$ means the predicted probability that the label c is the observation o. The loss function in the training process adopts the dice loss function, and the expression of the coefficients of the dice loss function is shown in Equation (9).

$$s = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \delta} \quad (9)$$

In Equation (9), p_i and g_i denote the predicted and the actual probability that a pixel occurs in a representative FAU region, respectively. δ denotes the smoothing parameter, and N denotes the product of the number of sample channels, the number of sample frames, and the image width and height. The dice loss function can be expressed in Equation (10).

$$L_{Dice} = 1 - s \tag{10}$$

When classifying micro-expressions, the softmax function is used to process the prediction results of the samples, at this time, it is necessary to ensure that the sum of the output probabilities of each category is 1, and then calculate the cross-entropy loss with Equation (11).

$$H(p, q) = -\sum_{i=1}^{n} p(x_i) \log(q(x_i)) \tag{11}$$

In Equation (11), p denotes the original category; q denotes the predicted category of the sample; n means the amount of categories; $p(x_i)$ denotes the probability that the sample belongs to the i th category in the original category; $q(x_i)$ denotes the probability that the sample belongs to the i th category in the predicted category.

B. MICRO-EXPRESSION RECOGNITION BASED ON FAU-GCN PROTOTYPE NETWORK

After using FAU-GCN based feature extraction for task facial micro-expressions, it needs to classify the facial micro-expressions of the characters based on different feature representations [19], [20]. These micro-expression features are then recognized with a sample less network based on the classification. Prototype network is a method of less sample learning. Sample less learning is the process of allowing a model to learn a new task or category with only a few samples [21], [22], [23]. Prototype networks aim to classify or compare between categories by learning to compute a prototype for each category. This prototype of a category can be understood as the mean of all sample features under that category. When predicting a new sample, the prototype network compares it to the prototypes of all categories and then categorizes it to the one that is closest to it. In practice, prototype networks usually need to compute prototypes for each category. For a given test sample, it is necessary to use the embedding function to compute its features, and then measure its distance from the prototypes of each category, and categorize the test sample into the category of prototypes closest to it, the specific framework and process is shown in Figure 5.

The study saves the categorized micro-expression features and other parameters obtained from FAU-GCN, as prototype data, and calculates the Euclidean distance between the prototype data and the test image of the recognition network. The adaptive loss function of the prototype network is obtained by combining the Triplet loss function with Equation (11), and the distances between the micro-expression prototypes of different categories are calculated, and then the MER can

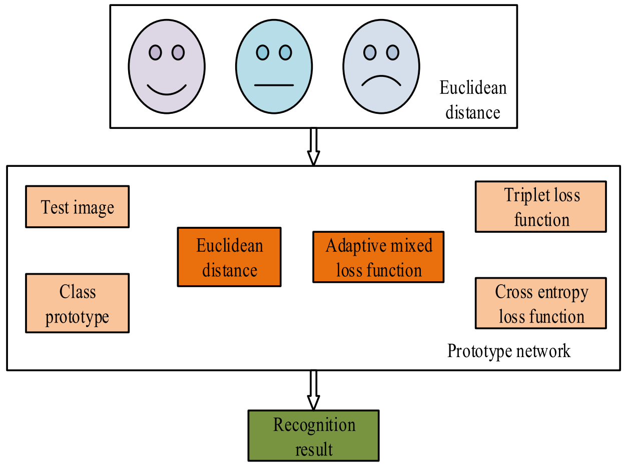

FIGURE 5. Basic framework and process of micro-expression recognition.

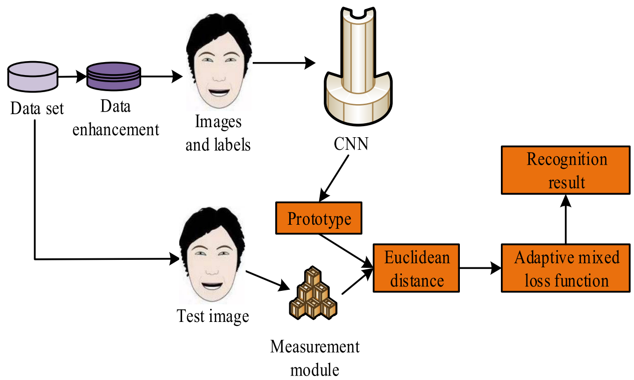

FIGURE 6. Overall structure of prototype network.

be accomplished by decreasing the distance of the samples in the same category, enlarging the distance of the samples in different categories, and then decreasing the percentage of the loss of the samples with too large distances. The prototype network used in the study is a method to migrate the information already learned by the network to the new domain, and the structure of this network includes the feature extraction module, the metric module, and the classification and recognition module, in which the feature extraction module is the FAU-GCN model established in Chapter 3.1, and the specific structure of this network is expressed in Figure 6.

After obtaining the facial expression features using FAU-GCN, these features are mapped to the embedding space, and each feature corresponds to a matrix, the weighted average of the embedding vectors of each image in the training sample is calculated, and then the corresponding covariance matrices of the individual images are clustered, at which point the prototype of each category can be represented by Equation (12).

$$c_k = \frac{1}{|S_k|} \sum_{(x, y \in S_k)} f_\phi(X_i) y_i \tag{12}$$

In Equation (12), $f_\phi(X_i)$ denotes the embedding function, S_k is the sample set containing k categories, $|S_k|$ denotes the

TABLE 1. Network structure parameters.

GCN structure parameters			Prototype network structure parameters		
Layer	Output Size	Convolution kernel size	Layer	Output Size	Convolution kernel size
Con_1	T×56×56	3×7×7,64	Con_0	64×46×46	3×3
Res_block1	T×56×56	3×3×3,64	Pool_0	64×23×23	2×2
Res_block2	T/2×28×28	3×3×3,128	Con_1	128×20×20	4×4
Res_block3	T/4×14×14	3×3×3,256	Pool_1	128×10×10	2×2
Res_block4	T/8×7×7	3×3×3,512	Con_2	64×8×8	3×3
/	/	/	Pool_2	64×4×4	2×2
/	/	/	Con_3	96×3×3	2×2

number of sample sets in k categories, X_i denotes the feature vector of the i th sample in S_k, and y_i denotes the label of the i th sample in the sample set S_k. The Euclidean distance between each class prototype is calculated in Equation (13).

$$d_k(i) = \sqrt{(x_i - c_i)^T M_k (x_i - c_i)} \tag{13}$$

In Equation (13), M_k denotes the inverse matrix of the covariance matrix of the category k. According to the Euclidean distance of the prototypes in each category, the study uses the softmax function to classify the micro-expressions, at which time the probability that the sample x belongs to the category k can be expressed in Equation (14).

$$p_\phi(y = k|x) = \frac{\exp(-d(f_\phi(x), c_k))}{\sum_k \exp(-d(f_\phi(x), c_k))} \tag{14}$$

In Equation (14), denotes the probability that the$x$ sample belongs to the category k. Finally, through the cross-entropy loss function, the network loss value can be calculated. The cross-entropy loss function only considers the correct label when calculating the network loss, and the non-correct label will be ignored, which leads to a decrease in the classification accuracy of network recognition. To address this problem, the study proposes an adaptive loss function that combines the Triplet loss function with the cross-entropy loss function. The Triplet loss function has high requirements for samples in the process of application and needs to consider the distance between different samples. In classification recognition, if we assume that the experiment selects a N image samples, each category in the training, randomly selected k sample pairs. The distance between N positive samples and $(K-1) \times N$ negative samples can be calculated according to Equation (15).

$$Loss_T = \sum_{i=1}^{N} \sum_{a=1}^{N} \left[m + \max\left(d\left(f_\theta\left(x_a^i\right)\right) - d\left(f_\theta\left(x_p^i\right)\right)\right)\right] \tag{15}$$

In Equation (15), $f_\theta(x_a^i)$ denotes positive samples; $f_\theta\left(x_p^i\right)$ denotes negative samples; $\max\left(d\left(f_\theta(x_a^i)\right) - d\left(f_\theta\left(x_p^i\right)\right)\right)$ denotes the maximum distance between positive and negative samples; and m denotes the hyperparameters of the function. The basic principle of the Triplet loss function is to compute the distance between the samples, and because of this, when the function is used as a loss function of the MER network, it is necessary to reduce the distance between the samples of the same category, and at the same time, to increase the distance between samples of different categories, but the distance between different samples must not be too large. samples must not be too large.

IV. ANALYSIS OF THE EFFECT OF MICRO-EXPRESSION FEATURE EXTRACTION RECOGNITION ANALYSIS BASED ON FAU-GCN

In Chapter 2, the study proposes a micro-expression feature extraction and recognition algorithm based on FAU-GCNN, so the main content of Chapter 3 is the validation of the algorithm, including the parameter setting of the algorithm, the performance comparison of the algorithm, and the analysis of the algorithm's feature extraction and recognition effect.

A. EXPERIMENTAL PARAMETER SETTINGS

The data sources for the study were CASME II dataset and SAMM dataset, 33 samples of pleasure, 25 samples of disgust, 27 samples of surprise and 60 samples of repression in CASME II dataset and 24 samples of pleasure, 13 samples of disgust, 8 samples of surprise and 20 samples of repression in SAMM dataset. The real-time operating system is windows 10, the CPU is Intel(R) Core(TM) i5-11800H @ 2.30GHz, the GPU is NVIDIA GeForce RTX 3070 Laptop, the RAM is 16GB, the video memory is 8GB, and the data analysis platform is Python 3.6 and PyTorch 1.8.0. The algorithm has a Batch size of 40, a learning rate of 0.0001, a training period of 100, a Dropout of 0.3, and a Momentum of 0.9. The parameter settings for the GCN and the prototype network are shown in Table 1.

The GCN model designed for research consists of five residual blocks and two stacked layers. The length of the input image sequence is T, and the resolution of the input image is 112 × 112. After pooling, the input image features are mapped to a size of 512 × 1. In this model, the output dimension of the first stack layer is 1024, and the output dimension of the second stack layer is 512. The prototype network consists of four convolutional layers, three pooling layers, and two fully connected layers. The convolutional

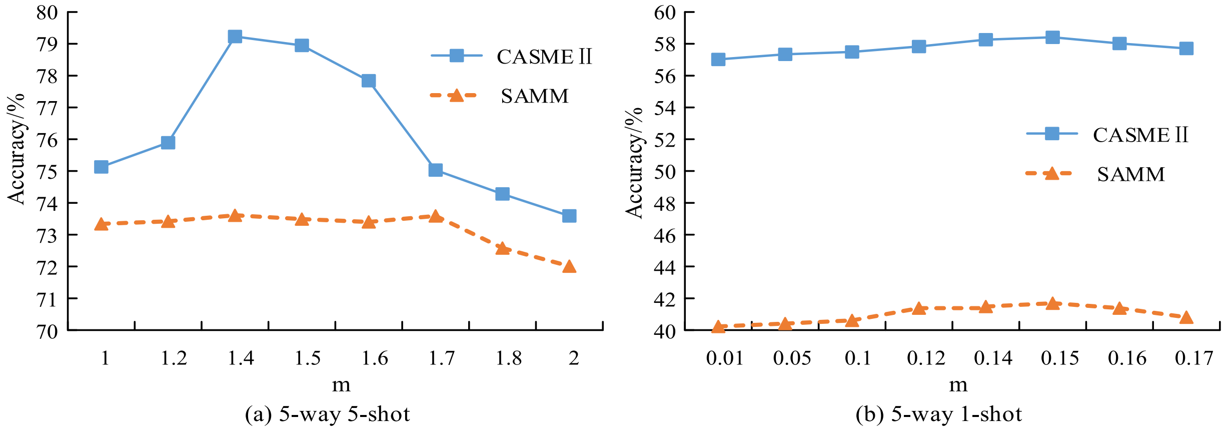

FIGURE 7. The impact of *m* on model performance during 5-way 5-shot.

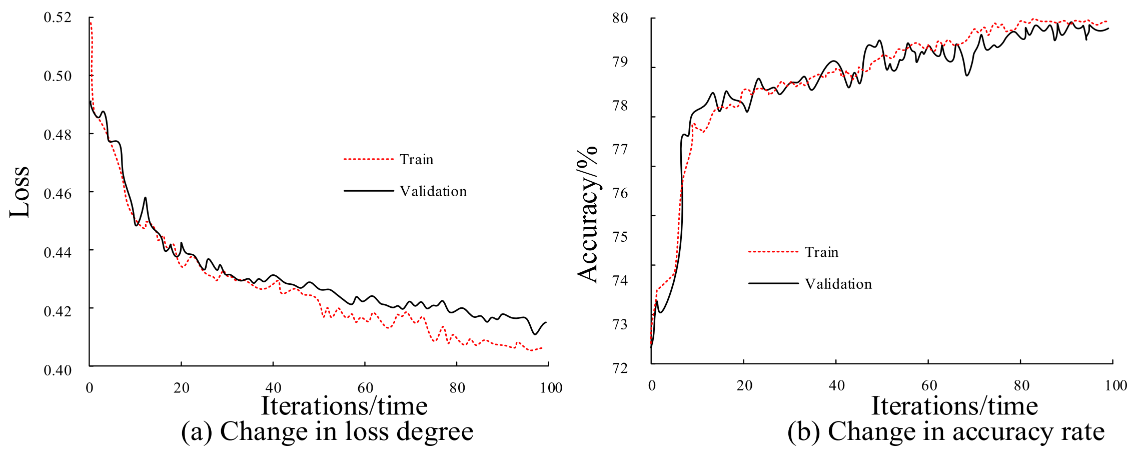

FIGURE 8. The change results of loss and accuracy of the 5-way 5-shot model on the CASMEII dataset.

layer has a step size of 1 and the pooling layer has a step size of 2. When conducting validation analysis on micro expression recognition models, recognition accuracy is the main indicator. The accuracy is the ratio of the correct samples in the micro expression classification results to the total classification samples, and the correctness of the classification results is tested by the experimental personnel. Samples with incorrect classification will be added to the model training set as training data to improve the accuracy of the model's micro expression recognition.

B. ALGORITHM HYPERPARAMETER SETTINGS

The loss function used in the prototype network constructed by the study is different from the ordinary loss function, the loss function used in the prototype network constructed by the study is an adaptive hybrid loss function, which has the hyperparameter m that will have an effect on the performance of the network, so the study compares the effect of m on the 5-way 5-shot network under different data, and the results are shown in Figure 7.

Figure 7(a) shows, the results of the effect of m on the performance of the model on the CASMEII dataset. The accuracy of the model on the CASMEII dataset takes the maximum value of 79.168% when the model is set to 5-way 5-shot and m is taken as 1.4. The accuracy of the model on the SAMM dataset takes the maximum value of 73.569% when the model is set to 5-way 5-shot and m takes 1.4. Figure 7(b) shows, the results of the effect of m on the model performance on the SAMM dataset. The accuracy of the model on CASMEII dataset is taken to the maximum value of 58.486% when the model is set to 5-way 1-shot and m is taken to 0.15. The accuracy of the model on the SAMM dataset takes the maximum value of 41.683% when the model is set to 5-way 1-shot and m is taken as 0.15. The results of the variation of the loss degree versus the correctness of the 5-way 5-shot model on the CASMEII dataset are shown in Figure 8.

Figure 8(a) shows the change in loss degree of the 5-way 5-shot model on the CASMEII dataset. The loss value of the model decreases as the number of iterations increases, and the training loss value reaches about 0.52 at one point in the

TABLE 2. Comparison results of recognition accuracy and F1 scores.

Method	SAMM		CASME II	
	F1	Accuracy	F1	Accuracy
FAU-GCN	0.718	0.738	0.748	0.795
DSSN	0.469	0.577	0.731	0.706
3D-FCNN	0.545	0.558	0.566	0.584
GTCN	0.698	0.751	0.724	0.743
MER-GCN	/	/	/	0.428
SRCN	0.711	0.766	0.722	0.785
Deep Coral	/	/	/	0.753

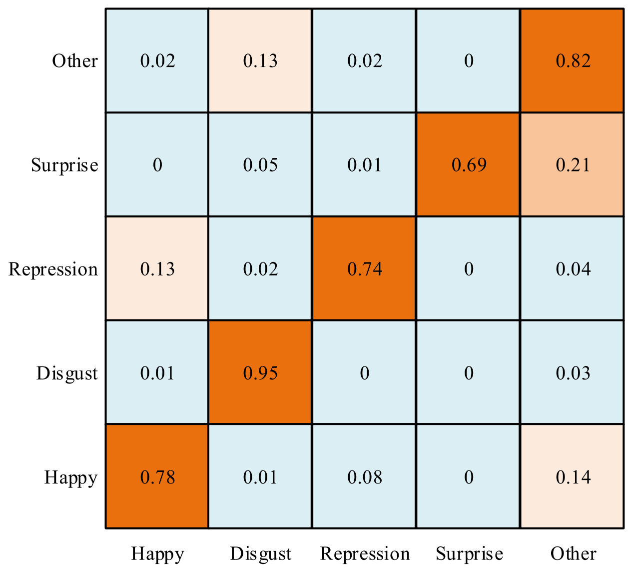

FIGURE 9. Figure 9 Confusion matrix of micro-expression recognition accuracy.

early stage of the model run. After 20 iterations, the loss value of the algorithm decreases dramatically, and the training loss value decreases from 0.5186 to 0.4317. After 100 iterations, the training loss value of the model decreases to around 0.40, while the model validation loss value is higher compared to the training set. Figure 8(b) shows the accuracy change of the 5-way 5-shot model on CASMEII dataset. at the beginning of the model run, the training and validation accuracy is only about 72%, after 10 iterations, the model accuracy increases dramatically, and after 100 iterations, the model's training and validation accuracies are around 80%. Based on the above results, the study plotted the classification accuracy confusion matrix during model MER with the CSAME II dataset, as shown in Figure 9.

In Figure 9, the dark area of the matrix is the accuracy of micro-expression recognition, and the light area is the probability that the category of micro-expressions is recognized as other categories. It can be seen that among the above five types of expressions, disgust micro-expressions have the highest recognition accuracy of 0.92, and surprise micro-expressions have the lowest recognition accuracy of only 0.69. Surprise micro-expressions are difficult to be recognized as happy, disgusted, or repressed expressions, but are extremely easy to be recognized as other expressions.

C. FEATURE EXTRACTION AND EXPRESSION RECOGNITION ANALYSIS

Deep CORAL algorithm is currently a common MER algorithm, the study compares the changes in the recognition accuracy of the two algorithms' feature extraction under different representative regions on CASME II dataset and SAMM dataset, as shown in Figure 10.

Figure 10(a) shows the effect of the number of representative regions on the accuracy of the FAU-GCN algorithm. The accuracy of the FAU-GCN algorithm is only 50.6% on the SAMM dataset and 53.2% on CASME II when the number of representative regions for feature recognition is only 1. FAU-GCN has the highest accuracy when the number of representative regions for feature extraction is 5, and at this time the accuracy is 75.2% on the SAMM dataset and 79.3% on the CASME II dataset. Figure 10(b) shows the effect of the number of representative regions on the accuracy of the Deep CORAL algorithm. Similar to Figure 10(a), the Deep CORAL algorithm has the highest accuracy when the feature recognition representative regions are four. The study also compares the F1 scores and accuracy of other existing methods on CASME II and SAMM datasets, which include, Dual-stream Shallow Networks (DSSN), 3D-Facial Convolutional Neural Network (3D-FCNN), Graph Temporal Convolutional Neural Networks (GTCN), Micro-expression Recognition-Graph Convolutional Networks (MER- GCN), and Spatiotemporal Recurrent Convolutional Neural Networks (SRCN). The comparison results of each model are shown in Table 2.

In Table 2, the proposed FAU-GCN algorithm of the study has an accuracy of 0.795 on the CASME II dataset, with an F1 score of 0.748, and the rest of the algorithms have the highest accuracy of 0.785, and the highest F1 score of 0.724. On the SAMM dataset, the accuracy of the FAU-GCN algorithm has an accuracy of 0.738, with an F1 score of 0.718, and the rest of the algorithms have the highest accuracy of 0.766 and the highest F1 score of 0.711. On the SAMM dataset, the FAU-GCN algorithm does not have the highest accuracy, but it has the highest F1 score, and the F1 scores that have higher accuracy than the FAU-GCN algorithm are lower. To verify the practicality of the proposed model, a human-machine

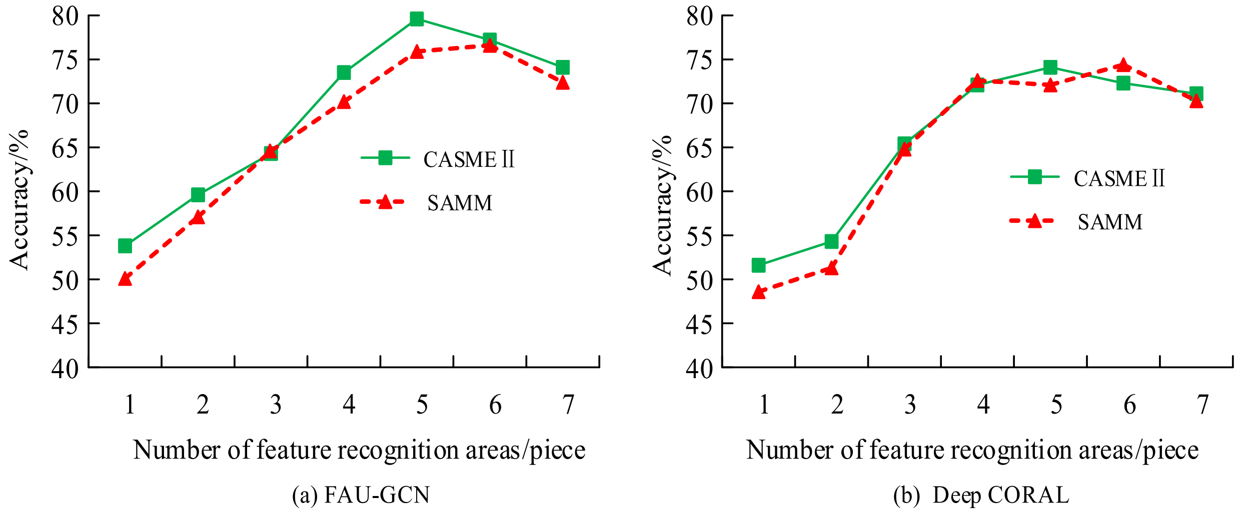

FIGURE 10. Effect of the number of representative regions on the accuracy rate.

TABLE 3. Practical application results analysis.

Number of experiments (times)	3D-FCNN			FAU-GCN		
	Number of emotions (times)	Number of detections (times)	Accuracy (%)	Number of emotions (times)	Number of detections (times)	Accuracy (%)
1	20	15	87.3	20	19	96.3
2	20	16	85.2	20	20	95.5
3	20	14	86.5	20	18	96.6
4	20	16	86.6	20	19	97.5
5	20	16	87.4	20	19	96.3

interaction scenario was constructed, with a duration of 10 minutes. The study conducted 10 repeated experiments, recording the number of times the model detected the experimenter's micro expressions and the classification accuracy of the experimenter's micro expressions in each experiment. In this experiment, the recognition effects of 3D-FCNN and FAU-GCN on facial micro expressions of experimenters were compared, and the results are shown in Table 3.

As shown in Table 3, the 3D-FCNN model has poor performance in recognizing micro expressions of experimental personnel in practical application scenarios of human-computer interaction. About 20% of micro expressions are not recognized and detected. And the model also has poor accuracy in classifying detected micro expressions, with a maximum of only 87.4%. The FAU-GCN model can achieve a maximum recognition accuracy of 100% for facial micro expressions of experimenters, and the model also has a high classification accuracy for micro expressions. The accuracy of micro expression classification in all five experiments has reached over 95%.

V. CONCLUSION

To realize the feature extraction and recognition of characters' micro-expressions, the study proposes a GCN model incorporating FAU feature extraction, which divides the characters' facial expression features into seven representative regions and, with the OFM, characterizes the characters' facial expressions as a way to strengthen the feature extraction capability of GCN. The findings denoted that the proposed FAU-GCN algorithm has an accuracy of 0.795 and an F1 score of 0.748 on the CASME II dataset, while the remaining algorithms have the highest accuracy of 0.785 and the highest F1 score of 0.724. The FAU-GCN has the highest accuracy when the number of representative regions of the feature extraction is five, at which point the SAMM dataset the accuracy was 75.2% and 79.3% on the CASME II dataset. Surprise micro-expressions have the lowest recognition accuracy of only 0.69. Surprise micro-expressions are difficult to be recognized as happy, disgusted, or repressed expressions, but are very easy to be recognized as other expressions. The model is set to 5-way 5-shot and when 1.4 is taken, the accuracy of the model is taken to a maximum of 79.168% on CASMEII dataset and 73.569% on SAMM dataset. The proposed algorithm of the study performs well in MER and outperforms multiple existing machine learning and deep learning algorithms. However, GCNs are computationally intensive, which affects the execution efficiency of the algorithm when processing large-scale datasets, and the model, in practice, fails to complete the continuous microexpression recognition task separated by intervals within 5s. Future work will attempt to use a more efficient CNN model to reduce the computational workload of the model and improve the task processing efficiency of the model.

REFERENCES

[1] T. Ahmad, L. Jin, X. Zhang, S. Lai, G. Tang, and L. Lin, "Graph convolutional neural network for human action recognition: A comprehensive survey," *IEEE Trans. Artif. Intell.*, vol. 2, no. 2, pp. 128–145, Apr. 2021, doi: 10.1109/TAI.2021.3076974.

[2] M. R. Mahmood, M. B. Abdulrazaq, S. R. M. Zeebaree, A. K. Ibrahim, R. R. Zebari, and H. I. Dino, "Classification techniques' performance evaluation for facial expression recognition," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 21, no. 2, p. 1176, Feb. 2020, doi: 10.11591/ijeecs.v21.i2.pp1176-1184.

[3] M. R. Ali, T. Myers, E. Wagner, H. Ratnu, E. R. Dorsey, and E. Hoque, "Facial expressions can detect Parkinson's disease: Preliminary evidence from videos collected online," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–4, Sep. 2021, doi: 10.1038/s41746-021-00502-8.

[4] A. V. Savchenko, L. V. Savchenko, and I. Makarov, "Classifying emotions and engagement in online learning based on a single facial expression recognition neural network," *IEEE Trans. Affect. Comput.*, vol. 13, no. 4, pp. 2132–2143, Oct. 2022, doi: 10.1109/TAFFC.2022.3188390.

[5] Y. Fang, B. Luo, T. Zhao, D. He, B. Jiang, and Q. Liu, "ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting," *CAAI Trans. Intell. Technol.*, vol. 7, no. 4, pp. 744–757, Nov. 2022, doi: 10.1049/cit2.12145.

[6] H. Ge, Z. Zhu, Y. Dai, B. Wang, and X. Wu, "Facial expression recognition based on deep learning," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, Art. no. 106621, doi: 10.1016/j.cmpb.2022.106621.

[7] P. Liu, Y. Lin, Z. Meng, L. Lu, W. Deng, J. T. Zhou, and Y. Yang, "Point adversarial self-mining: A simple method for facial expression recognition," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12649–12660, Dec. 2022, doi: 10.1109/TCYB.2021.3085744.

[8] G. Muhammad and M. S. Hossain, "Emotion recognition for cognitive edge computing using deep learning," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16894–16901, Dec. 2021, doi: 10.1109/JIOT.2021.3058587.

[9] A. S. Cowen, D. Keltner, F. Schroff, B. Jou, H. Adam, and G. Prasad, "Sixteen facial expressions occur in similar contexts worldwide," *Nature*, vol. 589, no. 7841, pp. 251–257, Dec. 2020, doi: 10.1038/s41586-020-3037-7.

[10] S. Li, W. Li, S. Wen, K. Shi, Y. Yang, P. Zhou, and T. Huang, "Auto-FERNet: A facial expression recognition network with architecture search," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2213–2222, Jul. 2021, doi: 10.1109/TNSE.2021.3083739.

[11] T. Chen, T. Pu, H. Wu, Y. Xie, L. Liu, and L. Lin, "Cross-domain facial expression recognition: A unified evaluation benchmark and adversarial graph learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9887–9903, Dec. 2022, doi: 10.1109/TPAMI.2021.3131222.

[12] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng, and Y.-J. Liu, "Video-based facial micro-expression analysis: A survey of datasets, features and algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5826–5846, Sep. 2022, doi: 10.1109/TPAMI.2021.3067464.

[13] R. Levie, W. Huang, L. Bucci, M. Bronstein, and G. Kutyniok, "Transferability of spectral graph convolutional neural networks," *J. Mach. Learn. Res.*, vol. 22, no. 1, pp. 12462–12520, Jan. 2021.

[14] T. Zhao, Y. Hu, L. R. Valsdottir, T. Zang, and J. Peng, "Identifying drug–target interactions based on graph convolutional network and deep neural network," *Briefings Bioinf.*, vol. 22, no. 2, pp. 2141–2150, Mar. 2021, doi: 10.1093/bib/bbaa044.

[15] H. Tong, R. C. Qiu, D. Zhang, H. Yang, Q. Ding, and X. Shi, "Detection and classification of transmission line transient faults based on graph convolutional neural network," *CSEE J. Power Energy Syst.*, vol. 7, no. 3, pp. 456–471, May 2021, doi: 10.17775/CSEEJPES.2020.04970.

[16] J. Y. Choi, P. Zhang, K. Mehta, A. Blanchard, and M. Lupo Pasini, "Scalable training of graph convolutional neural networks for fast and accurate predictions of HOMO-LUMO gap in molecules," *J. Cheminformatics*, vol. 14, no. 1, pp. 1–10, Oct. 2022, doi: 10.1186/s13321-022-00652-1.

[17] T. Nguyen, G. T. T. Nguyen, T. Nguyen, and D.-H. Le, "Graph convolutional networks for drug response prediction," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 19, no. 1, pp. 146–154, Jan. 2022, doi: 10.1109/TCBB.2021.3060430.

[18] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, "Impact of deep learning approaches on facial expression recognition in healthcare industries," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5619–5627, Aug. 2022, doi: 10.1109/TII.2022.3141400.

[19] K. Kottursamy, "A review on finding efficient approach to detect customer emotion analysis using deep learning analysis," *J. Trends Comput. Sci. Smart Technol.*, vol. 3, no. 2, pp. 95–113, Jul. 2021, doi: 10.36548/jtcsst.2021.2.003.

[20] S. Zhao, G. Jia, J. Yang, G. Ding, and K. Keutzer, "Emotion recognition from multiple modalities: Fundamentals and methodologies," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 59–73, Nov. 2021, doi: 10.1109/MSP.2021.3106895.

[21] Y. Zhao and J. Xu, "A convolutional neural network for compound micro-expression recognition," *Sensors*, vol. 19, no. 24, p. 5553, Dec. 2019, doi: 10.3390/s19245553.

[22] H. Pan, H. Yang, L. Xie, and Z. Wang, "Multi-scale fusion visual attention network for facial micro-expression recognition," *Frontiers Neurosci.*, vol. 17, Jul. 2023, Art. no. 1216181, doi: 10.3389/fnins.2023.1216181.

[23] R. Soni and B. Mehta, "Diagnosis and prognosis of incipient faults and insulation status for asset management of power transformer using fuzzy logic controller & fuzzy clustering means," *Electr. Power Syst. Res.*, vol. 220, Jul. 2023, Art. no. 109256, doi: 10.1016/j.epsr.2023.109256.

XULIANG YANG received the bachelor's degree in management from the Anyang Normal College, in 2009, specializing in information management and information systems, and the master's degree in software engineering from Huazhong University of Science and Technology, in 2016. He has been an Associate Professor with Dongguan City University, since 2021. His research interests include artificial intelligence models and digital application research.

YONG FANG received the bachelor's degree in computer science and technology from Changjiang University, in 2005, and the master's degree in computer application from Wuhan University of Science and Technology, in 2008. He is currently an University Teacher with research interests include big data, cloud computing, and artificial intelligence.

C. RAGA RODOLFO JR. is currently with the Faculty College of Computing and Information Technologies, National University.