

## Research Article

# Footprint Extraction and Sports Training Action Recognition Based on Wireless Network Communication

Lu Jiang 

*Physical Culture Institute, Hunan University of Humanities, Science and Technology, Loudi, 417000 Hunan, China*

Correspondence should be addressed to Lu Jiang; 3435@huhst.edu.cn

Received 19 October 2021; Revised 26 November 2021; Accepted 1 December 2021; Published 18 January 2022

Academic Editor: Deepak Kumar Jain

Copyright © 2022 Lu Jiang. 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.

The combination of scientific and technological achievements and sports has found new opportunities to change people's sports habits. Sport training takes up an increasing proportion of people's lives. In order to improve the efficiency of sports training and standardize the training actions of players, this article is based on wireless network communication and uses different types of recognition methods in the field of action recognition to build basic classifications. Iterative mutual training to improve generalization performance can reduce the cost of labeling and realize the complementary advantages of different recognition methods, thereby improving the recognition accuracy of human actions. Finally, the algorithm is used to recognize human movements. This method can effectively overcome the problem of differential degradation of base classifiers in the iterative process of collaborative training and further improve the accuracy of human action recognition. The experimental results prove that the motion recognition of wireless network communication proposed in this paper can effectively improve the accuracy of athletes' movements, which is more than 20% higher than traditional methods, and, under the guidance of standardized movements, can reduce athletes' sports injuries.

## 1. Introduction

The ubiquitous wireless network establishes a connection between everything and everything, so that the interaction of the information network can be realized. In addition to being able to complete traditional information transmission tasks, wireless networks can perceive the activity information of targets within their coverage areas and have received extensive attention from researchers in recent years. With the popularity of network deployment, no matter where we are, we are surrounded by wireless signals. These wireless signals form a huge wireless network that senses and monitors all activities in the network coverage area. Through wireless network communication, the shielding effect of the target on the wireless network can be used to intelligently perceive the human body, and the traditional wireless network for communication can be evolved into an intelligent network with human body position and motion recognition capabilities.

In response to the strong development of the sports industry and the rapid development of technological innova-

tion, various wearable devices and smart products have been continuously applied in the sports industry, and somatosensory technology and VR technology have been deeply explored in the sports field. The high-tech products brought by technology to the development of the smart sports industry have quietly changed people's lives and the trajectory of sports. We designed a motion recognition algorithm which was based on finite state machine based on the evaluation of the demand analysis of motions training traffic, coupled with existing algorithms, in extracting the communication features of wireless network, and we also implemented and improved the client and server side to create a motion recognition system.

The movements of sports athletes are complex and difficult to recognize with current intelligent technology. Therefore, in order to improve the recognition rate of sports athletes, Xu and Yan analyzed the motion recognition system based on cluster regression and improved ISA deep network. Through literature survey, this article chooses ISA neural network as the basis of the algorithm. At the same time, this paper analyzes the shortcomings of the traditional

ISA neural network, combines the athletes' movement recognition needs to improve the traditional ISA neural network, and builds an athlete's movement recognition system based on the improved ISA neural network algorithm. In addition, this article uses the network data collection method to construct the athlete's action video library, analyzes the basketball event as an example, and recognizes it through feature judgment. Finally, this article establishes experiments for model performance analysis [1]. Ashwan et al. proposed a new framework for human action recognition based on salient object detection and a new combination of local and global descriptors. We first detect the salient objects in the video frame and only extract the features of these objects. Then, we use a simple strategy to identify and process those video frames that contain salient objects. Processing salient objects instead of all frames not only makes the algorithm more efficient, but more importantly, it also suppresses the interference of background pixels. We combine this method with a new combination of local and global descriptors, namely, 3D-SIFT and histogram of oriented optical flow (HOOF). The resulting saliency-guided 3D-SIFT-HOOF (SGSH) features are used together with multiclass support vector machine (SVM) classifiers for human action recognition. Experiments conducted on the standard KTH and UCF-Sports action benchmarks show that our new method is superior to the most advanced human action recognition methods based on spatiotemporal features [2]. Sun and Ma proposed a dynamic template mechanism for the problem of uneven brightness and sudden changes in sports competitions that cause target recognition errors. In the target recognition algorithm, the correlation degree of the data feature change is fully considered, and the time control factor is introduced when using SVM for classification. At the same time, this study uses the unsupervised clustering method to design the classification strategy to realize the target identification when the environment brightness changes quickly. Improve recognition accuracy. In addition, the machine learning method uses the Adaboost algorithm, which optimizes the algorithm in terms of fast feature selection and dual-threshold decision-making, which effectively improves the training time of the classifier. Finally, for complex human postures and partially occluded human targets, this article proposes to express the entire human body through multiple parts. Experimental results show that this method can be used to detect multiposture and partially occluded sports athletes in complex backgrounds, and it provides an effective technical means for detecting the action characteristics of sports games in complex backgrounds [3]. These studies have a certain reference effect for this article, but the data samples of these studies are insufficient, and the research time and research methods are too narrow to be widely applicable in reality.

In the process of motion recognition algorithm research, this paper adopts a finite state machine model based on template matching feature description and proposes a general motion recognition algorithm model, which can be reused to add new motion recognition, reducing repetitive work and increasing the extensibility of the system. The precise model design for each standard movement in sports recogni-

tion has scientific guiding significance for the requirements of sports training movement specifications during use. At the same time, the collection and sorting of sports training movement conditions can scientifically formulate sports training prescriptions to maximize the benefits of exercise on physical health.

## 2. Footprint Extraction and Sports Training Action Recognition Method

*2.1. Action Recognition.* Early motion recognition research mainly used some mobile devices to obtain information. Although a large amount of motion data can be obtained from the device, this is due to efficiency, cost, environmental constraints, and other reasons [4]. There are major flaws, and the scope of research is limited. With the popularization and update of video equipment, it is now mainly to collect motion data through vision, that is, to conduct motion recognition research through video data.

Information processing refers to the processing of collected related action information data, mainly using related machine learning algorithms to create related data models. The standard model is to learn data, that is, to update the model parameters to give it a humanized appearance. Cognitive experience is able to discern new data. Motion recognition can conduct professional analysis of human motion [5, 6]. For example, analyzing the sports behavior of an athlete can identify the wrong sports behavior and then correct his behavior. In terms of intelligent monitoring, such as taking care of the elderly and children, analyzing whether the elderly's daily behavior is normal, it can detect abnormal behaviors of the elderly, such as physical discomfort or fainting after a long time. It is very important to predict whether children's behaviors and actions are dangerous in combination with the scene they find themselves in, to send omen information to the outside world in time, and to monitor the safety of human life in real time [7].

In action recognition, action data is a kind of data that changes over time. An action sequence contains dozens of gesture frames. The gesture sequence is an action [8]. Actions and actions are continuous actions and more complex interactive actions, which become more and more difficult. The framework of the motion recognition algorithm is shown in Figure 1.

Recognition of actions is the studying of human-centered, natural, instinctual, and interactive methods that conform to human habits. And this has become a hot topic of research. Considerable research progress with human action recognition techniques is being achieved at all levels, the emergence of many excellent learning algorithms, but there are still many problems on how to design an efficient and robust recognition scheme [9]. The main difficulties and challenges are

- (1) Changes within and between action types. For the same action category, due to zooming, even if it is done by the same person, the video image rendering size may be different. For example, you can walk in different scenes. Walking speed directly affects step

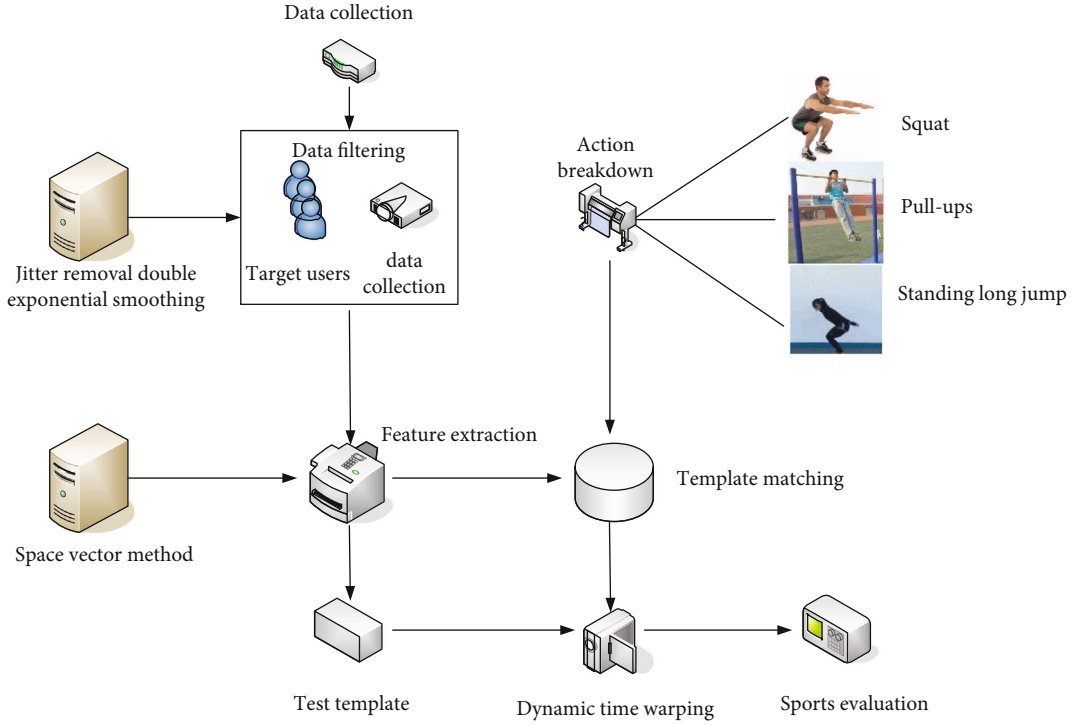


FIGURE 1: Motion recognition algorithm framework.

length, and from different angles, it also shows different temporal and spatial characteristics. Therefore, this is also an urgent research topic in the field of human behavior recognition

- (2) The ability to generalize the model [10]. At present, in the field of action recognition, most actions are recognized through a single method. Both model-based methods and statistical probability methods have their own limitations, which limits the possibility of model generalization to a certain extent. In the standard method, for example, there was significant performance degradation when there were insufficient labels, whereas statistical probability methods were significantly degraded when classes overlapped

**2.2. Action Characteristics.** The human body is made up of multiple frames of one energy. Each frame was a single attitude. In stopping sequence, the whole action is described, which is processed to characterize the attitude and then completes the representation of the whole action [11, 12]. These can describe the action well and match the human action. In the case of static features, global information about the human subjects can be better in representation and provide valid information for determining the action.

A certain frame in the human body motion sequence data is represented by the three-dimensional coordinates of 20 joint points in space, where the parent joint point of a certain limb is  $(x_f, y_f, z_f)$ , and the child joint point corresponding to it is  $(x_i, y_i, z_i)$ . The coordinate system where they are located is the world coordinate system, with the parent joint point as the origin of the local spherical coordinate

system, and the original coordinate system is translated to the coordinate system with the parent joint point as the origin. Then, the coordinate transformation formula is as follows.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} - \begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix}. \quad (1)$$

Among them,  $(x, y, z)$  is the converted coordinate of the child joint point, the connection line between the child joint point and the coordinate origin, that is, the parent joint point, and the direction is from the coordinate origin to the child joint point. The included angle between it and the positive Z axis is  $\phi$ , the value range of the included angle is  $(0^\circ, 180^\circ)$ , and its calculation formula is as follows:

$$\phi = \arccos \left( \frac{z}{\sqrt{x^2 + y^2 + z^2}} \right). \quad (2)$$

In practical applications, when the subjoint points are around the positive axis, the amplitude of actions performed by different people is different, which results in an excessive sudden change in the angle and affects the recognition effect. The calculation method of the included angle  $\theta$  is

$$\theta = \begin{cases} \arctan \left( \frac{y}{x} \right) + \pi & x < 0, y > 0, \\ \pi - \arctan \left( \frac{y}{x} \right) & x < 0, y < 0. \end{cases} \quad (3)$$

After calculating the human action model, an overall posture of the human body can be expressed as

$$p_t = \{(\theta_i, \phi_i) | i = 1, 2, \dots, 10\}. \quad (4)$$

In order to control the classification noise of the generated pseudolabeled samples, we start with improving the accuracy of the data, first calculate the prediction confidence of the unlabeled data, then sort it, and select the higher ones to add to the labeled samples to expand the training set. For unlabeled data, the classifier can calculate a predicted probability for each action recognition category:

$$p(y = c_j | x_i) = \frac{d_m(x)}{k}. \quad (5)$$

Among them,  $k$  is the number of neighbor samples set, and  $d_m$  is the number of neighbors of different categories included in the test sample. It is sometimes not reasonable to consider the maximum predicted probability of a class as a degree of confidence. Therefore, we adopt a new confidence calculation standard: the calculated maximum predicted probability value minus the second largest predicted probability value to express the size of the confidence.

$$C(x_i) = P(y = c_{\max j} | x_i) - P(y = c_{\text{sub-max } j} | x_i). \quad (6)$$

Second, for the base classifier SVM, since it classifies samples through the hyperplane, it does not calculate the predicted probability of the class. Further, we have improved the strategy of adding pseudolabels. Traditional collaborative training often adds pseudolabeled data to labeled samples in a single manner based on the confidence level. We estimate  $e(h_i)$  error rate:

$$e(h_i) = \sum_{j=1}^k |P_U(|h_i) - P_L(y = j)|. \quad (7)$$

The trained model is returned to the frame-level labeling of the training data, and each frame of the action sequence is labeled at the pose level, which is the state labeling in the model. The node of the path with the highest probability among all single paths is

$$\psi_t(i) = \arg \max [\delta_{t-1}(i) a_{ij}]. \quad (8)$$

Initialize first

$$\delta_1(i) = \pi_i b_i(o_1), (1 \leq i \leq N). \quad (9)$$

Each new time corresponds to a new stage in the algorithm. Then, the recurrence formula for time point is as follows

$$\psi_t(j) = \arg \max [\delta_{t-1}(i) a_{ij}]. \quad (10)$$

Optimal path backtracking:

$$q_T^* = \psi_{t+1}(q_{t+1}^*), t = T - 1, T - 2, \dots, 1. \quad (11)$$

For every situation being dealt with, it is a node and is considered to be the leader of the world's best route. All these points will be deleted after the backtracking operation, except for the last one.

The customization of data is based on the premise of which distribution it conforms to, and then training and analysis are performed according to the hypothetical distribution model. Therefore, learning the distribution of feature data according to the energy model can solve all the above problems.

$$E(v, h | \theta) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j. \quad (12)$$

Among them,  $\theta$  is the parameter model,  $a_i$  is the bias of the visible layer unit,  $b_j$  is the bias of the hidden layer unit, and  $W_{ij}$  is the connection weight between the visible layer and the hidden layer. The joint probability distribution that can be obtained according to the energy function  $(v, h)$  is as follows

$$Z\theta = \sum_{v, h} e^{-E(v, h | \theta)}. \quad (13)$$

Which  $Z(\theta)$  is represents the normalization factor in the calculation of joint probability. The likelihood function solved through specific calculations can be expressed as

$$p(v | \theta) = \frac{\sum_h e^{-E(v, h | \theta)}}{Z(\theta)}. \quad (14)$$

According to the state of the hidden layer unit, the formula for obtaining the visible layer unit in reverse is

$$P(v_i = 1 | h, \theta) = \sigma \left( a_i + \sum_j W_{ij} h_j \right). \quad (15)$$

The specific solution algorithm of the function  $P(v | \theta)$  is to use the contrast divergence algorithm, and the specific solution process will be described below in conjunction with specific applications.

Evidence theory is a reasoning method for solving uncertain problems. It is a theory based on a nonempty recognition framework (or hypothesis space). For any hypothesis in the problem domain, it satisfies

$$f(\sigma) = 0, \quad (16)$$

$$\sum_{A \in \Theta} f(A) = 1. \quad (17)$$

The synthetic formula of evidence theory is

$$f(A) = \frac{1}{1-k} \sum_{A_1 \cap A_2} f_1(A_{i1}) * f_2(A_{i2}). \quad (18)$$

Finally, given a set  $H$ , for each category  $j$ , use the evidence fusion formula to synthesize the evidence to obtain an evidence function, as shown in the following formula:

$$f_j(x|H) = \frac{a_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N a_t(i)a_{ij}b_j(o_{t+1})}. \quad (19)$$

$\beta$  is the backward variable obtained through the backward algorithm. The user behavior model parameter expression obtained by using the intermediate variable is as follows:

$$\bar{a}_{ij} = \frac{\sum_{t=lando_t}^{T-1} \gamma^t(j)}{\sum_{t=1}^T \gamma^t(i)}. \quad (20)$$

**2.3. Wireless Network Communication.** By the development of wireless networks, the environment we live in is filled at the moment with signals such as WiFi, FM, and Zigbee. They have realized the coverage of the whole city and promoted the development of a series with new emerging technologies, like the Wireless Network, Smart Space, and Smart City. These emerging applications meet the growing needs of people, make their lives more convenient and continuously improve their quality of life. Penetrating wireless networks enable connections being made between everything so that information network interaction can be achieved [13, 14].

In addition to being able to complete traditional information transmission tasks, wireless networks can perceive the activity information of targets within their coverage areas, and they have received extensive attention from researchers in recent years [15]. With the popularity of network deployment, no matter where we are, we are surrounded by wireless signals. These wireless signals form a huge wireless network that senses and monitors all activities in the network coverage area. When the target moves within the range of the wireless network, it will affect the signal such as occlusion, refraction, and reflection, which will change the propagation mode of the signal. When the target is in a different position, the affected wireless link is different; when the target performs differently, the mode of influence on the wireless link signal is also different [16].

Target activities in the wireless network deployment area will inevitably have an impact on the wireless link. The target is in a different location or performs different behaviors, and the impact on the wireless link is also significantly different [17]. The position of the human body is different, the wireless link covered by the target is different, the motion amplitude and frequency domain performed by the human body are different, and the influence on the wireless link signal is also different, which is reflected in the average size of the link, variance fluctuation, and energy concentration,

the frequency band, and other aspects [18]. Current research has explored the corresponding relationship between these characteristic changes of wireless signals and the target state, but the current research results cannot meet the needs of large-scale implementation of wireless networks. As a basic technology, the wireless network-based footprint extraction and sports training action recognition technology have made the traditional wireless network used for transmission and communication evolved into an intelligent network with comprehensive perception capabilities such as location and behavior.

Wireless network communication is a distributed network system that collects information by a series of miniature sensor nodes and transmits information through wireless communication. Sensor networks are widely used in industry, medical and health, agriculture, transportation, and other fields [19], such as monitoring various physiological data of the human body and hospital drug management. Special-purpose sensor nodes are installed on the patient's body, such as heart rate and blood pressure monitoring equipment. The doctor can check the patient's condition at any time through the sensor network, and if there is an abnormality, it can be rescued in time. Because the wireless sensor network has the characteristics of low cost, flexibility, ease of use, reliability, and security, it can bring a good user experience [20].

Behavior recognition research can be roughly divided into two directions: one is the recognition method based on image vision, which uses image equipment to generate pictures or videos and monitor user behavior from them. Taking the video data provided by the image equipment as input, a moving human segmentation algorithm based on the inter-frame difference and an improved CV model is proposed. This algorithm can solve the problem of moving human behavior recognition in complex situations and can provide real-time alarms to the surveillance security system [21]. The other is to monitor user behavior through sensors. The sensor monitoring process is completely transparent to the user and will not affect the daily life and work of the monitored person. The sensor readings are collected by a computer network and stored in a database, which is used for pattern recognition and prediction.

The method of enhancing user behavior characteristics, in view of the concurrency of behavior and the diversity of peer behavior, proposes an algorithm to enhance the characteristics of sensor data and uses the feature enhancement coefficient to express the feature recognition ability of behavior, in order to highlight the characteristics of behavior data. Characteristics, time sequence characteristics are added to behavior characteristics, and the physical meaning of behavior data is improved to make it easier to understand [22, 23]. Although this method solves the problem of concurrency of behavior, it ignores the problem of similarity of behavior. Some studies use Hidden Markov Model to segment the user's daily behavior sequence. There are studies using Hidden Markov Model to classify the user's daily behavior sequences, proposed a method which based on Hidden Markov Model with dynamic segmentation of continuous action sequences for the existence of continuity for actual

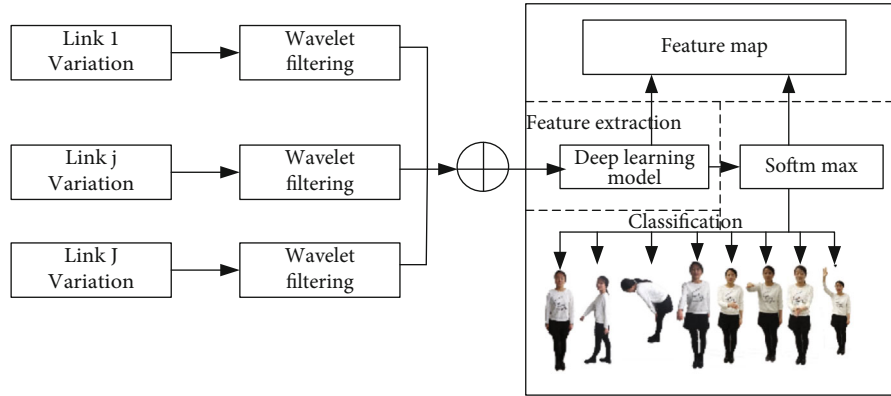


FIGURE 2: The overall framework of the system.

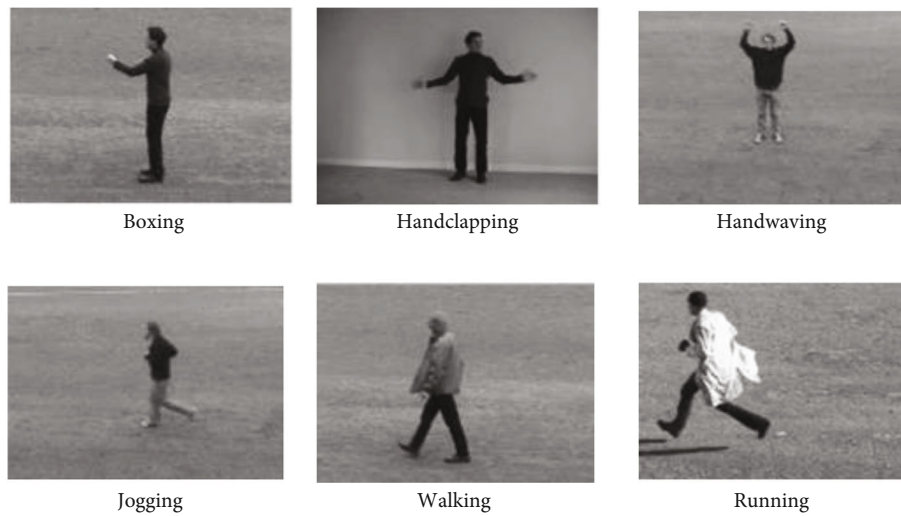


FIGURE 3: KTH sample diagram.

actions for the characteristics of multisensor data fusion, the results of which directly affect the accuracy of action recognition, and at the end, the average recognition accuracy of 89% was obtained using SVM [24, 25].

As shown in Figure 2, the system architecture mainly includes the following main modules: the filter module based on wavelet transform, apply wavelet transform on the RSS variation, which retains the data in the low-frequency portion of the signal and discards to the high-frequency band data, to effectively remove the multifrequency noise; feature extraction module based on deep learning, using sparse autoencoding network to automatically learn complete features from the input signal; system fine-tuning and classification module based on softmax, input the learned features into the softmax system, using supervised. The training method fine-tunes the system parameters and realizes the estimation of the target position and the recognition of behavior and gestures [26, 27].

### 3. Action Recognition Experiment and Results

*3.1. Database Display.* In this article, we use KTH and the sports movement database collected by our university, which

is a human action data widely used in human action recognition, where each person in the video performs multiple repetitive actions in different scenes to generate video samples, which are adapted to the study of various human action problems, such as the study of human action methods under cross-view, human action recognition methods, and action recognition methods between human interaction. The collected database samples are shown in Figures 3 and 4.

In the multi-instance and multifeature learning method, I used the multi-instance learning method based on the RBF kernel function. Each sample was used as a package, and the 1000-dimensional feature vectors of 3 different features of each video were used as examples in the package. Multi-instance learning is applied to the multifeature representation of samples. In the training process, all samples of the current action are taken as positive packets, and samples of other actions are taken as negative packets, and finally, a classifier for each action type is obtained. The experimental results are shown in Figure 5.

It can be seen from the figure that the feature vector consumed by the wireless network communication classifier used in this article is generally lower than other methods in the recognition of various actions, which shows that



FIGURE 4: Sample data collected by our school.

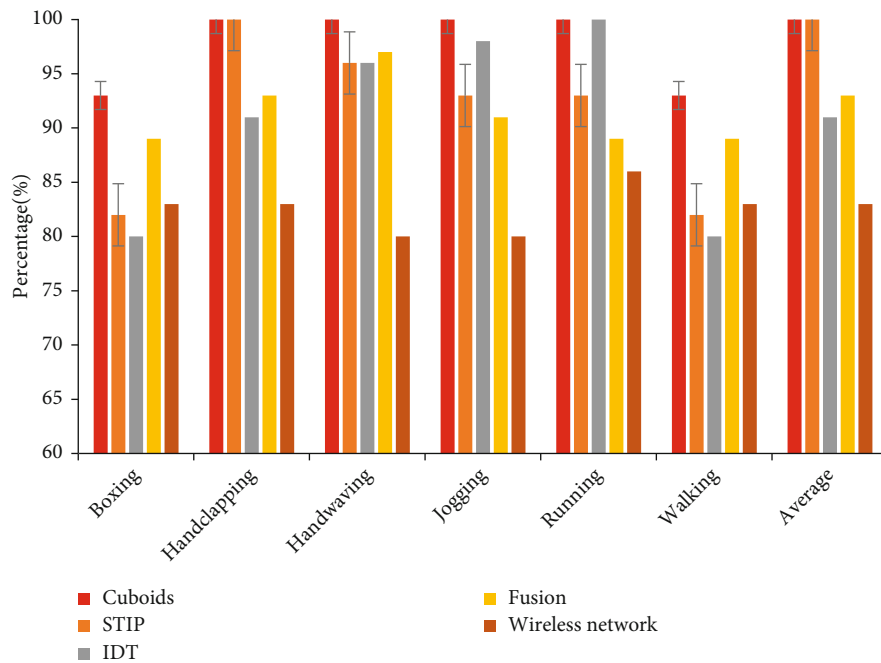


FIGURE 5: Database feature experiment results.

wireless network communication is superior to other methods [28–31].

**3.2. Comparison of Experimental Results.** In order to compare the recognition effect between different methods more intuitively, we compare the recognition effect of the method in this article with the traditional method, as shown in Figure 6.

It can be seen that on the average, the network communication method used in this article is better than the traditional recognition method when squatting and standing, while it is slightly lower than the traditional method when the arm is swinging, and the recognition when moving.

Above, there is little difference between the two methods. In order to better determine the difference between the different methods, we make statistics on the difference in the closure value between the different methods, as shown in Figure 7.

It can be seen from the figure that the movement of walking is faster than other movements, so the corresponding zero-crossing point value is relatively large, and the peak value can reflect the magnitude of the target action. The target’s behavior during the squatting process is relatively large. The variance is also relatively large, reflecting the degree of signal convergence in a certain frequency band. Let us take the football and basketball actions in our school’s database

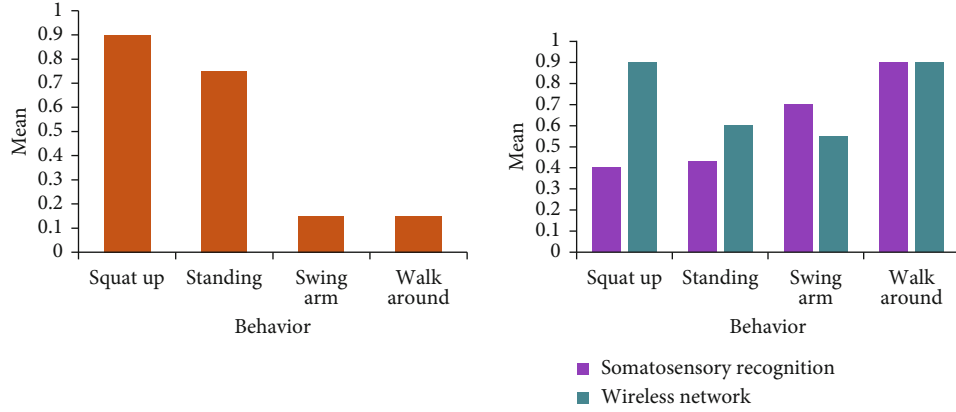


FIGURE 6: Feature performance comparison.

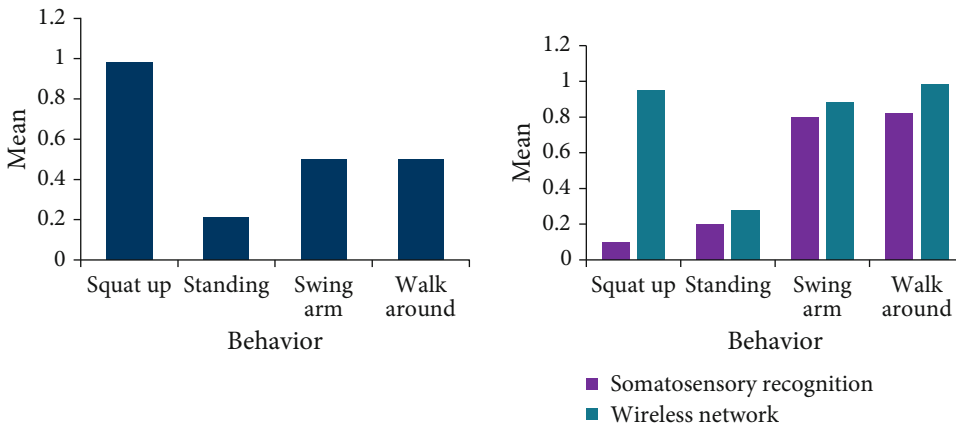


FIGURE 7: Peak feature difference.

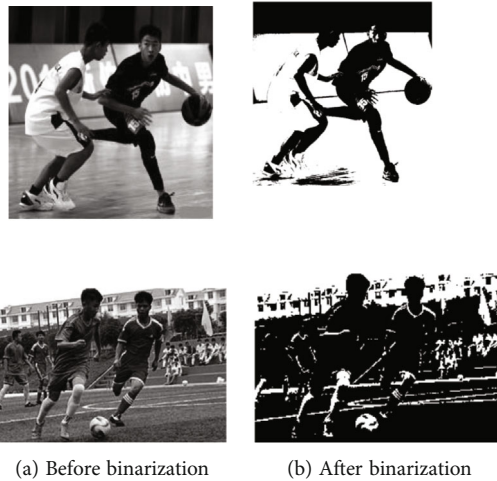


FIGURE 8: The effect of image binarization.

as an example to compare the recognition effects between them. We first binarize the pictures, as shown in Figure 8.

After the picture is binarized, we use the wireless communication in this article to process and recognize and compare the recognition results with traditional recognition methods, as shown in Figures 9 and 10.

It can be seen from the recognition results that Figures 9(b) and 10(b) are the results of using traditional feature recognition, and both recognition uses the selection model of the school's data. The comparison of the recognition results shows that the traditional recognition results are significantly lower than the recognition results based



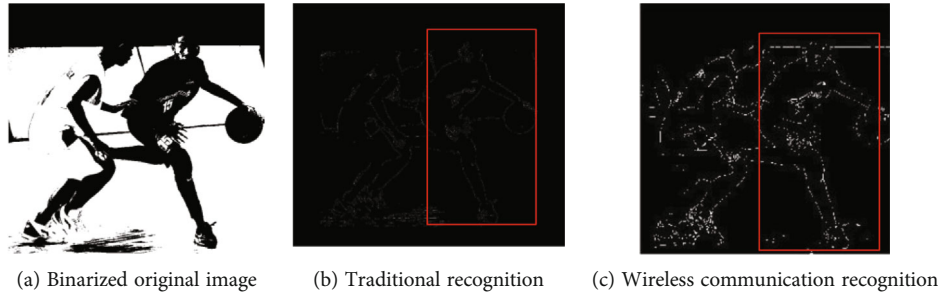


FIGURE 9: Basketball recognition results.

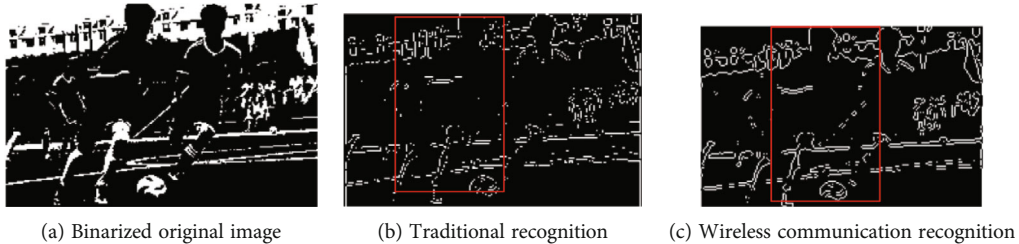


FIGURE 10: Football recognition results.

on the combination of characteristics of wireless communication used in this article. The addition of wireless communication can improve the recognition effect and is more suitable for the recognition of actions. This is because it predicts and selects pseudolabeled samples based on multiple base classifiers. Through the introduction of wireless communication, the accuracy of labeled samples is improved while the classification noise of training samples is better controlled. But the traditional method does not effectively solve the problem of distributed noise, so the classification result is slightly lower than the method we proposed.

At the same time, we also added the prediction results of the selected base classifiers KNN and SVM for comparison, which themselves are supervised learning methods. It can be seen from the table that the accuracy of our method has been significantly improved under different number of labeled samples, which further verifies the point of view in the previous section, as shown in Table 1.

It can be seen from the table that when the number of labeled samples is 220, the performance improvement of the classifier is the fastest. When the number of samples is 660, the existing labeled sample data can be used to train a classifier with better performance. By continuing to add unlabeled data, the performance of the model cannot be effectively improved.

In the process of system testing, many people have done multiple system tests for the motion recognition algorithm. 15 boys and 5 girls were invited to use a test experiment lasting 5 days each day, and 200 test samples were obtained. The recognition rate of exercise and the accuracy rate of exercise result evaluation are statistically shown in Table 2. Among them, the standing long jump is defined that the difference between the measured distance and the true distance is accurate within 2 cm.

TABLE 1: Comparison between the method in this paper and other methods.

Number of marks	SVM	KNN	Co-RAINING	MCM	Method of this article
220	0.47	0.51	0.60	0.75	0.77
440	0.64	0.66	0.69	0.77	0.80
660	0.73	0.74	0.76	0.77	0.79
Accuracy	0.61	0.66	0.69	0.76	0.79

TABLE 2: Experimental test analysis table.

Exercise type	Number of exercises	Motion recognition rate	Accuracy
Pull-ups	197	100%	99.2%
Squat	199	98%	97.3%
Stand too far	593	97.7%	88.7%

The test results show that the system has a very high recognition rate for sports, and the motion recognition function is very reliable, but the accuracy rate, except for the standing long jump, has reached the level of consumer-grade products, including pull-ups and squats. In order to explore the problem of low accuracy of standing long jump, a more detailed investigation was carried out. The comparison table of standing long jump test and actual distance was obtained by jumping from near to far within the effective visual range, and the results are shown in Table 3.

Analyzing the following experimental data, it is found that there are hidden dangers in the measurement distance accuracy of less than 1.4m and greater than 2.2m. There

TABLE 3: Standing long jump distance test table.

Test distance (m)	0.62	0.81	1.08	1.17	1.21	1.28	1.33	1.45	1.52	1.68
Actual distance (m)	0.55	0.74	0.99	1.08	1.18	1.26	1.27	1.43	1.54	1.68
Test distance (m)	1.71	1.77	1.87	1.94	2.07	2.09	2.21	2.27	2.27	2.31
Actual distance (m)	1.74	1.76	1.88	1.95	2.09	2.09	2.11	2.20	2.29	2.36

TABLE 4: Comparison of recognition rates of different methods on the database.

	Cross subject test			Test one			Test two		
	AS1	AS2	AS3	AS1	AS2	AS3	AS1	AS2	AS3
Hu	72.8	71.8	78.1	89.4	88.9	96.2	92.3	91.8	95.2
Cao	86.8	84.4	62.4	97.4	95.6	92.4	97.5	96.8	93.8
Shi	87.6	86.6	87.2	—	—	—	97.7	97.2	94.7
Our	88.1	79.4	85.6	93.8	89.4	95.5	97.5	95.6	94.6

may be several reasons for this problem: due to the visual detection of the somatosensory technology sensor, the detection distance of the ambient light will affect the accuracy of the environment detection, resulting in errors, and the information received by the sensor will cause a certain point collected by the camera to cause a certain image. The distortion of the image, especially the extreme position of the image, coupled with the height of the installation, will cause system errors [32]. In the algorithm, the complexity of long jump is much greater than the other two sports. Among them, the rapid movement of the long jump has affected the algorithm's estimation of the foot size and the accuracy of the calculated distance. Compare the parameters under different algorithms, and the results are shown in Table 4.

Comparing the test results of the two databases and other algorithms, it can be seen that the model action recognition in this article can achieve higher recognition results. Comparison of test results in the AS1 data set. The result of the first test is lower than the traditional model, while the results of the other two tests are higher. This is because in the Test One test, only 1/3 of the training data is used for training, and there are fewer training samples. The general deep learning model generally has inadequacy when processing data with fewer training samples. We compare it with other algorithms, as shown in Table 5:

From the summary of the table, it can be seen that in most cases, the speed and accuracy of the algorithm in this article are higher than those of other methods.

## 4. Discuss

*4.1. Algorithm Comparison.* The recognition rate of the algorithm proposed in this paper is higher than that of the traditional algorithm, and the recognition accuracy is increased by 1.3% and 2.6%, respectively. The reason is that the former first generates multiple classifiers, then predicts the unlabeled number, and selects the highly confident ones to be added to the training set for further training of the base classifier, during which no new classifiers are generated. There-

TABLE 5: Comparison of recognition rates of different methods on the database.

Action	Hu	Cao	Shi	Our
Walk	95.5	89.9	91.9	100
Sit down	90.4	99.8	95.9	99.9
Stand up	92.4	99.9	99.9	97.8
Pick up	96.4	100	98.7	99.7
Carry	97.4	68.5	72.4	92.6
Throw	58.9	94.9	87.8	89.9
Push	81.4	89.5	74	93.5
Pull	92.4	98	97	100

fore, as the number of cycles increases, the differences between classifiers gradually disappear. And if the initial sample data is not sufficient, it will also affect the accuracy of the classifier. The latter is based on the boosting method. Although a new classifier is generated during the iterative training process, and pseudolabeled data is selected by combining the confidence level and the similarity between the samples, it uses all the classifiers in the set to label the samples. Because the performance of earlier classifiers is often poor, the prediction effect of this method is not as good as that of selective recognition. In essence, these two methods are self-training methods. Our proposed wireless network communication retains the newly generated intermediate classifiers in the iterative training process and then predicts and selects unlabeled samples through different screened base classifier groups, which expands the difference between classifiers. Finally, the base classifiers are integrated based on evidence theory to improve the recognition performance of the system.

The accuracy of the traditional algorithm improves faster in the initial stage of iterative training, but the latter is relatively flat compared to the algorithm we proposed. The main reason is that the difference of the classifier is reduced. The method we propose is essentially the collaborative training of two different base classifier groups. By selecting a classifier with good performance based on the pseudovalidation set, it can not only reduce the cumulative effect of errors in the self-training process but also effectively slows down in the iterative training processes, the difference of the base classifier is reduced.

*4.2. Human Action Recognition.* This paper proposes wireless network communication for sports training action recognition. This method uses two different classifiers of KNN and SVM to construct a base classifier. The difference

between the largest and the second largest probability of the predicted sample category is used to define the sample prediction as the category with the largest predicted probability. For a subset of unlabeled samples in the training set, use a classifier to predict unlabeled samples, take a certain percentage of samples before the confidence of the predicted category as pseudolabel samples, and add them to the training set of another classifier. Design the classification decision of the base classifier combination to realize the classification and recognition of test samples. This method is used for video action recognition. The experimental results show that this method can effectively improve the accuracy of video action recognition and reduce the workload of class labeling of action samples when the number of labeled action samples is insufficient.

In view of the increasing number of iterations of collaborative training, the difference of the base classifier gradually degenerates, inspired by ensemble learning, the use of ensemble learning to create the difference of the base classifier. This method divides the base classifier into two sets and selects different sample sets for initial training of the classifiers in the two sets. Defines a confidence calculation method based on the edge of the largest evidence, selects a certain proportion of unlabeled samples by category, and uses these two sets of classifiers to identify these unlabeled samples. If the two sets of classifiers identify the category with the highest total confidence of the sample category, if they are consistent, the sample is used as the pseudolabel sample of this category and added to all training sets for iterative training of each classifier. After the iterative training is over, the base classifier group is selected according to the performance and prediction closeness of the newly generated base classifier on the pseudovalidation set. Use the confidence of the sample class based on the edge of the largest evidence to predict the class of the test sample. This method is used for video action recognition, and the experimental results show that this method effectively reduces the differential degradation of the base classifier and further improves the accuracy of human action recognition.

## 5. Conclusions

Human action recognition is currently a hot and challenging research topic. Although this article has achieved certain research results in semisupervised recognition methods, there are still many problems in the field of human action recognition that need to be resolved. In real scenarios, under the influence of complex factors such as noise, occlusion, and shadow, the recognition effect is often not satisfactory. Moreover, due to the limitation of the storage capacity of the actual storage system, low-resolution human motion videos are often recorded. Although many scholars have conducted research on these issues, there is still a lack of systematic and universal methods. Therefore, more research is needed to solve these practical problems. Of course, there are some shortcomings in the research of this paper. The data used is only from two databases, and there may be some missing samples. Moreover, due to the limitation of the storage capacity of the actual storage system, low-resolution

human motion videos are often recorded. Therefore, more research is needed to solve these practical problems.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The author states that this article has no conflict of interest.

## References

- [1] H. Xu and R. Yan, "Research on sports action recognition system based on cluster regression and improved ISA deep network," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 4, pp. 5871–5881, 2020.
- [2] A. Abdulmunem, Y.-K. Lai, and X. Sun, "Saliency guided local and global descriptors for effective action recognition," *Media*, vol. 2, no. 1, pp. 97–106, 2016.
- [3] C. Sun and D. Ma, "SVM-based global vision system of sports competition and action recognition," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2265–2276, 2021.
- [4] Y. L. Shih and C. Y. Lin, "The relationship between action anticipation and emotion recognition in athletes of open skill sports," *Cognitive Processing*, vol. 17, no. 3, pp. 259–268, 2016.
- [5] R. Vemulapalli, F. Arrate, and R. Chellappa, "R3DG features: relative 3D geometry-based skeletal representations for human action recognition," *Computer Vision & Image Understanding*, vol. 152, pp. 155–166, 2016.
- [6] Y. Yang, C. Deng, S. Gao, W. Liu, D. Tao, and X. Gao, "Discriminative multi-instance multitask learning for 3D action recognition," *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 519–529, 2017.
- [7] P. Wang, W. Li, Z. Gao, J. Zhang, C. Tang, and P. O. Ogbona, "Action recognition from depth maps using deep convolutional neural networks," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 4, pp. 498–509, 2016.
- [8] D. Das Dawn and S. H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector," *The Visual Computer*, vol. 32, no. 3, pp. 289–306, 2016.
- [9] Y. Du, Y. Fu, and L. Wang, "Representation learning of temporal dynamics for skeleton-based action recognition," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3010–3022, 2016.
- [10] C. Li, Y. Hou, P. Wang, and W. Li, "Joint distance maps based action recognition with convolutional neural networks," *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 624–628, 2017.
- [11] M. R. Malgireddy, I. Nwogu, and V. Govindaraju, "Language-motivated approaches to action recognition," *Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2189–2212, 2017.
- [12] Q. Ke, S. An, M. Bennamoun, F. Sohel, and F. Boussaid, "SkeletonNet: mining deep part features for 3-D action recognition," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 731–735, 2017.
- [13] X. Wang, L. Gao, J. Song, and H. Shen, "Beyond frame-level CNN: saliency-aware 3-D CNN with LSTM for video action recognition," *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 510–514, 2017.

- [14] D. Carbonera Luvizon, H. Tabia, and D. Picard, "Learning features combination for human action recognition from skeleton sequences," *Pattern Recognition Letters*, vol. 99, pp. 13–20, 2017.
- [15] K. Charalampous, I. Kostavelis, and A. Gasteratos, "Robot navigation in large-scale social maps: an action recognition approach," *Expert Systems with Applications*, vol. 66, pp. 261–273, 2016.
- [16] H. Xu, Q. Tian, Z. Wang, and J. Wu, "A survey on aggregating methods for action recognition with dense trajectories," *Multimedia Tools & Applications*, vol. 75, no. 10, pp. 5701–5717, 2016.
- [17] H. Kataoka, Y. Satoh, Y. Aoki, S. Oikawa, and Y. Matsui, "Temporal and fine-grained pedestrian action recognition on driving recorder database," *Sensors*, vol. 18, no. 2, pp. 627–629, 2018.
- [18] Z. Zhao, H. Ma, and X. Chen, "Semantic parts based top-down pyramid for action recognition," *Pattern Recognition Letters*, vol. 84, pp. 134–141, 2016.
- [19] F. Murtaza, M. H. Yousaf, and S. A. Velastin, "Multi-view human action recognition using 2D motion templates based on MHIs and their HOG description," *IET Computer Vision*, vol. 10, no. 7, pp. 758–767, 2016.
- [20] Y. Kong, B. Satarboroujeni, and Y. Fu, "Learning hierarchical 3D kernel descriptors for RGB-D action recognition," *Computer Vision and Image Understanding*, vol. 144, pp. 14–23, 2016.
- [21] X. Wang and C. Qi, "Action recognition using edge trajectories and motion acceleration descriptor," *Machine Vision and Applications*, vol. 27, no. 6, pp. 861–875, 2016.
- [22] S. Y. Chun and C. S. Lee, "Human action recognition using histogram of motion intensity and direction from multiple views," *IET Computer Vision*, vol. 10, no. 4, pp. 250–257, 2016.
- [23] B. Victor, A. Nibali, Z. He, and D. L. Carey, "Enhancing trajectory prediction using sparse outputs: application to team sports," *Neural Computing and Applications*, vol. 33, no. 18, pp. 11951–11962, 2021.
- [24] J. P. Pluijms, R. Cañal-Bruland, M. J. M. Hoozemans, M. W. van Beek, K. Böcker, and G. J. P. Savelsbergh, "Quantifying external focus of attention in sailing by means of action sport cameras," *Journal of Sports Sciences*, vol. 34, no. 16, pp. 1588–1595, 2016.
- [25] R. Prerna and C. K. Singh, "Evaluation of LiDAR and image segmentation based classification techniques for automatic building footprint extraction for a segment of Atlantic County, New Jersey," *Geocarto International*, vol. 31, no. 6, pp. 694–713, 2016.
- [26] S. E. Jóhannesson, B. Davíðsdóttir, and J. T. Heinonen, "Standard ecological footprint method for small, highly specialized economies," *Ecological Economics*, vol. 146, pp. 370–380, 2018.
- [27] A. Milosavljević, "Automated processing of remote sensing imagery using deep semantic segmentation: a building footprint extraction case," *International Journal of Geo-Information*, vol. 9, no. 8, pp. 486–488, 2020.
- [28] C. Gong, Y. Hu, J. Gao, Y. Wang, and L. Yan, "An improved delay-suppressed sliding mode observer for sensorless vector-controlled pmsm," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5913–5923, 2019.
- [29] Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, and Y. Pan, "Generative adversarial networks: a survey toward private and secure applications," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–38, 2021.
- [30] S. Yang, B. Deng, J. Wang et al., "Scalable digital neuromorphic architecture for large-scale biophysically meaningful neural network with multi-compartment neurons," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 148–162, 2019.
- [31] H. Wang, C. Dong, and Y. Fu, "Optimization analysis of sport pattern driven by machine learning and multi-agent," *Neural Computing and Applications*, vol. 33, no. 4, pp. 1067–1077, 2021.
- [32] T. Sui, D. Marelli, X. Sun, and M. Fu, "Multi-sensor state estimation over lossy channels using coded measurements," *Automatica*, vol. 111, article 108561, 2020.