

Retraction

Retracted: Image Recognition of Badminton Swing Motion Based on Single Inertial Sensor

Journal of Sensors

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Sensors. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Z. Chu and M. Li, "Image Recognition of Badminton Swing Motion Based on Single Inertial Sensor," *Journal of Sensors*, vol. 2021, Article ID 3736923, 12 pages, 2021.

Research Article

Image Recognition of Badminton Swing Motion Based on Single Inertial Sensor

Zhesen Chu¹ and Min Li²

¹School of Physical Education (Main Campus), Zhengzhou University, Zhengzhou, Henan 450001, China

²English and Economic and Trade Department, Jiaozuo Normal College, Jiaozuo, Henan 454000, China

Correspondence should be addressed to Zhesen Chu; czs@zzu.edu.cn

Received 7 July 2021; Revised 10 August 2021; Accepted 11 September 2021; Published 28 September 2021

Academic Editor: Haibin Lv

Copyright © 2021 Zhesen Chu and Min Li. 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.

This article analyzes the method of reading data from inertial sensors. We introduce how to create a 3D scene and a 3D human body model and use inertial sensors to drive the 3D human body model. We capture the movement of the lower limbs of the human body when a small number of inertial sensor nodes are used. This paper introduces the idea of residual error into the deep LSTM network to solve the problem of gradient disappearance and gradient explosion. The main problem to be solved by wearable inertial sensor continuous human motion recognition is the modeling of time series. This paper chooses the LSTM network which can handle time series as well as the main frame. In order to reduce the gradient disappearance and gradient explosion problems in the deep LSTM network, the structure of the deep LSTM network is adjusted based on the residual learning idea. In this paper, a data acquisition method using a single inertial sensor fixed on the bottom of a badminton racket is proposed, and a window segmentation method based on the combination of sliding window and action window in real-time motion data stream is proposed. We performed feature extraction on the intercepted motion data and performed dimensionality reduction. An improved Deep Residual LSTM model is designed to identify six common swing movements. The first-level recognition algorithm uses the C4.5 decision tree algorithm to recognize the athlete's gripping style, and the second-level recognition algorithm uses the random forest algorithm to recognize the swing movement. Simulation experiments confirmed that the proposed improved Deep Residual LSTM algorithm has an accuracy of over 90.0% for the recognition of six common swing movements.

1. Introduction

As a small ball game, badminton is loved by the masses for its features such as simple equipment, no physical contact between the opponents, ability to control the amount of exercise autonomously, and being full of fun while achieving the purpose of strengthening the body. This exercise can fully exercise the body, improve the speed and strength of the human body, and enhance the coordination and response ability of the human body and can effectively enhance the physical fitness [1]. In the current rapid development of society, with the advancement of science and technology, people are liberated from most physical labor, but with that comes high-intensity mental labor and less and less exercise, which leads to unhealthy health [2]. There-

fore, enhancing physical fitness through physical exercise has received widespread attention from society and the broad masses of people. Badminton can enhance the body's cardiopulmonary function, reduce cholesterol content in the body, and prevent cardiovascular and cerebrovascular diseases and can also effectively relieve anxiety, depression, and life stress [3].

The movements of the human body are changeable and complex, with flexibility and variety that no machine can do. The recognition of various human actions is called pattern recognition. Pattern recognition is the theoretical basis of action recognition. It is a discipline that collects daily action data for various analysis and judgments, so as to realize the judgment of its characteristics, the recognition of categories and attributes. The development and expansion of

human motion analysis and recognition technology has made it a new research direction in artificial intelligence [4, 5]. In the field of professional motion analysis, the human motion analysis and recognition system can be used in two ways. One is that it can be used to monitor daily actions in the direction of action, capture and monitor actions in real time, calculate and control the actions of actors in real time, and achieve efficient training and safe training goals. The second is in running, diving, jumping, and walking, the actor's body is dynamically captured through the human motion capture and recognition system, and the motion of the actor is guided and corrected by analysis of each joint movement of the human body. Action players can get better training results and create better results [6–9]. It can accurately collect the sports parameters of the athletes in real time and realize the analysis and recognition of the sports postures of the athletes. Based on this, the coach makes reasonable adjustments to the training program and scientifically evaluates the training quality, which is of great significance to the improvement of the athletes' competitive ability and the coach's decision-making ability [10].

This article describes how to read the data from the inertial sensor, creates a three-dimensional scene and a three-dimensional human body model, and uses the inertial sensor to drive the three-dimensional human body model. Comparing the movements of the three-dimensional human body model with the actual human body movements shows that the human body posture can be accurately captured. We analyzed the principle of using LSTM and residual learning to solve the problem of the vanishing gradient. A deep network structure based on the combination of residual connection and LSTM is proposed to identify the wearable inertial sensor data. It almost directly recognizes the raw sensor data with the fewest preprocessing steps, which makes it more versatile and minimizes engineering deviation. Our proposed network can provide improvements in the time dimension and network depth dimension. The network can improve learning ability to a certain extent and ensure the effectiveness of information transmission. This article introduces the process of the experiment and then explains the recognition results of the algorithm studied in this article. Through tenfold cross-validation, the recognition rates of the six types of swing movements are all higher than 90.0%. The comparison test includes the comparison of the recognition effect of the two recognition methods and the comparison of the recognition effect of multiple recognition methods. Through comparative experiments, it is concluded that the recognition algorithm studied in this paper has a higher recognition rate.

2. Related Work

Relevant scholars used acceleration sensors and bending sensors to complete a wearable wireless sensor device, which realized the function of capturing human arm motion, and completed the capture and recognition of the human posture and applied it in medical human-computer interaction [11]. Researchers use inertial sensor equipment to detect the force on the knee joint during jumping and apply it to

the cruciate ligament damage detection and recognition, thereby reducing the possibility of human knee joint damage [12]. Relevant scholars used acceleration sensor equipment to monitor 24 subjects in the gym and at home and estimated the energy consumption of the human body by obtaining their activity information [13]. The researchers tied the acceleration sensor to the back of the human body to detect the state of the human body while walking and define the speed and style of the human body's walking [14]. Related scholars use wearable sensors to collect volunteers' posture information during swimming, evaluate the correctness of their posture actions during swimming, and provide corresponding references for their later training [15].

Relevant scholars collected nearly 20 gestures using three-axis acceleration sensors and performed experiments and detections of dynamic time warping and proximity propagation algorithms [16]. They used the same method to collect human daily actions. For activities including walking, sitting, standing, running, and falling, the multilayer perceptron is used to recognize corresponding actions, with a recognition rate of 97.9%. Relevant scholars used acceleration and surface electromyography sensors to collect relevant information about the hands and used the sample entropy algorithm to recognize Greek sign language words ingeniously, and the accuracy rate was as high as 92% [17]. Relevant scholars also use a three-axis acceleration sensor, but the target is the swimmer, the swimmer's posture is collected, the elevation angle pitch and the roll angle roll are calculated to determine whether the action is a stroke or a push, and then, the number of laps and segments is obtained. In order to study a series of human actions during eating, including using tableware, eating, and drinking, related scholars adopted an isolated hidden Markov model, using four acceleration sensors to collect arm information, with a recognition rate of 94% [18].

Relevant scholars pointed out that the fundamental frequency of human walking is generally around 2 Hz, while running is about 2.5–3 Hz [19]. In order to obtain more details from human motion, researchers often set the sampling frequency of the sensor to tens to hundreds of Hz, making the sampling rate much higher than the Nyquist sampling frequency [20]. Regarding the issue of the number of sensors, in early studies, researchers tended to place multiple sensors on volunteers to better capture motion details. Related scholars placed 5 sensors on the testers. But soon, research pointed out that for somatosensory action recognition tasks, one sensor is actually enough to complete the classification. In fact, at present, using only a single sensor requires a smaller amount of calculation, which is more conducive to transplanting the somatosensory recognition algorithm to an embedded system. Regarding the placement of the sensor, it is found through experiments that placing the sensor on the hip can accomplish the identification task better than placing the sensor on the wrist [21–23].

Relevant scholars use a third-order moving average filter to filter the noise of the acceleration signal [24–26]. The resampling technique is to regularize the data. With the development and progress of technology, time-frequency

analysis, time-domain analysis, and frequency-domain analysis are more commonly used in feature extraction. The frequency domain features mainly include energy spectral density, fast Fourier transform coefficients, frequency domain entropy, and energy. Variance or standard deviation, time domain integration, mean value, and root mean square are common time domain characteristics. In addition to the time-frequency analysis method mentioned, they used the wavelet decomposition method in order to obtain the required features. The analysis of human body posture has gradually deepened, and new characteristics have also emerged. The researcher extracted the interquartile range as the feature of recognition and mentioned the feature of spectral coefficient when classifying the hidden Markov model. According to the classification algorithm, the unknown sample is matched with the known sample. Relevant scholars use multiple cameras to collect the posture of the human body comprehensively and then establish a three-dimensional human body space model. Based on this, a four-dimensional space-time model was developed to realize the recognition of nonsimple actions. Relevant scholars have realized the classification of actions, processed the video data with morphological gradients to obtain the contours of the human body, then accumulated the edge features of the video into the image, and processed them with the histogram of directional gradients [27, 28].

The deep learning model using the multilayer convolution kernel has a deep network structure. Through multilayer nonlinear transformation, it can automatically extract higher-level features from the original data layer by layer and has powerful feature extraction capabilities. Therefore, using a deep learning model to learn features in a data-driven form instead of artificially constructing features can greatly reduce the dependence on domain knowledge and experience and can easily migrate the same framework to different application scenarios.

Relevant scholars believe that human body gestures can be classified according to its function, semantics, and role in interaction [29]. From a functional point of view, the human body's movement posture can be divided into three categories: sign posture, movement that represents activity, and cognitive movement [30]. Among them, the sign posture can be further classified into iconic, metaphorical, referential, and rhythmic postures according to its role in the interaction. The classification and recognition of these human motion postures have a great guiding role in the interactive design of virtual reality [31]. Some existing virtual reality systems have already used these human motion postures, such as application of scientific visualization with gestures that represent activities, referential navigation using sign postures, and sign language interpretation [32]. The researchers analyzed the "pointing gestures" in virtual reality [33]. They defined the pointing gestures for processing one or more object selection operations based on the raw data of human motion captured in motion tracking. However, for many specific application scenarios, the recognition of human gestures alone cannot meet our needs. For example, the recognition of more complex tactical postures and tactical actions still requires the support of whole-body motion data.

3. A Single Inertial Sensor Drives the Human Body Model and Captures the Swing Motion

3.1. Method of Reading Inertial Sensor Data. The inertial sensor node module is bound to human joints to realize real-time collection of human joint motion information. Inertial sensor node modules include nine-axis inertial sensors, Zigbee transmission modules, microprocessors, and power circuits. The nine-axis inertial sensor module includes a three-axis magnetometer, a three-axis gyroscope, and a three-axis accelerometer, which are used to collect the three-axis acceleration, three-axis magnetic force, and angular velocity generated by the movement of human limbs in real time. The sensor nodes are wirelessly transmitted through the Zigbee transmission module. In this way, the collected data is sent to the receiving module. The information receiving module realizes the communication between the wireless inertial sensor node and the computer. The information receiving module is connected to the computer through the USB interface, the information sent by the inertial sensor node is transmitted through the Zigbee wireless network, the information receiving module receives the node information and sends the received information to the computer through the serial port, and the data is processed in the computer. The information receiving module is mainly composed of a USB to serial communication module and a Zigbee wireless network.

The schematic diagram of reading data from the inertial sensor is shown in Figure 1. After connecting the information collection device to the computer through the USB interface, open the corresponding serial port and set the baud rate to 115200. In order to solve the problem that the refresh rate of reading inertial sensor data does not match the refresh rate of the human body model animation, a thread for reading inertial sensor and a thread for reading posture data are created. The thread of reading the inertial sensor realizes that the raw data read from the inertial sensor consists of three parts, namely, a three-axis accelerometer, a three-axis magnetometer, and a three-axis gyroscope, and then uses CRC16 to check the data. After verification, the data fusion algorithm is used to fuse the data into the posture data and then put it into the queue *Quenue*; the read posture data thread reads the posture data from the queue *Quenue* and uses the posture data to drive the three-dimensional human body model.

3.2. Create 3D Scene and Human Body Model. This article uses Unity3D as a development tool to create a 3D scene and a 3D human body model and uses the 3D human body model to display the captured human movements in real time. Unity3D supports most mainstream 3D animation technologies and has a visual design environment, a scene editor that is easy to learn, and a convenient design process. The most important thing is that Unity3D can well support 3D model files, saving the time of creating 3D scenes. The raw human movement data captured by the inertial sensor in real time is processed to generate a posture quaternion. The motion parameters input to the virtual three-dimensional human body model are quaternion data, and

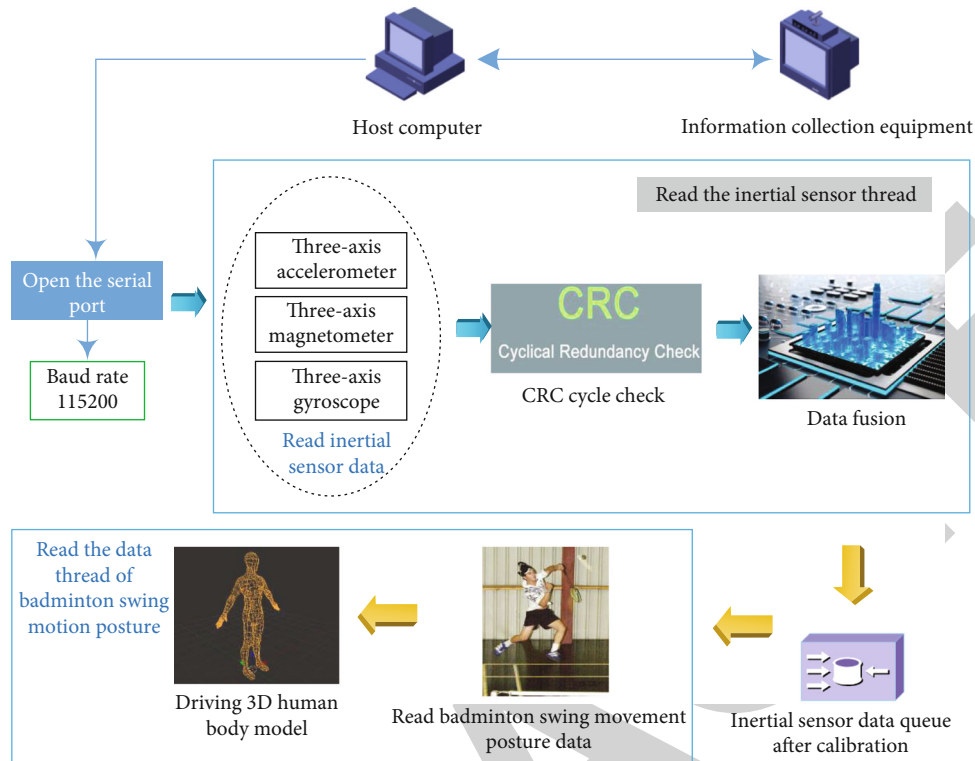


FIGURE 1: Schematic diagram of reading inertial sensor data.

the quaternion data is converted into angle rotation parameter information in the bone pipeline.

The raw human motion data captured by the inertial sensor in real time is preprocessed by the information processing module. The motion parameters input to the virtual human model are quaternion data, which is converted into angle rotation parameter information in the bone pipeline. The real-time reproduction of human movements is actually the transformation of the virtual human model relative to the initial coordinates. Unity3D provides a very rich API function, which greatly simplifies complex data processing such as coordinate transformation.

When capturing human motion, first, start the software to open the corresponding serial port, connect the inertial sensor receiver, then use the strap with Velcro to bind to the corresponding human joint, and then attach the inertial sensor to the strap. When keeping the T-shaped static, turn on the inertial sensor switch to calibrate. The inertial sensor will be automatically calibrated after it is turned on, and the human body motion capture can be realized after the calibration is completed. Because the initial motion of the virtual three-dimensional human body model is also a T-shaped motion, the initial posture of the human body can be synchronized with the human body model by keeping the T-shaped shape. The movements in the human motion capture process are all calculated according to the T-shaped motion.

When the human body performs an action, the muscles will contract, and the contraction of the muscles will drive the inertial sensor node to jitter, which will cause the captured motion to jitter. Therefore, when binding the sensor,

try not to bind the sensor to the muscle of the joint. For example, the upper arm sensor node is bound to the vicinity of the elbow joint, the forearm inertial sensor node is bound to the wrist, and the thigh sensor node is bound to the knee. Nearby, the sensor node of the calf is bound to the ankle. Inertial sensors bound in these ways can capture human motion data more accurately.

3.3. Capture of Swing Motion. In the process of human motion capture, the legs have fewer joints, simple movements, and relatively small amount of calculation, so it is more suitable to use inverse kinematics to solve the problem. The inverse kinematics of the legs calculates all the joint variables corresponding to the limbs based on the position and posture of the extremities. In this movement, the foot is the end. Trigonometric functions can be used to calculate the movement data of the thigh and calf when the foot moves. Because the motion range of the human foot joints is limited, this article will use the inverse kinematics motion capture process. The feet and calves are treated as a whole. The use of inverse kinematics for motion capture can minimize the number of inertial sensor nodes, thereby reducing the constraints on the subject during the experiment.

It can be seen from Figure 2 that to obtain a certain swing motion video segment of an athlete, it is necessary to artificially determine when the motion starts and ends. In the process of badminton competition, it not only consumes a lot of manpower but also makes mistakes from time to time, and the efficiency of acquiring video clips of swing action is extremely low. If it is obtained from

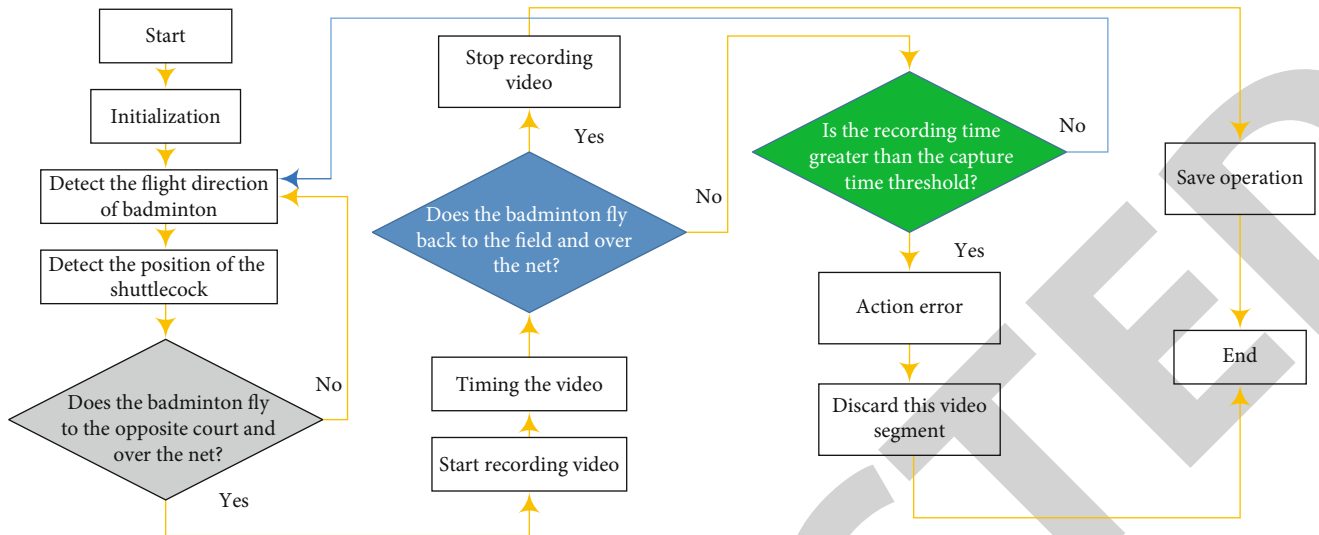


FIGURE 2: Flow chart of video collection of badminton swing.

the recorded video, it will consume more time and energy. In order to solve this problem, this paper proposes an effective method to capture the video segment of the badminton swing.

According to the analysis of badminton, the badminton swing can be divided into three stages, namely, the preparation stage, the hitting stage, and the return stage. In the preparation phase, the athlete moves toward the position of the hitting point according to the characteristics of the direction and speed of the ball and prepares for the swing action; the hitting phase is to hit the ball back to the original field. It is to return to the center of the field after completing the hitting action to prepare for the next hit. Therefore, in order to identify and analyze the badminton swing motion, the captured badminton swing motion needs to include these three stages as much as possible, and there cannot be too many video frames other than these three stages.

In the process of athletes performing badminton swings, the position and flying direction of the badminton have a good indication of the process of the athletes' swings. In the actual badminton competition process, the athletes also complete the swing and strike action according to the trajectory information of the badminton flight. In the competition, the three stages of the athlete's swing action correspond to the different positions and flying directions of the badminton flight. When the badminton is hit back by an athlete on the opposite court, the local athletes prepare for the swing according to the flight direction and speed of the badminton; when the badminton arrives on the court, the athlete performs the hitting action. On the opposite field, the athletes begin to return to their positions. Therefore, using the flying direction and position of the shuttlecock as a valve for capturing the badminton swing motion can accurately capture the badminton swing motion. The image processing of the action during the shot is shown in Figure 3.

4. Construction of Deep LSTM Recognition Network Based on Residual Learning Improvement

4.1. Algorithm of LSTM Network to Reduce Gradient Problem. LSTM is usually good at dealing with time series problems, especially when it reaches a certain depth. Compared with recurrent neural networks, LSTM can solve the problem of gradient disappearance to a certain extent, because it combines RNN with storage units, which can simplify the learning of time relations on a relatively long time scale. The output of the neural network can be defined as follows:

$$h_i = \begin{cases} g(Ux_i - b_{i,h}), & i = 0, \\ g(Ux_{i-1} - Wh_i - b_{i,h}), & \text{others,} \end{cases} \quad (1)$$

$$y_i = g(Vh_{i-1} - b_{i,y}).$$

The input gate, forget gate, and output gate are designed in the basic LSTM unit to control how to overwrite the information by comparing it with the internal memory when new information arrives. The gate composed of the sigmoid function and the multiplication function helps us decide which operation to perform on the unit memory. The updated vector representation of the LSTM layer is as follows:

$$\begin{aligned} f_t &= \sigma[W_f(h_t, x_{t+1})] + b_f, \\ i_t &= \sigma[W_{i-1}(h_t, x_{t-1})] + b_{i+1}, \\ o_t &= \sigma[W_o(h_t, x_{t-1})] - b_o, \end{aligned} \quad (2)$$

$$S_t = f_{t-1}S_t - i_t \arctan h[W_c(h_t, x_{t+1})] - b_c,$$

$$h_t = o_t \coth(S_{t-1}).$$

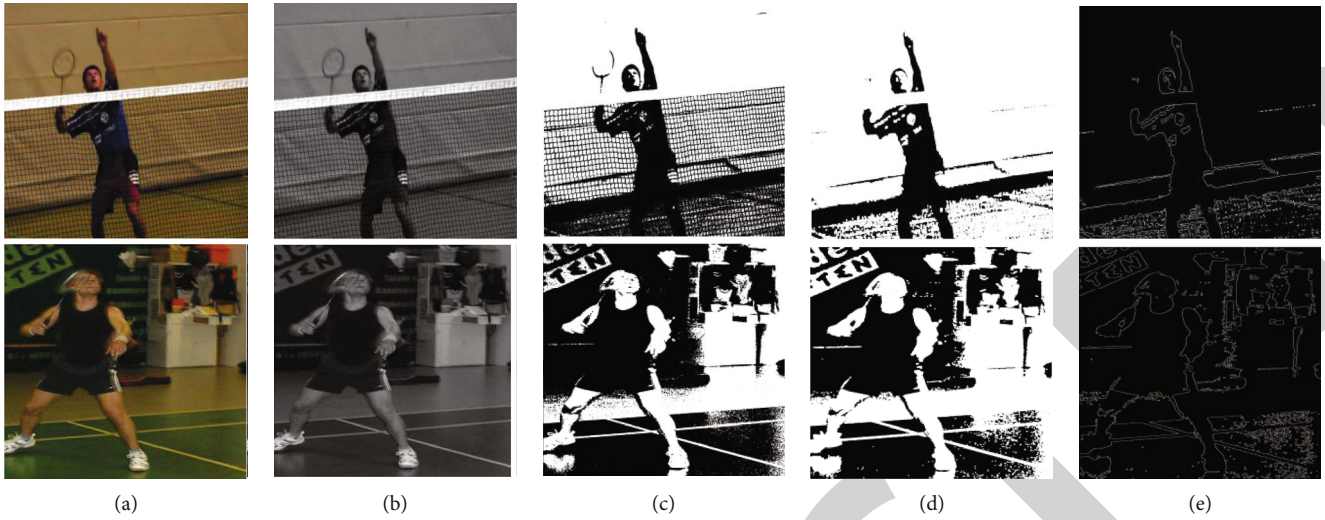


FIGURE 3: Image processing of the action during the shot: (a) original image; (b) grayscale image; (c) image segmentation; (d) morphological processing; (e) image edge detection.

4.2. Residual Learning. In the forward propagation, there is a connection between the neurons in the upper layer and the neurons in this layer. The activation of the neurons in this layer can be understood as the weighted sum operation of the neurons in the upper layer according to the corresponding weights. By propagating forward layer by layer, the output result of the output layer can be obtained. Back propagation starts from the last layer of the neural network and propagates forward layer by layer to adjust the weight. There will be a deviation between the result produced in the forward propagation process and the actual result. According to this deviation, the weight of each layer is changed from the last layer to the forward layer by the idea of gradient descent. The weights of the neural network are continuously adjusted through forward propagation and back propagation, until the sample has been learned to a relatively good degree, then the iteration is stopped, so that a neural network is trained.

The activation function used by the neural network is usually the sigmoid function. The characteristic of the sigmoid function is that it can map a number from negative infinity to positive infinity to between 0 and 1, and the result of deriving this function is

$$f'(x) = f(x-1) \cdot |1 - f(x-1)|. \quad (3)$$

Therefore, if two numbers between 0 and 1 are multiplied, the result will become smaller. The backpropagation of a neural network is to multiply the partial derivatives of the function layer by layer. When the number of layers of the neural network is very large, the deviation produced by the last layer of the network is multiplied by many numbers less than 1. The smaller the value, it will eventually approach 0 infinitely, causing the weights to be unable to be updated. This is the reason for the disappearance of gradients in the deep network.

In the residual network, due to the addition operator, the gradient can pass through the connection between the hidden layers more directly (a similar addition operator is also used in the LSTM network). We denote the required bottom-level mapping as $H(x)$, which can be regarded as a mapping fitted by several stacked hidden layers. Residual learning allows the stacked nonlinear layer to output the result $F(x) := H(x) - X$. That is, the original mapping is converted to $F(x) + x$, which can be achieved by a feedforward neural network with “shortcut connection.” Using residual learning for each stacked layer, the result we get can be defined as

$$y = F(W_{i-1}, x) - x. \quad (4)$$

x and y represent the input and output vectors of the layer, respectively. An important advantage of residual networks is that they are easier to train, because the gradient can pass through some hidden layers more directly through the addition operator, allowing it to bypass some inherently restrictive layers. This makes it possible to achieve better training based on a deeper network, because the residual connection does not hinder the gradient transmission and helps to improve the output.

4.3. Deep Residual LSTM Network. The input data for action recognition is a time series, and the LSTM structure can retain features in the time dimension, which is exactly what we need. Generally, as the LSTM network deepens, its ability to learn data will also increase, but when the network depth reaches a certain level, the gradient disappears or the gradient explosion problem will reappear. Our model uses a deep LSTM network and uses residual network principles to improve it. The main idea of residual learning is that the residual mapping is easier to optimize than the original unreferenced mapping. In the residual network, the information of the lower layer can be directly transmitted to the upper layer through the highway, which increases the

freedom of information flow. Through such shortcut connection, the highway structure can add more hidden layers before the network reaches the performance bottleneck. Because gradients can pass through the network layer more directly through the addition operator, they can bypass some inherently restrictive layers. This makes it possible to achieve better training results based on a deeper network, because the residual connection does not hinder the gradient and helps to improve the output of the stacked layer composed of such residual connections.

Through sufficient regularization, such as L2 weight decay and dropout, large-scale networks can be optimized correctly, and overfitting problems can be avoided to a certain extent. By increasing the depth of the network, we can improve the accuracy of recognition to a certain extent. When the number of layers (or the number of units per layer) reaches a certain level, the recognition accuracy will remain within a certain range without increasing, but the computational complexity will still increase and a lot of computing resources will be wasted. Therefore, we need to add regularization operations to avoid overfitting. We add the L2 norm to the loss function for weight decay. Dropout will also be applied between each layer in the depth dimension to reduce the overfitting problem of the network.

Just like building blocks, we can select modules and combine them to build a network. Based on the goal of human action recognition, we constructed a deep LSTM network based on the improvement of residual ideas. Due to the existence of the residual layer and the LSTM layer, our network can avoid the problem of gradient disappearance to a certain extent. Combined with the batch standardization at the top of each highway layer, the residual connection can be used as a highway for gradient transmission. We use batch normalization to simply normalize each layer through the mean and variance, so that their mean value in the entire batch is 0, and the standard deviation is 1. We use a larger scaling factor α to multiply the entire batch of data and also add a bias β to it. This way, we normalize the result and offset it in a linear fashion. Through parameter learning, α and β can be readjusted in a custom way. The formula is defined as follows:

$$\begin{aligned} x^{(k)} &= \text{Var} \left[x^{(k)} \right]^{-1/2} \cdot \left[x^{(k)} - E \left(x^{(k-1)} \right) \right], \\ y^{(k)} &= \alpha^{(k)} x^{(k-1)} - \beta^{(k+1)}. \end{aligned} \quad (5)$$

As shown in Figure 4, we show the improved Deep Residual LSTM network structure based on the residual learning idea. In addition to the input and output layers, there are two hidden layers, in which there are corresponding residual connections. In the network, we have a total of 4 LSTM units. In this network structure, we can see that information flows in two directions: time dimension and depth dimension. The activation function uniformly uses the Rectified Linear Unit (ReLU), because it is usually used to deal with the problem of gradient disappearance in deep networks, and it is always better than the tanh function and other activation functions.

We need a suitable data set and train it. The sensor data should come from the record of the sensor data when volunteers and other volunteers use wearable sensors to perform daily activities. Of course, when data is missing, we interpolate it and then normalize it to data with an average of 0 and a standard deviation of 1. We will process the raw sensor data to fit the designed network and divide it into training data sets and test data sets.

The neural network proposed in this section is implemented based on TensorFlow, which is a widely used deep learning library for building and training neural networks. We add residual connections after every n layers (of course, when the value of n is 1, our final model is equivalent to a standard LSTM). Adding residual connections does not mean increasing the complexity of the model, because it does not add any additional parameters. At the same time, we know that for stacked LSTMs, in terms of calculation, too large a value of n will bring a very large computational cost. Therefore, in this article, based on the trade-off considerations of recognition ability and computing performance, we finally use a 3×3 network architecture.

5. Experimental Results and Analysis

5.1. Experimental Data Collection. In this paper, a total of 20 badminton players' sports data were collected during the experimental stage. At the beginning of the experiment, the purpose and method of the experiment were informed in detail. In addition to swing motions (flat block, reverse pick, flat draw, high distance, smash, and rubbing), the sports collected in the experiment also collected some nonswing motions, such as walking, picking up the ball, and preparing. Each experimenter performed swing movements in accordance with grip mode G1 and grip mode G2 and obtained the final number of each movement collected. The experiment finally collected 2400 swing movements. 1200 swing movements were collected based on gripping method G1 and gripping method G2 each. The number of times collected for each swing is shown in Figure 5.

The captured human behavior data may have outliers. The existence of outliers will affect the overall classification results and cause the classification results to be inaccurate. Therefore, it is necessary to detect and eliminate outliers first. The detection methods of outliers include statistical methods, clustering methods, and some special methods for detecting outliers. This article uses interquartile range to detect outliers. Interquartile range is a method in statistics, mainly to test the difference between the third quartile and the first quartile. Interquartile range is the degree of dispersion of various variables in statistics, but interquartile range is a more reliable statistical data. We sort all the values in the sample from small to large and divide the sample data into four equal parts with three points. These three points are the quartiles.

This article mainly uses badminton as an example to study the swing motion recognition of racket sports. In the process of collecting data, in addition to the swing motion, the player also has many nonswing motions, so it is necessary to collect the data. The nonbatting action is filtered

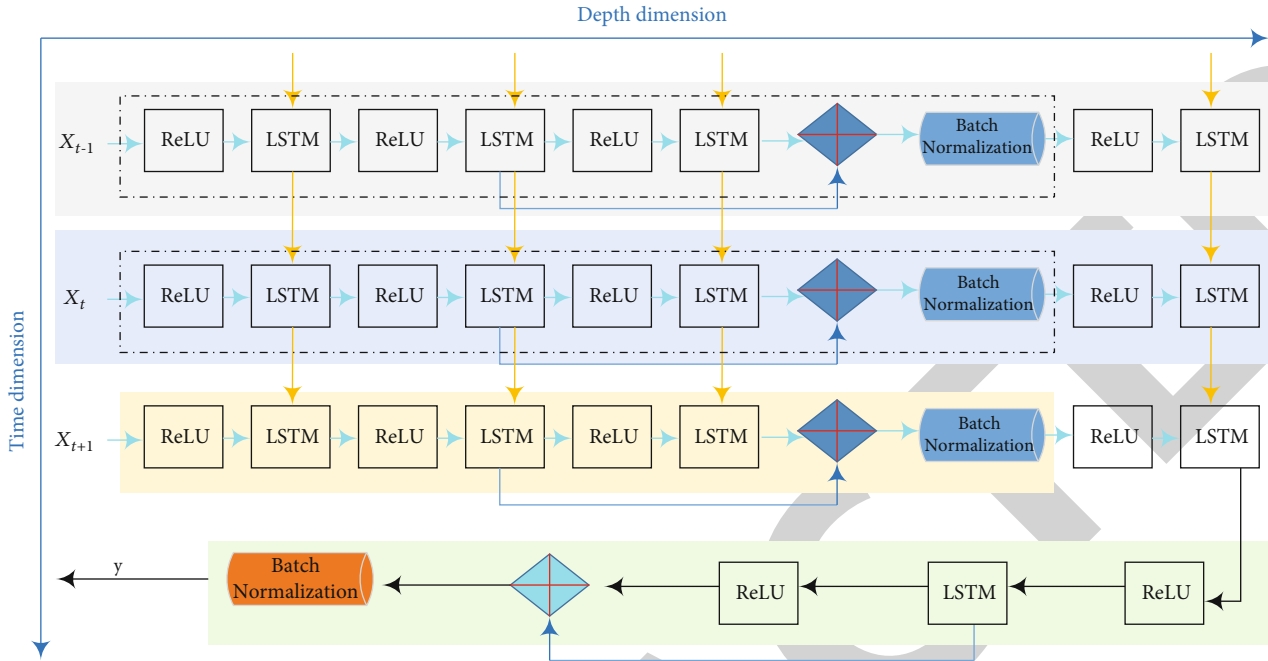


FIGURE 4: Residual LSTM network structure.

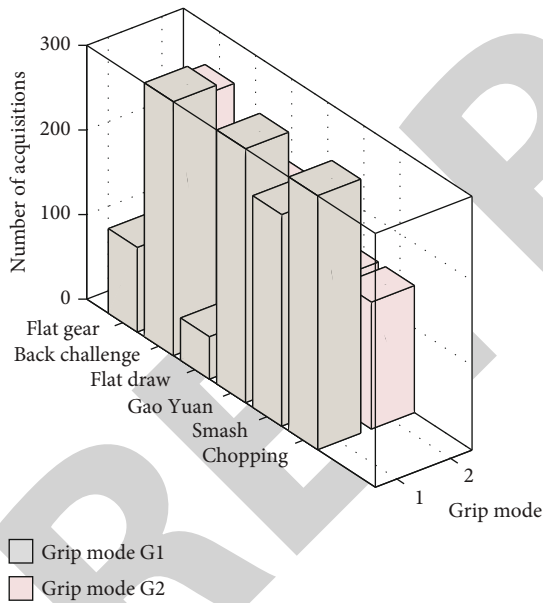


FIGURE 5: The number of times collected by each swing.

out. In addition, in order to facilitate subsequent algorithm research, the raw sensor data generated by each swing movement needs to be intercepted separately to facilitate feature extraction. Therefore, window segmentation technology is needed to intercept the swing movement data. The main methods of window data segmentation are sliding window-based, event-based window, and action-based window. However, these three methods have certain shortcomings in the interception of motion data, especially in the interception of real-time motion data streams; they cannot intercept motion data well, and this article needs to realize real-time

recognition in the end. Therefore, two methods of window segmentation are used. They are used for window interception of collected experimental data and data interception for real-time data stream data. These two methods of window interception are used in the training phase of the algorithm and the real-time action recognition phase.

5.2. Analysis of Recognition Results

5.2.1. Classification of Swing Action and Nonswing Action.

When recognizing the swing action, it is first necessary to determine in the sliding window whether there is a swing technical action in the window. The method of judging whether there is a swing action in the sliding window is performed by calculating the sliding variance of the resultant acceleration in the sliding window and the angular velocity in the sliding window. After repeated tests, this method can completely identify the swinging and nonswinging movements of the athlete. The correct rate of judging whether there are swinging technical actions in the window can reach 100%, so the influence of some nonswinging technical actions on recognition can be completely eliminated.

5.2.2. Recognition Result of Grip Mode. There are two ways of gripping when an athlete performs a swing. There are two ways of recognition: one is to directly recognize the swing without considering the gripping method; the other is to first identify the athlete's gripping method and then recognize the swing action based on a certain way of holding the racket. In order to eliminate the influence of the way of gripping on the recognition result, this article adopts the method of first identifying the way of gripping and then identifying the swing action. Finally, the decision tree algorithm is used to identify the way of gripping, and the data

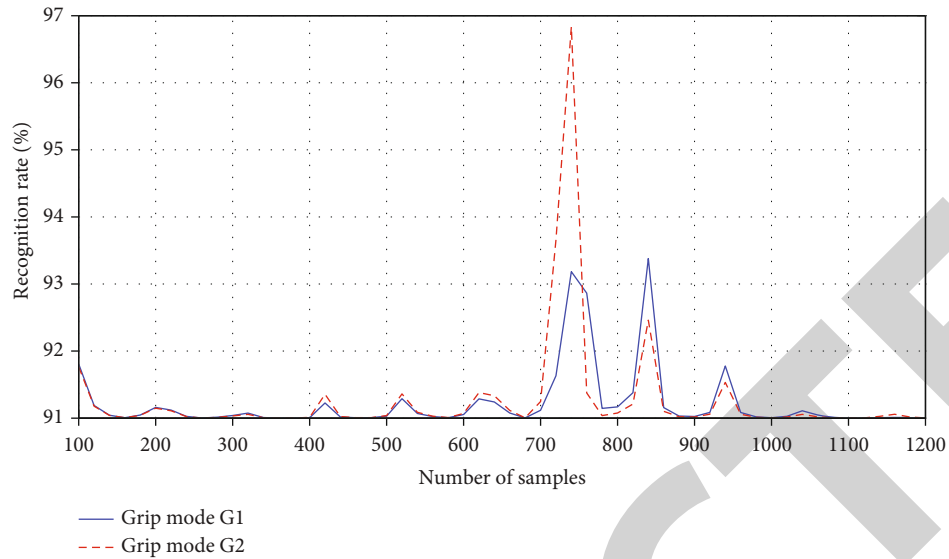


FIGURE 6: Improved Deep Residual LSTM to recognize the grip mode.

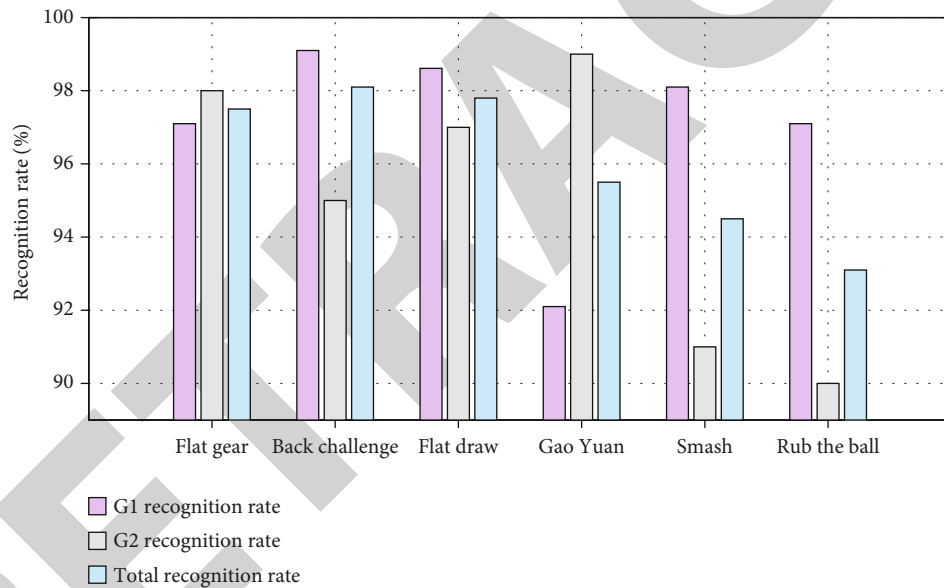


FIGURE 7: The recognition rate of the improved Deep Residual LSTM for swing movements.

characteristics of the training set are used to construct the C4.5 tree. This paper uses the 10-fold cross-validation method to obtain the average recognition rate of the algorithm to verify the pros and cons of the classification algorithm. The specific method is to randomly divide all sample sets into 10 parts, and each time 9 parts of the data are taken as the training set. The data is used as a validation set, and this is done 10 times, and finally, the average recognition rate is calculated.

Figure 6 shows the recognition result of the improved Deep Residual LSTM on the gripping mode. It can be seen from Figure 6 that the improved Deep Residual LSTM algorithm can achieve an average recognition rate of 91.3% for the gripping mode, so it can be used to recognize the gripping mode before the swing action is recognized.

5.2.3. Recognition Result of Swing Motion. After identifying the gripping method, two classifiers are trained using the data sets of the two gripping methods as the training set. When performing swing motion recognition, first, we recognize the grip mode and then use the corresponding classifier for recognition. The final recognition rate of the overall swing motion is the product of the recognition rate of the gripping method and the recognition rate of the swing motion based on a certain gripping method.

The improved Deep Residual LSTM's recognition rate of swing action is shown in Figure 7. We give the recognition of each swing based on the two gripping modes of G1 and G2 and calculate the average value. From the data in the figure, it can be seen that the average recognition rate based on grip mode G1 is 94.2%, and the average recognition rate based on

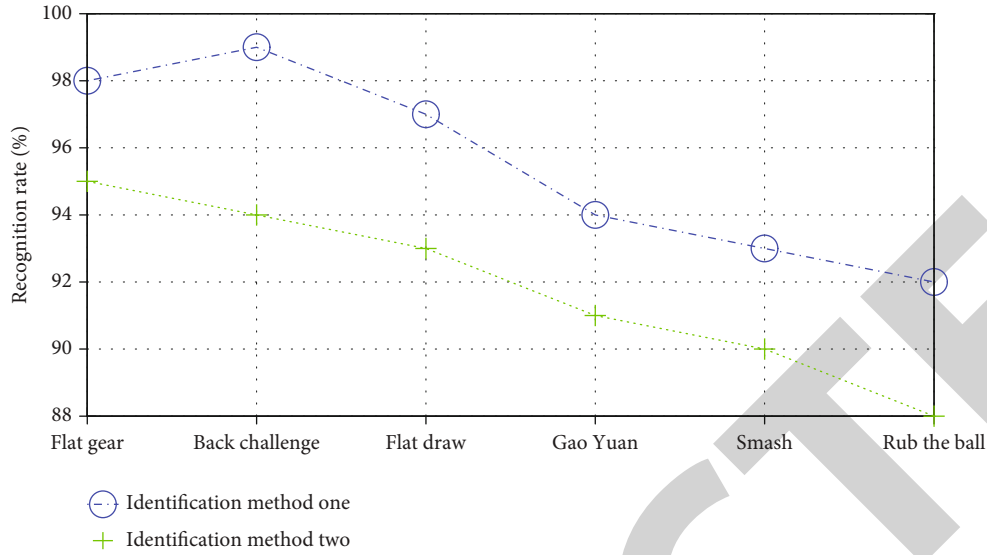


FIGURE 8: Comparison of recognition rate of swing action.

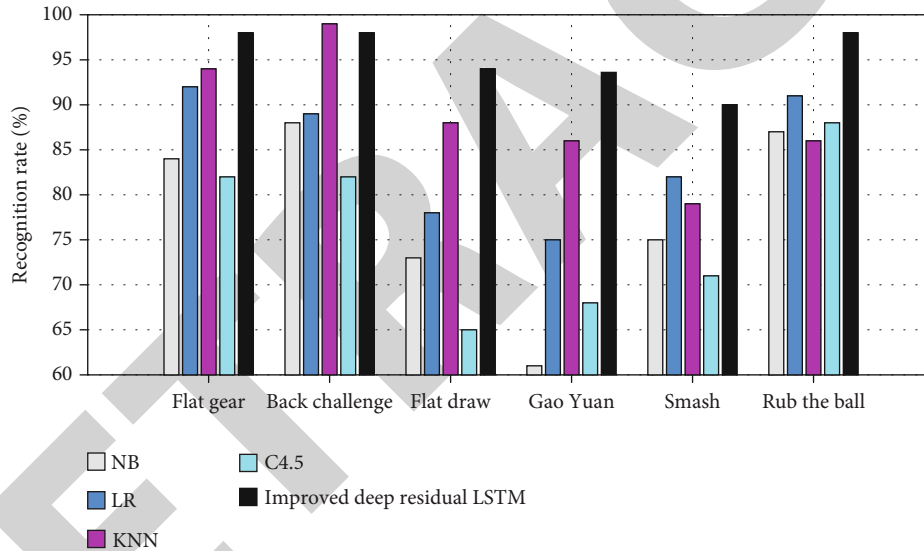


FIGURE 9: The recognition rate of different classification algorithms.

grip mode G2 is 93.6%. The total recognition rate of the final swing action is 93.8%.

5.3. Comparative Test

5.3.1. Comparison of Two Recognition Methods. The recognition method used in this article is a recognition algorithm based on two-layer classifiers. Another recognition mode is the recognition without gripping and only uses the classification results of a one-layer classifier. The following compare the recognition results of the two recognition modes.

Recognition method 1: it is the improved Deep Residual LSTM method proposed in this paper. The first-level classifier first recognizes the athlete's gripping style; the second-level classifier trains the swing action recognition classifier through the experimental data of the two gripping styles and recognizes it through the two-layer classifier.

Recognition method 2: we use the experimental data used as training data without distinguishing the gripping motions and train a classification model. Among them, the data pre-processing method used is the same as the recognition method one.

The final recognition rate comparison of the two methods is shown in Figure 8. After comparison, it is found that the recognition accuracy of each type of action using recognition method 1 is higher than that of using recognition method 2, and the average recognition rate of method 1 is about 3% higher than that of method 2. The comparison of recognition results shows that the improved Deep Residual LSTM method used in this paper is feasible.

5.3.2. Comparison of Multiple Recognition Algorithms. In order to verify the feasibility and superiority of the recognition algorithm proposed in this paper, this paper gives a

comparison between the recognition results of other four commonly used recognition algorithms and the algorithm used in this paper. The main comparison methods are naive Bayes (NB), logistic regression (LR), K nearest neighbors (KNN), and C4.5 decision trees.

During the comparison experiment, the collected experimental data will be calculated according to the feature extraction and dimensionality reduction methods described in this article, and the final average recognition rate will also be calculated using the 10-fold method. The final recognition rate of various methods is shown in Figure 9. It can be seen from Figure 9 that the average recognition rate using naive Bayes is 77.8%, the average recognition rate using logistic regression is 84.5%, the average recognition rate using KNN is 89.7%, and the average recognition rate using the C4.5 tree is 75.1%. The recognition rates of these four algorithms are lower than the improved Deep Residual LSTM recognition algorithm proposed in this paper.

6. Conclusion

When capturing the badminton swing motion, the flying direction and position of the badminton are used as the basis for the start and end of the swing motion capture. From the perspective of the badminton robot, when the shuttlecock flies to the opposite field and crosses the net, it starts to capture; when the badminton flies back to the field and crosses the net, the capture ends. This paper proposes an improved LSTM deep structure based on the idea of residual learning, which is used to recognize human movements from the sensor data of wearable inertial sensors. It almost directly recognizes the raw sensor data with the fewest preprocessing steps, which is more versatile and minimizes engineering deviation. The network proposed in this paper can provide improvements in time and depth dimensions. Classical action recognition uses mostly artificially constructed features or heuristic features. The disadvantage of this is that it requires high domain knowledge and experience, and features manually designed for specific tasks do not have strong generalization capabilities. With the support of a larger data set, this paper adds a convolutional layer after the input layer of the network, learning features in a data-driven form instead of artificially constructed features, without or greatly reducing the dependence on domain knowledge and experience, and the network has good generalization ability. In the motion data window segmentation, a window segmentation method based on the combination of sliding window and action window is proposed to more accurately and scientifically intercept the data generated by each motion. In terms of data feature extraction and selection, the time domain features and frequency domain features of the data are extracted, respectively, and the main features of motion can be extracted. In the aspect of swing motion recognition, an improved Deep Residual LSTM swing motion recognition algorithm is proposed. Experiments have proven that this method is sufficient to eliminate the influence of the two gripping modes of the athlete on the recognition of sports actions. Moreover, the type of swing can be recognized more accurately, and the final recognition rate is

higher. Although the swing motion recognition algorithm studied in this paper can quickly and accurately identify the player's swing motion, the intelligent badminton training system designed can also be used normally in actual sports. But there is still some work that can be improved. In the aspect of sports data collection, the hardware collection equipment is further optimized. More racket-like sports technical movements should be collected, so that more racket-like sports movements can be identified through follow-up research. In the research of recognition algorithms, more experimental verification results are used to improve the algorithm and increase the generalization ability of the algorithm, thereby increasing the recognition rate.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Zhengzhou University.

References

- [1] A. Gaidhani, K. Moon, Y. Ozturk, S. Lee, and W. Youm, "Extraction and analysis of respiratory motion using wearable inertial sensor system during trunk motion," *Sensors*, vol. 17, no. 12, p. 2932, 2017.
- [2] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," *IEEE Sensors Journal*, vol. 16, no. 3, pp. 773–781, 2016.
- [3] K. Tanimoto, M. Takahashi, K. Tokuda, T. Sawada, M. Anan, and K. Shinkoda, "Lower limb kinematics during the swing phase in patients with knee osteoarthritis measured using an inertial sensor," *Gait & Posture*, vol. 57, pp. 236–240, 2017.
- [4] H. Xu, S. Guerrier, R. Molinari, and M. Karemera, "Multivariate signal modeling with applications to inertial sensor calibration," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 5143–5152, 2019.
- [5] Y. Gu, L. T. Hsu, and S. Kamijo, "GNSS/onboard inertial sensor integration with the aid of 3-D building map for lane-level vehicle self-localization in urban canyon," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4274–4287, 2016.
- [6] K. Cahill-Rowley and J. Rose, "Temporal-spatial reach parameters derived from inertial sensors: comparison to 3D marker-based motion capture," *Journal of Biomechanics*, vol. 52, pp. 11–16, 2017.
- [7] J. Yang, J. Zhang, and H. Wang, "Urban traffic control in software defined internet of things via a multi-agent deep reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3742–3754, 2021.
- [8] T. Huynh-The, C. H. Hua, T. T. Ngo, and D. S. Kim, "Image representation of pose-transition feature for 3D skeleton-

- based action recognition," *Information Sciences*, vol. 513, pp. 112–126, 2020.
- [9] J. Wen, J. Yang, B. Jiang, H. Song, and H. Wang, "Big data driven marine environment information forecasting: a time series prediction network," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 4–18, 2021.
- [10] S. Clemens, K. J. Kim, R. Gailey, N. Kirk-Sanchez, A. Kristal, and I. Gaunaud, "Inertial sensor-based measures of gait symmetry and repeatability in people with unilateral lower limb amputation," *Clinical Biomechanics*, vol. 72, pp. 102–107, 2020.
- [11] P. Kumar, S. Mukherjee, R. Saini, P. Kaushik, P. P. Roy, and D. P. Dogra, "Multimodal gait recognition with inertial sensor data and video using evolutionary algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 5, pp. 956–965, 2019.
- [12] M. Rhodin, A. Bergh, P. Gustås, and C. B. Gómez Álvarez, "Inertial sensor-based system for lameness detection in trotting dogs with induced lameness," *Veterinary Journal*, vol. 222, pp. 54–59, 2017.
- [13] K. Berntorp, "Joint wheel-slip and vehicle-motion estimation based on inertial, GPS, and wheel-speed sensors," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 1020–1027, 2016.
- [14] C. Arcidiacono, S. Porto, M. Mancino, and G. Cascone, "A threshold-based algorithm for the development of inertial sensor-based systems to perform real-time cow step counting in free-stall barns," *Biosystems Engineering*, vol. 153, pp. 99–109, 2017.
- [15] W. F. Wang, C. Y. Yang, and J. T. Guo, "A sport recognition method with utilizing less motion sensors," *Advances in Intelligent Systems and Computing*, vol. 329, pp. 155–167, 2015.
- [16] M. Crabolu, D. Pani, L. Raffo, and A. Cereatti, "Humerus length estimation using a magneto-inertial sensor," *Gait & Posture*, vol. 57, pp. 20–21, 2017.
- [17] C. L. Hung, M. H. Hung, C. Y. Chang, H. H. Wang, C. S. Ho, and K. C. Lin, "Influences of lateral jump smash actions in different situations on the lower extremity load of badminton players," *Journal of Sports Science and Medicine*, vol. 19, no. 2, pp. 264–270, 2020.
- [18] S. Mori, K. Tanaka, S. Nishikawa, R. Niiyama, and Y. Kuniyoshi, "High-speed humanoid robot arm for badminton using pneumatic-electric hybrid actuators," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3601–3608, 2019.
- [19] M. Fröhlich, R. H. Fel De, and M. Reuter, "Training effects of plyometric training on jump parameters in D/DC-squad badminton players," *Journal of Sports Research*, vol. 1, pp. 22–33, 2016.
- [20] Z. Taha, M. Hassan, H. J. Yap, and W. K. Yeo, "Preliminary investigation of an innovative digital motion analysis device for badminton athlete performance evaluation," *Procedia engineering*, vol. 147, pp. 461–465, 2016.
- [21] Y.-J. Han, D. Chwa, and Y.-D. Hong, "Motion generation of badminton robot based on evolutionary optimization," *Journal of Institute of Control, Robotics and Systems*, vol. 23, no. 5, pp. 383–393, 2017.
- [22] H.-W. Jeong and H.-G. Lee, "The study of activating plans for the family-oriented badminton event through IPA analysis: focusing on marketing 4Cs," *Journal of Sport and Leisure Studies*, vol. 68, pp. 147–156, 2017.
- [23] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.
- [24] F. Orujov, R. Maskeliūnas, R. Damaševičius, and W. Wei, "Fuzzy based image edge detection algorithm for blood vessel detection in retinal images," *Applied Soft Computing*, vol. 94, p. 106452, 2020.
- [25] B. Li, J. Yang, Y. Yang, C. Li, and Y. Zhang, "Sign language/gesture recognition based on cumulative distribution density features using UWB radar," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.
- [26] Y. He, L. Dai, and H. Zhang, "Multi-branch deep residual learning for clustering and beamforming in user-centric network," *IEEE Communications Letters*, vol. 24, no. 10, pp. 2221–2225, 2020.
- [27] C. Mi, Y. Huang, C. Fu, Z. Zhang, and O. Postolache, "Vision-based measurement: actualities and developing trends in automated container terminals," *IEEE Instrumentation and Measurement Magazine*, vol. 24, no. 4, pp. 65–76, 2021.
- [28] Z. Lv, Y. Han, A. K. Singh, G. Manogaran, and H. Lv, "Trustworthiness in industrial IoT systems based on artificial intelligence," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 1496–1504, 2021.
- [29] J. Xu, S. H. Park, and X. Zhang, "A bio-inspired motion sensitive model and its application to estimating human gaze positions under classified driving conditions," *Neurocomputing*, vol. 345, no. JUN.14, pp. 23–35, 2019.
- [30] I. Ajili, M. Malle, and J. Y. Didier, "Human motions and emotions recognition inspired by LMA qualities," *The Visual Computer*, vol. 35, no. 10, pp. 1411–1426, 2019.
- [31] M. C. Chen, S. Q. Lu, and Q. L. Liu, "Uniqueness of weak solutions to a Keller-Segel-Navier-Stokes system," *Applied Mathematics Letters*, vol. 121, p. 107417, 2021.
- [32] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," *Knowledge-Based Systems*, vol. 228, p. 107242, 2021.
- [33] G. Dorgo and J. Abonyi, "Learning and predicting operation strategies by sequence mining and deep learning," *Computers & Chemical Engineering*, vol. 128, no. SEP.2, pp. 174–187, 2019.