

Research Article

Social Group Behavior Analysis Model Integrating Multitask Learning and Convolutional Neural Network

Chuan Zhou ^{1,2} **Suying Gui**^{1,3} **Gaoming Zhang** ² **Qirui Zhang**³ **Xudong Wang**⁴
and **Jianqing Wei**²

¹School of Microelectronics, Tianjin University, Tianjin 300072, China

²The Second Institute of China Aerospace Science and Technology Corporation, Beijing 100039, China

³Polytechnic Institute, Zhejiang University, Hangzhou, 310015 Zhejiang, China

⁴Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100195, China

Correspondence should be addressed to Chuan Zhou; zhouchuan@tju.edu.cn

Received 12 January 2021; Revised 9 March 2021; Accepted 28 July 2021; Published 15 August 2021

Academic Editor: Hasan Khattak

Copyright © 2021 Chuan Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Social group behavior analysis has always been the key research direction of sociologists and psychologists. With the rapid development of the Internet of Things and the proposal of deep learning theory, convolutional neural networks are also used in social group research. In the process of social development, group incidents continue to increase, and there are more and more studies on social group behavior analysis. Although the research content and research methods are also richer, the research that combines the Internet of Things, convolutional neural network, and group behavior is more. This article will specifically propose a social group behavior analysis model that combines multitask learning and convolutional neural networks. This paper deeply learns the research of convolutional neural network and group behavior-related theories, makes full use of the advantages of convolutional neural network algorithm and multitask learning mode, and builds a social group behavior analysis model based on multitask learning and convolutional neural network. The experimental results on different data sets are analyzed. The results show that the accuracy rate of the experimental algorithm of convolutional neural network is as high as 95.10%, and it is better than other algorithms in time complexity, which is very suitable for social group behavior analysis.

1. Introduction

In recent years, with the development of the economy, the adjustment of economic structure, and the acceleration of urbanization, various social collective actions have also increased, and their development has become more and more complex, which has had a huge impact on the construction of today's society and moral culture. Various group theories are becoming more and more abundant and perfect, exploring and explaining group processes or group dynamics and discussions of related issues. As an object of study, the group integrates multitask learning and convolutional neural networks and allows for more group behavior characteristics to be fully considered under the Internet's big data.

Multitask learning and convolutional neural network are both promising fields, and they are the development exten-

sion of deep learning. The role of multitask learning is to extract feature information to help improve learning more accurately. Its biggest characteristic is parallel transfer learning, which is different from the traditional progressive and procedural learning methods. In multitask learning, multitask learning is also called parallel migration learning because information is shared between tasks and is transferred between different tasks. It can realize the sharing and transmission of information between tasks. Convolutional neural network has the same characteristics as it; that is, it can classify the network data information by translation. It has the representation learning ability of artificial intelligence. At the same time, it can also be called "translation invariant artificial neural network."

In this paper, Mishkin et al. systematically studied the impact of a series of recent developments in the structure

and learning methods of convolutional neural networks (CNN) on object classification (ILSVRC). The evaluation tested the impact of the following architecture choices: non-linearity (ReLU, ELU, max-out, comparability of batch standardization), pool variables (random, maximum, average, mixed), network width, classifier design (convolution, fully connected, SPP), image preprocessing, and learning parameters: learning rate, batch size, data cleanliness, etc. The performance gain of the proposed modification is first tested individually and then tested in combination. When all the modifications are introduced, the sum of individual returns is greater than the observed improvement, but the “deficit” is small, which indicates that their returns are independent. They show that the use of 128×128 pixel images is sufficient to make a qualitative conclusion on the best network structure of the full-size Caffe and VGG network. The result is an order of magnitude faster than the standard. However, CNN is not good at feature understanding [1]. Guo et al. have extensively studied traditional artificial methods and intelligence-based methods, which extract effective features from vibration data to perform high-precision classification and diagnosis of various mechanical faults, such as support vector machines and backpropagation neural networks. He researched and proposed a new layered learning rate adaptive deep convolutional neural network based on an improved algorithm and studied its application in bearing fault diagnosis and determining its severity. In order to test the effectiveness of the proposed method, experiments were carried out on the bearing failure data samples obtained from the test bench. This method has achieved satisfactory performance in both fault pattern recognition and fault size evaluation. However, the problem of maximizing accuracy under overly complex situations has not yet been solved [2]. Yan et al., in recent years, head pose estimation (HPE) based on low-resolution monitoring data has received increasing attention. However, the monocular and multiview HPE methods still have poor results under target motion, because when a person moves, the facial appearance will be distorted due to changes in the camera’s perspective and scale. To this end, they proposed a new multitask learning- (MTL-) based framework FEGA-MTL, which is used to classify the head posture of people moving freely in an environment monitored by multiple large-field surveillance cameras. After FEGA-MTL divides the surveillance scene into dense and uniform spatial grids, it also learns the head pose classifier of a specific area and divides the grid into areas with similar facial features. In the learning phase, FEGA-MTL uses two pictures as a guide to perform a priori modeling of (1) the meshing based on camera geometry and (2) the similarity between head pose classes to obtain the optimal scene division and related pose classifiers. However, it has not been able to completely solve the problems arising from the movement of the target [3].

The innovations of this article are as follows: (1) This article uses a combination of quantitative and qualitative methods, which is well reflected in the fourth part of this article, the convolutional neural network model; (2) This article uses a combination of theoretical analysis and empirical research. This method uses experimental data to explain the

problem while establishing the model for analysis. This method runs through this article.

2. Method of Social Group Behavior Analysis Model Integrating Multitask Learning and Convolutional Neural Network

2.1. Multitask Learning Theory. Multitask learning and applications are also very promising. In machine learning, the useful information of historical data is used to analyze future data [1]. Usually, a lot of labeled data is needed for the next purpose for excellent training. The deep learning model is a typical model of machine learning. Because the model is a neural network with multiple levels and multiple parameters, it usually requires millions of data samples to obtain the correct parameters, and a lot of manual operations are usually required to label the data, so this data requirement cannot be met. Under the background of such data problems, the solution of MTL is to extract the feature information that can be used in other related tasks, so as to make up for the sparse data [2].

MTL research topics include multiple useful information, so that every student can get the correct goal [3]. Assume that all work (at least part) is related. Based on this, experimental and theoretical, I understand that learning multiple tasks at the same time is more effective than individual learning. According to the nature of the work, MTL can be divided into multiple levels of monitoring learning, multitasking nonmonitoring learning, multitasking nonmonitoring learning, multitasking and semimonitoring learning distribution, multitasking learning distribution, other learning, and other settings. One job and multiple jobs learn multiple perspectives [4, 5].

Based on this parameter, MTSL uses model parameters to learn various tasks in relation to each other. Five different operation modes include low-rank method, task clustering method, task relationship learning method, dirty method, and multilevel method [6]. Specifically, because it is assumed that the tasks are related, the parameter matrix W is likely to be of low rank, which may be due to the low motivation of the W class. The purpose of the combination method is to divide the operations into multiple groups, on the premise that all operations of each complex have the same or similar parameter models. The way to learn working relationships is to learn working relationships directly from data. The premise of the dirty method is that the parameter table W can be decomposed into two subtables. Here, each matrix is normalized with various types [7]. The multilevel method is a popular form of the dirty method, which decomposes the parameter list into more than 3 component lists to simulate the complex relationship between all tasks.

2.2. Convolutional Neural Network Structure. Convolutional neural network is a typical algorithm in the deep learning module. It uses the convolution calculation method in mathematical calculations and a neural feedforward network with a deep convolution structure [8, 9]. Therefore, it is also called “translation invariant artificial neural network” [10]. Since the beginning of the 21st century, with the advancement of

deep learning theory and the continuous improvement of computer equipment, collective neural networks have developed rapidly and are widely used in computer vision, language processing, and other fields [11]. The main neural network is composed of input layer, hidden layer, and output layer. The hidden layer includes various convolutional layers and pooling layers [12, 13].

2.2.1. Input Layer. Its input layer is capable of multidimensional processing of input information [14]. A one-dimensional input layer can usually accept one to two-dimensional values, such as time or spectral data collection; a two-dimensional neural network can receive two- to three-dimensional arrays, and so on, to receive input of multidimensional data information. Aggregate neural networks are widely used in the field of computer vision, so three-dimensional input data needs to be introduced in many studies, that is, the level of two-dimensional pixel and RGB channel structure [15].

The input layer expression of convolution neural network is as follows:

$$x_{ij}^l = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab} y_{(i+a)(j+b)}^{l-1}. \quad (1)$$

2.2.2. Convolutional Layer. The effect of the aggregation layer is to draw characteristics from the enter data [16]. Each factor of the core assembly homologous the same weight quotient and departure as the neural supply network. Each neuron cell in the assembly layer is linked to many nerve cells neighboring to the front floor layer. The size of this area varies according to the size of the core [17]. It is called "receptive field" in the literature, which depends on the receiving area of visual cortex cells. When fusing the core work, periodically scan the input characteristics, multiply the matrix elements, summarize the possibility of entering the acceptance field, and apply deviations.

$$Z^{l+1}(i, j) \left[Z^l \otimes \omega^{l+1} \right] (i, j) + b \sum_{k=1}^{k_l} \sum_{x=1}^f \sum_{y=1}^f \left[Z_k^l(s_0 j + x, s_0 j + y) \omega_k^{l+1}(x, y) \right] + b, \quad (2)$$

$$(i, j) \in \{0, 1, \dots, L_{l+1}\} L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1. \quad (3)$$

The sum of the formulas is equal to the analysis of the relationship, cross-correlation [18]. b is the amount of deviation, Z^l and Z^{l+1} represent the total of the input and output of the $l + 1$ level (also referred to as a feature map), L_{l+1} and Z_{l+1} are the same total size as the length and width of the feature map. Corresponding to the pixels in the feature map $Z(i, j)$, K is the number of channels in the feature map. f , s_0 , and p are the parameters of the entire layer, corresponding to the size of the connection core, the length of the connection step, and the number of filling layers.

Especially when the size of the convolution kernel is $f = 1$, the step size is $s_0 = 1$, and the filling unit of the convolution

kernel is not included. The interconnection calculation of the convolutional layer is the same as matrix multiplication, and the matching layer is fully connected to the network as follows:

$$Z^{l+1} = \sum_{k=1}^{k_l} \sum_{i=1}^L \sum_{j=1}^L \left(Z_{i,j,k}^l \omega_k^{l+1} \right) + b = \omega_{l+1}^T Z_{l+1} b, L^{l+1} = L. \quad (4)$$

The convolutional layer parameters include the size of the convolution kernel, the stepping, and the size of the overlay layer. These are the size of the output map of the aggregation layer and the hyperparameters of the convolutional neural network. The combined kernel size can be determined to be any value smaller than the input image size. The larger the interconnected core, the more complex the exportable entry function.

The folding layer embody activation functions to give assistance to express sophisticated properties [11], and its representation is as follows:

$$A_{i,j,k}^l = f \left(Z_{i,j,k}^l \right). \quad (5)$$

After exporting the aggregate level features to the aggregate level, the output feature matching is moved to a concentrated level, the function is selected, and the information is filtered. The pooling layer contains the default pooling function [19]. The function of this function is to use the mapping of adjacent regions to create statistics. The steps of selecting the pooling area are the same as the mapping characteristics of the convolution kernel and are controlled by the concentration, step length, and filling.

Lp pool is a type of pooling model which is aroused by the hierarchical construction in the sense of sight cerebral cortex, and its form is as follows:

$$A_k^l(i, j) = \left[\sum_{x=1}^f \sum_{y=1}^f A_k^l(s_0 i + x, s_0 j + y)^p \right]^{1/p}. \quad (6)$$

In the formula, the step size s_0 and pixel (i, j) have the same meaning as the convolutional layer, and p is a prespecified parameter. When $p = 1$, Lp pooling takes the average value in the pooling area, which is called mean pooling; at that time, Lp pooling takes the maximum value in the area, which is called maximum pooling.

Random pooling and blending pooling are all development of the concept of pooled Lp. Random pooling stochastic chooses values within the pooling range on the basis of a specially appointed probability distribution, so that the specific nonmaximum signal stimulus enters the next stage of construction. Pooling concentration may be indicating a linear combination of maximum and average pooling.

$$A_k^l = \lambda L_1 \left(A_k^l \right) + L_\infty \left(A_k^l \right), \lambda \in [0, 1]. \quad (7)$$

According to research, compared with average pooling and maximum pooling, mixed pooling and random pooling have a normalization function, which is beneficial to avoid overconfiguration of the group neural network.

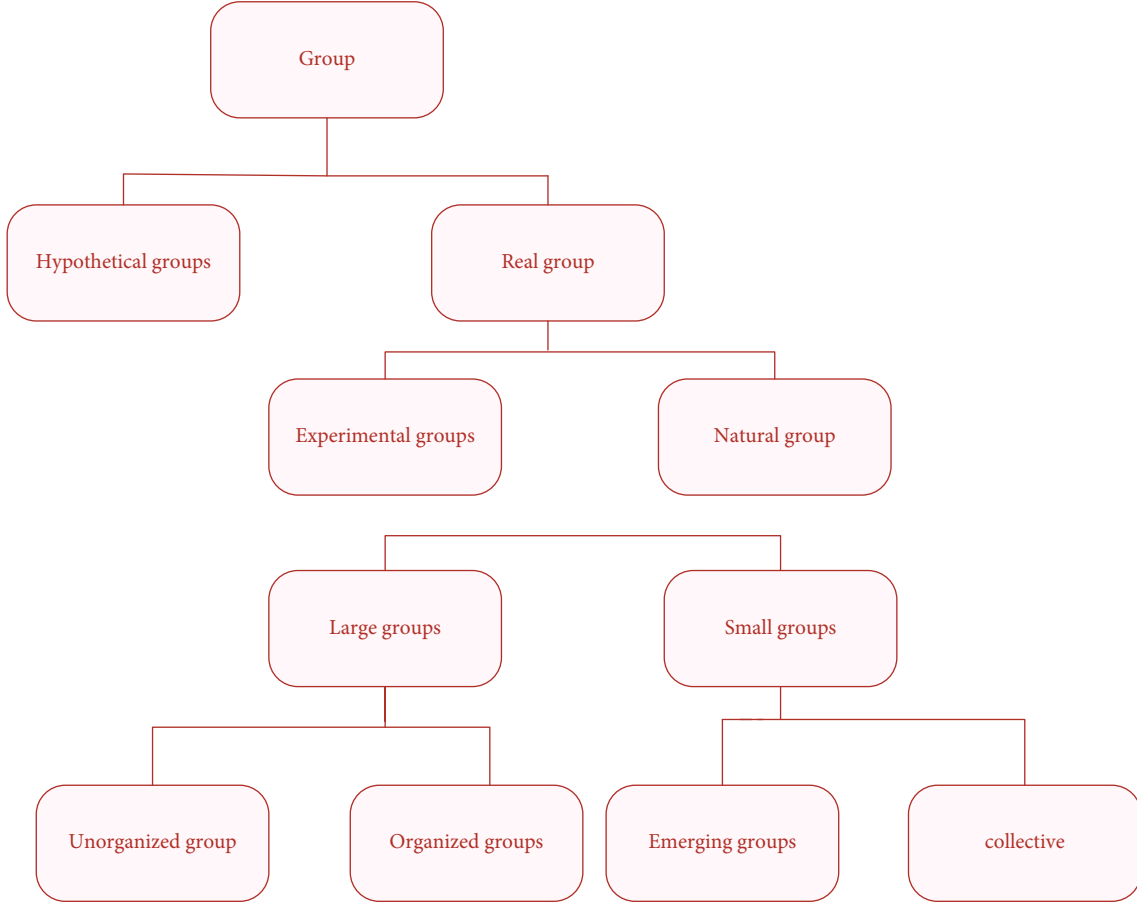


FIGURE 1: Structure of group theory.

2.2.3. *Output Layer.* In a convolutional neural network, the upper end of the output layer is generally a fully connected layer, which is the same as the traditional neural network algorithm. In the process of image classification, it is generally necessary to use mathematical logic functions or use normalization methods to add complete classification labels; in the field of object recognition, the design of the output layer is very different, which can be divided into coordinates, dimensions, etc.; in semantics segmentation, only pixel classification is required [20].

The output layer expression of convolution neural network is as follows:

$$y_{ij}^l = \sigma(x_{ij}^l). \quad (8)$$

If the input is represented by x_1, x_2, \dots, x_n , the detailed calculation process of convolution neural network is as follows:

$$\left\{ \begin{array}{l} o_i^{(l)} = x_i, l = 1 \\ O_i^{(l)} = (\omega_{ij}^{(l)}) * o_i^{(l-1)} + (b_i^{(l)}) \\ O_i^l = \sigma(o_i^l), 2 \leq l \leq n \end{array} \right\}. \quad (9)$$

2.3. *Sociological Theory of Group Behavior.* Sociologists and social psychologists have conducted in-depth research on group behavior very early and have reached systematic conclusions as shown in Figure 1. The French sociologist Gustave Le Pen proposed the infection theory in 1896. Regarding a group as an individual with a collective will, the ability to stimulate an individual group is incompatible; that is to say, gaining power from a simple person, infection, rapid spread of new ideas, and similar infections are three factors. It spreads between people and is easy to get infected. According to Mr. Rupe, in a group, people's thinking can easily be reduced to low-level activities. It is easy to accept the actions and attitudes of other people in the group and passively imitate [21].

In abnormal action theory, this theory was considered by the participants to be a group action caused by the participants' violation of social rules and the belief that there was no normal action and was praised by the participants. At the same time, as a member of a group, individual abnormal behavior can avoid severe punishment. Such a social environment provides a destructive environment for action, and many people will inevitably gather together to produce the same reaction, leading to an explosive increase in collective action [22].

In emergency theory, this theory believes that the actions of a group are generated for the purpose of discovering rules.

Someone guides their actions and unifies the actions of the entire group. Such rules are not general social rules, but they will be temporarily overwhelmed when the masses become emotional. These rules will guide people to deal with unexpected situations at the time. This theory denies the infection theory and believes that group actions occur based on the knowledge of participants. People recognize the code of conduct in emergency situations and replace the expansion of feelings with common sense [11].

In convergence theory, this theory believes that people in this group will see things the same and have the same tendency to act. This tendency also contacts [23]. In social comparative theory, this theory is the main benchmark for others to evaluate and verify their own abilities and opinions. This is a simple concept and an overall relationship of use within a group. Actions between social groups are more scientific.

3. Experiment on Social Group Behavior Analysis Model Integrating Multitask Learning and Convolutional Neural Network

3.1. Construction of a Social Group Behavior Analysis Model Based on Convolutional Neural Networks. The linear convolutional layer and the multilayer MLP receptor together form a mlpconv layer, and only a part of the input information in the range domain is needed to obtain the corresponding feature vector [24]. The process of acquiring the feature information of the target by the mlpconv layer is to use a large number of nonlinear activation functions to perform algorithmic calculations, to integrate the feature information and output it into a feature map, and then to perform the next feature input cycle process and then uses function mapping until it enters the next level.

One of the biggest advantages of convolutional neural network is that it has local characteristics. It can effectively separate in a complex background to obtain effective feature information of the target, and it can automatically complete multitask learning and deep learning goals and respond to background changes. The same is true for aggressive goals.

When training the plug-in network model, the weights of the mlpconv full neural network model including a single layer are first divided, and then, the entire neural network is trained. The entire PV update process ends. Then, connect to the second layer of mlpconv. The entrance of the second layer of mlpconv is the exit of the first layer of mlpconv. First, initialize the weights of the second layer of mlpconv and then train the entire neural network. After the entire training process is completed, lpconv will be notified of the second layer weight. When adding a new mlpconv level, perform weighted initialization, comprehensive neural network training, and weight update according to the above procedures.

In addition, the combination of convolution calculation and BN technology can make the nonlinear unit generate a relatively stable distribution, thereby achieving the combined effect. Add the BN operation to the nested mlpconv layer, and the calculation method of the feature map in the model

is as follows:

$$f_{i,j,k_1}^1 = \max \left(\text{BN} \left(W_{k_1}^{1T} x_{i,j} + b_{k_1} \right), 0 \right), \quad (10)$$

$$f_{i,j,k_2}^2 = \max \left(\text{BN} \left(W_{k_2}^{2T} f_{i,j}^1 + b_{k_2} \right), 0 \right), \quad (11)$$

$$f_{i,j,k_n}^n = \max \left(\text{BN} \left(W_{k_n}^{nT} f_{i,j}^{n-1} + b_{k_n} \right), 0 \right). \quad (12)$$

In formulas (10)–(12), BN (g) is the BN layer, the position of the pixel in the feature map is represented (i, j) , the input $x_{i,j}$ blocks with the pixel as the center (i, j) , etc., are the channel numbers in the feature map, and n is the number of MLP layers. Figure 2 is a flow chart of a convolutional neural network.

3.2. Pooling Model Design

3.2.1. Classic Pooling Model. The most commonly used in classic pooling models is average pooling and maximum pooling models. Average pooling is to take the average value on the calculation result of the algorithm and use it as the feature value in the subsampling; the maximum pooling is to take the maximum value to complete the pooling process.

The algorithmic expressions of average pooling and maximum pooling are as follows:

$$S_{ij} = \frac{1}{c^2} \left(\sum_{i=1}^c \sum_{j=1}^c F_{ij} \right) + b_2, \quad (13)$$

$$S_{ij} = \max_{i=1,j=1}^c (F_{ij}) + b_2. \quad (14)$$

In the above formula, F is the feature map matrix, the pooling area range is $c*c$, the offset is b_2 , and the final subsampling feature map is S .

3.2.2. Improved Intermediate Model. The classic pooling model has several disadvantages. For example, it cannot best extract the features of the target, and the use of the maximum value instead of the feature value will also have a certain weakening effect, which is not conducive to improving the accuracy of the model. Therefore, in view of the shortcomings of the classic pooling model, two improved models are proposed, namely, the maximum two-mean pooling method and the median pooling method.

The formula of the maximum two-mean pooling method is as follows:

$$S_{ij} = \frac{1}{2} \sum_{i=1,j=1}^c M2(F_{ij}) + b_2. \quad (15)$$

This formula can extract the two largest values from the pooling domain for summing.

The formula of the median pooling method is as follows:

$$S_{ij} = \frac{T}{2} + b_2, \quad (16)$$

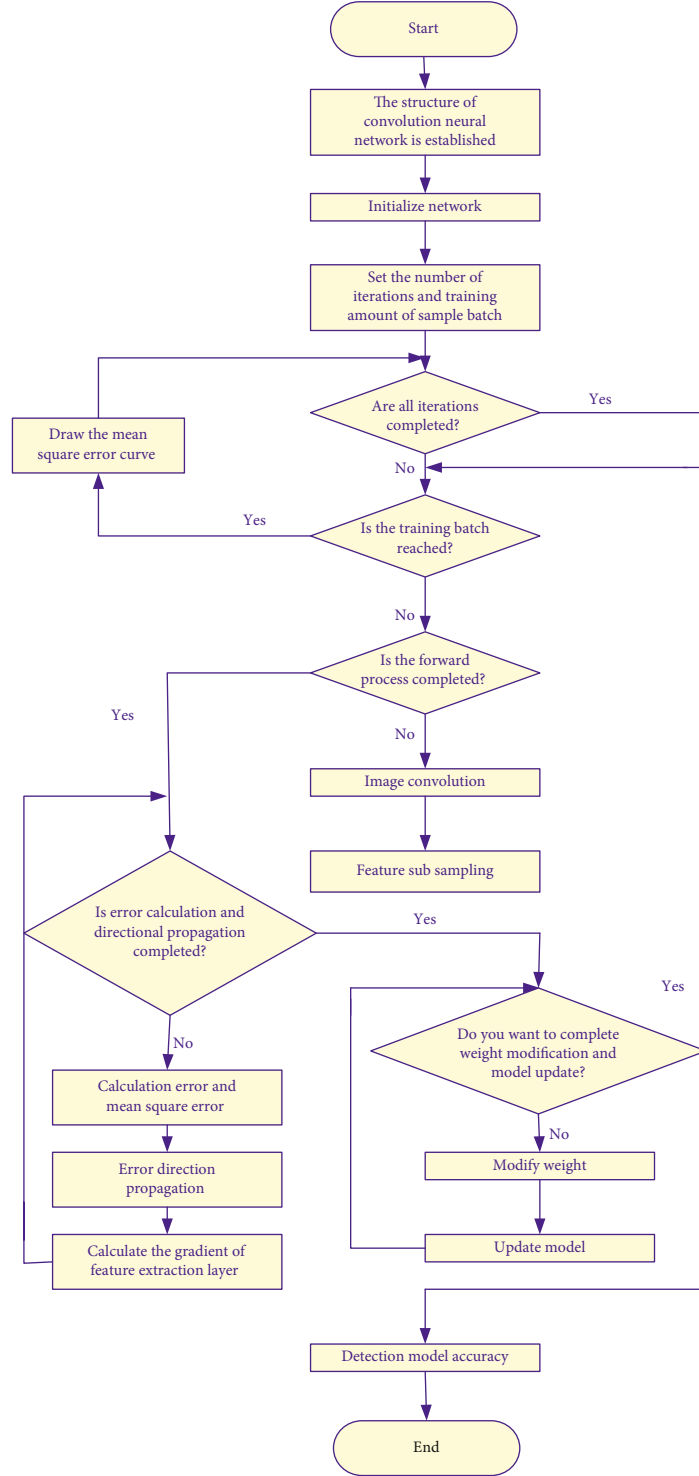


FIGURE 2: Flow chart of convolution neural network.

$$T = \frac{1}{c^2} \left(\sum_{i=1}^c \sum_{j=1}^c F_{ij} \right) + \max_{i=1,j=1}^c (F_{ij}). \quad (17)$$

This algorithm is a compromise algorithm, which can minimize the damage of model accuracy and reduce errors. It is suitable for general image algorithms.

3.2.3. Dynamic Adaptive Pooling Model. The purpose of optimizing output characteristics is to improve output pooling. In the whole process of learning neural network union, various feature maps and concentrated areas will be created. This is a feature that is difficult to focus on patterns and patterns, and it is difficult to achieve a satisfactory result [25].

In order to be further optimized on the traditional pooling model, this paper proposes a dynamic adaptive pooling model based on the maximum pooling algorithm. The advantage of this model is that compared with the traditional pooling model, it is more flexible and can make dynamic adjustments according to different feature values and different pooling content to adapt to the pooling model and optimize the pooling process. If there is only one pooling value in the existing pooling area, then this value is the maximum value, which can indicate its characteristics; if all the eigenvalues in the pooling area are the same, the maximum value is still taken as the eigenvalue. Therefore, after understanding the calculation process of the maximum pooling algorithm, a corresponding mathematical function model can be made.

Set as the pooling factor, the dynamic adaptive pooling algorithm function can be expressed as follows:

$$S_{ij} = \mu \max_{i=1,j=1} (F_{ij}) + b_2. \quad (18)$$

This type is a basic expression of the possibility of adaptive algorithms. The essence is to use the pooling coefficient m to optimize the maximum concentration algorithm. Through optimized performance, functions can be expressed more accurately. Other parameters are set according to the parameters of the maximum pooling model.

$$\mu = \rho \frac{a(v_{\max} - a)}{v_{\max}^2} + \theta. \quad (19)$$

In formula (14), a represents the average value of all elements except the maximum value, which V_{\max} is the maximum value, θ represents the correction error term, and ρ represents the characteristic coefficient. The specific mathematical formula is as follows:

$$\rho = \frac{c}{1 + (n_{epo} - 1)c^{n_{epo} + 1}}, \quad (20)$$

where n_{epo} is the number of iterations during training.

According to the maximum pooling to improve and optimize the dynamic adaptation pooling, the input part of the maximum pooling model is the core of the two-dimensional item feature table and correlation. Use 4 different convolution kernels, such as matrices with weights $A =$

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \text{ and } D = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

to convolve the input feature maps to obtain the convolution results corresponding to 4 different values in the pooling domain.

TABLE 1: Model parameter design.

Parameter	Value
Learning rate	0.001
Learning strategy	“inv”
Energy	0.75
γ	0.01
Maximum number of iterations	2000000
Impulse	0.9
Weight attenuation	0.05
Optimization mode	Adam

TABLE 2: AUC and EER are frame and pixel level comparisons on ped1 (UCSD data set).

Algorithm	Ped1 (frame)		Ped1 (pixel)		Ped2 (frame)	
	EER/%	AUC	EER/%	AUC	EER/%	AUC
TCP	8	0.967	38	0.632	18	0.817
AMDN	17	0.922	40.2	0.662	18	0.918
Motion energy	23	0.811	56	0.590	22	0.913
ST-CNN	24.5	0.861	38.6	0.881	25.6	0.853
Ours	7.6	0.961	35.3	0.686	16.3	0.932

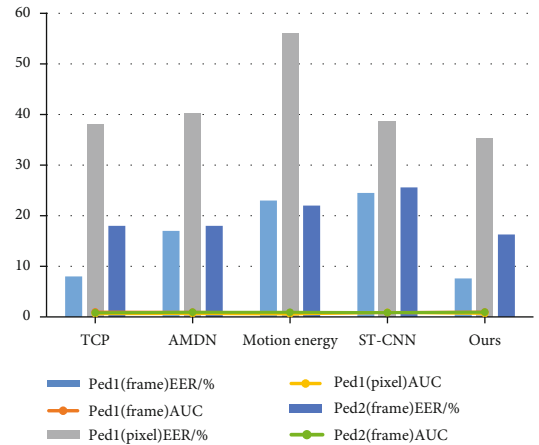


FIGURE 3: AUC and EER are frame and pixel level comparisons on ped1 (UCSD data set).

4. Social Group Behavior Analysis Model Integrating Multitask Learning and Convolutional Neural Network

4.1. Social Group Behavior Model Based on Multitask Learning and Convolutional Neural Network. This experiment was carried out in the MATLAB environment. Related procedures include the process of data acquisition system, data processing process based on a complete neural network, and data classification process.

Model data parameters are shown in Table 1.

In 3.0 GHz CPU, 64-bit Windows 7 software operating system, MATLAB2016a, and Open CV were selected as the development tools, and then, the simulation experiment

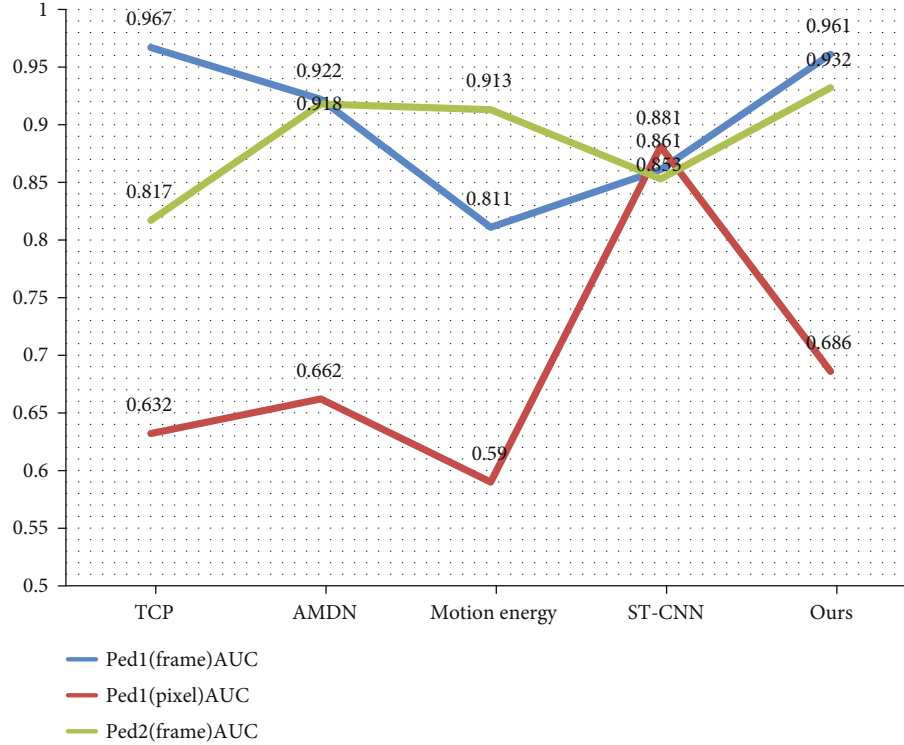


FIGURE 4: AUC comparison of five algorithms on data subset.

TABLE 3: AUC and EER for frame and pixel level comparisons on UMN data sets.

Algorithm	EER/%	Frame-level	
		AUC	Time/s
TCP	3.3	0.987	0.33
Motion energy	4.2	0.988	0.08
ST-CNN	3.8	0.997	0.48
Commotion	3.2	0.987	0.35
Ours	3.2	0.991	0.22

was carried out. In order to achieve the purpose of verifying the effectiveness of the algorithm, the reference data sets commonly used in the recognition investigation of multiple actions are selected in this article, namely, the UCSD data set and the UMN data set. These two data sets cover all extractable group behavior actions. In the simulation experiment, a quantitative evaluation method is used, and the evaluation indicators select AUC, EER, and calculation time parameters.

4.1.1. Experimental Results on the UCSD Data Set. The UCSD data set was produced by the University of California, San Diego. The data set is collected through the camera as the medium.

The camera is selected to be able to observe the sidewalk at a specified height, mainly to collect social group behaviors that occur under natural conditions. In this paper, we choose TCP model, AMDN model, energy motion model, spatial neural network model, chaos model,

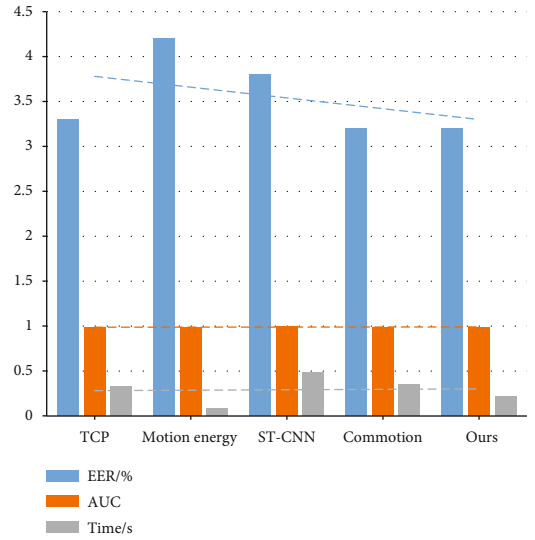


FIGURE 5: AUC and EER for frame and pixel level comparisons on UMN data sets.

TABLE 4: The relationship between the number of prototypes and accuracy.

Number of prototypes	Accuracy	Precision	Recall	F1	TNR
50	98.6	98.6	98.6	98.6	99.7
100	99.1	99.1	99.1	99.1	99.9
150	98.6	98.7	98.6	98.6	99.7
200	98.6	98.6	98.6	98.6	99.7
250	98.1	98.1	98.1	98.1	99.6

TABLE 5: Comparison of deep learning method and nondeep learning method on classified data sets.

		Accuracy	Precision	Recall	F1	TNR
Nondeep learning method	LapRLS	0.635	0.534	0.548	0.667	0.917
	KSRC	0.831	0.792	0.853	0.814	0.954
	mSRC	0.877	0.844	0.923	0.872	0.963
	MCMi-AB	0.618	0.595	0.574	0.564	0.898
	mcSVM	0.684	0.698	0.578	0.577	0.922
	ESRC	0.801	0.731	0.885	0.778	0.941
Deep learning methods	CNN + SVM	0.969	0.969	0.969	0.969	0.993
	CNN + mi-SVM	0.988	0.989	0.988	0.988	0.998
	CNN + MI-SVM	0.982	0.982	0.982	0.982	0.996
	K-means	0.991	0.991	0.991	0.991	0.998
	Random	0.976	0.976	0.976	0.975	0.995

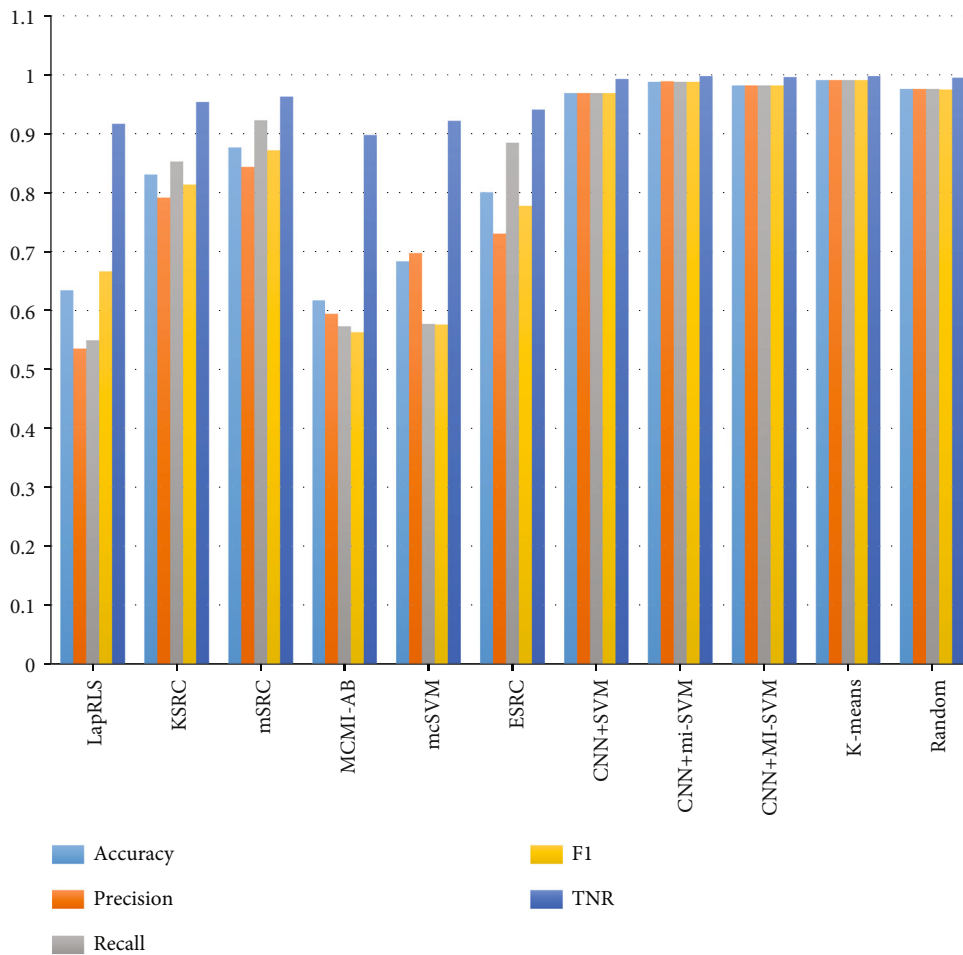


FIGURE 6: Comparison of different algorithms in data classification.

and other algorithms to achieve a better recognition rate in the above database. The source of the total data is divided into ped1 and ped2. These two subdata sets store 100 scene videos. Each video is subdivided into approximately 200 frames of video clips with a pixel resolution of 160×249 and 250×368 pixel.

The effectiveness of the algorithm needs to be verified through the UCSDped1 and UCSDped2 data sets. It can be seen from Table 2 that on the UCSDped1 data set, when frame-level metrics are used, the EER of the ST-CNN algorithm used in this paper is 24.5% and 38.6%, and the AUC values are 0.861 and 0.881; when pixel-level metrics are used,

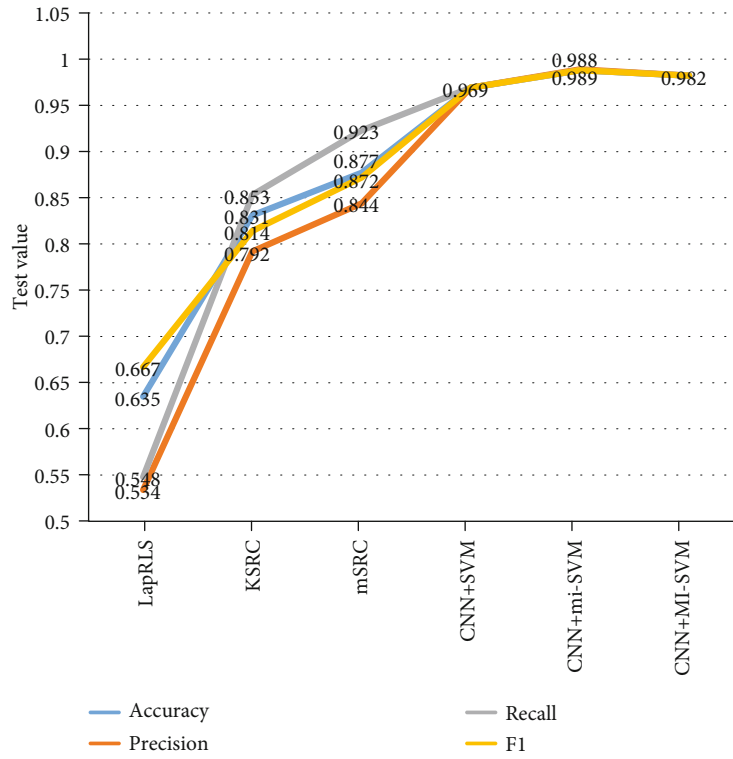


FIGURE 7: Line chart of comparison of deep learning method and nondeep learning method on classified data sets.

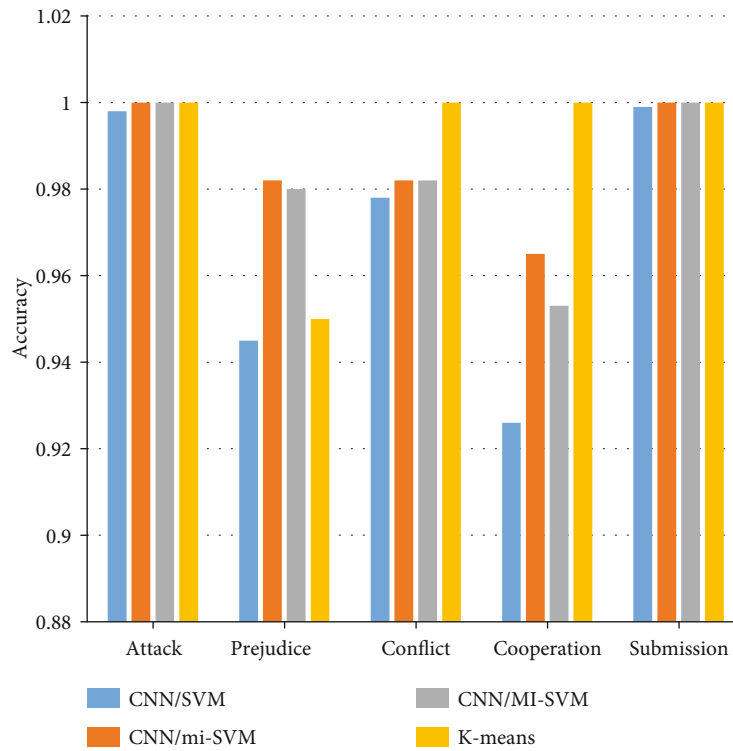


FIGURE 8: Accuracy comparison of five social group behaviors using CNN algorithm.

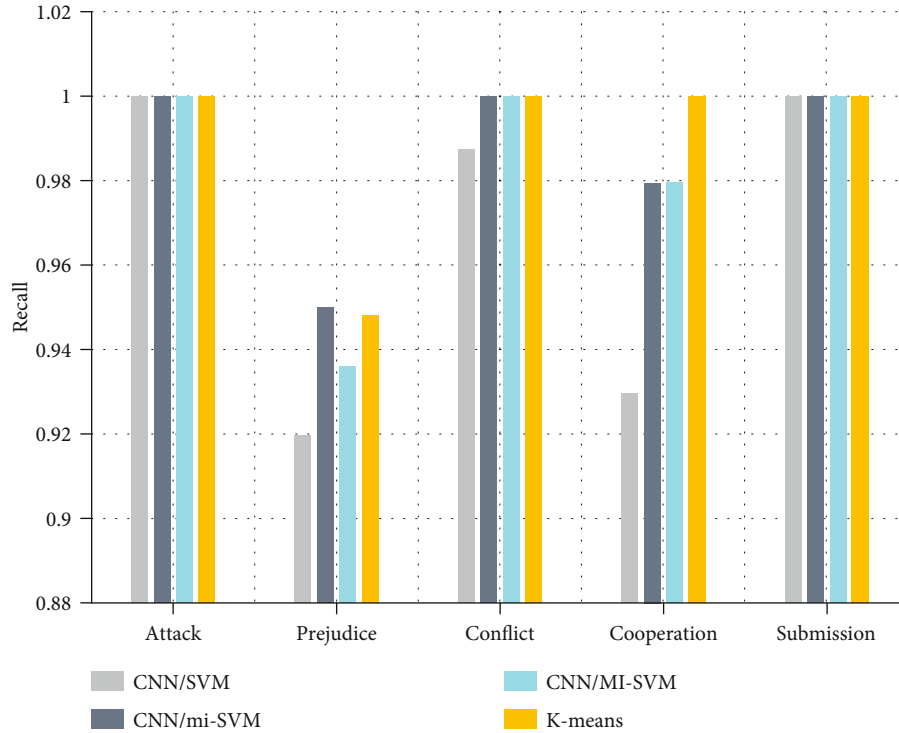


FIGURE 9: Comparison of recall rate data of CNN algorithm in social group behavior analysis.

the EER index of CNN is 25.6%, and the AUC evaluation index is 0.853.

It can be seen from Figures 3 and 4 that in the UCSDped1 data set, if framework-level metrics are used, the EER of the paper's algorithm will decrease, and the AUC evaluation will be greatly improved. When using pixel level measurement, the EER and AUC values show that the index improvement effect is not good, but it is still higher than the other two index algorithms. In the UCSDped2 data set, the frame-level measurement scale is used for testing. The algorithm in this paper is good for EEA and AUC scores, and the AUC score is increased to 0.11.

4.1.2. Experimental Results on UMN Data Set. In addition, the UMN data set has also been experimentally verified. Data set is used in the first half and the second half. In this data set, framework-level metric EEA and AUC evaluation indicators are used to evaluate the performance of the algorithm. The verification results are shown in Table 3.

As can be seen from Figure 5, the results of the algorithm performance test using the EER and AUC evaluation indicators below the framework level of the UMN data set are as follows. The algorithm of this paper has the same performance as the existing algorithm of AUC rate index and other algorithms with higher performance than EEA index. The algorithm time spent has been improved.

4.1.3. The Effect of Adjusting the Number of Prototypes on the Results. The research method used in this article needs to verify the influence of the number of prototypes on the effect of group behavior division. The relationship between the num-

ber of prototypes and accuracy is shown in Table 4. The number of prototypes in the experiment is set to 80, 100, 150, 200, and 250; it can be seen from Table 4 that the number of experiments is 100, which has the highest accuracy. Therefore, in the subsequent experiments, the number is set to 100.

4.2. Group Behavior Recognition Model Fusing Multitask Learning and Convolutional Neural Network. We compare and test the algorithm proposed in this paper and related methods in five aspects: accuracy, accuracy, recall, F1 value, and TNR. As shown in Table 5, they are all methods of manually extracting features, as shown in Table 5. The data shows that the highest accuracy index of traditional nondeep learning methods is 87.7%, which is much lower than the accuracy of deep learning methods. In the deep learning method, CNN combined with multitask learning methods are used, and the accuracy values are greater than 95%.

It can be clearly seen from Figures 6 and 7 that compared to traditional learning methods, the fusion multitask learning method and convolutional neural network learning method used in this paper have achieved better results. The algorithms in the table are all for target monitoring, and for the target recognition operation, the training mode adopts the fully supervised form of multitask learning. From the data results, the results of the new algorithm are all above 0.95, and the effect is very good.

We divide social group behaviors into five types of behaviors: aggression, prejudice, conflict, cooperation, and obedience. The following three graphs are the comparison diagrams of the classification data effects of the five types of group behaviors on various indicators. As shown in

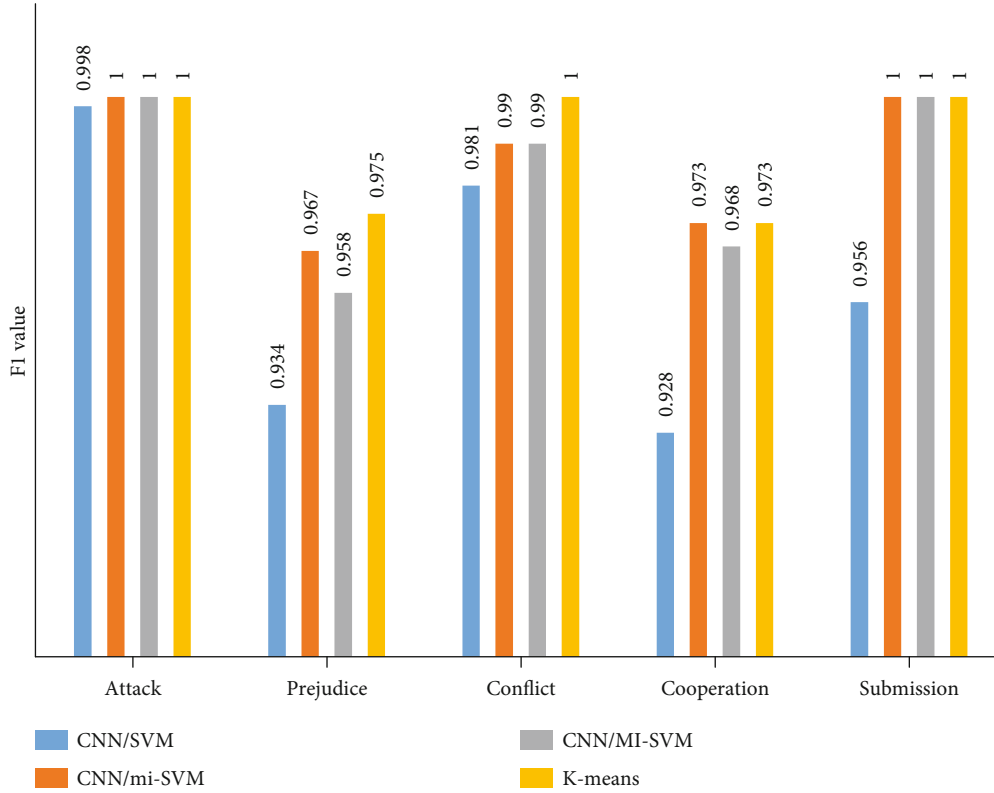


FIGURE 10: Comparison of F1 values of different group behaviors in four algorithms.

Figure 8, the accuracy of the five behaviors in each test method is greater than 85%. Among them, the accuracy of the attack behavior and the compliant behavior using the CNN combination method is higher, reaching more than 95%.

As can be seen in the above two figures, using the CNN joint method, the recall rate and F1 value of the observation results of the five group behaviors are clearly contrasted. In Figure 9, the recall rate of each method of attack and obedience has reached 100%, but the overall recall rate of biased behavior is very low, of which CNN/SVM is only 92%. In Figure 10, the F1 values of prejudice, conflict, and cooperation are continuously improved by using four methods. The conflict is extremely obvious. The F1 value of CNN/SVM is 0.981, and the value of CNN/mi-SVM and CNN/MI-SVM is 0.981. The values are the same; both are 0.985, and the F1 value of *K*-means reaches 100%. It can be seen that the algorithm used in this paper has achieved good results in the study of social group behavior.

It can be seen from Table 6 that the overall accuracy of the recognition of group behaviors integrated with deep learning is very high, and the corresponding reaction behaviors can be fully made in the process of cross-experiment.

Table 6 shows the comparison between the traditional neural network model, the MVP model, and the convolutional neural network model in terms of testing algorithm accuracy and cross-validation accuracy. Comparing the accuracy of the experimental algorithms of the three models, it can be seen that the experimental accuracy of the convolutional neural network method is as high as 92.10%, which is

TABLE 6: Accuracy comparison of three common behavior recognition models.

Machine learning algorithm	Test set accuracy (%)	Cross-validation accuracy (%)
Traditional neural network	52.12	65.82
MVP	58.19	75.54
Convolution neural network	92.10	84.5

much higher than the other two experimental methods. Therefore, modeling based on convolutional neural network is more suitable for multifeature-based modeling. Through the model analysis of group behavior. The cross-validation accuracy is also significantly higher than traditional neural networks and MVP, reaching 84.5%.

5. Conclusion

This paper mainly studies the social group behavior analysis model that combines multitask learning and convolutional neural network and consults a large number of references and in-depth study of convolutional neural network, multitask learning, and group theory-related content. This paper constructs a social group behavior analysis model and a dynamic adaptive pooling model based on convolutional neural networks and makes full use of the advantages of convolutional neural network algorithms to analyze social group

behaviors. Combining multitask learning and convolutional neural networks can already extract the deep-level features of foreground moving targets, and the improved convolutional neural network reduces the acquisition of redundant information.

The innovation of this paper is that it adopts a combination of quantitative and qualitative methods, which is well reflected in the convolutional neural network model in the fourth part of this paper. The method of combining theoretical analysis and empirical research is used to establish model analysis. While using experimental data to explain the problem, this method runs through this article. The complete convolutional neural network of this experiment can obtain effective feature information of social group targets and avoid the collection of redundant information. The experiments of UCSD and UMN show that the convolutional neural network algorithm can optimize the calculation time and learning time to the greatest extent, which is very suitable for social group behavior analysis. The innovation of this paper is to combine multitask learning and convolutional neural networks with group behavior research in the field of psychology, which shows the development and application of the Internet of Things in behavior research cognition.

The disadvantage of this article is that due to the limitations of actual conditions, the number of samples collected is small, and the sample results need to be more standardized; the convolutional neural network itself has the defects of translation invariance and back propagation, which has a certain impact on the data parameters.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Demonstration Research on Intelligent Application of Regional Politics and Law (Project number 2020YFC0833407).

References

- [1] D. Mishkin, N. Sergievskiy, and J. Matas, "Systematic evaluation of convolution neural network advances on the Imagenet," *Computer vision and image understanding*, vol. 161, pp. 11–19, 2017.
- [2] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [3] Y. Yan, E. Ricci, R. Subramanian, G. Liu, O. Lanz, and N. Sebe, "A multi-task learning framework for head pose estimation under target motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 6, pp. 1070–1083, 2016.
- [4] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang, "CU partition mode decision for HEVC hardwired intra encoder using convolution neural network," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5088–5103, 2016.
- [5] Y. Tang and X. Wu, "Scene text detection and segmentation based on cascaded convolution neural networks," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1509–1520, 2017.
- [6] K. H. Cha, L. M. Hadjiiski, R. K. Samala et al., "Bladder cancer segmentation in CT for treatment response assessment: application of deep-learning convolution neural network-a pilot study," *Tomography a Journal for Imaging Research*, vol. 2, no. 4, pp. 421–429, 2016.
- [7] Y.-x. Li, Y.-y. Pu, D. Xu, W.-h. Qian, and L.-p. Wang, "Image aesthetic quality evaluation using convolution neural network embedded learning," *Optoelectronics Letters*, vol. 13, no. 6, pp. 471–475, 2017.
- [8] R. Mena-Yedra, J. Casas, and R. Gavaldà, "Assessing spatio-temporal correlations from data for short-term traffic prediction using multi-task learning," *Transportation Research Procedia*, vol. 34, pp. 155–162, 2018.
- [9] X. Zhu, H.-I. Suk, S.-W. Lee, and D. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 607–618, 2016.
- [10] X. Lu, X. Li, and L. Mou, "Semi-supervised multitask learning for scene recognition," *IEEE Transactions on Cybernetics*, vol. 45, no. 9, pp. 1967–1976, 2017.
- [11] R. Mogan, R. Fischer, and J. A. Bulbulia, "To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect," *Journal of Experimental Social Psychology*, vol. 72, pp. 13–20, 2017.
- [12] J. Yu, B. Zhang, Z. Kuang, D. Lin, and J. Fan, "iPrivacy: image privacy protection by identifying sensitive objects via deep multi-task learning," *IEEE Transactions on Information Forensics & Security*, vol. 12, no. 5, pp. 1005–1016, 2017.
- [13] Z. Zhang, Y. Xie, W. Zhang, Y. Tang, and Q. Tian, "Tensor multi-task learning for person re-identification," *IEEE Transactions on Image Processing*, vol. 29, pp. 2463–2477, 2020.
- [14] A.-A. Liu, N. Xu, W.-Z. Nie, Y.-T. Su, and Y.-D. Zhang, "Multi-domain and multi-task learning for human action recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 853–867, 2019.
- [15] Z. Wen, K. Li, Z. Huang, C.-H. Lee, and J. Tao, "Improving deep neural network based speech synthesis through contextual feature parametrization and multi-task learning," *Journal of Signal Processing Systems*, vol. 90, no. 7, pp. 1025–1037, 2018.
- [16] H. D. Vargas Cardona, M. A. Alvarez, and A. A. Orozco, "Multi-task learning for subthalamic nucleus identification in deep brain stimulation," *International Journal of Machine Learning & Cybernetics*, vol. 9, no. 7, pp. 1181–1192, 2018.
- [17] A. A. Sandel, J. A. Miller, J. C. Mitani, C. L. Nunn, S. K. Patterson, and L. Z. Garamszegi, "Assessing sources of error in comparative analyses of primate behavior: intraspecific variation in group size and the social brain hypothesis," *Journal of Human Evolution*, vol. 94, pp. 126–133, 2016.
- [18] G. Robert, "Animated images in the analysis of zebrafish behavior," *Current Zoology*, vol. 63, no. 1, pp. 35–44, 2017.
- [19] F. C. Mace, "Tony Nevin: the embrace of translational work by a basic scientist," *Journal of the Experimental Analysis of Behavior*, vol. 109, no. 1, pp. 56–65, 2018.

- [20] A. T. Stull, L. Fiorella, and R. E. Mayer, "An eye-tracking analysis of instructor presence in video lectures," *Computers in Human Behavior*, vol. 88, pp. 263–272, 2018.
- [21] M. Sutter, D. Glätzle-Rützler, L. Balafoutas, and S. Czermak, "Cancelling out early age gender differences in competition: an analysis of policy interventions," *Experimental Economics*, vol. 19, no. 2, pp. 412–432, 2016.
- [22] J. M. Garcia, J. R. Sirard, N. L. Deutsch, and A. Weltman, "The influence of friends and psychosocial factors on physical activity and screen time behavior in adolescents: a mixed-methods analysis," *Journal of Behavioral Medicine*, vol. 39, no. 4, pp. 610–623, 2016.
- [23] A. Rana, M. Hamid, and Y. Qin, "Understanding the joint behavior of temperature and precipitation for climate change impact studies," *Theoretical and Applied Climatology*, vol. 129, no. 1, pp. 321–339, 2016.
- [24] M. L. Chang, "On the relationship between intragroup conflict and social capital in teams: a longitudinal investigation in Taiwan," *Journal of Organizational Behavior*, vol. 38, no. 1, pp. 3–27, 2017.
- [25] I. Tung and S. S. Lee, "Latent trajectories of adolescent antisocial behavior: serotonin transporter linked polymorphic region (5-HTTLPR) genotype influences sensitivity to perceived parental support," *Development and Psychopathology*, vol. 29, no. 1, pp. 185–201, 2016.