J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 9577--9597

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

A deep learning-based action recognition and confrontation analysis system for sparring players

Ranfeng Fan[1, ✉]

[1] Police Combat Teaching and Research Department, Guangxi Police College, Nanning, Guangxi, 530000, China

ABSTRACT

There is an increasing demand for assisted training techniques in the sport of sparring. In this paper, a sparring multiple recognition and analysis system is designed and fabricated for the movements of sparring sports and used to recognize and analyze the players' technical movements using the collected data and the model built using deep neural networks. The CNN-LSTM network is applied to extract the feature classification of the preprocessed sparring inertia data, and then the DTW algorithm is combined with the spatial distance classification method to realize the matching and recognition of sparring behaviors by stretching and compressing transformations of the time axis, effectively eliminating the distortion error in the time domain and obtaining the similar path with the shortest cumulative distance of the effective matches between different sequences. Experiments on the application of this paper's system were conducted in two groups of sparring players, and after 12 weeks of training intervention, the average confrontation striking speed of the experimental group progressed from 0.36 seconds before the experiment to 0.32 seconds after the experiment, and the average performance of the control group progressed from 0.38 seconds before the experiment to 0.36 seconds after the experiment, which indicates that although the traditional resistance training also has a positive impact on the training effect of sparring training, the training effect of this paper's system is more obvious The systematic training effect of this paper is more obvious. This paper makes an innovative exploration for the combination of sports programs such as sparring and cutting-edge information technology.

Keywords: CNN, DTW, LSTM, feature extraction, action recognition

1. Introduction

The rapid development of artificial intelligence promotes the research of human action recognition technology, which is widely used in the fields of smart home, intelligent security, and motion analysis. In the field of motion analysis, the motion recognition algorithm is applied to the action training of

✉ Corresponding author.
E-mail address: he1231512315@163.com (R. Fan).
Received 10 January 2024; Revised 15 February 2024; Accepted 25 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-528
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

sparring athletes, through the detection and recognition of various types of movements of sparring athletes, the movements of sparring athletes are analyzed, which helps sparring athletes accurately assess the shortcomings of their own movements and optimize the training system to improve the training effect [1-3].

At the beginning of the development of action recognition technology, the feature extraction of data was mainly performed manually. In recent years, with the expansion of data size, the popularity of high-performance computing devices and the development of neural network technology, the use of deep learning to replace the traditional manual extraction of features has become a mainstream trend [4-5]. Compared with manual feature extraction, feature extraction can be performed automatically using deep learning to process large-scale data, which greatly simplifies the process of manual feature extraction [6-9]. In addition, the attention mechanism in the field of deep learning can determine the key information more accurately and reduce the attention to other information by learning the feature weights from the feature graph and redistributing the weights to enhance the effective features and suppress the ineffective features [10-13]. There are two main methods for human movement recognition, one is to use wearable devices to obtain human movement parameters to recognize human movements, this method relies on hardware devices and is costly [14-16]. The other is to recognize human movements by double streaming method or 3D convolution method, but the recognition effect is easily affected by the environment. Therefore, the deep learning-based approach to recognize sparring movements has become a current research hotspot [17-19].

Wearable device human behavior recognition systems can rely on microsensors to collect inertial information from different nodes of the human body during motion. Literature [20] investigated a human behavior recognition system based on mobile and wearable sensors, which can significantly improve the system's performance in recognizing human activities by incorporating deep learning and artificial intelligence techniques that can automatically learn features from a large amount of multimodal and high-dimensional sensor data compared to traditional manual feature extraction methods. Literature [21] shows that the rough feature extraction process in traditional machine learning classifiers is unable to recognize complex human activities, whereas the integration of deep learning techniques with smartphone and wearable sensor recognition systems can efficiently retrieve and classify data features and improve the accuracy of human activity recognition. Literature [22] evaluates the advantages and challenges of convolutional neural networks applied to human activity recognition systems for wearable or stationary devices, which can effectively and accurately improve the human activity recognition performance, but still face challenges in terms of data availability, high computational resource requirements, data privacy issues, and edge computing limitations. Literature [23] developed a hybrid convolutional neural network incorporating a channel attention mechanism, which is capable of acquiring deep spatio-temporal features in sensor behavioral data to distinguish between similar or overlapping human activities. Literature [24] showed that a sensor action recognition system fused with machine learning techniques can provide rich contextual information for real-world applications, and experiments on human behavior recognition in heterogeneous datasets based on acceleration sensors showed that the fused action recognition system achieves highly accurate results in sensor data classification.

The essence of dual-stream convolutional neural network is to divide the captured video into spatial and temporal streams, so that the model learns the spatial information along with the changes in the spatial information to achieve action recognition. Literature [25] examined the performance of dual-stream convolutional neural network model in human behavior recognition, using the model of LSTM in the spatial stream to extract the spatial and temporal features of the RGB image, and introducing DenseNet in the temporal stream for high recognition accuracy, which strengthens the model's ability of extracting the temporal features of the behavioral data, and makes the dual-stream model have good recognition performance. Literature [26] pointed out that making full use of the temporal information of behavioral data and the acquisition of a large amount of data for training optimization are difficult challenges for deep learning techniques, based on which the optical flow and convolutional neural network migration learning method is proposed, which not only successfully resolves the spatio-temporal information of human movement, but also realizes the migration learning of convolutional neural network. Literature [27] elucidated that the two-stream convolutional neural network can extract behavioral information from two dimensions of spatial and temporal flow for recognition, based on which the introduction of bidirectional gated loop unit can give the model to extract the appearance consistency characteristics of the action, and the introduction of the attention mechanism based on neuroscience theory can improve the recognition accuracy and stability of the model.

On the basis of 2D convolution, 3D convolution method introduces the temporal dimension, which can simultaneously acquire both temporal and spatial feature information from consecutive video frames as a way to realize action recognition. Literature [28] developed a hybrid convolutional tube (MiCT) that combines 2D convolutional modules with 3D convolutional modules, and based on this, designed an end-to-end deep 3D network, MiCT- Net, which is capable of fully recognizing the temporal and spatial information of human behaviors in videos. Literature [29] proposed an asymmetric unidirectional convolutional model based on local 3D convolutional network for the high computational cost, storage cost and learning difficulty in 3D convolutional neural network behavioral recognition methods, and the recognition effect and efficiency of the asymmetric 3D-CNN model were effectively improved in the recognition of complex actions. Literature [30] proposes a real-time action recognition architecture based on temporal convolutional 3D network (T-C3D), which recognizes behavioral actions in video through a hierarchical and multi-granularity approach, and the combination with temporal coding method makes the model have real-time processing capability, which significantly improves the speed of action reasoning in the process of behavior recognition.

This paper constructs a sparring action recognition and analysis system based on CNN-LSTM network. Firstly, the sparring action data are collected and preprocessed, and the inertia wearing nodes are calibrated according to the human skeleton model. In order to eliminate the influence of noise and smooth the data, error model correction and wavelet denoising are performed on the collected data, and the data segmentation method is designed. Then according to the theoretical knowledge of CNN and LSTM network, design the automatic feature extraction and recognition network based on CNN-LSTM network, take advantage of CNN's feature extraction and Bi-LSTM temporal modeling to extract the spatio-temporal and temporal features of sparring sports targets, train the behavior classifier, and finally use DTW algorithm combined with spatial distance classification method to realize sparring behavior matching. A controlled test is conducted to examine the effect of this system's action recognition and analysis function on the enhancement of sparring players' confrontation speed and strength.

2. Deep Learning-based Spread Sparring Action Recognition and Analysis System

2.1. Data Acquisition and Preprocessing

2.1.1. Information Acquisition and Processing System for Sporadic Sports. Human motion information acquisition will obtain the inertial motion data storage for the development of follow-up work, but its placement position needs to be uniformly planned and configured, the data acquisition system obtains data in addition to the effective action segment, will always contain all kinds of noise, in order to be able to provide correct and effective action data for the subsequent action recognition algorithm, it is inseparable from some signal processing fields such as data conversion, data analysis and data processing and other operations. The data processing stage of this paper can be divided into three levels: data acquisition, data processing, result analysis and display, the data acquisition function as a whole is completed independently in the module, data pre-processing is required after data acquisition, and then enter the formal algorithm design stage, feature extraction and classification are generally carried out in the upper computer, data acquisition and classification and identification take a separate design, and classification is placed on the PC side can effectively Reduce the power consumption of the data acquisition module. The system structure is shown in Figure 1.

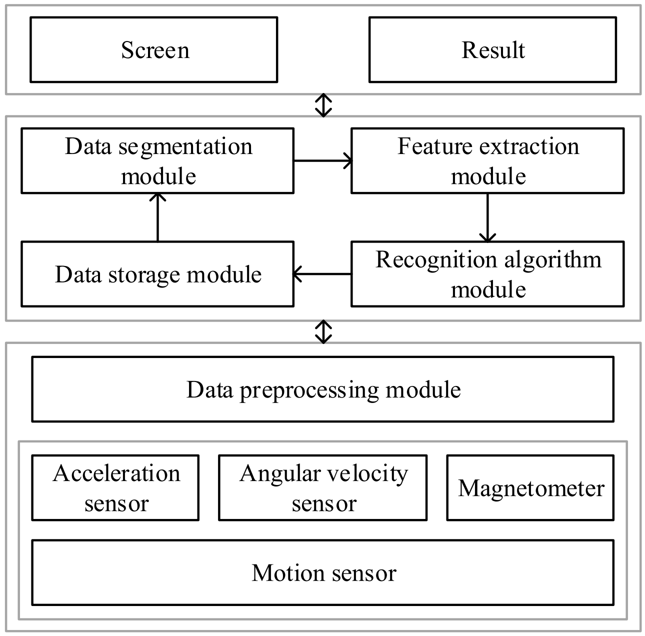

Fig. 1. Structure of data acquisition and processing system

The system includes the following steps:

1) Data acquisition. By implanting the inertia nodes into the guard, the inertia motion data of different sparring techniques are collected to provide the system with a raw database.

2) Data transmission. The 9-axis inertial data collected by the inertial node is transmitted to the PC by wireless communication, and then processed in the next step. Here, WiFi wireless transmission is adopted, which automatically connects with the data transceiver node after each node is powered up to realize the real-time transmission of data from the acquisition module.

3) Data frame processing. This stage includes noise reduction and data standardization of motion data, and then the motion data is segmented with a sliding window to obtain effective action segments.

4) Feature extraction. The pre-processed inertial data for feature extraction, the acceleration, angular velocity, magnetic field generated by different actions are not the same, but the original signal can not be intuitively and effectively differentiate between the different actions, the extraction of various types of action does not move to the action of the amount of features used to distinguish.

5) Classification identification. Use the extracted feature vectors to train the model, and finally get the trained classification model, then you can use the model to classify the data samples.

2.1.2. Data acquisition. After the establishment of the data acquisition system is completed, the data acquisition can be carried out on the sparring athletes. Before the data acquisition work, first of all, it is necessary to clarify the type of sparring action that needs to be collected, and then according to the type of action to select the sampling environment, determine the position of the sensor wear, etc. Good preparation can provide a scientific data basis for the next action classification and recognition algorithm. In order to effectively and accurately collect the action data, wear the sensor with reference to the human skeleton model, consider the human skeleton as a tree, and the root node is the root of the tree, in the human body model, according to the principle that each joint of the human body exists independently and can be acted upon, the human body's own motion state is constrained by the joint degrees of freedom and is always composed of a few key joints, therefore, wearing the inertia node at the key joints can effectively and reasonably collect the required action data, and then determine the sampling environment based on the action type. Therefore, wearing inertial nodes at the key joints can effectively and rationally collect the required motion data, and the human body joint model is shown in Figure 2.

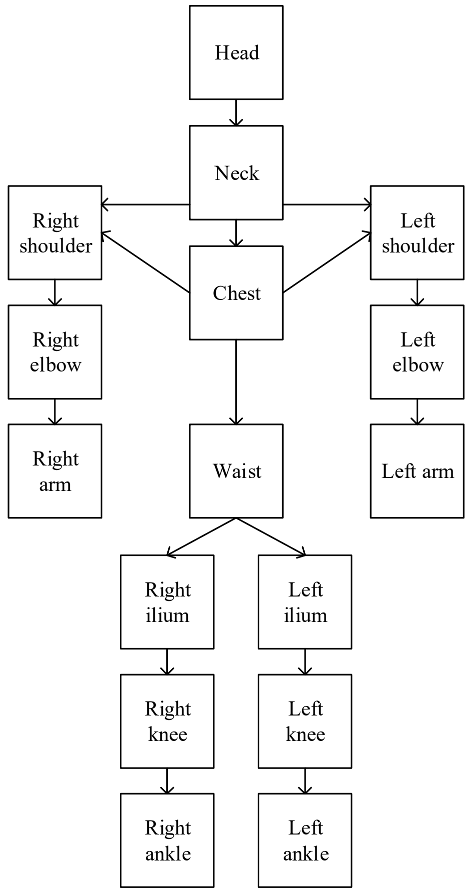

Fig. 2. Human joint structure

2.1.3. Data processing:

1) Data denoising. The sensor used has its own industrial design, manufacturing system comes with the noise, at the same time in the actual data acquisition process, MEMS is subject to electromagnetic interference in the natural environment, the data in the wireless transmission will also be interfered with, and the final cache of data will inevitably contain noise, so in the need to establish the sensor's error correction model, in order to reduce the impact of the measurement error on the subsequent classification and identification.

The error model of the accelerometer is shown in equation (1):

$$a_i(t) = \hat{a}_i(t) + g_i(t) + b_{ai}(t) + \varepsilon_{ai}(t) \tag{1}$$

where, $a_i(t)$ for the accelerometer in t moments of the output value of the i-axis, $\hat{a}_i(t)$ for the acceleration generated by external forces, in this topic for the acceleration generated by the movement of the athlete, $g_i(t)$ for t moments of the acceleration of gravity in the i-axis component, $b_{ai}(t)$, $\varepsilon_{ai}(t)$, respectively, for the accelerometer's own offset and the Gaussian white noise.MEMS accelerometers have a better static performance, with time offset is not large.

The process of constructing the error model of the gyroscope is analogous to that of the accelerometer, but the effect of gravity on the gyroscope is negligible, and the error equation of the gyroscope is shown in Equation (2):

$$\omega_i(t) = \hat{\omega}_i(t) + b_{\omega i}(t) + \varepsilon_{\omega i}(t) \tag{2}$$

where $\omega_i(t)$ is the actual output value of the gyroscope, $\hat{\omega}_t(t)$ is the real value of the gyroscope, and $b_\omega i(t), \varepsilon_\omega i(t)$ is the static offset and Gaussian white noise of the gyroscope, respectively. The static offset of the gyroscope has a large impact, in order to reduce its impact, the sensor will be placed in the stationary state, calculated the average output of the gyroscope in the stationary state of the internal axes of the output $\overline{\omega_t}$, and this value is regarded as the static offset of the gyroscope, in the calculation of any moment after the calculation from the have been subtracted from the $\overline{\omega_t}$, so as to obtain a more accurate value of the gyroscope, as shown in equation (3):

$$\hat{\omega}_t = \omega_i(t) - \overline{\omega_t} \tag{3}$$

Magnetometers are more susceptible than the previous two to magnetic environmental disturbances in the environment, including hard and soft magnetic disturbances, resulting in a larger deviation of the measured value from the true value, whose simplified error model is described by equation (4).

$$h = AH + B + \varepsilon \tag{4}$$

where A is soft magnetic interference, B is hard magnetic interference, and ε is Gaussian white noise in the real environment. Ideally, the values measured by the magnetometer should be distributed on a sphere with the radius of the local magnetic field strength, while the magnetometer output data is distributed as an ellipse due to the presence of environmental interference and self-turbulence.

In the decomposition process of wavelet denoising, the original signal S is wavelet filtered to obtain the low-frequency signal A and high-frequency signal D [31]. In order to get better results, the low-frequency components of the decomposed signal can be subdivided again in multiple scales, and finally the processed individual components are wavelet reconstructed, and the final signal S can be expressed as equation (5).

$$S = A_n + D_1 + D_2 + K + D_n \tag{5}$$

where n denotes the number of decomposition layers, and the desired signal details are continuously focused through multiple multi-scale subdivisions until the desired signal reconstruction effect is achieved.

2) Data Standardization. Acceleration data, angular velocity data and magnetic intensity data are used in this paper. Since these three types of data contain different physical meanings and their numerical differences are large, the feature values extracted when feature extraction is performed on the three types of data are also very different, which will cause some interference in the classification of actions. At the same time, the data acquisition node will bring noise in the external small vibration and so on, which will make the samples produce values different from the normal value, which will have an impact on the recognition results.

In this paper, we adopt the method of standardization using the mean and standard deviation of the original data for mathematical operations.

Formula (6) ~ formula (8) for the acceleration x, y, z -axis data standardization calculation process.

$$X_{normalize} = \frac{X - \mu_x}{\sigma_x} \tag{6}$$

$$Y_{normalize} = \frac{Y - \mu_y}{\sigma_y} \tag{7}$$

$$Z_{normalize} = \frac{Z - \mu_z}{\sigma_z} \tag{8}$$

where: $X_{normalize}, Y_{normalize}, Z_{normalize}$ is the normalized value; X, Y, Z represents the original value; μ_x, μ_y, μ_z represents the mean value of the data in the x, y, z -axis, respectively; $\sigma_x, \sigma_y, \sigma_z$ represents the variance of the data in the x, y, z -axis, respectively.

3) Data segmentation. Collecting time series data such as sparring action data often takes a long time and has much redundant data, in order to get effective action data segments for feature extraction and classification, the data need to be segmented by windows [32].

In order to be able to fully describe each sparring technique action, on the basis of the traditional sliding window, according to the characteristics of the sparring movement, overlapping sliding between adjacent windows. Adjacent windows have a certain proportion of data duplication, which can better describe the movement characteristics. Sliding window segmentation is easy to implement, simple to operate, after experimental verification, the final determination of the window length of 128 sampling points, 64 sampling data points for the sliding length, that is, 50% overlap rate, where S that is the sliding step length of 64. the inertia of the standardized nodes of each node to take the sliding window segmentation, the data will be truncated into an equal-length sequence, and then each small section of the data sample to do a good job of classifying labels will become a complete time series sample.

2.2. Deep Learning Algorithm Preparation

2.2.1. CNN networks. The input layer in the CNN network structure is the input data, which can be image data or serialized data and so on. If there is a linear mapping relationship between the data in the upper and lower layers in the neural network, it is easy to lead to data overfitting, which affects the detection performance of the model. Therefore, the emergence of activation layer solves the problem of linear mapping between upper and lower layer data. The activation layer is a nonlinear mapping of data between the upper and lower layers of the network structure through the activation function to improve the problem of data overfitting.

The convolution layer extracts the feature information through the convolution operation between the convolution kernel and the input data, and generates the feature map as the output data of the network layer.

The feature map is generated as the output data of the network layer. The calculation method of feature map is shown in equation (9).

$$O_{i,j} = f\left(\sum_{m=0}^{l}\sum_{n=0}^{t} w_{m,n}\, x_{i+m,j+n} + b\right) \tag{9}$$

where the height and width of the convolution kernel are $1, t, O_{i,j}$ to represent the element value out of the feature map point (i, j), the activation function is f, the bias value is b and the weights are w.

The pooling layer replaces the region by the maximum value of the region or the average value of the region, which reduces the computational complexity of the data on the basis of retaining the information of the data and reduces the size of the feature map. The calculation method is shown in equations (10) and (11).

$$W_o = \frac{(W_i - F + 2P)}{s} + 1 \tag{10}$$

$$H_o = \frac{(H_i - F + 2P)}{s} + 1 \tag{11}$$

where the width and height of the input data before the convolution operation are W_i, H_i, the width and height of the feature map obtained by convolution W_o, H_o, F represents the size of the convolution kernel, the step length is denoted by S, and P is the number of fills.

The fully connected layer integrates all the feature information extracted by the neural network by connecting all the neurons of the previous layer. If the number of neuron connections in the fully connected layer is too large, a Dropout method can also be introduced, where some of the neuron connections will be randomly deleted to alleviate data overfitting. The fully connected layer will update the weights of each neuron connection by calculating the amount of error between the output and the expected result, and backpropagation is a commonly used weight training method. Where the error $e_i(n)$ of the i th data connection node is calculated as shown in equation (12).

$$e_i(n) = y_i(n) - p_i(n) \tag{12}$$

where $y_i(n)$ is the actual value and $p_i(n)$ is the predicted value. Then the error correction method is shown in equation (13).

$$\varepsilon(n) = \frac{1}{2}\sum_i e_i^2(n) \tag{13}$$

Gradient descent is a commonly used parameter updating method in deep learning networks, which is calculated by iteratively calculating the eigenvalues to find the optimal weight matrix, as shown in equation (14).

$$\Delta w_{i,j}(n) = -\eta \frac{\partial \varepsilon(n)}{\partial v_j(n)} y_i(n) \tag{14}$$

where y_i is the neuron output of the previous layer and η is the learning rate.

2.2.2. LSTM network theory. RNN is a type of network that connects network nodes according to time series ordering, and its structure is schematically shown in Fig. 3.RNN is organized as a continuous network layer by connecting neuron nodes into a network, and the upper and lower two nodes are connected to each other by unidirectional connection.

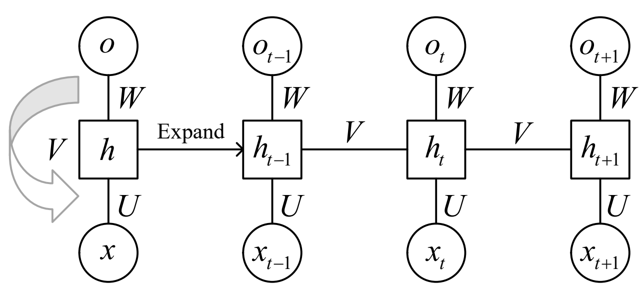

Fig. 3. RNN recurrent neural network

In Fig. 3, h is the value of the hidden layer, U is the weight matrix from the input layer x to the hidden layer h, o is the output value of the network, and W is the weight matrix from the hidden layer h to the output layer o. The value of the hidden layer h of the RNN is correlated by both the current input x and the hidden layer v of the previous layer. Therefore, the output layer of the RNN is calculated as shown in Eq. (15) and Eq. (16).

$$o(t) = f(Wh_t) \tag{15}$$

$$h_t = f(Ux_t + Vh_{t-1}) \tag{16}$$

The content of the LSTM state cell C_t is jointly controlled by the forgetting gate f_t and the input gate X_t. Where input gate X_t guarantees the moment when valid data content enters the memory cell; forgetting gate f_t is used to determine the number of previous moment cell states C_{t-1} that need to be saved into the current moment cell state C_t; and output gate o_t controls the number of output values h_t that need to be output to the LSTM cell by controlling the cell state C_t. LSTM avoids the problems of gradient vanishing and gradient explosion that occur in RNNs due to long term dependency by setting these three gates.

The mathematical model of the input gates is:

$$i_t = \sigma(W_i x_t + W_{hi} h_{t-1} + b_i) \tag{17}$$

The mathematical model of the forgetting gate is:

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f) \tag{18}$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \tag{19}$$

The output gate is mathematically modeled as:

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o) \tag{20}$$

$$h_t = o_t \tanh(c_t) \tag{21}$$

In Eqs. (17)-(21), matrix W is the weights and variable b is the bias.

2.3. Design of CNN-LSTM based network structure

In this paper, we will design the algorithm for sparring action detection and behavior recognition based on deep learning theory, and the algorithm flow is shown in Fig. 4. After video preprocessing, CNN network is used to extract spatio-temporal features of the data, and Bi-LSTM network is used for temporal modeling of the feature vectors to extract temporal features [33].

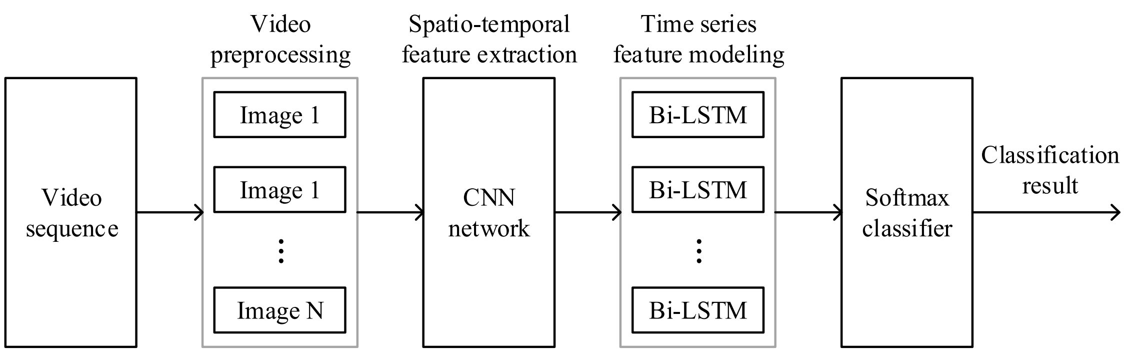

Fig. 4. Fall detection and behavior recognition process

2.3.1. CNN low-level spatio-temporal feature extraction. CNN networks have important applications in image data and speech signal processing through the mechanism of weight sharing and local connectivity.CNN networks have 4 convolutional layers, where the input data is a 224*224*3 RGB image sequence.

1) Batch Normalization Layer. Because the deep learning network model is an interconnected structure of upper and lower layers, the input data of the next layer will differ due to the change of neuron parameters in the previous layer, and the deeper the network layers are, the greater the difference. This phenomenon is known as internal covariate bias. When the internal covariates are

distributed differently in the training and test data, it can lead to problems of vanishing gradients, slow model convergence and overfitting.

Batch normalization is a method used to normalize the input layers by re-centering and scaling, resulting in faster and more stable convergence of artificial neural networks. Batch normalization allows the distribution of covariates of the inputs to each network to be maintained in a stable state, thus solving the problems of vanishing gradients and overfitting of data.

Suppose the input value of the layer i network in the deep learning network model is I^i and the output value is O^i, i.e:

$$O^i = f(I^i) = f(w_i I^i + b_i) \tag{22}$$

where f is the activation function, w_i is the weights, and b_i is the bias.

The problem of internal covariate bias is solved by normalizing the input values of each layer of the network using the standard normal distribution method to ensure that the covariate distribution remains in a stable state.

$$\overline{I^i} = \frac{I^i - E(I^i)}{\sqrt{\operatorname{var}(I^i) + \varepsilon}} \tag{23}$$

$E(I^i)$ and $\operatorname{var}(I^i)$ are the expectation and variance of I^i under the current parameters.

2) ReLu activation function layer. If the activation function layer is missing in the deep network model, the input of each layer between the network structure is a linear function of the output of the upper network layer, then the feature representation of the model is limited.The ReLU function is a commonly used activation function in deep learning, through which a nonlinear mapping is done to the output data of the network layer, which is defined as shown in equation (24).

$$\operatorname{Re}LU(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} = \max(0, x) \tag{24}$$

2.3.2. LSTM high-level timing modeling. CNN network followed by cascaded LSTM network for temporal modeling, using the hidden layer states in the LSTM network to further extract the features of the sequence data, after which the features are combined using the fully connected layer to output the predicted categories. The traditional LSTM can only learn from a single direction, i.e., it learns the motion features from the beginning to the end of the occurrence of the motion behavior, but it ignores the contextual information, i.e., the future information, which has certain limitations.

Bidirectional Long Short-Term Memory Network (Bi-LSTM) improves the traditional LSTM algorithm by designing two LSTM layers with different directions.Bi-LSTM can utilize the two propagation directions, front and back, to learn the past information and the uncollected information, respectively, and model the temporal features of the data. Therefore, Bi-LSTM has excellent temporal processing capability.

The video sequences will be fed into the CNN-LSTM network designed in this paper after preprocessing. Where the CNN network will extract the spatio-temporal features of the data to get the feature vector $X = \{x_1, x_2, \ldots, x_n\}$. Then the Bi-LSTM network is used to learn the temporal feature information of the vector. It is assumed that after the feature vector X is input to the Bi-LSTM network, the corresponding hidden layer state H is obtained, and the hidden layer states of the upper and lower layers are correlated. The timing modeling calculation method is shown in equation (25).

$$H_n = f(Ux_n + Wx_{n-1} + b) \tag{25}$$

W and U are the weights and b is the bias value. According to the parameter sharing mechanism of Bi-LSTM network, each neuron W, U, b has the same value.

2.3.3. Characterization. Finally, high-level temporal features need to be classified. The casual action detection and behavior recognition studied in this paper is a multi-classification task, so it is especially important to choose an appropriate classifier. In the current deep learning, softmax classifier is the most widely used classifier. Its core principle idea is briefly summarized as follows: Assume that for a multi-classification task, the category label $y \in \{1, 2, \ldots, C\}$ can have C values. Given a sample x, the conditional probability of belonging to category c is predicted by softmax:

$$p(y = c \mid x) = softmax(w_c^T x) = exp(w_c^T x) / \sum_{c=1}^{C} exp(w_c^T x) \tag{26}$$

where w_c is the weight vector of category c.

After passing through the Bi-LSTM temporal modeling network designed in this paper, the output will be obtained as the temporal feature h_n of the video image, after which the softmax classifier will analyze and make the category decision for h_n. Suppose Y is the predicted category corresponding to this video. Then the formula is:

$$Y = softmax(Vh_n + c) \tag{27}$$

where V represents the weights and c represents the bias.

2.4. DTW-based behavioral feature matching

The human daily activity behavior is different from the static normal action, as a continuous sequence related to time, its effective characteristics can be used in accordance with the method of time series analysis for research. The daily dynamic behaviors under study are characterized by specific actions as templates, which involves the problem of individual variability of experimental personnel, i.e., the same action in the process of repetition by different personnel, due to the different speed and frequency of individuals, the resulting action sequence is not completely uniform, and individual differences in the length of time are also manifested. In order to solve the problem of mutual symmetry of time axis and consistency of action frequency well, m uses dynamic time regularization algorithm to match the personnel action behavior characteristics.

DTW was firstly used to solve the problem of time-length know-it-all and language speed inconsistency in speech recognition.DTW, as a kind of nonlinear planning method, provides the method of time and distance unification to study the similarity matching of the information of random-length time series.

For the given time series $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$, their corresponding lengths are m and n, and the sequences X and Y are set as the reference and matching terms, respectively. When calculating the similarity of element mapping in two sequences, the Euclidean distance is generally used to represent the corresponding element distance in two sets of sequences with the following expression:

$$d(x_i, y_j) = (x_i - y_j)^2 \tag{28}$$

Define the regularized matching path $W = (w_1, w_2, \ldots, w_l)$ as the reference for the similarity between sequences X and Y, where $w_l = (i, j)_l$ is used as the matching index to indicate the degree of matching of the mapped elements in the two sequences.

In order to ensure that the obtained path is the optimal path, the starting point and the end point of the path are defined as (1, 1) and (M, N), respectively, and the path is defined to satisfy the following constraints:

$$w_{l+1} - w_l \in \{(0,1), (1,0), (1,1)\} \quad l = 1, 2, \ldots, L \tag{29}$$

$$\max(M,N) \le L < M + N - 1 \tag{30}$$

The completeness and coherence of the paths are ensured by the above constraint function to prevent the occurrence of sequence crossing and element loss.

After establishing the path constraint specification, the minimum matching path $DTW(X,Y)$ for the two sets of sequences X and Y is obtained by accumulating the distances of each element as follows:

$$DTW(X,Y) = \min\left(\frac{\sqrt{\sum_{l=1}^{L} W_l}}{L}\right) \tag{31}$$

3. Shotgun movement recognition and confrontation analysis

3.1. *Action Recognition Evaluation Experiment*

Movement evaluation aims to compare the test movement with the standard movement in order to judge how well the test movement performs. The traditional method of movement evaluation is manual observation and assessment by experienced personnel such as coaches and referees, but the results of this evaluation may be affected by subjective factors. In order to achieve fair and impartial scoring, computer-assisted scoring can be considered, and the method used is the principle of similarity. The DTW algorithm was utilized to calculate the distance between the eight joint angles of the test movement and the standard movement sequence as a similarity assessment parameter. The front pendulum punch in the ten movements of the data set was selected as the evaluation object, and a total of 20 samples were chosen as the test sequence, and the movement video data of the sports majors' classmates was used as the template of the standard movement sequence. The distribution of joint angle distance was observed through multiple calculations.

Figure 5 shows the DTW distance distribution of the left shoulder joint angle, and Figure 6 shows the DTW distance distribution of the left elbow joint angle. From the figure, it can be seen that most of the DTW distances of the left shoulder joint angle are distributed between 650 degrees and 1250 degrees, and a few are distributed between 1250 degrees and 1300 degrees, which will be discarded for a small portion; most of the DTW distances of the left elbow joint angle are distributed between 600 degrees and 1150 degrees, and a few are distributed between 550 degrees and 600 degrees, which will be discarded for a small portion.

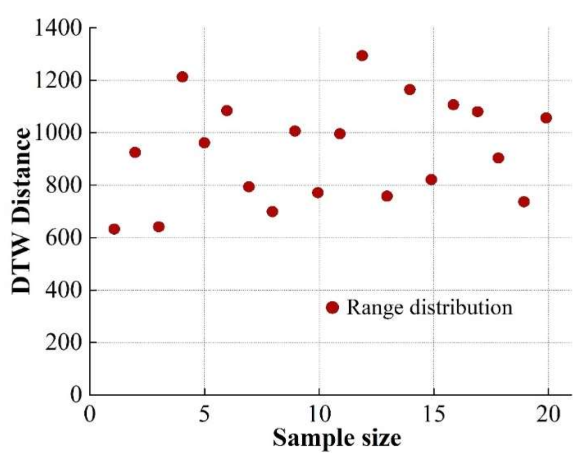

Fig. 5. The DTW distance distribution of the Angle of the left shoulder

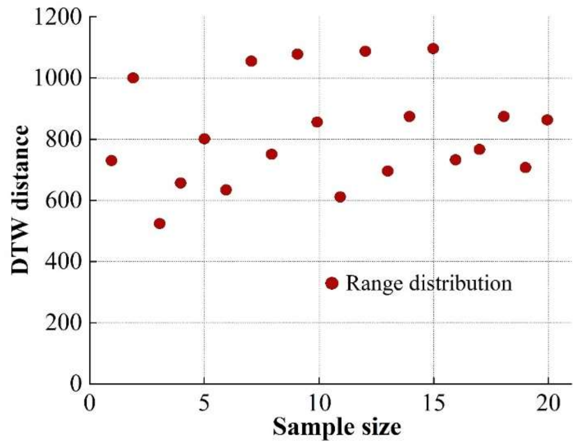

Fig. 6. The DTW distance distribution of the Angle of the left elbow

Figure 7 shows the DTW distance distribution for the right shoulder joint angle and Figure 8 shows the DTW distance distribution for the right elbow joint angle. From the figure, it can be seen that the DTW distance distribution of the right shoulder joint angle is between 90 degrees and 250 degrees; the DTW distance distribution of the right elbow joint angle is between 100 degrees and 240 degrees.

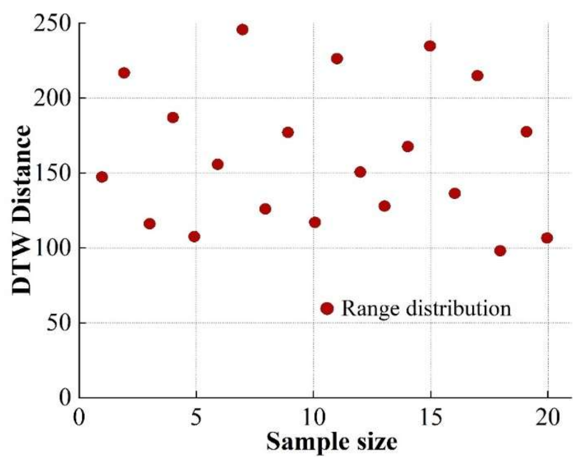

Fig. 7. The DTW distance distribution of the Angle of the right shoulder

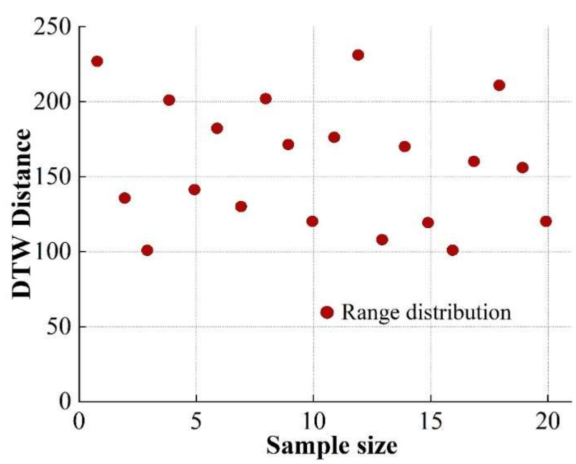

Fig. 8. The DTW distance distribution of the Angle of the right elbow

The DTW distance distributions for the left knee and left hip joint angles are shown in Figures 9 and 10, respectively.

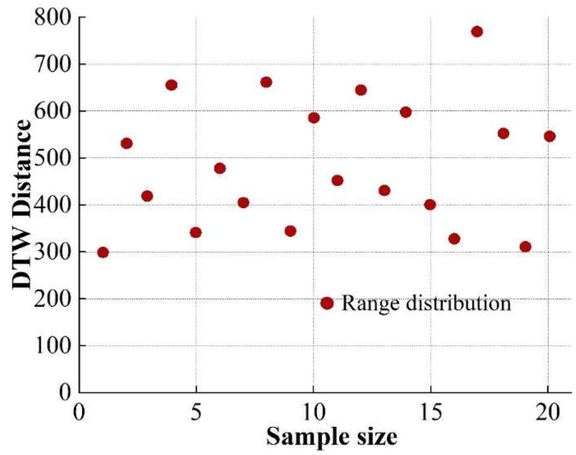

Fig. 9. The DTW distance distribution of the left knee Angle

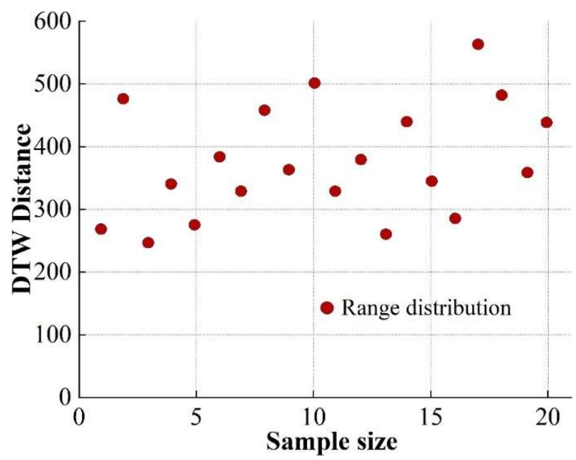

Fig. 10. The DTW distance distribution of the Angle of the left crotch

The DTW distance distributions for the right knee and right hip joint angles are shown in Figures 11 and 12, respectively.

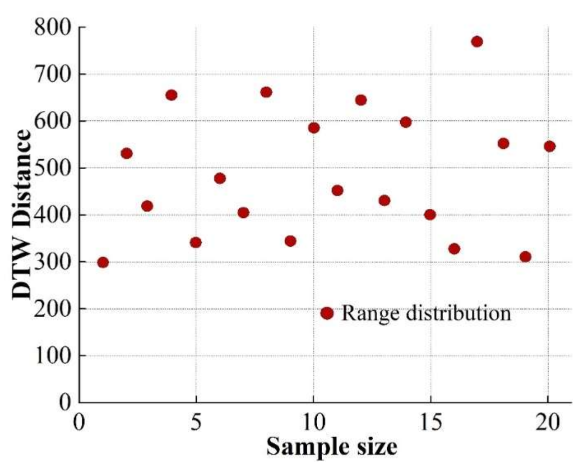

Fig. 11. The DTW distance distribution of the right knee Angle

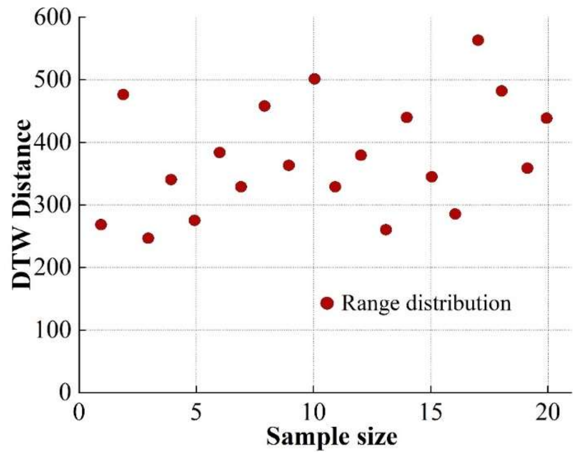

Fig. 12. The DTW distance distribution of the Angle of the right crotch

Analyzing the above experimental data, there are DTW distance distribution intervals for each joint angle, and the minimum values within the effective intervals for different joint angles and the loss parameters for joint angles are obtained after several experiments, as shown in Table 1.

Table 1. DTW distance effective interval minimum value and loss parameter value

Joint Angle name	DTW minimum	Loss function
Left shoulder	640	0.43
Left elbow	620	0.356
Right shoulder	110	0.663
Right elbow	110	0.691
Left knee	320	0.619
Left crotch	560	0.403
Right knee	330	0.551
Right crotch	170	0.438

The results of action evaluation are shown in Table 2, which lists the DTW distance values corresponding to the eight joint angles, as well as the evaluation scores of this paper and the professional scoring results. Summing the evaluation scores, we can get the total score of this paper as 95.4 points, and the total score of the professional evaluation as 95.9 points, with a difference of 0.5 points, which shows that the method of this paper can be effective in identifying and matching the sporadic fighting actions, and the validity of the algorithm of this paper can be verified. The validity of the algorithm of this paper is verified.

Table 2. Action evaluation results

Joint Angle name	d	Evaluation score	Professional rating
Left shoulder	820	12	12
Left elbow	1030	11	11.1
Right shoulder	270	11.6	11.6
Right elbow	150	12.5	12.6
Left knee	420	11.8	12
Left crotch	310	12.2	12.3
Right knee	430	12	12
Right crotch	170	12.3	12.3

3.2. Confrontation enhancement effect analysis for sparring players

3.2.1. System application experiment setup. Experiments on the application of the system of this paper were carried out in two groups of sparring players, and the experimental group applied the system to identify and analyze the movements, and developed and modified the training plan accordingly. The control group carried out traditional resistance training. The effect of the players' sparring backhand straight punch strikes was observed before the beginning and after the end of the experiment, respectively.

1) The subjects recruited in the experiment were all healthy sparring athletes with no history of sports injuries or illnesses within half a year, and all of them were able to participate in training normally;

2) Communicate with the recruited subjects before the beginning of the experiment to ensure that they can participate in the training normally and actively, and can cooperate with the completion of the tasks in the process, and make it clear in advance that they need to maintain a normal and healthy work and rest habits, and do not carry out other regular training activities during the period;

3) The pre- and post-tests of the two groups of subjects were conducted with the same personnel, equipment, methods and standards to exclude measurement errors;

4) During the experiment, the experimental subjects were not informed of the group to which they belonged (experimental and control groups), and the training of the two groups was the same except for whether or not the system in this paper was applied for the evaluation of action recognition;

5) Attendance of the experimental subjects' participation in the training was taken, and in case of absenteeism by leave, make-up training was conducted on the same day or the same week. The experiment lasted for 12 weeks.

3.2.2. Pre-test of the effect of a backhanded straight punch. The backhand straight punch striking effect is mainly reflected by three dimensions: strength, speed and accuracy, but considering the fit of accuracy and strength training, only strength and speed were analyzed in this study, and all of them used a single strike to reflect the striking effect of strength and speed. Table 3 shows the pre-test results of the experimental group and the control group.

There is no significant difference between the two groups of experimental subjects' backhand straight punch hitting effect, the mean values of the hitting strength index are 83.38 kg and 87.12 kg, respectively, $P = 0.523 > 0.05$, and the mean values of the hitting speed index are 0.38 seconds, $P = 0.284 > 0.05$, and the P-values of the strength and speed indexes are all greater than 0.05, which indicates that before the intervention training, the two groups of experimental subjects' The effect of backhand straight punch striking is basically the same, which meets the requirements for further experiments.

Table 3. Experimental group and control group

	Experimental group	Control group	t	P
Single strike strength (kg)	86.38±2.71	87.12±2.69	0.458	0.523
Single hit speed ((s))	0.38±0.05	0.38±0.05	0.913	0.284

3.2.3. Hitting speed test results. In Sanda competition, athletes only in the rules of the allowed range, the use of boxing technology clear and powerful to hit the opponent's scoring parts, can be called an effective punch, and get a point, and the backhand straight punch in the ever-changing game to seize the moment and effective, the formation of the “effective punch”, athletes need to be punched out of the box with sufficient The faster the speed, the shorter the reaction time given to the opponent, and the more likely that the opponent will be difficult to form an effective dodge and block. Table 4 shows the results of the experimental group and the control group.

According to the data in Table 4, after 12 weeks of training intervention, the single striking speed of the backhand straight punch of the two groups of experimental subjects had changes, and the average performance of the experimental group was improved from 0.38 seconds before the experiment to 0.32 seconds after the experiment, and the results of the paired-sample t-test of the average performance before and after the experiment were P<0.001, which was highly significant, indicating that the experimental group under the application of the action recognition and analysis system of the present paper The average performance of the control group after the experiment was improved from 0.38 seconds before the experiment to 0.36 seconds after the experiment, and the paired-sample t-test result of the average performance before and after the experiment was P<0.01, which is a highly significant difference, indicating that the traditional resistance training of the control group can also improve the single striking speed of the backhand straight punch of the sparring athletes, but the result was is poorer.

Table 4. The experimental group and the control group single hit speed test

	Before(s)	After(s)	T test	
			T value	P value
Experimental group	0.38	0.32	6.351	<0.001
Control group	0.38	0.36	7.682	<0.001
t	0.913	8.628		
P	0.284	0.008		

In order to reflect the difference between the two groups' achievements after the experiment, the achievements of the experimental group and the control group after the experiment were compared horizontally, and the results are shown in Table 5. Its independent samples t-test results show P < 0.01, highly significant difference, indicating that after the experimental intervention the two groups' performance has produced obvious differences, and the training effect of applying the systematic movement recognition and analysis in this paper is higher than the effect of traditional resistance training. The experimental group changes 0.06, while the control group changes 0.02, P < 0.01, highly significant difference, it can be considered that this paper's system sparring players backhand straight punch hitting speed enhancement effect is higher than the traditional resistance training enhancement effect.

Table 5. Test of single hit speed of the experimental group and control group

	Difference mean	Difference and 95% confidence interval	T test	
			T value	P value
Experimental group	-0.06	-0.019	-4.38	<0.001
Control group	-0.02			

In summary, both systematic training and traditional resistance training in this paper can enhance the striking speed of backhand straight punch of sparring athletes, but compared with the enhancement effect, the enhancement effect of composite training is more obvious, which is higher than the enhancement of traditional resistance training.

3.2.4. Hitting strength test results. In the competition, after the sanda player's straight punch touches the opponent's scoring part, the referee needs to determine whether there is a clear sound during the strike, whether the hit part has been displaced, etc., so as to determine whether the straight punch of the back hand is "powerful" and whether it can be judged as an "effective punch". Therefore, the striking effect of backhand straight punch needs to be supported by the strength, and at the same time, the strong and powerful backhand straight punch is also an important guarantee for knocking down the opponent to force the opponent to read the seconds or even directly KO the opponent to get the victory. Table 6 shows the comparative results of the single blow strength test results.

After 12 weeks of training intervention, the strength of the backhand straight punch of the two groups of experimental subjects in a single hit has been improved, and the average strength of the experimental group before and after the experiment has been improved from 86.38 kg to 95.83 kg, and the paired samples t-test result of the performance before and after the experiment is $P < 0.01$, which is highly significant difference, indicating that the systematic movement recognition training of the experimental group has an improvement effect on the strength of the backhand straight punch of the sparring athletes, and the improvement is extremely obvious. The average performance of the control group after the experiment was increased from 87.12 kg to 88.47 kg, and the paired-sample t-test result of the performance before and after the experiment was $P=0.002<0.01$, which was significant, indicating that the traditional resistance training of the control group could also improve the hitting strength of the backhand punch of the sparring athletes.

Table 6. Comparison results of single hit test results

	Before(s)	After(s)	T test	
			T value	P value
Experimental group	86.38	95.83	-16.803	<0.001
Control group	87.12	88.47	-4.452	0.002
t	0.458	6.554		
P	0.523	0.017		

Further comparison of the difference in the performance of the two groups after the experiment was carried out and the results are shown in Table 7. The mean of the difference between the experimental group before and after the experiment was 9.45 kg, while the mean of the difference in the control group was 1.35 kg, $P < 0.001$, highly significant difference, which is further supported by the fact that the experimental group in this paper's systematic movement recognition and evaluation training for the enhancement of the strength of the striking force is superior to the enhancement of the control group of the traditional resistance training.

Table 7. The difference between the results of the results

	Difference mean	Difference and 95% confidence interval	T test	
			T value	P value
Experimental group	9.45	5.42	7.82	<0.001
Control group	1.35			

In summary, after 12 weeks of training intervention, both the systematic training in this paper and the traditional resistance training were able to improve the single blow strength of the backhand straight punch of the sparring athletes, but there was a difference in the improvement effect, and the improvement effect of the systematic training in this paper was higher than the improvement effect of the traditional resistance training.

4. Conclusion

In this paper, deep learning algorithms are applied to the analysis of sparring sports, and a sparring action recognition and analysis system based on CNN-LSTM network is constructed. Using the DTW algorithm to calculate the distance between the eight joint angles of the test action and the standard action sequence as a similarity evaluation parameter, it is found that the evaluation scores of the system in this paper are basically the same as the professional scoring results, with a difference of only 0.5 points. The system can identify and scientifically and objectively evaluate the sparring movements in training, which can help targeted personalized training. In the control test of sparring action confrontation analysis, the experimental group in the application of this paper's system for action recognition and analysis of the conditions, sparring confrontation speed and strength of strikes than the control group have significant progress.

References

[1] Zhou, B. (2024). Sanda training program based on multi-sensor data fusion in the internet of things environment. Internet Technology Letters, 7(2), e470.

[2] Li, S. (2021, June). Analysis of Sanda action based on CCNN model of IBS distance measurement. In Journal of Physics: Conference Series (Vol. 1941, No. 1, p. 012055). IOP Publishing.

[3] Lei, Z., & Lv, W. (2022). Feature Extraction-Based Fitness Characteristics and Kinesiology of Wushu Sanda Athletes in University Analysis. Mathematical Problems in Engineering, 2022(1), 5286730.

[4] Wu, D., Sharma, N., & Blumenstein, M. (2017, May). Recent advances in video-based human action recognition using deep learning: A review. In 2017 International joint conference on neural networks (IJCNN) (pp. 2865-2872). IEEE.

[5] Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. ACM Computing Surveys (CSUR), 54(4), 1-40.

[6] Kumar, P., Chauhan, S., & Awasthi, L. K. (2024). Human activity recognition (har) using deep learning: Review, methodologies, progress and future research directions. Archives of Computational Methods in Engineering, 31(1), 179-219.

[7] Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. Sensors, 22(4), 1476.

[8] Zhang, H. B., Zhang, Y. X., Zhong, B., Lei, Q., Yang, L., Du, J. X., & Chen, D. S. (2019). A comprehensive survey of vision-based human action recognition methods. Sensors, 19(5), 1005.

[9] Kong, Y., & Fu, Y. (2022). Human action recognition and prediction: A survey. International Journal of Computer Vision, 130(5), 1366-1401.

[10] Kamel, A., Sheng, B., Yang, P., Li, P., Shen, R., & Feng, D. D. (2018). Deep convolutional neural networks for human action recognition using depth maps and postures. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 49(9), 1806-1819.

[11] Koohzadi, M., & Charkari, N. M. (2017). Survey on deep learning methods in human action recognition. IET Computer Vision, 11(8), 623-632.

[12] Andrade-Ambriz, Y. A., Ledesma, S., Ibarra-Manzano, M. A., Oros-Flores, M. I., & Almanza-Ojeda, D. L. (2022). Human activity recognition using temporal convolutional neural network architecture. Expert Systems with Applications, 191, 116287.

[13] Al-Faris, M., Chiverton, J., Ndzi, D., & Ahmed, A. I. (2020). A review on computer vision-based methods for human action recognition. Journal of imaging, 6(6), 46.

[14] Dai, W., Chen, Y., Huang, C., Gao, M. K., & Zhang, X. (2019, July). Two-stream convolution neural network with video-stream for action recognition. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[15] Dai, C., Liu, X., & Lai, J. (2020). Human action recognition using two-stream attention based LSTM networks. Applied soft computing, 86, 105820.

[16] Javed, M. H., Yu, Z., Li, T., Rajeh, T. M., Rafique, F., & Waqar, S. (2022). Hybrid two-stream dynamic CNN for view adaptive human action recognition using ensemble learning. International Journal of Machine Learning and Cybernetics, 1-10.

[17] Wang, L., Ge, L., Li, R., & Fang, Y. (2017). Three-stream CNNs for action recognition. Pattern Recognition Letters, 92, 33-40.

[18] Ding, Z., Wang, P., Ogunbona, P. O., & Li, W. (2017, July). Investigation of different skeleton features for cnn-based 3d action recognition. In 2017 IEEE International conference on multimedia & expo workshops (ICMEW) (pp. 617-622). IEEE.

[19] Soo Kim, T., & Reiter, A. (2017). Interpretable 3d human action analysis with temporal convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 20-28).

[20] Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Systems with Applications, 105, 233-261.

[21] Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. IEEE Sensors Journal, 21(12), 13029-13040.

[22] Islam, M. M., Nooruddin, S., Karray, F., & Muhammad, G. (2022). Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects. Computers in biology and medicine, 149, 106060.

[23] Mekruksavanich, S., & Jitpattanakul, A. (2023). Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition. Scientific Reports, 13(1), 12067.

[24] Janidarmian, M., Roshan Fekr, A., Radecka, K., & Zilic, Z. (2017). A comprehensive analysis on wearable acceleration sensors in human activity recognition. Sensors, 17(3), 529.

[25] Zhao, Y., Man, K. L., Smith, J., Siddique, K., & Guan, S. U. (2020). Improved two-stream model for human action recognition. EURASIP Journal on Image and Video Processing, 2020, 1-9.

[26] Xiong, Q., Zhang, J., Wang, P., Liu, D., & Gao, R. X. (2020). Transferable two-stream convolutional neural network for human action recognition. Journal of Manufacturing Systems, 56, 605-614.

[27] Wang, Z., Lu, H., Jin, J., & Hu, K. (2022). Human action recognition based on improved two-stream convolution network. Applied Sciences, 12(12), 5784.

[28] Zhou, Y., Sun, X., Zha, Z. J., & Zeng, W. (2018). Mict: Mixed 3d/2d convolutional tube for human action recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 449-458).

[29] Yang, H., Yuan, C., Li, B., Du, Y., Xing, J., Hu, W., & Maybank, S. J. (2019). Asymmetric 3d convolutional neural networks for action recognition. Pattern recognition, 85, 1-12.

[30] Liu, K., Liu, W., Gan, C., Tan, M., & Ma, H. (2018, April). T-C3D: Temporal convolutional 3D network for real-time action recognition. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).

[31] Akshay Rajendra Patil ,Sandaram Buchaiah& Piyush Shakya. (2024). Combined VMD-Morlet Wavelet Filter Based Signal De-noising Approach and Its Applications in Bearing Fault Diagnosis. Journal of Vibration Engineering & Technologies(prepublish),1-25.

[32] Jonas Kohler,Thomas Bielser,Stanislaw Adaszewski,Basil Künnecke & Andreas Bruns. (2024). Deep learning applied to the segmentation of rodent brain MRI data outperforms noisy ground truth on full-fledged brain atlases..NeuroImage120934.

[33] Nguyen Quoc Minh, Nguyen Trong Khiem & Vu Hoai Giang. (2024). Fault classification and localization in power transmission line based on machine learning and combined CNN-LSTM models. Energy Reports5610-5622.