International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 3829-3841

https://doi.org/10.70517/ijhsa463325

Research on action recognition and tactical confrontation decision-making of sparring players based on artificial intelligence optimization algorithm

Ranfeng Fan[1,*]
[1] Police Combat Teaching and Research Department, Guangxi Police College, Nanning, Guangxi, 530000, China
Corresponding authors: (e-mail: he1231512315@163.com).

Abstract Sanshou is becoming more and more popular as a comprehensive sport that integrates traditional Chinese martial arts and modern fighting techniques. In this paper, a CNN-LSTM model fused by a convolutional neural network (CNN) and a long-short-term memory network (LSTM) is proposed, which is capable of capturing global information and sequence relationships to realize accurate and real-time recognition of sparring techniques. A dynamic time planning algorithm is utilized to calculate the difference in movement joint angles between the standard and test movement sequences, and a movement evaluation formula is defined to evaluate the sparring movements, which provides intuitive and reliable training suggestions for the trainees, thus improving their learning efficiency. The research results show that the CNN-LSTM model can recognize six kinds of sparring actions with 98.69% recognition accuracy, while the action scoring model proposed in this paper can realize the effective evaluation of sparring actions and achieve fair and impartial auxiliary scoring. Finally, the sparring tactics confrontation decision is proposed to improve the tactical quality and effect of the students during the competition.

Index Terms action recognition, convolutional neural network, tactical decision making, DTW

I. Introduction

In recent years, the sport of Sanshou has been gaining attention and popularity worldwide, and its popularity has been rising [1]. Sanshou originated from Chinese combat sports, which integrates traditional Chinese martial arts with modern combat techniques, and as a kind of fighting confrontation program that is won through scoring, its competition decision-making is the fundamental guarantee for the success of the competition, leading the development direction of Sanshou competitions [2], [3]. As the traditional scoring method in sparring matches is manual scoring by five side judges, this method may have different scoring standards, and may also be interfered by off-court shouts, which is very easy to omit judgments, misjudgment phenomenon, and even the use of the dark score system, so that the sparring matches lack of fairness [4]-[6]. In order to effectively avoid the negative impact caused by subjective factors in the match, improve the fairness and impartiality of the sparring match, and promote the benign development of Chinese Wushu sparring program, an auxiliary scoring system integrating motion analysis technology is proposed [7], [8].

The rapid development of artificial intelligence promotes the research of human movement recognition technology, which is widely used in the fields of smart home, intelligent security, and motion analysis [9]. In the field of motion analysis, the combination of human action recognition technology and sparring action can realize the recognition of sparring action, and then by comparing the similarity, the effect of assisted scoring can be realized [10]-[12]. Intelligent auxiliary scoring system can not only improve the fairness of the sparring match, but also promote, promote the analysis of the coach's competition tactical confrontation decision-making, and promote the development of intelligent sports equipment technology [13], [14].

In the study, eight persons are firstly selected for sparring action recognition, and effective data are obtained through the collection of sensors worn on the body. Based on the preprocessed action data, a multi-sensor data fusion approach is adopted to combine convolutional neural network and long short-term memory network, and a deep learning neural network model - CNN-LSTM is proposed to realize accurate and real-time recognition of sparring techniques. Then an auxiliary training system based on Kinect is designed and developed, while the joint angle features of the sparring movements are selected, the distance between the joint angles of the standard action sequence and the test action sequence is calculated by using the dynamic time planning algorithm, and the action evaluation formula is defined to evaluate the sparring movements.

II. Sparring player action recognition based on artificial intelligence optimization algorithm

II. A. Data Acquisition and Processing

II. A. 1) Data acquisition

The sensors used in this paper to acquire the signals of the casual hitting action are mainly composed of a board switch, power supply, radio transceiver DA14583, and a motion sensing chip BMI160. The whole inertial sensor has the dimensions of 10.3mm×8.7mm×2mm, the weight of 2g, and the acquisition frequency of 50Hz, which is worn above the ankle to acquire the three-axis acceleration and angular velocity data generated during the striking action.

Data acquisition was conducted in a sparring competition venue with 8 participants, including 4 sparring professionals and 4 sparring amateurs.The 8 participants wore the sensors on their left and right wrists, and each of them performed 150 punches and kicks, respectively. After data collection and deletion of invalid action data, a total of 2975 valid sample data were collected from the eight participants for both actions.

II. A. 2) Data pre-processing

The action signals acquired through the sensors usually contain a certain amount of noise interference, and these interference signals mainly come from the experimenter's own shaking, the inherent noise of the sensor hardware, and the noise generated during the transmission of signal data. Before the signal analysis, the acquired signal needs to be filtered to reduce the noise interference on the experiment.

The experiments in this paper choose the commonly used sliding mean filtering method for signal preprocessing, the principle of this method is to use the mean value of the signal in the window at a certain moment to represent the signal at that moment, by using this method can effectively solve the problem of interference with abnormal signals, and the application of this method in this paper is as follows.

Given the original timing signal $F=\{f_1,f_2,\cdots,f_n\}$, let the size of the sliding window be w_s, and the data $F'=\{x_1,x_2,\cdots,x_n\}$ after sliding mean filtering, x_m is the average of all the data in the window, i.e.:

$$x_m=\frac{1}{w_s}\sum_{i=m-d}^{m+d}f_i \tag{1}$$

In this experiment, the sliding window size w_s is set to 7, where $d=3$. After the sensor data is processed by sliding filter, each channel is normalized, and finally the data is segmented with a fixed-width window of 1.2s (the window size is 70), and the labels are labeled.

Since the sensor signal is a one-dimensional structure, a certain time dependence can be obtained by the traditional one-dimensional convolution, but the spatial dependence for different sensors and different time nodes of the same sensor cannot be fully utilized. Therefore, in this paper, the six-axis sensing data within each window is converted into a six-dimensional action graph to more fully utilize the temporal and spatial dependencies between different axes to improve the recognition performance.

II. B. CNN-HALSTM Model Recognition Framework

II. B. 1) CNN Neural Lattices

The essence of convolutional neural network is artificial neural network, which has the advantages of local connection and weight sharing on the basis of artificial neural network. Together with the characteristics of sampling dimensionality reduction, it can effectively reduce the complexity of the network, which further consolidates the centrality and representativeness of convolutional neural networks in the field of deep learning, and makes them more suitable for processing image tasks. Compared with simple artificial neural networks, the convolutional neural network model adds featured category layers, which are mainly composed of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. The number of model layers of the convolutional neural network can be adjusted according to the needs and purpose of the task. By constructing a multi-layer network, convolutional neural networks enable data abstraction and deep feature extraction [15].

II. B. 2) Long- and short-term memory networks

The Long Short-Term Memory Network (LSTM) is a special kind of recurrent neural network that controls the data flow by introducing gating units, thus effectively solving the gradient vanishing and gradient explosion problems encountered by traditional recurrent neural networks when dealing with long sequential data. EMG signals usually contain complex temporal dynamics that may exhibit correlations on longer time scales, and the unique structure of LSTM enables it to learn these long-term dependencies.

The working mechanism of the LSTM consists of processing the current input x_t and the hidden state h_{t-1} of the previous time step through a sigmoid function to generate the oblivious gate output f_t, which varies between 0 and 1. An output value of 0 means “completely forgotten” and 1 means “completely retained”. The formula is:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{2}$$

In Eq. (2): w_f is the weight; σ is the sigmoid activation function; b_f is the bias.

The input gate plays a key role in the process of cell state update. First, the output i_t of the input gate is generated by determining the part of information to be updated via sigmoid. Second, a new candidate value $\tilde{C}_t$ is generated through the tanh layer, which is a vector representing the new information that may be added to the state. Finally, the new cell state C_t is updated by multiplying the output f_t of the forgetting gate with the previous cell state C_{t-1} to discard useless information and multiplying the output i_t of the input gate with the candidate value $\tilde{C}_t$. The computational formula is:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

where: $\tilde{C}_t$ denotes a candidate value; $f_t * C_{t-1}$ denotes selective forgetting of irrelevant information; and $i_t * \tilde{C}_t$ denotes retaining useful information.

In LSTM, the output gate goes through a sigmoid layer to determine the cell state that should be output, and the cell state is adjusted through a tanh layer and multiplied with the output of the sigmoid layer to produce the final output h_t. The calculation formula is:

$$O_t = (W_O[h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = O_t * \tanh(C_t) \tag{7}$$

where: O_t denotes the activation function of the output gate.

II. B. 3) Multi-Layer Attention Model HALSTM

In this paper, there are 2 layers of HALSTM in the model, and each layer introduces a hierarchical attention mechanism, which is the word-level attention as well as the document-level attention firstly, the feature vector x_{iN} extracted from each joint by the CNN module is inputted into the joint coordinate encoder, and the LSTM is used to encode the bi-directionally of x_{iN} to get the bi-directionally encoded vectors $h_{iN} = \left[\overrightarrow{h_{iN}}, \overleftarrow{h_{iN}}\right]$. where $\overrightarrow{h_{iN}} = (x_{i1}, x_{i2}, \cdots, x_{iT})$, $\overleftarrow{h_{iN}} = (x_{iT}, x_{i,T-1}, \cdots, x_{i1})$; $[\cdot,\cdot]$ denotes the splicing of 2 vectors. In order to enable the neural network to automatically focus its “attention” on the coordinates of the joints that are more relevant for action recognition, an attention layer based on the coordinates of each joint is designed. The attention mechanism assigns different words according to a certain sparse weight and then recombines them. Firstly, the tanh activation function is used to nonlinearly transform h_{iN} to get μ_{iN}; then the Softmax function is used to calculate the importance of each joint point in the representation of the action information to get the weight coefficient a_{iN} of each joint point:

$$a_{iN} = \frac{\exp(\mu_{iN}^T \mu_w)}{\sum \exp(\mu_{iN}^T \mu_w)} \tag{8}$$

$$\mu_{iN} = \tanh(W_w h_{it} + b_w) \tag{9}$$

where: μ_w denotes the random initialization parameters; W_w, b_w are the hidden layer weights and biases. The weighted sum of each joint point feature encoding vector and the corresponding weight coefficients is performed to obtain the i th joint coordinate sequence vector S, which is used as the input to the joint coordinate sequence encoder.

The process of encoding the sequence of joint coordinates is similar to the process of encoding the coordinates of each joint point. Firstly, we encode S in both directions to get $h_i = \left[\overrightarrow{h_i}, \overleftarrow{h_i}\right]$, and then we linearly transform h_i by the tanh activation function to obtain the linear representation μ_i, and then the Softmax function is used to calculate the importance of each coordinate sequence for the representation of the action information to obtain the weight coefficient a_i for each joint coordinate sequence:

$$a_i = \frac{\exp(\mu_i^T \mu_s)}{\sum \exp(\mu_i^T \mu_s)} \tag{10}$$

$$\mu_i = \tanh\left(W_s h_i + b_s\right) \tag{11}$$

The feature vector v for each action category is obtained by weighted summation, i.e., the high-level features of the skeleton joint coordinate sequences with weights, and then the Softmax classifier is directly applied to classify the actions to obtain the action label P. By designing the joint coordinate-level “attention” model, the importance of each joint coordinate and each coordinate sequence is taken into account, which makes the model more effective in mining the features that contribute to the action classification, and better learns the covariance features.

II. B. 4) CNN-HALSTM modeling

The extracted skeleton joint point data is saved as a CSV file.Each row of data in the CSV dataset consists of 18 skeletal joint point coordinates of the human body, and each row of the joint point coordinate sequence represents an action, which transforms the behavior recognition problem into a joint coordinate classification problem based on action labels. Drawing on the idea of text classification, local high-level features between neighboring n joint coordinates and local correlations between joint point coordinates are extracted by CNN. HALSTM is used to mine the contextual links between joint coordinates and extract the time-series features of the skeleton sequence. To this end, the CNN-HALSTM network architecture is designed by combining CNN and HALSTM's own network structure characteristics, with convolution followed by cyclic recursion, so that the model can not only capture the high-level features around each joint coordinate, but also obtain the long-sequence dependencies between joint coordinates.

II. C.Experimental results and analysis

II. C. 1) Data sample set

The data are first processed for noise reduction and standardization, and after repeated comparison, the Coif5 wavelet basis function is finally selected to perform 5-layer multi-scale decomposition of the data. x-axis acceleration data wavelet denoising comparison curve is shown in Figure 1.

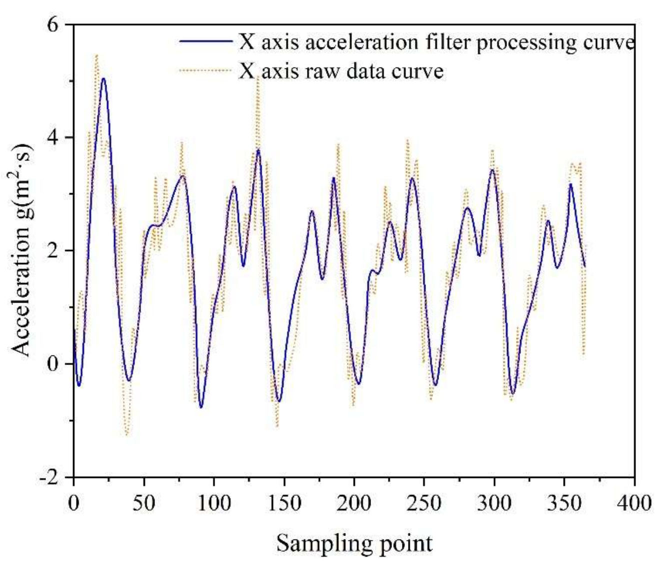
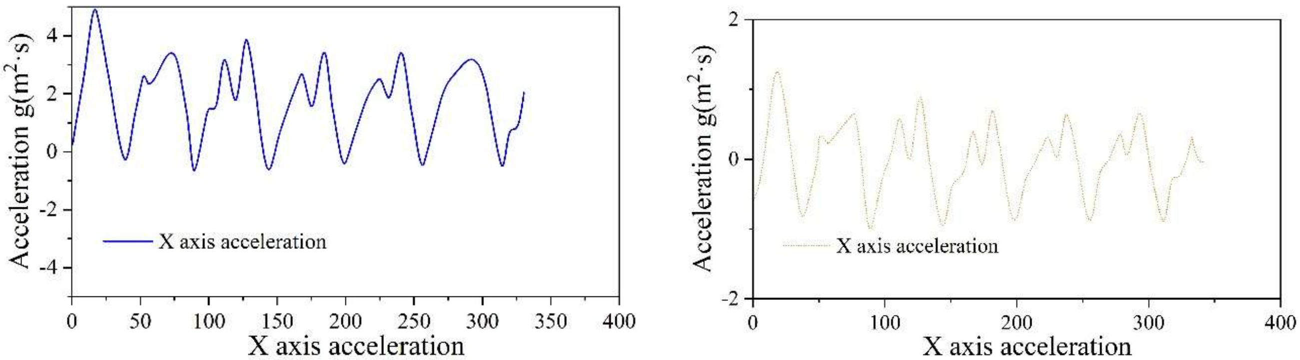


Figure 1: Sensor X-axis acceleration

The normalized data is shown in Figure 2, where it can be seen that the processed data maintains the original characteristics while reducing the magnitude.



(a) Raw data curve (b) Standardized data curve

Figure 2: Standardized data is compared to the original data curve

The sliding window splits the data into fixed lengths and becomes a sample after adding the label. According to the technical characteristics of the sparring sport and the sampling frequency of the sensor, after controlled experiments, the window width of 132 was finally selected, and sliding once every 67 sample points, i.e., 50% overlap rate, as shown in Fig. 3. Where S represents the sliding step size of 67.

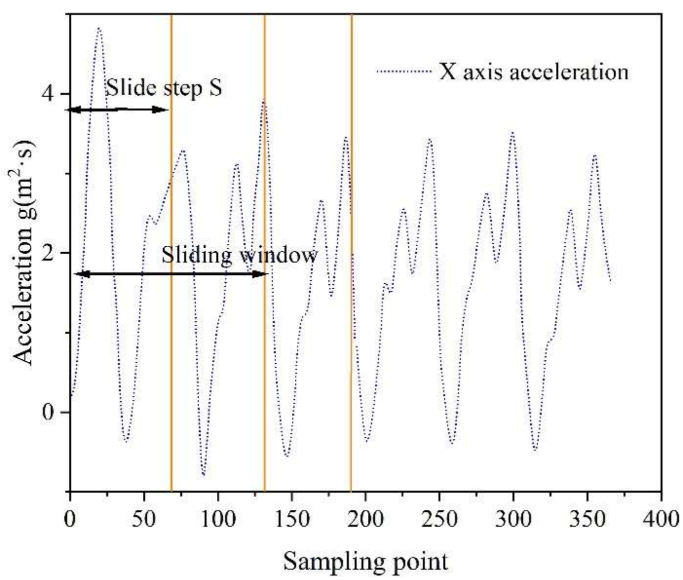


Figure 3: The sliding window log is processed by slicing

After data processing and sliding window segmentation, the final sample set was obtained as shown in Table 1. Use 67% quantity used for training and 33% for testing.

Table 1: Six kinds of loose action sample Numbers

Action pattern	Training set sample	Test set sample
Left hook fist	398	179
Left jab	423	147
left straight fist	288	128
Right hook	147	55
Right uppercut	429	101
Straight right	317	90
Total	2002	973

II. C. 2) Experimental results

In order to reduce the randomness, the model validation is averaged after 15 experiments. For the hyperparameter selection of the model, the CNN-LSTM model is analyzed by comparing different parameters, recording the accuracy and loss values of the test data and training data, and then the best parameters are selected.

(1) Effect of window segmentation length on model performance

According to the hitting rhythm of the sparring sport, the classification and recognition results of three sliding window segmentation methods were selected for comparison as follows:

a) The window length is 67 sampling points, and there is no overlap of neighboring windows.

b) Window length of 132 sampling points, no overlap of neighboring windows.

c) Window length of 132 sampling points, neighboring windows with 50% overlapping data.

Using the above three window segmentation methods, followed by the same processing and parameter configuration, the recognition results are shown in Figure 4. It can be seen that the 67 window length has the lowest recognition rate due to its inability to contain the complete interval of the sparring action data, while the 132 window length with no overlap contains incomplete information about the action, and the recognition rate is lower than that of the overlapping window segmentation method. So the segmentation window with a length of 132 and 50% overlap of neighboring window data is selected.

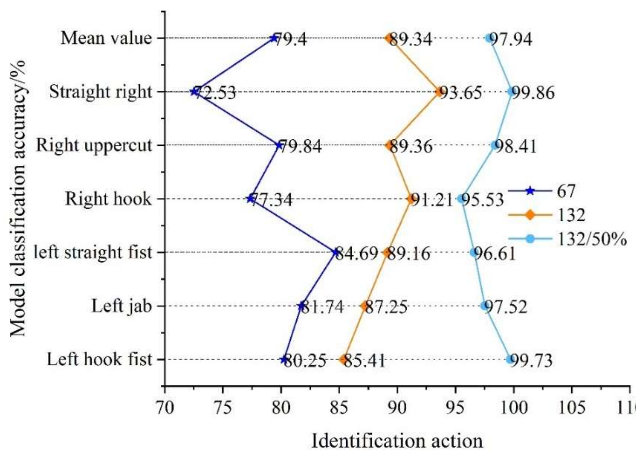


Figure 4: Comparison of different window segmentation methods

(2) Influence of neuron parameters on model performance

The neuron parameters directly affect the accuracy of the neural network model, comparing the accuracy of the model when filters are 8, 19, 34, 67, and 131, respectively, and the results are shown in Figure 5. It can be seen that the best classification accuracy is achieved when filters=67. In addition with the growth of parameters, the training parameters and training time of the model surge. Considering the recognition rate and training time, the number of neurons is chosen to be 67.

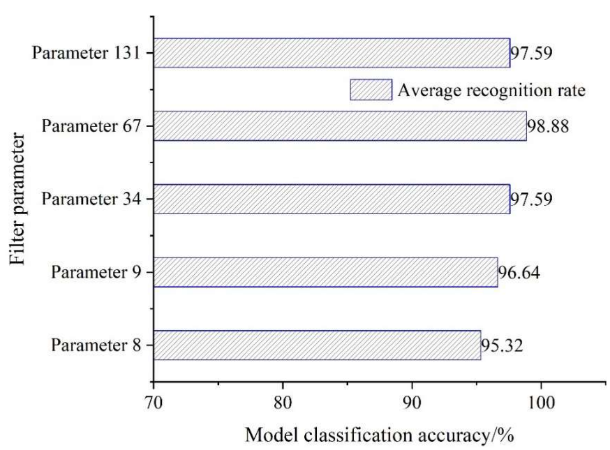


Figure 5: Classification accuracy of different filters parameters

The size of the Kernel is another important parameter of the neural network. Training was performed using the same experimental setup and testing a different set of kernels values: [2, 3, 5, 7, 11]. The final training results are shown in Figure 6. The results show that when the kernel size is 2, it possesses better performance and stability, so the kernel size of 2 is chosen as the kernel parameter of the model.

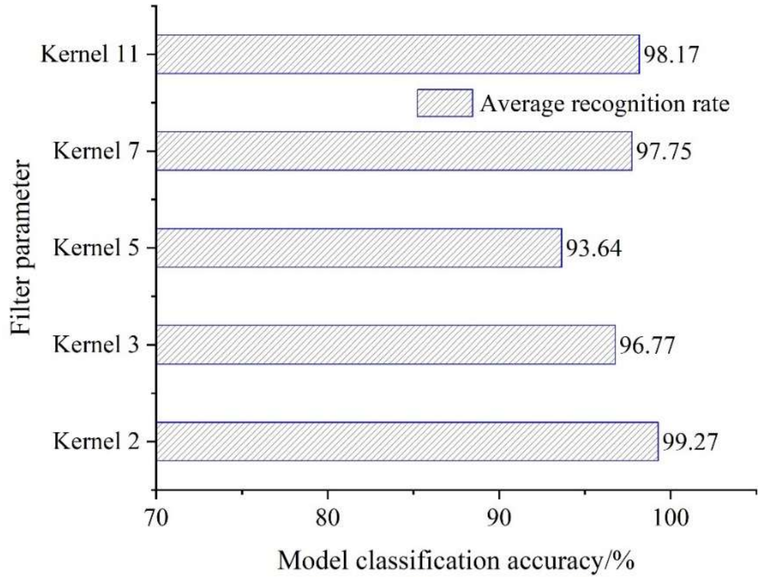


Figure 6: Different kernel size classification accuracy

(3) Model Evaluation

After determining the relevant parameters of the model (time step n=132, Dropout parameter is 0.4, the number of filters is 67, and the kernel size is 2), training is carried out, and the training process curve is obtained as shown in Fig. 7. It can be seen that, as the iteration continues, the model recognition accuracy continues to improve, and the model converges quickly; the 10th iteration tends to stabilize, and the average training time for each epoch is 1.55s, and for the test set, the recognition time is 0.0059s, and the accuracy rate reaches 98.69%.

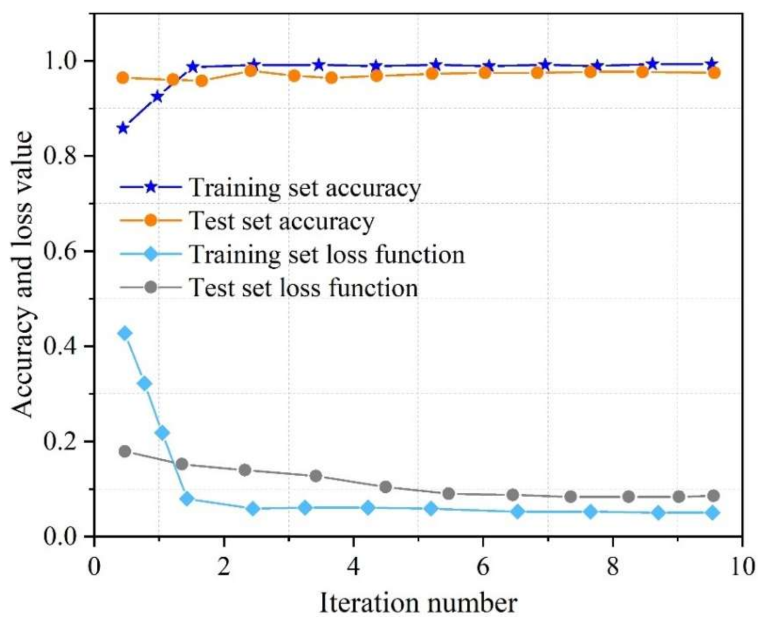


Figure 7: Model training process curve

The confusion matrix is a visualization method for evaluating classification effectiveness, which provides the overall classification rate of the model. Within the variance range, the confusion matrix drawn from the results of one experiment closest to the mean is selected here, as shown in Fig. 8, as a way to observe the accuracy of the model in classifying different categories of sparring moves. The number of samples judged correctly is indicated in the diagonal line in the figure, and it can be seen that the model can recognize all six categories of sparring moves well, but there are recognition errors for the two movement patterns of left swing and left hook, and the misclassification of the two is mainly due to the fact that the athletes have made the moves with high similarity of individual inertia information, which leads to recognition errors.

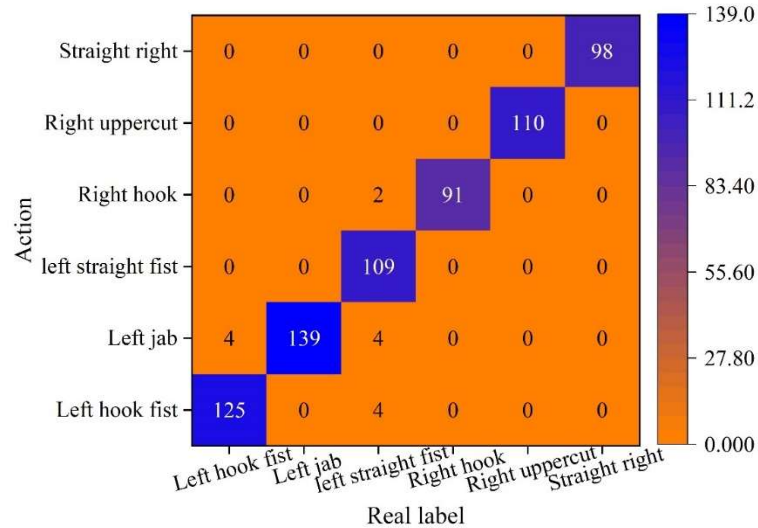


Figure 8: The confusion matrix of the model

III. Scoring and Confrontation Decision Making Based on Sparring Action Recognition

III. A. DTW-based action scoring design

III. A. 1) Extraction of angular features

The Kinect can be utilized to obtain the coordinate information of the joint points of the movement, which contribute to the state description of the movement to different degrees. According to the sports characteristics of sparring, the spatial position of the limbs determined by the rotational characteristics of the joints plays a key role in the action representation during the whole process. Therefore, this paper focuses on the joint angle characteristics of the main joint points of shoulder, elbow, hip and knee in the static posture.

III. A. 2) Extraction rate characteristics

When acquiring the skeleton sequence, the time interval of acquiring each frame is fixed, so the magnitude of the instantaneous velocity of the joint point in each frame is equal to the difference of the joint point motion from that frame to the previous frame. The calculation formula is shown in equation (12):

$$S_{Joint,i} = d\left(P_{Joint,i}, P_{Joint,i-1}\right) \tag{12}$$

where: Joint denotes any one of the 20 joints in the human skeleton, $P_{Joint,i}$ denotes the spatial coordinate of the joint in the i th frame, $d\left(P_{Joint,i}, P_{Joint,i-1}\right)$ denotes the Euclidean distance of the 2 vectors, and $S_{Joint,i}$ denotes the instantaneous velocity of the Joint joint in the i th frame.

Figure 9 shows a schematic diagram of the computation of the instantaneous velocity of the joint. The skeleton of the previous frame is represented by a solid line, and the skeleton of the next frame is represented by a dashed line, and the instantaneous velocity of a particular joint is the connecting line between the corresponding joints in the figure.

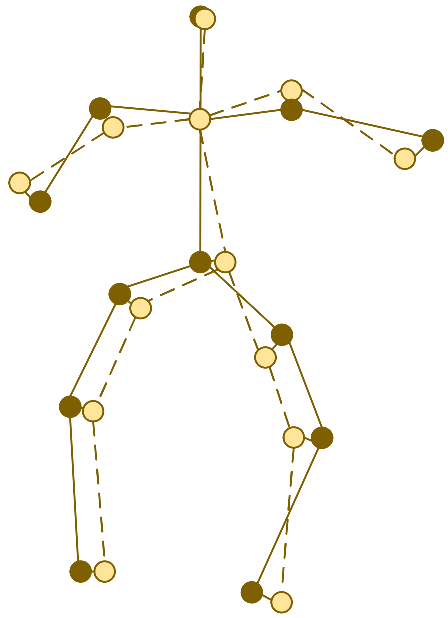


Figure 9: Diagram of the instantaneous velocity of the joint points

III. A. 3) DTW calculation of the corresponding frame

Regarding the corresponding frame problem, in order to solve the computational efficiency problem, equal interval uniform sampling is performed to find the key frames without considering the corresponding accuracy. The system adopts the DTW algorithm to calculate the corresponding frames based on the angular features.The DTW algorithm calculates the similarity between the test sequences and the standard sequences by extending and shortening the time sequences, and establishes a time-calibrated matching path between the test sequences and the standard sequences. The path that minimizes the cumulative distance between the two sequences during the matching process is solved as the optimal path.

The following is the process of DTW calculation.

Kinect collects data at a rate of about 30 frames/s, and 1 action sequence can be viewed as a collection of consecutive multiple frames of skeleton data. Define 2 action sequences $U=(U_1,U_2,\cdots,U_i,\cdots,U_m)$ and $V=\left(V_1,V_2,\cdots,V_j,\cdots,V_n\right)$, U is the test sequence and V is the standard sequence with lengths of m and n frames, respectively. U_i denotes the angular characterization of the human body posture in the i th frame of the action sequence U.

Typically $m\neq n$, a matrix grid of $m\times n$ needs to be constructed to define the matrix D.

$$D=\begin{bmatrix} d(U_1,V_1) & d(U_1,V_2) & \cdots & d(U_1,V_j) & \cdots & d(U_1,V_n) \\ d(U_2,V_1) & d(U_2,V_2) & \cdots & d(U_2,V_j) & \cdots & d(U_2,V_n) \\ d(U_i,V_1) & d(U_i,V_2) & \cdots & d(U_i,V_j) & \cdots & d(U_j,V_n) \\ \vdots & \vdots & & \vdots & & \vdots \\ d(U_m,V_1) & d(U_m,V_2) & \cdots & d(U_m,V_j) & \cdots & d(U_m,V_n) \end{bmatrix} \tag{13}$$

where $d(U_i,V_j)$ denotes the Euclidean distance of the corresponding angular feature of U_i,V_j.

The goal is to find 1 shortest path in this grid matrix that passes through a number of grid points, which are the points of the corresponding frames of the 2 sequences. Define this shortest path as a regularized path $W_k=\{\omega_1,\omega_2,\cdots,\omega_k\}$, a set of contiguous matrix elements that are used to map the correspondences of the angular features of the action sequence U,V. The optimal path of the distance matrix is computed, and if a frame U_i in the sequence U and a frame V_j in the sequence V, correspond to the same element in the optimal path, then U_i and V_j are the corresponding frames.

III. A. 4) Calculation of scores

Suppose that 2 action sequences $A=(A_1,A_2,\cdots,A_i,\cdots,A_m)$ with lengths of m -frames and n-frames respectively and $B=(B_1,B_2,\cdots,B_i,\cdots,B_n)$ have k corresponding frames. A is the test sequence and B is the standard sequence. Assume that B_i denotes any frame in sequence B, and the frame corresponding to frame B_i in sequence A is denoted by B_i'. The angular feature corresponding to frame B_i is θ_i, and the angular feature corresponding to frame B_i' is denoted by θ_i', then the distance between the angular feature of frame B_i' and the angular feature of frame B_i is denoted by the Manhattan distance of θ_i and θ_i' (e.g., D_{Angle} in Eq. (14)), and the distance between the test action sequence A and the standard action sequence B, i.e., summing and averaging the distances between each frame in A and its corresponding frame (i.e., D_{Angle}).

$$D_{Angle}(A,B)=\frac{1}{n}\sum_{i=1}^{n}D_{Angle}^{i}(B_i',B_i) \tag{14}$$

The principle of calculating the velocity difference between the corresponding frames of sequence A and sequence B is similar. Let the velocity feature of B_i frame be v_i and the velocity feature of B_i' frame be v_i'. The velocity characteristic distance between sequence A and sequence B is shown in equation (15).

$$D_{Speed}(A,B)=\frac{1}{n}\sum_{i=1}^{n}D_{Speed}^{i}(B_i',B_i) \tag{15}$$

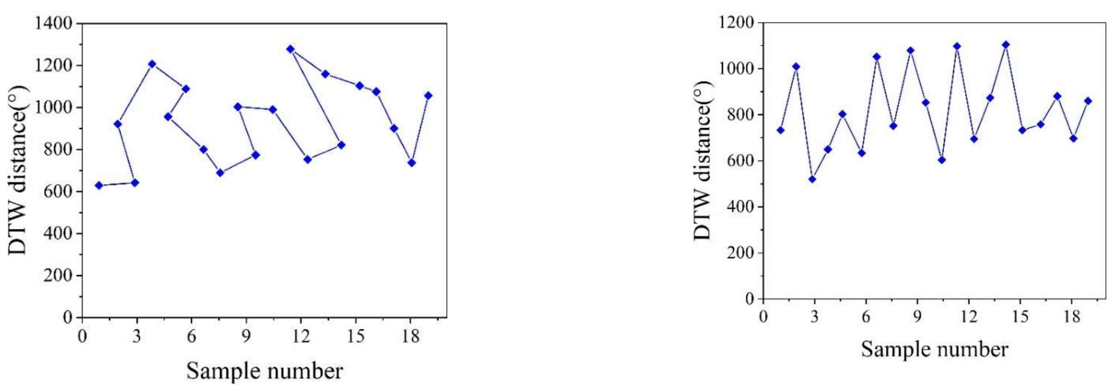
Equation (15) calculates the distance between 2 action sequences, i.e., the more similar the 2 action sequences are, then the smaller the distance value is, and the less similar they are, the larger the distance value is. The following mapping is done by mapping the distance value into a score, i.e., if the larger the distance, the smaller the score, and the smaller the distance, the larger the score. The mapping formula is shown in equation (16):

$$S=\alpha(D-\beta) \tag{16}$$

where: S denotes the angle or speed score; D denotes the characteristic distance of angle or speed; α and β are mapping parameters, which should be selected according to the actual test situation, so that the score can be mapped to a reasonable range.

III. B. Scoring of Sporadic Moves

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.

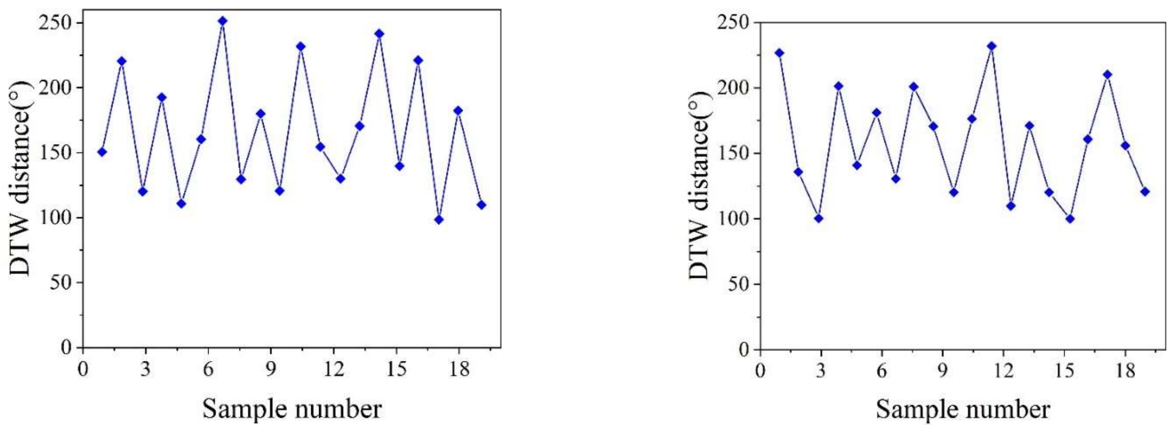


(a) Left hook fist (b) Left jab

Figure 10: Distribution of DTW distances for left hook and left swing

The DTW algorithm was utilized to calculate the distances of the 8 joint angles of the test movement and the standard movement sequence as a similarity assessment parameter. Observe the distribution of joint angle distances through multiple calculations, Figure 10 shows the distribution of DTW distances for the joint angles of the left hook and the left swing, most of the DTW distances for the joint angles of the left hook are distributed between 630 and 1200 degrees; most of the DTW distances for the joint angles of the left swing are distributed between 610 and 1100 degrees, and a small number of them are distributed between 500 and 600 degrees, which will be a small number of discarded.

Figure 11 shows the distribution of DTW distances for the joint angles of the left straight punch and the right hook, from which it can be seen that the DTW distances for the joint angles of the left straight punch are distributed between 100 and 250 degrees; and for the joint angles of the right hook, the DTW distances are distributed between 100 and 230 degrees.



(a) left straight fist (b) Right hook

Figure 11: The dtw distance distribution of the right straight punch and the right hook punch

The distribution of DTW distances for the right swing and right straight jab joint angles is shown in Figure 12.

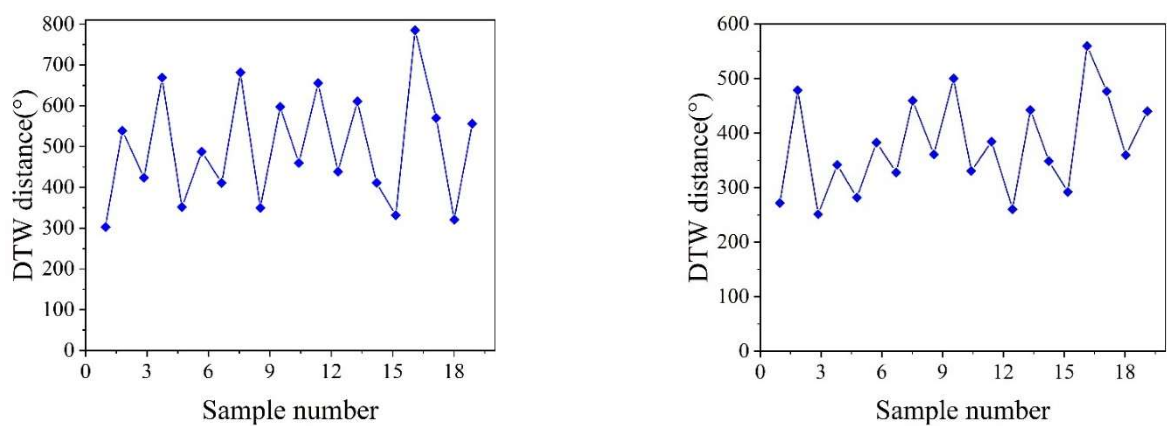


Figure 12: Distribution of dtw distance between right swing punch and right straight punch

The above experimental data were analyzed to define the front swing evaluation method as shown in equation (17).

$$e_a = e_c - \frac{f_c}{100} \times (d - d_{\min}) \tag{17}$$

In Eq. (17), e_a is the score for a single angle feature. e_c is the score assigned to a single angle, and since there are 8 joint angles, e_c takes the value of 12.5 points, making a full score of 100. d_1 is the DTW distance value, $d_{\min}$ is the minimum value within the valid interval of the DTW distance, and f_c is the loss parameter.

The sum of the 8 joint angle scores is represented by E , which is expressed by equation (18).

$$E = \sum_{a=1}^{8} e_a \tag{18}$$

There are DTW distance distribution intervals for each joint angle, and the minimum values within the valid intervals for different joint angles and the loss parameters for the joint angles are obtained after several experiments, as shown in Table 2.

Table 2: DTW range minimum and loss parameter values

Action name	$d_{\min}$	f_c
Left hook fist	637	0.447
Left jab	615	0.358
left straight fist	105	0679
Right hook	105	0.712
Right uppercut	322	0.615
Straight right	252	0.431

Substituting the action samples into Eq. (17), the scores of action evaluation can be obtained, and the results of action evaluation are shown in Table 3, which lists the DTW distance values corresponding to the joint angles of the six actions, as well as the evaluation scores of this paper and the results of the professional scoring, and the evaluation scores can be summed up to get the total score of this paper's evaluation of 70.7 points, and the professional evaluation of the total score of 72.2 points, with the difference between the two scores of 2.9 points, which indicates that the evaluation formula defined in this paper can effectively assist in scoring the sparring actions fairly and equitably. This result shows that the evaluation formula defined in this paper can be effective in scoring sparring movements in a fair and impartial way.

Table 3: Action evaluation results

Action name	d	Evaluation score	Professional rating
Left hook fist	810	12.1	12.3
Left jab	1058	11.2	11.2
left straight fist	264	11.7	11.9
Right hook	137	12.5	12.6
Right uppercut	336	11.3	12.2
Straight right	415	11.9	12.0

III. C. Sporadic Tactical Confrontation Decision Making

(1) Active Strike Tactics

When the opponent is in a trance and unprepared, suddenly attack the opponent to achieve the purpose of preemptive strike. Surprise and unpreparedness are the main features of this tactic. In order to achieve the effect of surprise and unpreparedness in a match, one must give full play to one's own strengths, fight for the initiative, use unique moves that are difficult for the opponent to anticipate, and attack before the opponent does.

The attacker must be prepared for two kinds of thoughts.

When you meet an opponent who is hesitant or indecisive, you can use fast, accurate and ruthless direct attack methods. Once you have won, you can attack in a chain, not letting your opponent have a chance to catch his breath, especially at the beginning of the match, when the opponent is too cautious and uncertain, you can take the initiative and overwhelm the opponent, i.e., “the first to strike is the strongest, and the second to strike is doomed”.

(2) Static control, rear-guard tactics

This method is a kind of defensive counterattacking after the use of technology. There are two kinds of post attack, one is “post attack first”; the other is “post attack later”. The so-called “after the first” that is, when the opponent wants to attack but has not yet attacked, but there are signs of exposure of the moment, with rapid action to grab the opponent before the opponent to hit the opponent. The so-called “after the attack after the system” that is, the opponent in the attack to me, I quickly side flash or back flash, so that the opponent's attack falls through, in the opponent's offensive action is not yet fully recovered, rapid progress to counterattack each other. Defensive counterattack, the focus is on counterattack, defense is only for better offense.

(3) Bait and Switch Tactics

Baiting the attack is a kind of anticipatory defense-a pre-planned defensive move. That is, consciously show a part of the body, so that the opponent is wrongly perceived as a breakthrough, taking advantage of the opportunity to attack, while I plan the defensive action, the opponent will be attacked to prevent the counterattack after the counterattack. The other type of defense is non-anticipatory defense - that is, the opponent's attack method to do temporary emergency defense, or automated by the action of the completion of the defense and counterattack. Attack is the expected defense counterattack, the key in this tactic is to entice and false, only to make the illusion realistic, in order to achieve the enticement of the enemy, so that they enter into their own ambush.

IV. Conclusion

In this paper, CNN neural network is combined with LSTM to propose a method for recognizing sparring techniques based on CNN-LSTM model. The experimental results show that the CNN-LSTM model achieves 98.69% accuracy on the test data. The Kinect-based assisted training system designed and developed includes action acquisition module and action scoring module, and the action scoring module is based on the angular features of the action joint points and the instantaneous velocity features of the joint points, and adopts the dynamic time planning algorithm for fair and impartial assisted scoring of the sparring action.

References

[1] Jin, M. (2017). The promotion and development of martial arts sanda. In 2017 2nd international conference on education and education research (Eduer 2017) (pp. 504-507).

[2] Jingwen, L. I., Zheng, L. I., & Weixi, S. H. I. (2014). Experimental Study on Chinese Athletes’ Physical Distribution in Sanda Competition. Higher Education of Social Science, 6(1), 15-19.

[3] Qu, R. (2015). To discuss the development of Sanda from safety. Biomedical Engineering and Environmental Engineering, 17.

[4] Cheng, N. (2022). Sports fear psychology in martial arts athletes' physical conflict psychology. Revista de Psicología del Deporte (Journal of Sport Psychology), 31(2), 48-56.

[5] Sui, S., & Leng, Z. (2022, June). Research on The Restrictive Factors of The Development of Wushu Sanda. In Of Papers Presented at 2022 3rd Asia Sport Science Conference (ASSC).

[6] 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.

[7] Zhou, Y., Du, Y., Lu, X., & Chen, H. (2022). Learning of Martial Arts Action Decomposition Method Based on Image Recognition. Scientific Programming, 2022(1), 1054294.

[8] Wu, B., & Zhou, J. (2024). Video-Based Martial Arts Combat Action Recognition and Position Detection Using Deep Learning. IEEE Access.

[9] Hui, B. (2023). Visualization system of martial arts training action based on artificial intelligence algorithm. Soft Computing, 1-12.

[10] Huang, L., Song, J., Lin, X., Du, Y., & Yang, S. (2022). Research on kick motion before Sanda based on 3D wireless sensor network image. Concurrency and Computation: Practice and Experience, 34(13), e5894.

[11] Zhao, H. (2021). Design of sanda action reconstruction model based on 3d images. Wireless Communications and Mobile Computing, 2021(1), 6668612.

[12] Fu, Z., & Wu, Y. (2024). The application of artificial intelligence in sports competition scoring. Journal of Computational Methods in Sciences and Engineering, 14727978241307151.

[13] Chang, Y., & Su, M. (2021). Study on 3D Image Reconstruction Model of Sparring Action Based on Graph Neural Network (GNN). Computational Intelligence and Neuroscience, 2021(1), 6882467.

[14] Fan, K., & Gu, X. (2021). Image quality evaluation of Sanda sports video based on BP neural network perception. Computational Intelligence and Neuroscience, 2021(1), 5904400.

[15] Zhilong Liu,Qinchao Wang,Tongxi Li,Yanping Huang,Changhua Nie & Tianzhou Xie. (2025). State identification of floating reactor control Rod drive mechanism based on similar state signal reconstruction and multiple convolutional neural networks. Measurement,252,117359-117359.