Group Non-Critical Behavior Recognition Based on Joint Attention Mechanism of Sensor Data and Semantic Domain

Chen Li, Baoluo Liu*
School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, 471023, China

***Abstract*—As science and technology continue to advance, sensor technology is being used in more and more industries. However, traditional methods have the problem of ignoring the semantic information of individual behavior and the correlation between individuals and groups. Based on this, the study proposes a new method for group behavior recognition. The process of feature extraction is performed on group behavior by collecting sensor data and combining a semantic domain joint attention mechanism. This is achieved through the construction of a recognition method based on a data domain and semantic domain joint attention mechanism, which enables the accurate identification of non-critical behaviors in the group. The findings showed that, when the group members are constant, the hybrid network based on a convolutional neural network and bi-directional long and short-term memory network improved the F1 by 0.2% and the accuracy by 0.19%. Moreover, the hybrid network combining graph neural network, bi-directional long and short-term memory network, and convolutional neural network improved results. In group behavior recognition, group relationship modeling based on graph convolutional network improved F1 by 0.17% and accuracy by 0.17% compared to the hybrid network, indicating that group relationship modeling better captures group interaction features and improves recognition. The method is highly effective in the field of group behavior recognition and is expected to provide a new idea for monitoring and managing group behavior in practical scenarios.**

Keywords—Sensor data; attention mechanisms; semantic domains; non-critical; group behavior

I. Introduction

A. Research Background

With the continuous development and application of intelligent technology, group behavior (GB) recognition is becoming increasingly important for understanding human social behavior patterns, intelligent security surveillance, intelligent traffic management, and other fields. In previous research, most of the work focuses on identifying critical behaviors in groups, and relatively little research has been done on the identification of non-critical behavior (NCB) in groups [1]. Traditional GB recognition methods are mainly based on sensor data (SD) or video data for analysis, but these methods tend to ignore the semantic information of individual behaviors (IBs) and the correlation between individuals and groups [2]. However, there is a complex semantic relationship between IBs and GBs, and the understanding of IBs can provide important clues for the inference of GBs. On the other hand, SD and semantic domain joint attention mechanism (SDJAM), as an innovative approach, combines the behavioral data of group members captured by sensors with semantic domains (SeD) information in order to improve the recognition of group NCB [3]. Due to the disparate sizes of GB data and the fact that input individual behavior data does not always align with the same group characteristics, non-critical individual behaviors can readily impede the recognition of GB, resulting in a reduction in the recognition rate of network models for GB. Joint attention mechanism (JAM) enables the model to selectively focus on the information that is more crucial to understanding GB by integrating the contextual meaning of SeD information with the individual characteristics of SD. This improves the model's comprehension of the relationship between IB and GB [4].

B. Research Methods and Significance

The study suggests a group NCB recognition technique based on SD and SDJAM in light of this. The method combines SD with SeD information, and effectively captures the correlation and semantic information between IBs and GBs through JAM, so as to realize the accurate recognition of NCBs in groups. The research aims to deeply explore the application of SD and SDJAM in group NCB recognition, and to provide new solutions and technical support for intelligent surveillance, security management and other fields. To better capture the correlation and semantic information of individual actions in GB and increase the identification accuracy of NCB, the research creatively mixes SD and SeD information through JAM.

C. Article Structure

There are six sections to this study. The background of the research, issues, and solutions related to group NCB recognition are covered in Section I. Section II reviews the previous research results on group NCB recognition and summarizes the difficulties and shortcomings of the method. Section III describes the improvement of the SD method by combining SDJAM. The study's suggested strategy is tested against alternative approaches in a comparative experiment designed in Section IV. Discussion is given in Section V and finally, Section VI concludes the paper.

II. Related Work

GB is an important carrier in promoting economic development and social and cultural exchanges. Therefore, the research on GB recognition is a very important task and has triggered many scholars to study it. Lu et al. suggested a graph neural network model based on multimodality combined with semantic context-awareness to address the problems of single mode and ignoring labeling relationships in GB recognition method. The method used an aggregator that fuses attention to refine the nodes in the model, which effectively improves the robustness of the recognition of group activities [5]. To solve the issue that individual differences and other factors can easily influence GB recognition, Tang et al. suggested a residual aggregation network based on group photos. Moreover, the study also designed a weighted aggregation strategy to recognize multilevel spatio-temporal features, which can provide a comprehensive characterization of group activities and effectively infer group activities [6]. Challa et al. proposed a hybrid recognition method that combines bi-directional long and short-term memory (BLSTM) and convolutional neural networks (CNN) using filters of different sizes to capture a variety of temporal local dependencies, which helps to improve the process of group feature extraction. This helps to address the issues of difficult feature extraction and large data bias in GB recognition [7]. To address the issue of balancing resource consumption and accuracy in human activity recognition, Tang et al. proposed an improved deep CNN that uses segmentation. This improved deep CNN is able to capture a wider range of human activity sensory fields in a single feature layer, improving the ability of multi-scale feature representation and improving the recognition effect [8].

In GB recognition, the use of sensors is usually effective in obtaining human movement information, which is of great help in behavior recognition, therefore, many scholars at home and abroad focus on sensors for behavior recognition. Teng et al. proposed a layer-wise CNN with local loss in order to address the shortcomings in sensor-based human activity recognition. This CNN combines wearable sensors with local loss and exhibits superior recognition performance when compared to global loss, offering new approaches to human activity recognition [9]. In order to overcome the lack of cross-interaction between various dimensions of sensor-acquired data in behavior recognition, Tang et al. proposed a triple cross-dimensional attention method [10]. The study conducted this by establishing three attention branches to capture the cross-interaction between sensor dimension, time dimension, and channel dimension, thereby improving the recognition effect. Abdel-Basset et al. proposed a supervised two-channel model to address the importance of wearable devices in behavioral recognition. The study introduced adaptive channel filtering operations to enhance the behavioral feature extraction capability, so as to effectively take into account the spatio-temporal information and improve the recognition rate [11]. For the heterogeneity of data from multimodal sensors and various human activities, Islam et al. proposed a guided multimodal fusion method based on cooperative multitask learning. This method was able to extract complementary multimodal representations, offering an efficient and better solution for behavioral recognition in real-world environments [12].

In summary, scholars around the world have conducted different studies on GB recognition and solved many problems in the recognition process. However, there are fewer existing studies on the interference of NCB on GB recognition for groups of varying sizes, and the presence of interfering behaviors makes it difficult to improve the recognition accuracy. As a result, the study suggests using SDJAM in conjunction with group NCB identification based on SD. First, the GB is detected using SD, and then the behavioral characteristics of key characters are obtained through the construction of a hybrid network model and the introduction of attention mechanism. Second, the joint SeD and data domain networks are explored to suppress NCB features and achieve the prediction of GB features. Finally, training tests were conducted using individual behavior datasets and group sizes to compare the performance of different recognition methods, in order to demonstrate the effectiveness of group NCB recognition methods and achieve technical support for intelligent surveillance and crowd management.

III. Methods for Recognizing Group Non-Critical Behaviors

Firstly, for the GB recognition problem, the study is based on SD to recognize the GB recognition problem. The study employs a hybrid network model of CNN and BLSTM to obtain IB features that contain group relationships through graph convolutional networks (GCN) and inference of GBs through attentional fusion. The study then considers how the method has an impact on the learning of attention weights, and further proposes a data-domain and SDJAM based method for NCB suppression in groups.

A. GCN-Based Group Behavior Recognition Method

GB recognition is a method of inferring the characteristics and intentions of people by observing and analyzing their collective behavior. Information about an individual or a group can be obtained by observing people's actions, language, facial expressions, etc. in different scenarios. Facing the problem that GB is difficult to characterize, the research proposed a GB recognition method based on GCN and group relationship modeling [13]. Fig. 1 shows the GB recognition process.

In Fig. 1, this GB recognition process first receives continuous SDs collected by different individual sensors and performs data segmentation through a sliding window. Using a hybrid CNN and BLSTM network, the segmented sequences are utilized to extract the spatiotemporal behavioral aspects of the individuals. In contrast to conventional approaches, this method mines the behavioral correlations and location correlations of people to compute IB associations while taking into account the interaction among individuals within a group [14]. Then, the individual spatio-temporal features and group relationship maps are inputted into the GCN to capture the behavioral interactions under global activities. In order to acquire the recognition results of GB, the features of various persons are finally fused through the attention module (A-Mod), and Softmax classification is carried out by the fully connected layer (FCL). Considering that the behavioral data captured by the sensors are time-series data, it is more suitable to be processed using recurrent neural network (RNN) to better characterize the dynamics of the behaviors. In order to extract well-characterized IB features, the study uses a hybrid BLSTM network and CNN to extract temporal and spatial behavioral features, and the flowchart of this process is shown in Fig. 2.

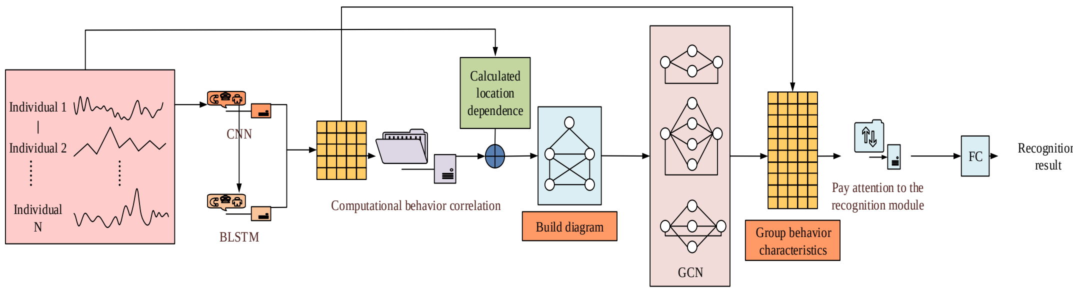


Fig. 1. Flow chart of group behavior recognition.

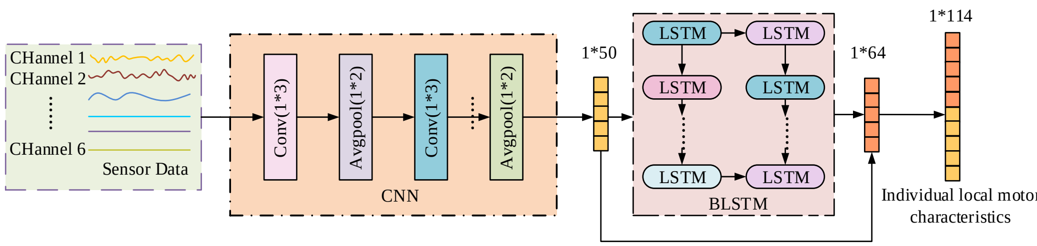


Fig. 2. Flow chart of extracting individual local features by CNN and BLSTM.

In Fig. 2, SD is convolved and pooled by CNN to get high-dimensional feature vectors, keeping the temporal order, and then input to BLSTM to extract its temporal features [15]. The feature vectors undergo an average pooling layer to obtain saliency features, which are then spliced with CNN features to generate individual features. For SDs at different local body locations, different CNN and BLSTM networks are trained independently to obtain more representative local features, which are then spliced into complete individual features. The flow of individual local features spliced into individual feature maps is shown in Fig. 3.

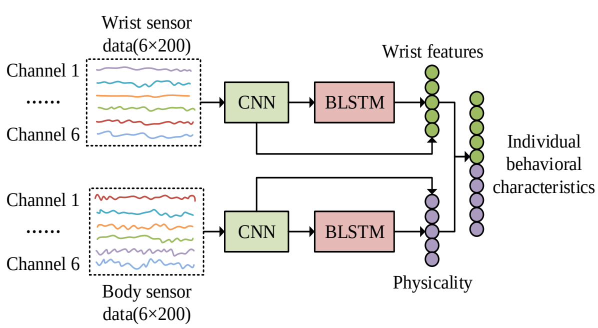


Fig. 3. Individual local features are spliced into individual feature maps.

In Fig. 3, individual local feature splicing refers to combining or connecting the local features of an object or an individual in some way to form an overall feature representation of that individual [16]. First, IB correlation is measured by analyzing the similarity of behavioral features of individuals over time. To simplify the process, the study uses the behavioral feature vectors within a specific sliding window to calculate the behavioral correlation of individuals i and j within the k th sliding window by using Eq. (1).

$$\begin{cases} \mathrm{SimFea}_{i,j}^{k} = \mathrm{Corr}_{i,j}^{k}\left(F_i^k, F_j^k\right) \\ \mathrm{Corr}_{i,j}^{k} = \dfrac{\mathrm{Cov}\left(F_i^k, F_j^k\right)}{\sqrt{D\left(F_i^k\right)}\sqrt{D\left(F_j^k\right)}} \end{cases} \tag{1}$$

In Eq. (1), F_j^k and F_j^k denote the feature vectors within the sliding window of individuals i and j. $D\left(F_i^k\right)$ denotes the standard deviation of F_j^k, and $\sqrt{D\left(F_j^k\right)}$ denotes the standard deviation of F_j^k. Using this method, the behavioral feature vectors of two individuals are converted to behavioral relevance measures [17]. To further standardize the correlation metric, the study maps the correlations to the same scale. The formula is displayed in Eq. (2).

$$\mathrm{Norm}_{i,j}^{k} = f\left(\mathrm{Corr}_{i,j}^{k}\right) \tag{2}$$

The study further calculates the Euclidean distance between the two bodies in order to obtain the positional relationship matrix, which is calculated mathematically in Eq. (3).

$$d_{ij}(t) = \sqrt{\left(x_i^t - x_j^t\right)^2 + \left(y_i^t - y_j^t\right)^2} \tag{3}$$

In Eq. (3), $d_{ij}(t)$ represents the distance between the two individuals at the moment t, and when this distance exceeds a certain value, then there will be no interaction between the two. In this study, the study set this value to 5. The distance d_{ij}^k in the k th sliding window represents the average value of $d_{ij}(t)$ during this period of time. Considering that the degree of interaction of an individual is inversely proportional to the distance, the study further measured the positional relationship between the individuals i and j to ensure that the individual has the maximum positional weight with itself. The mathematical formula for measuring the individuals is shown in Eq. (4).

$$\mathrm{Sim}\,Loc_{ij}^{k} = 1 / e^{d_{ij}^{k}} \tag{4}$$

In Eq. (4), e denotes a natural constant. Then through Eq. (5), the IB feature correlation is fused with the location relationship to obtain the individual interaction relationship vector.

$$G_{ij}^{k} = \lambda \cdot \mathrm{SimFea}_{i,j}^{k} + (1-\lambda) \cdot \mathrm{SimLoc}_{i,}^{k} \tag{5}$$

In Eq. (5), λ denotes the hyperparameter. G_{ij}^{k} denotes the pairwise interaction between individuals. $F_{per,i}$ is the behavioral characteristics of individual i. In GB recognition, IBs and the relationship between them have an impact on GB. In order to introduce GCN into GB recognition, the study uses two different relationship modeling methods, behavioral correlation graph and location correlation graph [18-19]. The fusion of these two graphs forms the final group relationship graph. The creation of graphs that are related to location and behavior facilitates the capturing of the features of various interpersonal connections. Fig. 4 illustrates the construction of the relationship graph.

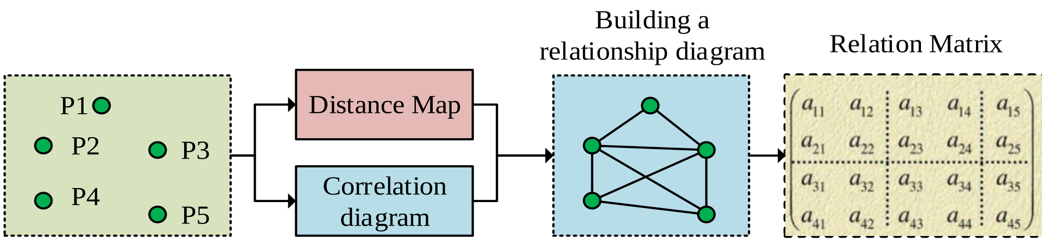


Fig. 4. Builds the diagram.

In Fig. 4, the nodes represent the behavioral characteristics of the individuals, while the edges represent the weights of the interactions between the individuals. The distance and correlation graphs together go through the construction of the relationship graph to obtain a relationship matrix. For the input of a set of N individuals, the behavioral features of individual i are extracted by BLSTM and CNN, and the spatial location features are $D_{per,i}$. The study uses one layer of GCN, and the inputs are the group relationship graph and the N individual features $F_{per,i}$ [20]. Eq. (6) displays the formula for graph convolution.

$$F_{int} = \sigma\left(\hat{D}^{-1/2} G \hat{D}^{-1/2} F_{pern} W_{g}\right) \tag{6}$$

In Eq. (6), σ is the activation function and W_g denotes the weight matrix. To improve the hierarchy of behavioral information, the study inputs the group interaction features ($F_{\mathrm{int},n}$) recovered by the GCN into the A-Mod after splicing the IB features ($F_{per,n}$) extracted by the hybrid CNN and BLSTM network in the spatiotemporal domain. In the A-Mod, the study calculates the correlation between individual feature vectors and normalizes it using the softmax function. Eq. (1) of the attention calculation is shown in Eq. (7).

$$Q^{d} = W_{q}^{d} \times \left[F_{per,n}, F_{int,n}\right] \tag{7}$$

In Eq. (7), σ denotes the weight matrix to be learned, and Eq. (2) for attention calculation is shown in Eq. (8).

$$K^{d} = W_{k}^{d} \times \left[F_{per,n}, F_{int,n}\right] \tag{8}$$

In Eq. (8), σ denotes the matrix to be learned, K_n^d denotes the matrix, and $(K_n^d)^T$ denotes the transpose of matrix K_n^d. Eq. (2) of the attention calculation is shown in Eq. (9).

$$\begin{cases} V^{d} = W_{v}^{d} \times \left[F_{per,n}, F_{int,n}\right] \\ s_{n}^{d} = Q^{d}\left(K_{n}^{d}\right)^{T} \\ \alpha_{n}^{d} = \exp\left(s_{n}^{d}\right) / \sum_{j=1}^{N} \exp\left(s_{j}^{d}\right) \\ v^{d} = \sum_{n=1}^{N} \alpha_{n}^{d} \cdot V^{d} \end{cases} \tag{9}$$

In Eq. (9), V^d is obtained by weighted fusion of attention, and V^d is maximally pooled to obtain F_{gro} for describing the GB feature. F_{gro} is computed as shown in Eq. (10).

$$F_{gro} = \max v^{d} \tag{10}$$

The study further takes the GB feature F_{gro} as an input to the FCL, which is processed through the FCL to finally obtain the GB recognition result. The corresponding mathematical expression is shown in Eq. (11).

$$\mathrm{output} = W_{f} \times F_{gro} + b_{f} \tag{11}$$

In Eq. (11), W_f and b_f are the weight coefficients and deviations of the feature vectors in the FCL, respectively, and output represents the output after the FCL. After processing through the FCL, the vectors are operated by the Softmax function, and the prediction results are normalized to obtain the prediction probability. The specific mathematical expression of the Softmax function is shown in Eq. (12).

$$\sigma(\mathrm{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{12}$$

In Eq. (12), $\sigma(\mathrm{z})_j$ and z_j denote the size of the probability of the j th category and the size of the fully connected output, respectively. $\sum_{k=1}^{K} e^{z_k}$ is the value of the sum of all fully connected output probabilities.

B. Semantic Domains Federated Attention Mechanism's Method for Suppressing Non-Critical Behaviors

The suggested approach makes use of a teacher-student network design, wherein the teacher network directs the student network to concentrate more successfully on important group members in order to find better attention weights. The network architecture of the GB recognition method based on preserving semantic attention is shown in Fig. 5.

In Fig. 5, the teacher-student network structure is introduced, and the teacher network, i.e., the SeD network, directly uses IB labels to learn a good attentional weight representation of IB semantics. In the data-domain network (DDN), SD is used as input to extract IB features through a hybrid network of CNN and BLSTM, and then GB features are obtained and classified through GCN and A-Mod aggregation. By integrating SeD and data domain network through loss function, attention knowledge and knowledge distillation techniques acquired from SeD are utilized to direct the learning of attention weights for IB characteristics in the data domain. The network simply considers the positional relationship between individuals and builds the graph using the distance relationship because the correlation of IB labels is not very high. The calculation formula is shown in Eq. (13).

$$O_s = \sigma\left(\hat{D}^{-1/2} A_s \hat{D}^{-1/2} F_{oh,n} W_s\right) \tag{13}$$

In Eq. (13), A_s is computed to obtain the degree matrix D, and W_s represents the weight matrix for SeD graph convolution. $F_{oh,n}$ undergoes graph convolution to obtain O_s. The O_s calculation formula is shown in Eq. (14).

$$Q^s = W_q^s \times F_{oh,n} \tag{14}$$

A weight distribution is learned through the self-attention mechanism. Eq. (15) illustrates how the attention is computed using the dot product in order to specifically focus on the important human behavioral traits that contribute to the final GB recognition.

$$\begin{cases} K^s = W_k^s \times F_{oh,n} \\ V^s = W_v^s \times F_{oh,n} \\ s_n^s = Q^s \left(K_n^s\right)^T \\ \alpha_n^s = \exp\left(s_n^s\right) / \sum_{j=1}^{N} \exp\left(s_j^n\right) \\ v^s = \sum_{n=1}^{N} \alpha_n^s \cdot V^s \end{cases} \tag{15}$$

The A-Mod yields the attention-weighted GB feature v^s, which is then processed via the Softmax activation function and FCL to predict the final GB label. In the study's suggested method, the SeD network and the data domain network are concurrently trained, in contrast to the conventional teacher-student network architecture. Semantically guided DDN training is achieved by learning all DDN parameters using a temporal backpropagation technique. The particular procedure is depicted in Fig. 6.

The study's suggested methodology is shown in Fig. 6, where two graph convolution modules are initially used to record the unique interaction information of features in the data and SeD domains, respectively. Then, the learned attentional knowledge of the SeD guides the attentional weights of the features in the data domain and aggregates them for final prediction.

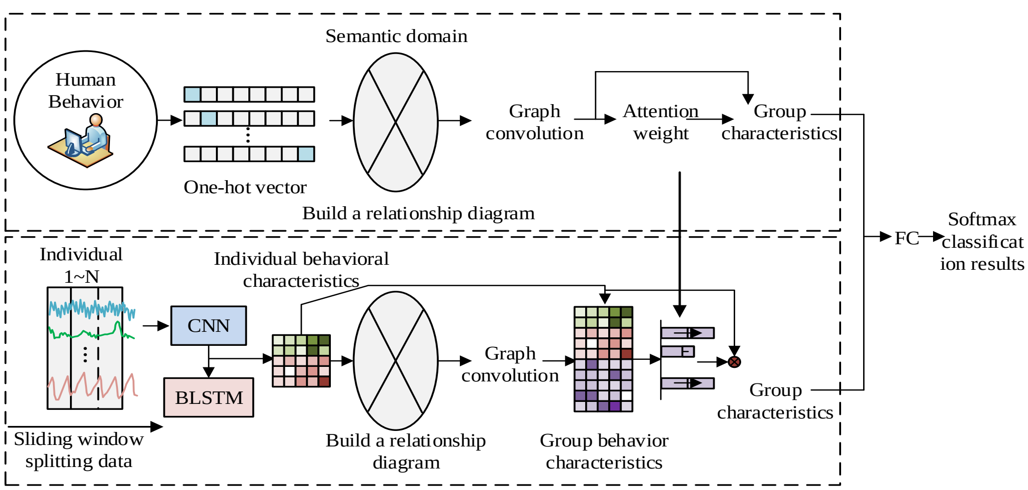


Fig. 5. Architecture of group behavior recognition method based on semantic attention retention.

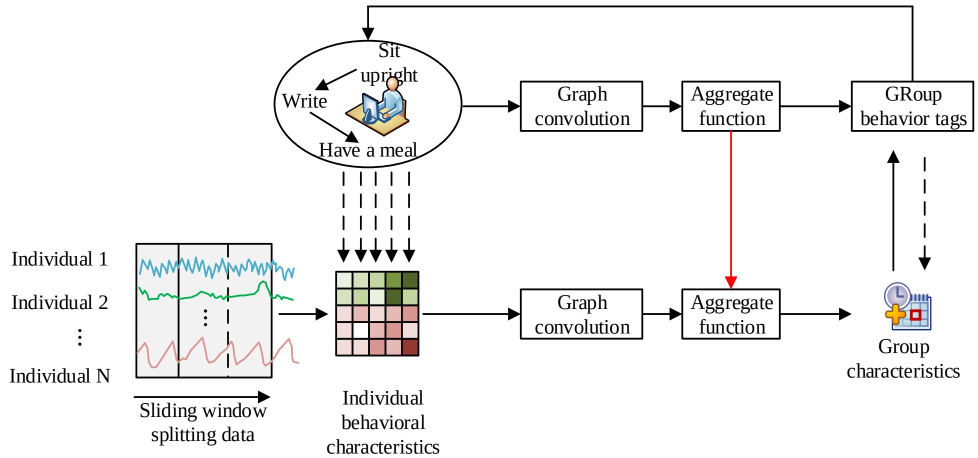


Fig. 6. The semantic domain guides the training process of the data domain.

IV. Performance Analysis of Group Non-Critical Behaviors Recognition Methods

The study sets up a comparison experiment in this chapter to compare the suggested technique with CNN BLSTMGCN, CNN+BLSTM, and CNN+BLSTM+GCN methods in order to assess the effectiveness and performance of the suggested approach. This is done in order to confirm the suggested method's efficacy.

A. Performance Validation Of GCN-Based Group Behavior Recognition Method

For this study, Python 3.7 is used as the development tool, Pytorch 1.4.0 is used as the deep learning framework, and Ubuntu 16.04 is served as the operating system. The research employs identical hyper-parameter configurations, network architectures, and experimental feature extraction tactics, with learning rates of 0.001, weight decays of 0.001, training batch sizes of 32, cross-entropy loss functions, Adam optimization algorithm, and thirty training iterations. Based on a publicly available IB dataset, the study creates a sophisticated GB dataset. The dataset is sourced from the UT Data public dataset of complex individual behaviors at the University of Twente, which mainly includes six types of human interaction behaviors, with a total of 20 samples. All participants in the dataset test carry two cell phones, which are put on the left wrist and the left lower body pocket for data collection. The samples of the final created GB dataset are organized in each sliding window and contains 200 rows of data. Table I details the categories and composition of the GBs in the dataset.

In Table I, the dataset contains 10 GBs and 10 IBs, with each group containing up to five individuals. A sliding window segmentation process yields a sequence of 1780 GB samples. The GBs have a duration of six minutes and a sampling frequency of 50 Hz. 80% of the data are used as a training set and 20% are utilized as a test set in fifty-fold cross-validation. All of the group members' data is the input for fixed group sizes. Data on randomly chosen individuals between the ages of two and five is the input for random group size. The confusion matrix is used to calculate accuracy, precision, recall, and F1 score. The results are averaged using a 5-fold (10 times per fold) cross-validation to reduce experimental error and chance. To minimize error and chance, the mean and standard deviation of the 5-fold average findings of the 10 experiments are finally compared. Table II displays a comprehensive comparison of the outcomes.

TABLE I. Group Behavior Composition

Tag	**Category**	**Makeup**	**Test set**	**Training set**
0	Walk	Walking 5 people	36	142
1	Professional	Typing 3, writing 2	36	142
2	Canter	Jogger 5	35	143
3	Rest	Sit for 3, drink coffee for 2	36	142
4	Queue up	Standing 5	36	142
5	Smoking	5 smokers	36	142
6	Lecture	One speaker, two sitting, two writing	36	142
7	Dine together	Dinner for 3, coffee for 2	36	142
8	Examination	Standing one, writing four	35	143
9	Discuss	Lecture 2, sit 2, type 1	36	142
Sample count	/	1780	36	142

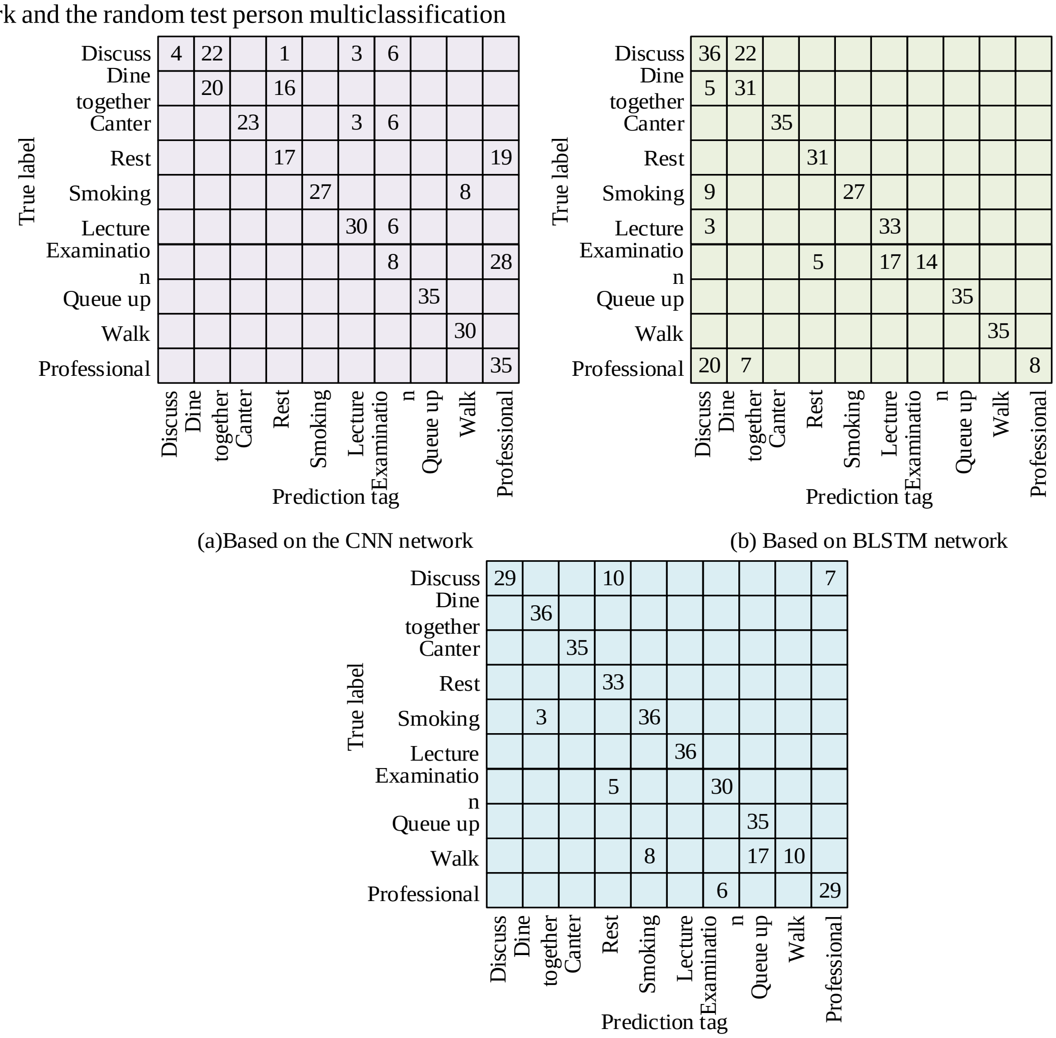
TABLE II. Comparison of Results

Method	**Recall rate (%)**	**F1 score (%)**	**Accuracy (%)**	**Accuracy rate (%)**
CNN	81.31±2.35	79.17±2.22	83.54±1.67	81.30±2.13
CNN+BLSTM+GCN	99.63±0.25	99.63±0.25	99.66±0.21	99.63±0.23
GCN	99.59±0.17	99.54±0.18	99.60±0.14	99.59±0.17
BLSTM	99.41±0.75	99.38±0.80	99.56±0.41	99.40±0.68
CNN+BLSTM	99.44±0.52	99.43±0.62	99.54±0.41	99.44±0.59
Research method	99.80±0.08	99.80±0.09	99.82±0.08	99.80±0.08

In Table II, when the number of groups is held constant, the model based on the hybrid network structure of CNN, BLSTM, and GCN demonstrates an improvement of 0.2% in F1 and 0.19% in accuracy in comparison to the model based on the hybrid network of CNN and BLSTM. This suggests that the addition of GCN is successful. In the GB recognition method, group relationship modeling based on GCN improves F1 by 0.17% and accuracy by 0.17% compared to the method based on the hybrid network structure of CNN, BLSTM and GCN. To verify the robustness of the network, a test with randomized group size is performed. During training, each group contained 5 people. However, in the test sample, the number of people in each group varied randomly between 2 and 5 people. The experimental setup is the same as when fixing 5 people. Confusion matrix results for some of the methods at the same folds are shown in Fig. 7.

In Fig. 7, Fig. 7(a) shows the random test person multiclassification mixing matrix based on CNN network, while Fig. 7(b) and Fig. 7(c) show the random test person multiclassification mixing and smoothing matrix based on BLSTM network and the random test person multiclassification mixing and smoothing matrix based on GCN and group relationship modeling, respectively. According to Fig. 7, more recognition errors are observed in the test case with random number of people in the group compared to the case with fixed five people in the group. This chapter does a 5-fold (10 times each) cross-validation to minimize experimental error and unpredictability. The precision, recall, F1 score, and accuracy of the average findings are computed. Table III displays the comparison of each method's mean and standard deviation.

Table III demonstrates that, when comparing the test case of a random number of groups to the fixed number of groups in the test set, all methods exhibited lower classification results. Additionally, the volatility of the results per fold increases significantly, suggesting that the GB recognition results are impacted by the change in group size. In terms of classification results, the GB recognition method based on GCN for group relationship modeling shows the highest accuracy, which is 0.84% higher in F1 score and 1.08% higher in accuracy than the model based on the hybrid network structure of CNN, BLSTM and GCN.



(a)Based on the CNN network

(b) Based on BLSTM network

(c) Modeling based on graph convolutional networks and group relations

Fig. 7. Confusion matrix results of some methods under the same fold.

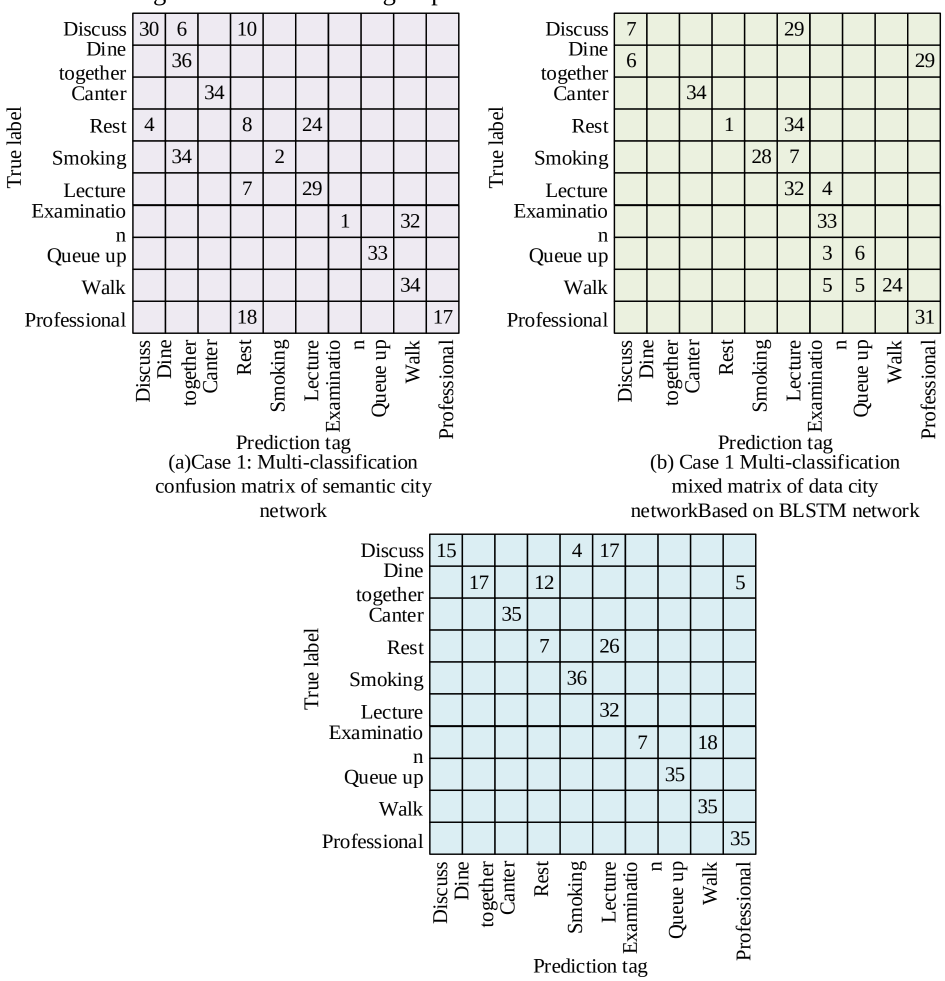
TABLE III. COMPARISON RESULTS OF THE AVERAGE VALUE AND QUASI DIFFERENCE OF EACH METHOD WHEN THE NUMBER OF PEOPLE IS RANDOM

Method	Recall rate (%)	F1 score (%)	Accuracy (%)	Accuracy rate (%)
CNN	64.78±2.09	61.63±1.50	64.77±1.92	69.05±2.07
GCN	77.05±2.96	76.52±3.20	77.06±2.67	81.82±4.00
CNN+BLSTM+GCN	82.20±1.25	81.40±1.21	82.18±1.08	83.17±1.29
CNN+BLSTM	82.04±0.71	80.65±0.60	82.06±0.72	86.86±0.70
BLSTM	80.47±2.45	79.61±2.64	80.46±2.18	88.76±0.82
Research method	83.25±0.37	82.24±0.40	83.26±0.39	87.88±0.71

B. Performance Analysis of Non-Critical Behaviors Suppression Methods for Semantic Domains Federated Attention Mechanism

Experiments are conducted using data from the same group of five people for training, and the effectiveness of the method is tested with a fixed group of five people and a randomized number of people to verify its ability to suppress NCBs in the group, respectively. The study analyzes the recognition effectiveness of three methods: the SeD network, the data domain network, and the data domain and SeD combined attention mechanism network. Fig. 8 demonstrates the confusion matrix under one-fold validation for the three methods when tested in the first case.

The multiclassification confusion matrix for the SeD network in case one is illustrated in Fig. 8(a). The multiclassification confusion matrix for the DDN, data-domain, and SDJAM networks in case one is presented in Fig. 8(b) and 8(c), respectively. The case one test in Fig. 8 shows that the inclusion of one or two NCBs in the input has the greatest impact on the data domain network. In contrast, NCBs had less of an effect on the data domain of the SDJAM network and the SeD network. The confusion matrices for the three approaches under one-fold validation in the second case test are shown in Fig. 9.

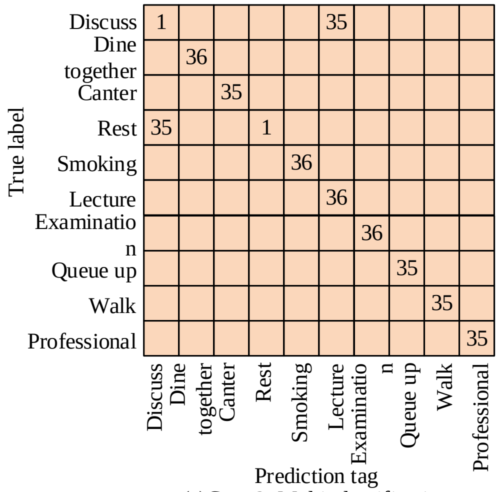


(a)Case 1: Multi-classification confusion matrix of semantic city network

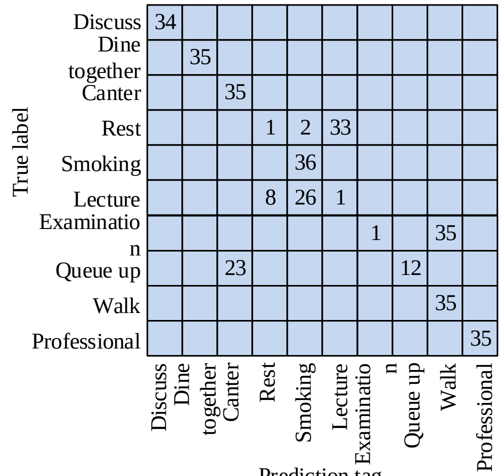
(b) Case 1 Multi-classification mixed matrix of data city networkBased on BLSTM network

(c) Case 1: Multi-classification confusion matrix of joint attention mechanism network between data domain and semantic domain

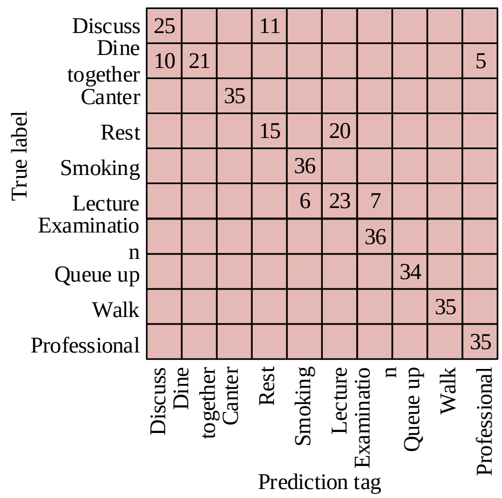
Fig. 8. Multi-class confusion matrix of semantic joint attention mechanism network.



(a)Case 2: Multi-classification confusion matrix of semantic city network



(b) Case 2 Multi-classification mixed matrix of data city networkBased on BLSTM network



(c) Case 2: Multi-classification confusion matrix of joint attention mechanism network between data domain and semantic domain

Fig. 9. The confusion matrix of the three methods tested in the second case is verified by one fold.

Fig. 9(c) displays the multiclassification confusion matrix for the SDJAM network under case II, while Fig. 9(a) and Fig. 9(b) display the multiclassification confusion matrices for the SeD network and the data domain network under case II. In Fig. 9, the SeD network has problems in misidentifying discussion as speech and misidentifying rest as discussion, while the other behaviors are classified well. The data domain network, on the other hand, has some errors in misidentifying breaks as speeches, and the classification confusion is more complex. In contrast, the data-domain and SDJAM networks performed more accurately in classifying each behavior, with improved overall correctness. In the end, it helps to lower the experimental inaccuracy and chance in the 5-fold findings by comparing the mean and standard deviation of the 5-fold average results of 10 studies. The suggested approach can still be processed well, and its recognition impact is superior to that of the data domain network and the single SeD network, according to the comparison results of the three ways in two test scenarios. The accuracy has increased by 18.75% and 21.96% when compared to the data domain technique using the suggested method, clearly demonstrating that the SeD network can effectively direct the learning of the data domain's attentional weights to deal with interference brought about by NCBs.

V. Discussion

A non-critical GB detection method based on sensors and GCN networks is studied and constructed for GB detection methods. In complex environments, GB characteristics cannot represent the behavior characteristics of key individuals or groups. Therefore, the collection of behavior data by sensors requires effective processing of global activity behavior interactions to ensure the extraction of the relationship between individual characteristics and GB. The construction of relationship graphs and the calculation of attention enable network models to accurately extract and predict key behavioral features. This study employs network structures of teachers and students to demonstrate the relationship between individual and GB. Additionally, it encompasses SeD networks and data domain networks for feature classification and joint training. Due to the excessive doping of NCB information in GB scenarios, this study employs semantic-guided data domain networks and graph convolution modules to extract key individual interaction information, suppress the representation of NCB, and achieve effective behavior recognition in group interactions. In the validation testing of public data, the accuracy of mixed network models for cross validation of fixed populations was above 99%, while the performance of models in random populations was significantly reduced. The accuracy, recall, and accuracy of CNN networks were reduced by about 16.5% compared to the former. However, the method proposed by the research still achieved a recognition rate of over 82% for GB. In conclusion, the network method proposed by the research, which combines SD and SeD information with a JAM, is capable of accurately extracting GB features and effectively suppressing NCB features in complex GBs.

VI. Conclusion

With the development of digitization, sensor technology has become one of the important tools for GB research. This research is to optimize the recognition method of NCB in the population. By constructing JAM, SD was combined with SeD in order to realize the accurate recognition of various types of NCB in the population. The outcomes of this study indicated that the GB recognition method based on GCN for group relationship modeling exhibited the highest accuracy, with 0.84% higher F1 score and 1.08% higher accuracy compared to the hybrid network. The impact of NCB on the SeD network and the data domain with the SDJAM network was small, and the research-proposed method was still able to handle it effectively. When compared to a single SeD network and a data domain network, the accuracy increased by 18.75% and 21.96%, respectively, demonstrating that the SeD network can successfully direct the learning of data domain attention weights to manage NCB interference. The results reveal that the proposed method of the study achieves significant performance improvement in GB analysis and provides a new technical path for the field of GB research. It is worth noting that for the complexity of GB, the current model still has some limitations in NCB recognition. In addition, the diversity of practical application scenarios is not fully considered in the study, and the future development direction can focus on applying the model to different environments and conducting more extensive empirical studies.

References

[1] Zulfiqar S, Khan M S. Organizational identification and knowledge sharing behavior: Mediating role of organiz

[2] ational citizenship behavior and moderating role of collectivism and leader–member exchange. Knowledge and Process Management, 2021, 28(4): 388-398.

[3] Groumpos P P. A Critical Historic Overview of Artificial Intelligence: Issues, Challenges, Opportunities, and Threats. Artificial Intelligence and Applications. 2023, 1(4): 197-213.

[4] Mokayed, H, Quan, T. Z., Alkhaled, L., & Sivakumar, V. Real-time human detection and counting system using deep learning computer vision techniques. Artificial Intelligence and Applications. 2023, 1(4): 221-229.

[5] Saminu S, Xu G, Zhang S, Kader IAE, Aliyu HA, Jabire AH, Ahmed YK, Adamu MJ. Applications of Artificial Intelligence in Automatic Detection of Epileptic Seizures Using EEG Signals: A Review. Artificial Intelligence and Applications, 2023,1(1): 11-25.

[6] Lu L, Lu Y, Yu R, Di H., Zhang L., Wang S. Gaim: Graph attention interaction model for collective activity recognition. IEEE Transactions on Multimedia, 2019, 22(2): 524-539.

[7] Tang Y, Lu J, Wang Z, Yang M., Zhou J. Learning semantics-preserving attention and contextual interaction for group activity recognition. IEEE Transactions on Image Processing, 2019, 28(10): 4997-5012.

[8] Challa S K, Kumar A, Semwal V B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. The Visual Computer, 2022, 38(12): 4095-4109.

[9] Tang Y, Zhang L, Min F, He J. Multiscale deep feature learning for human activity recognition using wearable sensors. IEEE Transactions on Industrial Electronics, 2022, 70(2): 2106-2116.

[10] Teng Q, Wang K, Zhang L, et al. The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition. IEEE Sensors Journal, 2020, 20(13): 7265-7274.

[11] Tang Y, Zhang L, Teng Q, Min F, Song A. Triple cross-domain attention on human activity recognition using wearable sensors. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 6(5): 1167-1176.

[12] Abdel-Basset M, Hawash H, Chakrabortty R K, Ryan M, Elhoseny M., Song H. ST-DeepHAR: Deep learning model for human activity recognition in IoHT applications. IEEE Internet of Things Journal, 2020, 8(6): 4969-4979.

[13] Islam M M, Iqbal T. Mumu: cooperative multitask learning-based guided multimodal fusion. Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(1): 1043-1051.

[14] Wang X, Cheng M, Eaton J. Fake node attacks on s. Journal of Computational and Cognitive Engineering, 2022, 1(4): 165-173.

[15] Oslund S, Washington C, So A, et al. Multiview Robust Adversarial Stickers for Arbitrary Objects in the Physical World. Journal of Computational and Cognitive Engineering, 2022, 1(4): 152-158.

[16] Niu J Y, Xie Z H, Li Y, Cheng S. J, Fan J. W.Scale fusion light CNN for hyperspectral face recognition with knowledge distillation and attention mechanism. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2022, 52(6): 6181-6195.

[17] Dongbo L I, Huang L. Reweighted sparse principal component analysis algorithm and its application in face recognition. Journal of Computer Applications, 2020, 40(3): 717-722.

[18] Khayam K N, Mehmood Z, Chaudhry H N, et al. Local-Tetra-Patterns for Face Recognition Encoded on Spatial Pyramid Matching. 2022, 0(3): 5039-5058.

[19] Fan Y. Face recognition algorithm of sprinters based on sliding data camera measurement. International Journal of Reasoning-based Intelligent Systems, 2023, 15(1): 79-85.

[20] Nam V H, Huong N M, Cuong P. Masked face recognition with convolutional neural networks and local binary patterns. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2022, 52(5): 5497-5512.

[21] Cao P, Zhu Z, Wang Z, Zhu Y, Niu Q. Applications of graph convolutional networks in computer vision. Neural Computing and Applications, 2022, 34(16):13387-13405.