

A Deep Learning Algorithm for Fusing Action Recognition and Psychological Characteristics of Wrestlers

Lianhong Yuan^{1*}, Yuan Yuan² and Jun Liu³

¹Hangzhou Polytechnic

²Zhejiang Guozi Robot Technology Co. LTD

³Zhejiang Tuofeng Intelligent Equipment Co. LTD

Hangzhou, 310014, China

[e-mail: ylh@mail.hzpt.edu.cn]*

* Corresponding author: Lianhong Yuan

*Received October 28, 2022; revised February 12, 2023; accepted February 20, 2023;
published March 31, 2023*

Abstract

Wrestling is one of the popular events for modern sports. It is difficult to quantitatively describe a wrestling game between athletes. And deep learning can help wrestling training by human recognition techniques. Based on the characteristics of latest wrestling competition rules and human recognition technologies, a set of wrestling competition video analysis and retrieval system is proposed. This system uses a combination of literature method, observation method, interview method and mathematical statistics to conduct statistics, analysis, research and discussion on the application of technology. Combined the system application in targeted movement technology. A deep learning-based facial recognition psychological feature analysis method for the training and competition of classical wrestling after the implementation of the new rules is proposed. The experimental results of this paper showed that the proportion of natural emotions of male and female wrestlers was about 50%, indicating that the wrestler's mentality was relatively stable before the intense physical confrontation, and the test of the system also proved the stability of the system.

Keywords: Deep Learning, Wrestling action, Action Recognition, Psychological Feature Analysis, Neural Network

This work is supported by the Key project of Hangzhou Polytechnic of Science and Technology (HKYZXZD-2022-1, HKYZXYB-2022-2), and Zhejiang Province New Talent Program for College Students (2021R457002, 2022R457A002).

1. Introduction

The emergence of wrestling is related to people's aggressive characters. Since the emergence of wrestling, it has gained a lot of attention with its extremely high interest and enjoyment. It is precisely with the increasing attention of wrestling, athletes' competitive level has attracted much attention, and thus bearing a great psychological burden. The current deep learning technology is becoming more and more mature, and there are more and more applications in sports, such as video analysis technology and action recognition technology, which are used in the training of athletes and can achieve good sports results. It is also the algorithms related to human action analysis and recognition that have broad application scenarios and needs, and have great significance and far-reaching influence on improving the level of artificial intelligence, human automatic control, and human living standards.

With the development of information technology and the arrival of the era of big data, the development and analysis of technical and tactical systems for various sports has attracted a lot of attention. Human recognition has a wide range of applications in medicine and biology, so there are many studies on it. In order to realize the recognition of human motion intention based on the model, Wang W studied the dynamic modeling and recognition of the lower limb rehabilitation robot iLeg [1]. Laraba S proposed a new motion sequence representation (seq 2 im-sequence-to-image), which projected motion sequences into the RGB domain. The 3D coordinates of the joints were mapped to red, green and blue values, thus, action classification became an image classification problem and algorithms from this field could be applied [2]. Kim YK conducted research on human motion recognition based on virtual reality (VR) technology [3]. Wu Q aimed to develop an effective method to identify upper limb movements based on EMG signals for community-based rehabilitation. The method can be applied to control systems in rehabilitation equipment to provide objective data for quantitative assessment. By decomposing the assessment activities of the Activities of Daily Living (ADL), an upper limb movement recognition target set was constructed [4]. He J compared the accuracy of four different classifiers in conservative motion recognition methods. Upper limb movements were monitored by using motion sensors and four conservative machine learning classifiers were selected for comparison, which mapped features to the Fugl-Meyer scale [5].

However, the research on the analysis of the psychological characteristics of athletes by the sports recognition and analysis platform is much less. H Kristjánsdóttir analyzed the differences in mental skills, mental toughness and anxiety of female football players according to their rank (national team, first division and second division); and predicted these three level (using a multivariate model) [6]. X Wu explored the psychological characteristics of athletes and their influencing factors in the context of COVID - 19. The results showed that negative psychological emotions such as anxiety and depression were common among athletes [7]. Through relevant research, it can be found that the analysis of the psychological characteristics of athletes has not attracted enough attention, and there are not many studies on it. Especially in combination with modern deep learning technology, the research on the analysis of the psychological characteristics of athletes is more scarce, so the research in this paper is very necessary.

The innovations of this paper are as follows: This paper develops a video-based wrestling match analysis system for the training of wrestlers, so that the coaches can manually label the key pose frames through the video of the match. The technical and

tactical statistical analysis report helps coaches intuitively and conveniently grasp the key points and points-loss movements of athletes in the game. In order to allow coaches to know the latest game dynamics anytime and anywhere through the Internet, check the technical and tactical analysis report of the game. The system develops the online retrieval subsystem of wrestling competition information, and realizes the interconnection with the wrestling competition video analysis subsystem. Finally, based on the deep learning network model, the technical key actions in the wrestling match are classified and identified.

2. Recognition of Wrestling Sports under Deep Learning

2.1 Wrestling and Deep Learning

Wrestling originated in Greece and has a long history [8]. Wrestling requires athletes to coordinate and cooperate with their legs, waist and hands, and fully display their strength and skills in the competition, which is conducive to the comprehensive development of human health and it can cultivate the tenacious and brave will and positive spirit of athletes, so it is loved by people all over the world and has developed rapidly around the world. Over the years, under the attention of the General Administration of Sports of the People's Republic of China and the strong leadership of various provinces and cities, Chinese classical wrestling has made great progress, and a number of outstanding athletes have emerged. Compared with the world wrestling power, there is still a big gap. In December 2013, the International Wrestling Federation revised the competition rules of wrestling competitions [9], among which the biggest modification was the competition rules of men's classical wrestling. The competition methods, scoring standards and negative restrictions of the current rules have been greatly revised, and higher requirements have been put forward for the standardization of athletes' competition movements. Competitive rules are an important basis for coaches to guide athletes' training. Therefore, people must use the latest competitive rules as the basis, rationally apply and effectively utilize the rules, do a good job in training and competition scientifically, and promote the improvement of athletes' performance.

In 2012, as the deep learning model AlexNet defeated the traditional machine learning algorithm, it became the champion of the ImageNet large-scale visual recognition challenge for image classification. This competition has a huge influence on the field of computer vision, and deep learning as a synonym of neural network is gradually familiar to people. Computer Vision is to let the computer simulate the human visual function, so that the machine can understand the objective world through observation like a human. As a branch of computer vision, deep learning has been developing rapidly and has become a field with many popular research topics and commercial applications. As a field in which deep learning has achieved theoretical and technological breakthroughs and applications earlier, computer vision brings great hope to the development of artificial intelligence, and it also marks the beginning of a new era of artificial intelligence. In the next few years, deep learning gradually expanded from the field of image recognition to various fields of machine learning [10-11]. Deep learning technology continues to break through traditional recognition methods and develops rapidly, breaking records and even far exceeding human capabilities in certain tasks, especially in image classification, target detection, speech recognition, language translation, and biological information processing and target tracking, etc., having made remarkable achievements. The research field of deep learning is shown in Fig. 1.

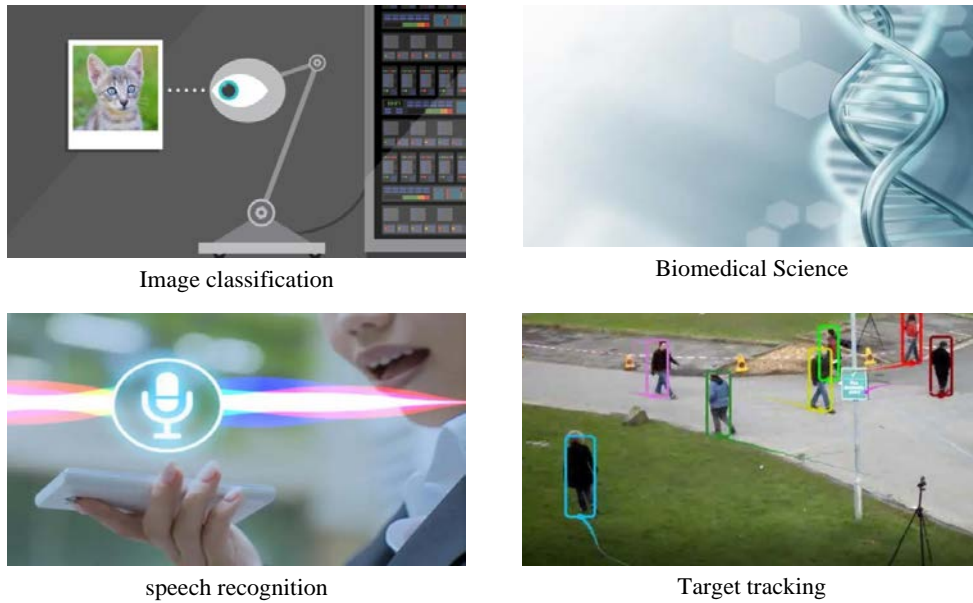


Fig. 1. Deep learning research field

In recent years, although human action recognition has achieved rapid development in the field of computer vision, it is still a fundamental and challenging task in computer vision due to differences in behavioral performance, environment and time. Human recognition technology and detection accuracy based on RGB video frame sequence have been greatly improved and significant research results have been achieved. However, due to the flexibility and diversity of human movements, it is affected by unavoidable factors such as light and dark conditions, image size, shooting angle, color and occlusion, and lack of three-dimensional spatial information. The impact of different environments on person recognition is shown in **Fig. 2** [12-13].



Fig. 2. Influencing factors of human body recognition

In wrestling, these influencing factors exist widely, so a more in-depth calculation method should be used for their identification. This paper adopts a human body identification method based on bone positioning.

2.2 LSTM Neural Network Bone Localization Algorithm

(1) Human skeleton sequence

The text is added as follows: The human skeleton is divided into left and right hands; left and right feet; left and right arms; left and right legs and trunk. Each part of the body obviously has a certain dependency or correlation, and the skeleton data can also be regarded as a sequence of certain dependencies between adjacent joint points. The human skeleton structure based on NTURGB+D is shown in [Fig. 3](#).

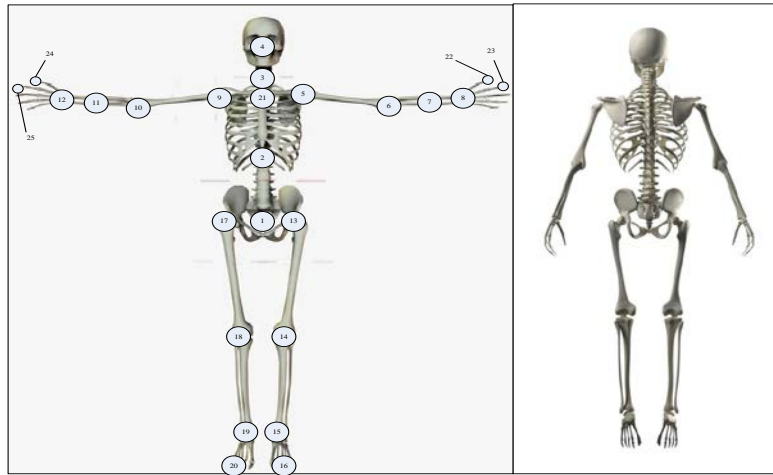


Fig. 3. Human skeleton structure

According to the temporal characteristics of skeleton sequences in the time domain, this chapter uses Long Short-Term Memory (LSTM) neural to extract skeleton temporal features from three branches: global, local and detail. Each branch consists of two layers of LSTMs. The first branch takes the overall skeleton point sequence [24, 25, 12, 11, 10, 9, 21, 5, 6, 8, 7, 8, 22, 23, 4, 3, 21, 2, 1, 17, 18, 19, 20, 13, 14, 15, 16] as input to a two-layer LSTM network, and the second layer of LSTM extracts the entire frame information. The second branch divides the body skeleton point sequence into the left branch [24, 25, 12, 11, 10, 9, 17, 18, 19, 20], the torso [1, 2, 3, 21, 4] and the right branch [22, 23, 8, 7, 6, 5, 13, 14, 15, 16], segmentation of body bones makes the model more sensitive to local body perception. The double-layer LSTM network is input, and the second layer of LSTM accepts the output of the first layer and processes it, making the judgment of actions more accurate. The third branch divides the body skeleton point sequence into the left arm [5, 6, 7, 8, 22, 23], the right arm [9, 10, 11, 12, 24, 25], the left leg [13, 14, 15, 16], right leg [17, 18, 19, 20] and torso [1, 2, 3, 21, 4], this branch temporal model is similar to the local temporal model to divide the human skeleton into more parts in order to better identify local detailed actions. It can effectively improve the accuracy of detailed action recognition.

(2) Time domain model

The first branch inputs the overall skeleton sequence into the double-layer LSTM network, and the double-layer LSTM performs feature extraction on the overall skeleton [14]. The second branch divides the human skeleton sequence points into three parts, the left branch, the trunk and the right branch. The segmentation of the human skeleton sequence makes the model more sensitive to local feature extraction, and the double-layer LSTM performs local feature extraction on the human skeleton sequence. The third branch divides the human skeleton sequence into five parts left and right arms, left and right legs and torso, which is similar to the second branch local time domain model. Dividing the human skeleton into more parts makes the model better for feature extraction of detailed actions [15-16]. The time network model structure is shown in Fig. 4.

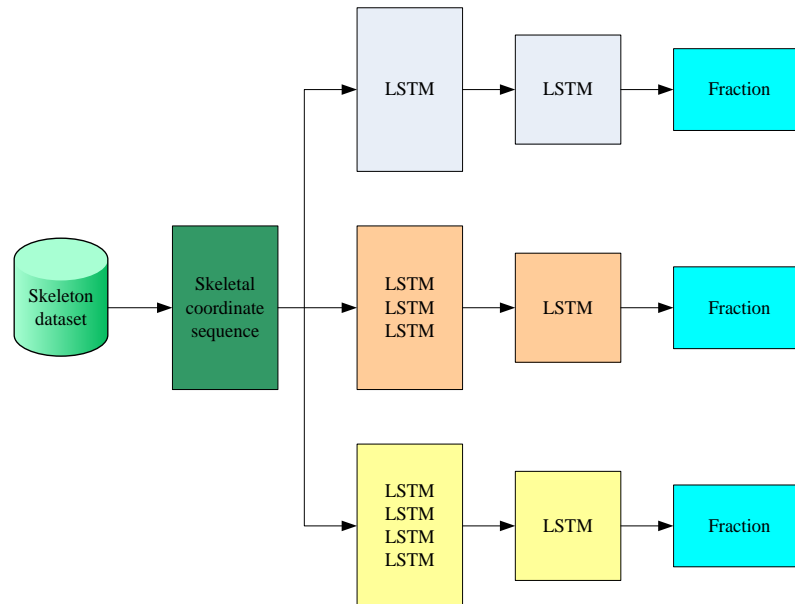


Fig. 4. Time domain model frame diagram.

Based on the NTURGB+D and UTD-MHAD datasets in the temporal model as shown in Fig. 4, the number of neurons in the global, local and detail networks are 128-256, 128-512 and 128-512. The number of neurons in the SBU interaction dataset are 64-64, 64-128 and 64-128. The Dropout of LSTM in the network model is 0.5.

The convolutional neural network adopts BP (Back Propagation) learning algorithm to learn the weight matrix and polarization parameters [17]. The idea of the BP algorithm is to update the parameters of each layer by passing the error between the current predicted result and the actual result in the direction of the fastest gradient descent to each layer of the neural network. After continuous iterative update, the optimal parameters are finally obtained. The mathematical basis of the BP algorithm is the chain rule of composite functions. In the process of forward propagation of the neural network, the output of each layer can be expressed as:

$$x^l = f(W^l x^{l-1} + b^l) \quad (1)$$

Where x^l represents the output of the W^l th layer, and b^l are the weight matrix and bias parameter of the l th layer network, respectively, and f is the activation function. The role of the activation function is to generate a nonlinear decision boundary by nonlinearly combining the inputs. Commonly used activation functions are sigmoid function and tanh function. The former converts the input real number to the $[0,1]$ interval, and the latter converts it to $[-1,1]$. Recently ReLU has become more and more popular as an activation function. When the input signal is less than 0, the output is 0, and when the input is greater than or equal to 0, the output is equal to the input. ReLU has the advantages of fast convergence speed and low computational complexity.

Convolution functions are widely used in the field of information processing. Continuous convolution is the fusion of two operations in the time dimension, expressed in mathematical formulas:

$$(f \cdot g)(\tau) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau \quad (2)$$

The convolution is extended to the discrete domain, and the corresponding mathematical expression is:

$$(f \cdot g)[\tau] = \sum_{m=-\infty}^{\infty} f(m)g(n - m) \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (3)$$

Hidden state is:

$$o_t = \sigma(W_0[h_{t-1}, X_t] + b_0) \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (4)$$

$$h_t = o_t * \tanh(C_t) \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (5)$$

When the data set has a total of N samples and is divided into c classes, its loss function is expressed by the mean square error as:

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (6)$$

t_k^n represents the real result of the k th dimension of the n th sample, y_k^n representing the prediction result. Among them, for one of the samples, its error can be expressed as:

$$E^n = \frac{1}{2} \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (7)$$

BP algorithm is based on the gradient descent strategy and adjusts the parameters in the direction of the negative gradient of the target. The algorithm is described as follows:

The number of training samples m ; the number of layers of the CNN model L ; the convolution kernel size K used by the convolution layer, the number of convolution kernels F , the padding size P , the step S ; the pooling size and step of the pooling layer, pooling method (average pooling or max pooling); the iteration step size in the gradient direction α , the maximum number of iterations MAX , are input, and the iteration will be stopped when the loss is lower than the threshold ϵ , and the activation function σ will be used.

The parameter matrix W, b of the network is output.

1) The parameter matrix of the network is randomly initialized.

2) Number of iterations are from 1 to MAX :

For $i = 1$ to m :

a) The x_i corresponding tensor is brought into a^l .

b) For $l = 2$ to $L-1$, the forward propagation is calculated according to the current layer type.

If it is currently a fully connected layer, there are:

$$a^{i,l} = \sigma(z^{i,l}) = \sigma(W^l a^{i,l-1} + b^l) \quad (8)$$

If it is currently a convolutional layer, there are:

$$a^{i,l} = \sigma(z^{i,l}) = \sigma(W^l a^{i,l-1} + b^l) \quad (9)$$

If the current pooling layer is:

$$a^{i,l} = \text{pool}(a^{i,l-1}) \quad (10)$$

If it is currently the output layer, there are:

$$a^{i,l} = \text{softmax}(z^{i,l}) = \text{softmax}(W^l a^{i,l-1} + b^l) \quad (11)$$

c) The output layer is calculated through the loss function $\delta^{i,L}$.

d) For $l=L-1$ to 2 , backpropagation is calculated according to the current layer type:

If it is currently a fully connected layer:

$$\delta^{i,l} = (W^{l+1})^T \delta^{i,l+1} \odot \sigma z^{i,l} \quad (12)$$

If it is currently a fully convolutional layer:

$$\delta^{i,l} = \delta^{i,l+1} * \text{rot180}(W^{l+1})^T \odot \sigma z^{i,l} \quad (13)$$

If the current full pooling layer:

$$\delta^{i,l} = \text{upsample} \delta^{i,l+1} \odot \sigma z^{i,l} \quad (14)$$

For $l = 2$ to L , update the first layer according to the type of the current layer W^l, b^l :

If it is currently a fully connected layer:

$$W^l = W^l - \alpha \sum_{i=1}^m \delta^{i,l} (a^{i,l-1})^T \quad (15)$$

$$b^l = b^l - \alpha \sum_{i=1}^m \delta^{i,l} \quad (16)$$

If it is currently a convolutional layer, for each convolution kernel there are:

$$W^l = W^l - \alpha \sum_{i=1}^m \delta^{i,l} * \text{rot180}(a^{i,l-1})^T \quad (17)$$

$$b^l = b^l - \alpha \sum_{i=1}^m \sum_{u,v} (\delta^{i,l})_{u,v} \quad (18)$$

Iterating will stop when all parameter changes are less than ϵ .

3) The parameter matrix is output.

(3) Airspace model

According to the spatial characteristics of the skeleton sequence, this chapter used the convolutional neural network to build the network model. The convolutional neural network model can effectively extract the spatial features of human skeleton sequences. The original skeleton point coordinates can only represent the absolute position of the skeleton points, but cannot represent the relative positional relationship between the skeleton points. The positional relationship between adjacent skeleton points has a strong correlation, so according to the traversal order of human skeleton points [24, 25, 12, 11, 10, 9, 21, 5, 6, 8, 7, 8, 22, 23, 4, 3, 21, 2, 1, 17, 18, 19, 20, 13, 14, 15, 16], and the relative positional relationship between adjacent skeleton points is calculated [18-19].

Assuming that the length of the human skeleton action sequence is T, including N joint points, the formula is as follows. The time length of an action skeleton sequence is $t=\{1,2,3,\dots,T\}$, $i=\{1,2,3,\dots,N\}$, and the i-th skeleton point at time t is expressed as V_i^t .

$$V_i^t = (x, y, z), x, y, z \in R \quad (19)$$

The relative position between the i-th skeleton point and the i+1-th skeleton point at time t is denoted by $V_{i,i+1}^t$.

$$V_{i,i+1}^t = V_{i+1}^t - V_i^t \quad (20)$$

Spatial model framework is shown in Fig. 5:

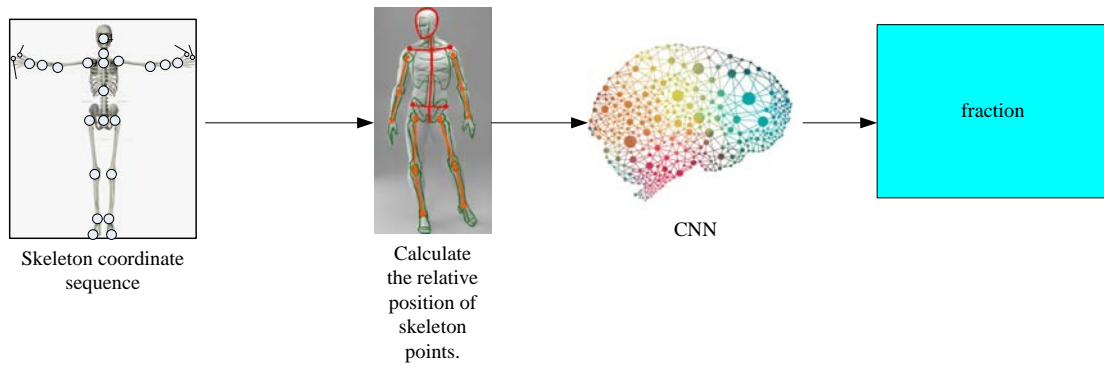


Fig. 5. Framework of the space model

Fig. 5 is composed of a convolutional neural network, which can effectively extract the spatial features of human skeleton sequences. The spatial domain model structure consists of four convolutional layers, three pooling layers and three fully connected layers.

Convolutional layer conv1: The size of the convolution kernel is (3, 3), and the number of channels is 32. Pooling layer pool1: Using pooling for dimensionality reduction, the pooling kernel is (2,2), and the sliding step stride is 2. Specific structural parameters are:

Convolutional layer conv1: The size of the convolution kernel is (3, 3), and the number of channels is 32. Pooling layer pool1: Using pooling for dimensionality reduction, the pooling kernel is (2,2), and the sliding step stride is 2.

Convolutional layer conv2: The size of the convolution kernel is (3, 3), and the number of channels is 64. Pooling layer pool2: The pooling kernel is (2,2), and the sliding step stride is 2. Convolutional layer conv3: The size of the convolution kernel is (2, 2), and the number of channels is 128.

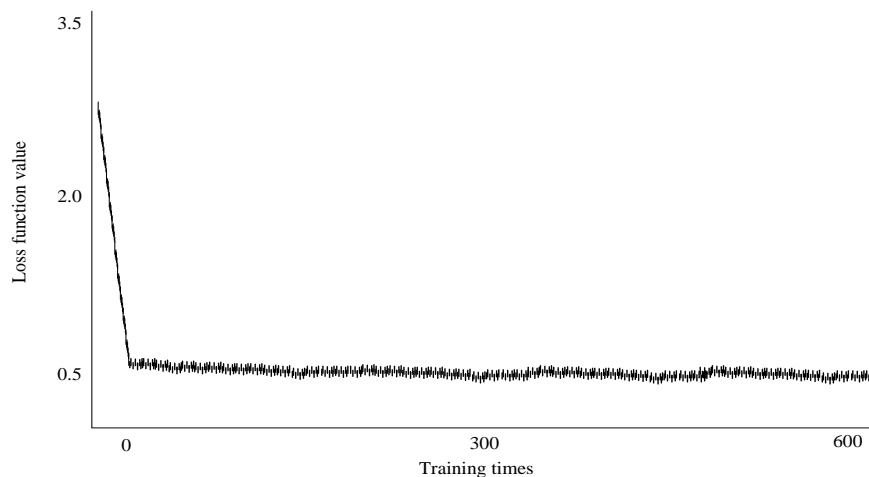
Pooling layer pool3: The pooling kernel is (2,2), and the sliding step stride is set to 2.

Convolutional layer conv4: The size of the convolution kernel is (2, 2), and the number of channels is 128.

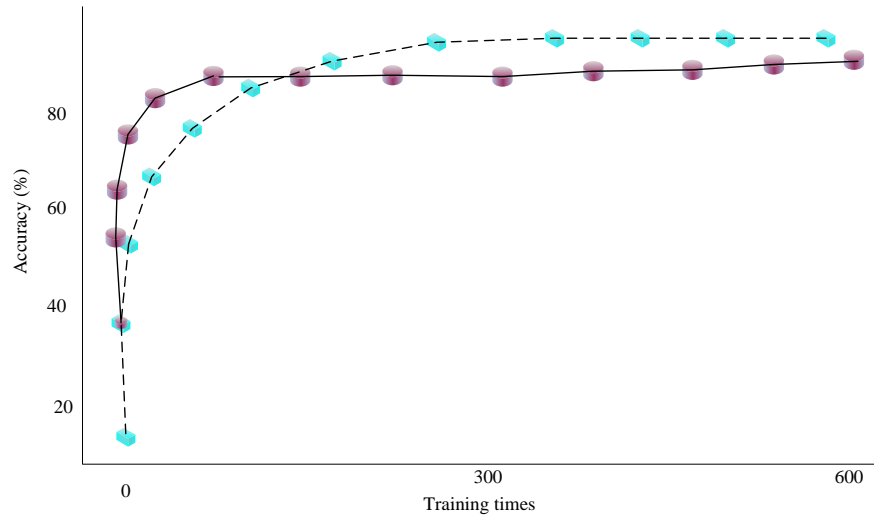
The last three fully connected layers have a loss rate of 0.5 during training.

2.3 Algorithm Identification

The network adopts the skeleton action modeling and the original data, time difference and space difference information Max fusion method, and the accuracy and loss function value change with the number of training times. The dotted line represents the value change of the training loss function and the change of the accuracy rate, and the solid line represents the change of the test loss function value and the change of the accuracy rate. When the number of training times reaches 600, the recognition accuracy and loss function value remain basically unchanged, and the human skeleton behavior recognition reaches convergence to complete the recognition task. As shown in [Fig. 6](#), the LSTM skeleton recognition method in this paper has good performance.



A) Loss function value change curve



B) Accuracy change curve

Fig. 6. Analysis of the task completion of the algorithm

3. Realization of Wrestling Video Platform

3.1 Realization of Wrestling Video Platform

The video analysis system of the wrestling game is composed of two subsystems: the video analysis subsystem of the wrestling game and the online retrieval subsystem of the information of the wrestling game. Among them, the video analysis subsystem of the wrestling game is based on the Microsoft NETFramework technology platform and uses the Winform control technology to realize the development of the local client part. The online retrieval subsystem of wrestling match information is based on the combination of Spring 3.2 and Hibernate 3.6 framework, which are very popular in current WEB development. And based on the design idea of MVC, it completes the development of the online retrieval system by completely separating the view display from the logical business control. Finally, in order to achieve data consistency between the two systems, the HTTP protocol is used to transmit the data collected by the local client system to the network database in the online retrieval subsystem of wrestling match information in real time. This chapter will introduce the key technologies used in system development in detail [20].

NETFramework is a new computing platform architecture that can support the generation and operation of next-generation applications, including XML, WebServices internal Windows components. Many programs and component applications require the support of this architecture to run.

The System.Windows.Forms.Control class directly or indirectly derived from the Winform control is a reusable component. Its base class System.Windows.Forms.Control provides the required methods for the appearance and display in the client application. Control provides window handles for handling message routing and provides keyboard events and mouse and other user-defined interface events. It not only encapsulates the user interface, but also can be easily applied to the client's Windows program development. On

the basis of providing many ready-made controls, the form also provides the infrastructure for users to develop controls by themselves. Developers can extend existing controls, combine existing controls, or customize controls as needed. The operating environment of the system is shown in [Table 1](#).

Table 1. System minimum configuration operating environment

CPU	INTER P4 1.8
RAM	512MB
HDD	40GB
Resolution	1024X 768
Operating System	WINDOWS

3.2 System Requirements

By watching the wrestling competition training activities on-site at the Olympic Wrestling Competitive Training Center, and deeply discussing the technical and tactical movements and rules in the competition with the coaches, the competition analysis system should implement the following functions: (1) Account login function; (2) game setting function; (3) video loading function; (4) initialization function; (5) game analysis function; (6) read analysis data function; (7) upload analysis data function; (8)) function of managing and analyzing data; (9) help function; (10) user registration function; (11) administrator review function; (12) member login function; (13) fuzzy query function for a single game; (14) multiple games Interval query function; (15) query function of key actions of the game; (16) analysis of technical and tactical analysis report of a single game; (17) analysis of technical and tactical report of multiple games; (18) expression analysis function.

In order to reduce the load of the local client system, it is more convenient for users to install and use the video analysis subsystem of the wrestling match. The user hopes to realize the data synchronization between the wrestling match video analysis subsystem and the wrestling match information online retrieval subsystem, that is, the game information and key action information after the analysis of the local client system is completed and sent to the wrestling match information online in real time in the form of data, instead of locally saving the data game information completed by the analysis, in order to reduce the load of the local system and avoid affecting the system startup efficiency due to the installation of the local database. In response to this requirement, in the local client system, the HTTP protocol is used to encapsulate the data analyzed by the system into the form of strings, and by sending a request, the data analyzed by the local client system is uploaded in real time to the online wrestling match information retrieval sub-system. In the database of the system, when the transmission is successful, the server will give a response, indicating that the upload is successful.

3.3 Overall System Design

According to the functional requirement analysis, data requirement analysis and technical feasibility analysis of the system, the design of the wrestling match video analysis and retrieval system is divided into two subsystems: the wrestling match video analysis subsystem and the wrestling match information online retrieval subsystem. The frame diagram is shown in [Fig. 7](#). The wrestling match video analysis subsystem includes four

parts: video playback module, analysis list module, operation module, and system interaction module. The video analysis subsystem of the wrestling match is developed as a client application instead of being integrated into the background management website as an operation, mainly considering its portability and operability. As a client application, all its operation processes and operations data are local, and there is no data interaction with the network to avoid the danger of data being intercepted and copied during the transmission process, making the operating environment more secure.

The integrated development environment NetFramework3.5 is used for the development of the video analysis subsystem of the wrestling match, and VS2013 is used for the development of the front-end interface and the development tool of the logic layer in the background. The system is mainly used in game videos, and timely display and feedback management of the data collected from the game videos. It is based on the research on the statistics of the technical and tactical movements of the latest wrestling competitions and new rules, counting the number of technical and tactical appearances, the success rate, the key points and moves, etc., and displaying the strengths and weaknesses of athletes with digital information to provide a reference for the improvement of athletes' movement techniques.

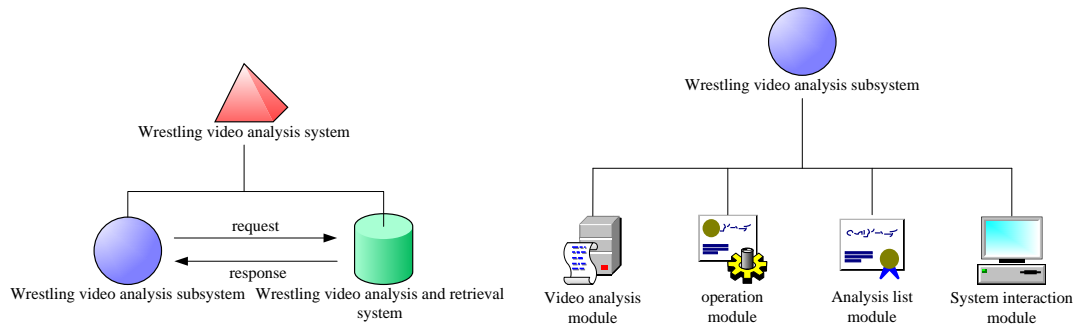


Fig. 7. Analysis of the task completion of the algorithm

Finally, the local client system (wrestling match video analysis subsystem) encapsulates the data analyzed by the local client into a string through the http protocol, and interacts with the online wrestling match information retrieval subsystem in the form of a request request. When the online retrieval system successfully receives the data, it will prompt feedback information to the local client system in the form of a response. The network architecture diagram is shown in **Fig. 8**.

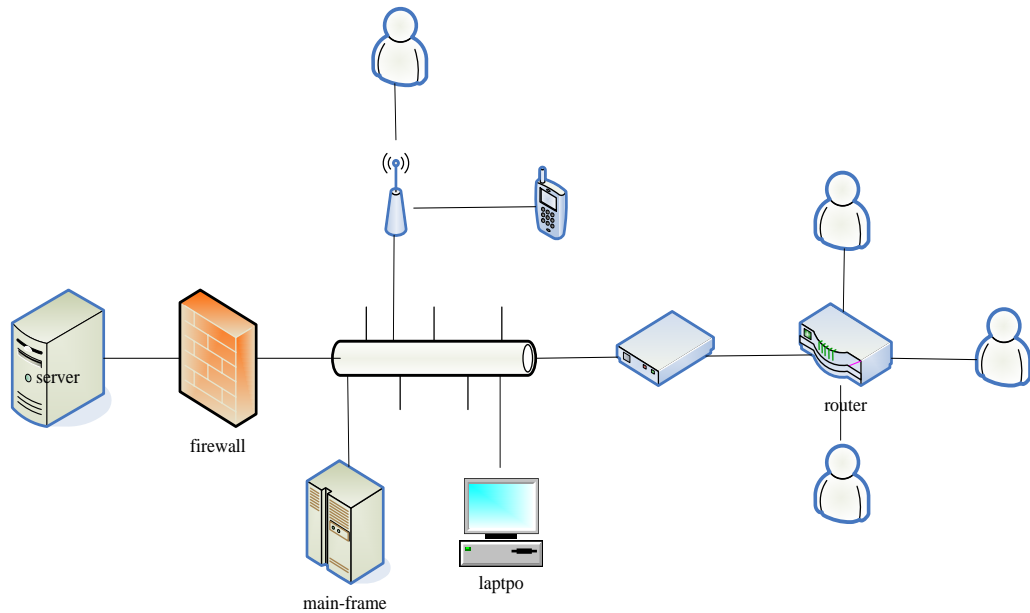


Fig. 8. Network architecture diagram

(1) Realization of video analysis module

In addition to action recognition, video analysis also recognizes the facial expressions of athletes to achieve the effect of psychological feature analysis. The video playback module adopts the combination of C# and EmguCV, uses the panel container in winform, and combines with the external interface libvlc library in VLCmediaplayer. It supports many audio and video decoders and file formats, and supports DVD hard disk, VCD hard disk and various streaming protocols. The video analysis platform also has cross-platform features and can be used under Windows, Linux, MacOS and other operating systems. The system uses VLC to create a player. VLC adopts a multi-threaded parallel decoding architecture. A separate thread is used to control the status of all threads between threads. The decoder adopts filter mode. The organization is a module architecture, mainly composed of libvlc, interface, Playlist, Video_output, Stream_output and other modules. The libvlc module is mainly used in the system, which is a library that provides interfaces, such as providing VLC with access to functional interface shunting, audio and video output, plug-in management, and threading system.

(2) Operation module implementation

The design of this module is based on the previous user demand research, based on WinForm control and layout technology, from the perspective of UI layout, it adopts the combination of Panel family control and Button control. Panel is an abstract class, which is the base class of all Panel panel controls. Panel does not inherit from the Control class, but directly from FrameworkElement. This module is developed using Canvas, which is the specific implementation class of the Panel control. Canvas defines a coordinate system in an area, and Child can determine the absolute position in the layout according to the coordinate system. The reason why Canvas is chosen is because Canvas has the best Arrange performance among all controls and also has good performance in the calculation (Measure) step.

(3) Analysis list module implementation

The design of the analysis list module is mainly based on the binding of the DataGridView control and the list, textbox and button controls in the Winform form. The reason why the DataGridView control is chosen is based on the data operability and flexibility of the powerful display of the control. The DataGridView has similar functions to the DataGrid control in VB and VC, but it is more powerful and more flexible in operation. There are two ways to operate the DataGridView. One is to use the control binding method. During the operation, it is only needed to change the data set of the DataSet control to realize the flexible display of the DataGridView data; the other way is to use the code to manually operate the display of DataGridView control, use as few controls as possible to make the code look more coherent and the operation more flexible.

(4) Realization of system interaction module

The wrestling match video analysis subsystem will continuously generate a large amount of real-time data in the process of analyzing the wrestling match video. In order to enable coaches to view the game analysis data of each game time period, it is necessary to provide the system with the encapsulation function of business data and status data. Since real-time data such as video data and game key actions will be continuously received at a fixed transmission rate, in order to avoid continuous IO operations during the data analysis process to block the functions and user responses of other modules. It is hoped that the analyzed data can be sent in real time. The independent database module is used for data reception, so that this module can complete the game data upload operation in parallel while other components are running. The design of the system interaction module adopts the method of triggering events from the Button button control in the Winform form to realize data interaction with the online retrieval subsystem of wrestling match information. The data collection part is that the coaches trigger the technical and tactical list module by clicking the mouse or space while watching the video, and add the key actions that appear in the current video to the list. In this paper, in order to make the coaches more convenient and quicker in the process of operation, the content of body position, action classification, and technical action is read in the form of xml document, and the cascade between lists is realized through index tags, which is convenient for users to mark the key actions of the current game. When the key actions are marked, the marked key actions need to be scored according to the latest competition rules. When the above work is completed, the coach needs to mark the winning method of the current game, and then calculate the score to complete the game action analysis. When the above work is completed, the coach can click the upload function to realize the local client system (wrestling technical and tactical competition analysis system) to complete the data upload to the online retrieval subsystem of wrestling competition information.

3.4 System Test

Only when the system can run stably can be regarded as the real completion of the design. This paper compared the running time and algorithm efficiency of the system. Under the minimum configuration requirements, it tested the response time of the system, clicked on the four modules of the system respectively, and recorded the response time. The test results are shown in [Fig. 9](#).

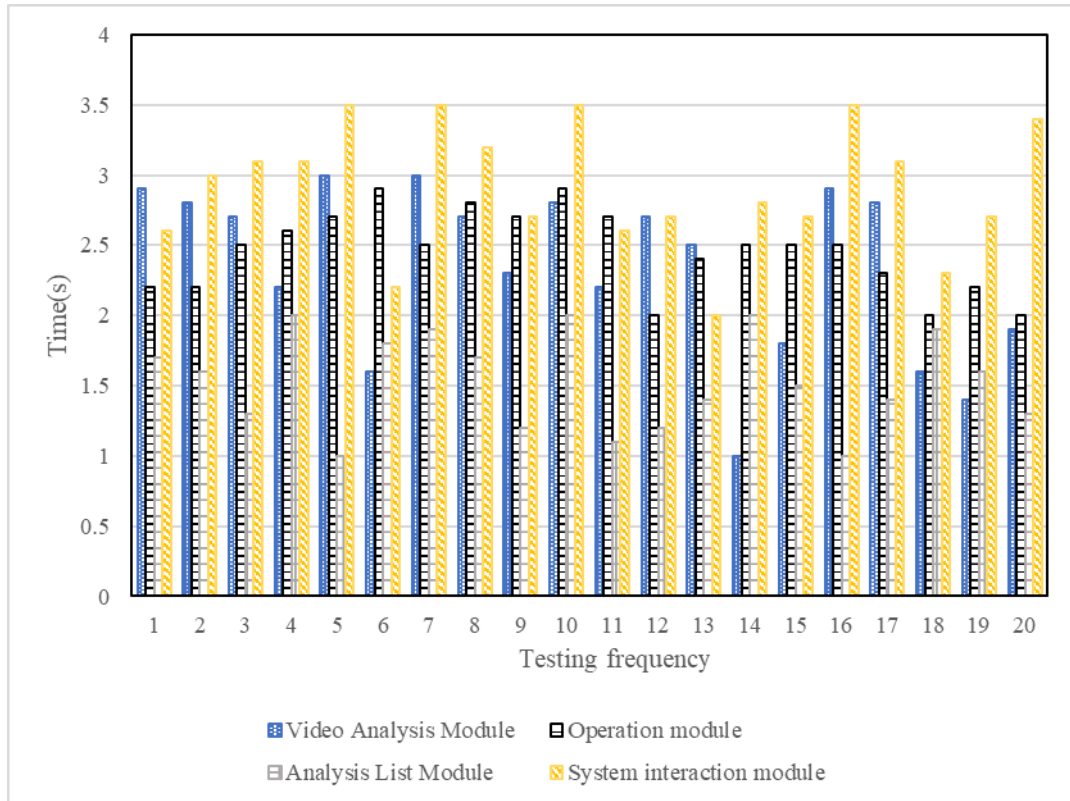


Fig. 9. Response time of different modules of the system

As can be seen from **Fig. 9**, the response time of the four modules of video analysis, operation module, analysis list module, and system interaction module are all below 3.5s, among which the average response time of the system interaction module is the highest, and the response time of the analysis list module is the lowest, which is related to the information processing speed of the operating system. The system interactively designs the response of multiple modules, so the time will be longer. In general, the show time is within 4S, and it can be said that the system can meet the requirements of use without being stuck.

Performance analysis was performed on the system's video facial expression recognition, using the 2008 Olympic wrestling video as data for analysis. And the expressions was divided into happy, surprised, disgusted, sad, scared, angry, natural, and other eight characteristics. After analysis in the wrestling match, the two expressions of happiness and fear did not appear, so the remaining 6 expressions were identified with high accuracy.

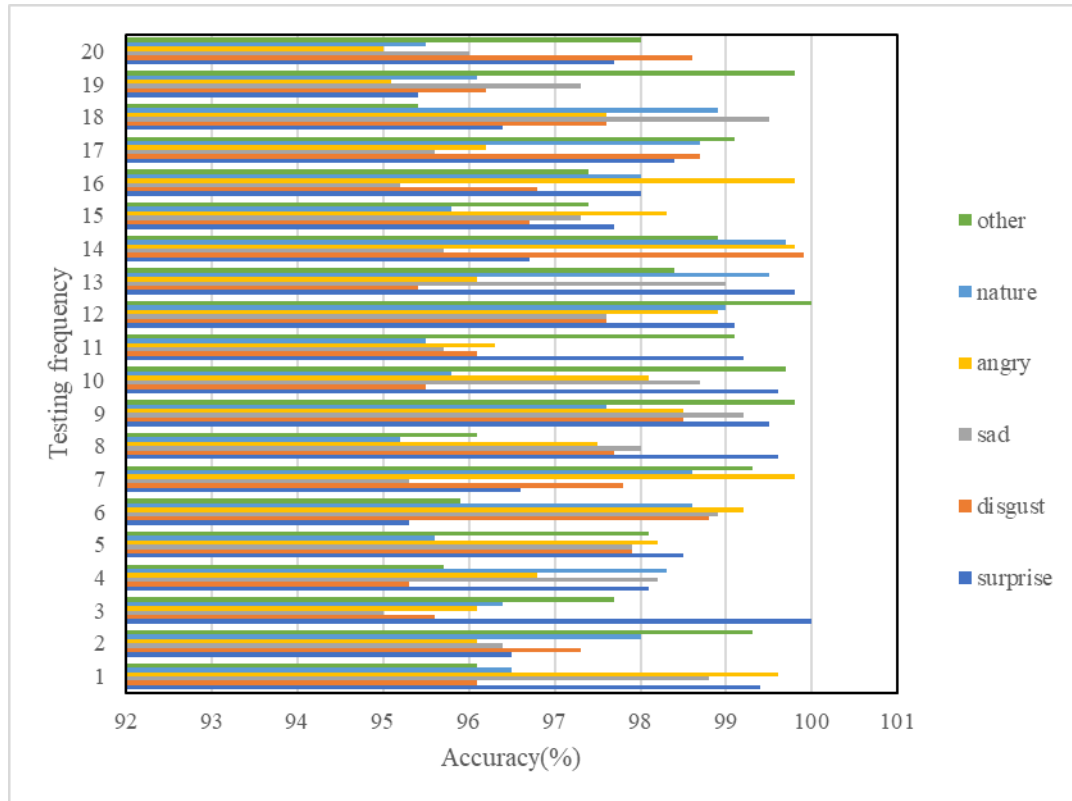


Fig. 10. Accuracy of facial expression recognition

As shown in **Fig. 10**, the recognition accuracy rate for various expressions is above 95%, and the average accuracy rate is 98%. Because wrestling often strikes at a very fast speed, plus boxing, elbow, kick and other actions make facial videos blurry and difficult to distinguish, with less than 100% accuracy.

3.5 Facial Expressions of Wrestling Finalists

A total of 18 gold medals are set for wrestling at the Beijing Olympics, including 7 levels for men's classical wrestling and freestyle wrestling, and 4 levels for women's freestyle wrestling. The facial expressions of 24 athletes in the gold medal finals of 12 levels in 4 levels are used to study the psychological state of high-level wrestlers on the field.

From the comparative analysis of the facial expressions of male and female athletes in the wrestling finals of the Beijing Olympic Games, it can be seen that the facial expressions of high-level wrestlers are not happy or fearful during the match, indicating that the wrestlers take the competition seriously and dare to fight. There is an extremely significant difference in emotion ($P < 0.001$), and a significant difference in anger ($P < 0.05$).

Table 2. Comparison of facial expressions of male and female athletes

facial expression	men's wrestling	women's wrestling	t	df	P
Happy	0	0			
surprise	4.40 ± 6.89	18.40 ± 9.59	-4.115	twenty two	0.000
disgust	4.19 ± 7.07	3.43 ± 6.44	0.254	twenty two	0.802
sad	21.45 ± 10.66	23.21 ± 10.94	-0.377	twenty two	0.710
Fear	0	0			
angry	13.16 ± 10.24	2.58 ± 4.80	2.749	twenty two	0.012
nature	52.71 ± 11.06	49.72 ± 3.94	0.735	twenty two	0.470
other	4.06 ± 2.22	2.63 ± 0.82	1.736	twenty two	0.097
N	16	8			

The facial expressions of male and female wrestlers have disgust, indicating that both male and female wrestlers have some sense of superiority in the gold medal final, and the side also reflects the wrestlers' confidence and perseverance before the final. The proportion of natural emotions is around 50%, indicating that the wrestler's mentality is relatively stable before the intense physical confrontation; the wrestler's daringness requires the athlete to maintain a certain level of excitement and arousal before the competition, so as to further beneficial for athletes to enter the state of fighting immediately. The proportion of sadness of male and female wrestlers is above 20%. Sadness is highly correlated with anxiety, and an appropriate proportion of anxiety can keep athletes at a certain level of arousal.

Table 3. Comparison of facial expressions of men's wrestling champions and runners-up

facial expression	champion	runner up	t	df	P
surprise	4.16 ± 6.40	4.20 ± 7.79	0.116	14	0.910
disgust	5.86 ± 8.88	2.52 ± 4.69	0.940	14	0.363
sad	20.66 ± 11.73	22.25 ± 10.23	-0.288	14	0.777
angry	15.68 ± 10.56	10.63 ± 9.94	0.985	14	0.341
nature	48.75 ± 10.93	56.68 ± 10.32	-1.493	14	0.158
other	4.42 ± 2.03	3.70 ± 2.48	0.639	14	0.553
N	8	8			

Table 4. Comparison of facial expressions of women's wrestling champions and runners-up

facial expression	champion	runner up	t	df	P
surprise	22.55 ± 11.03	14.25 ± 6.85	1.278	6	0.248
disgust	3.92 ± 7.85	2.95 ± 5.90	0.199	6	0.849
sad	20.37 ± 9.02	26.05 ± 13.29	-0.706	6	0.506
angry	0	5.17 ± 5.99			
nature	49.97 ± 3.36	49.47 ± 4.98	0.166	6	0.873
other	3.17 ± 0.38	2.10 ± 0.81	2.380	6	0.055
N	4	4			

From the comparative analysis of the facial expressions of the men's and women's wrestling champions and runners-up at the Beijing Olympic Games in **Table 3 and Table 4**, it can be concluded that there is no statistical difference in the facial expressions of the men's and women's wrestling champions and runners-up. The competitive mental state of the runner-up athletes is very similar when they are on the spot, and the side also reflects that the gap between the champion and the runner-up is small. The rules of the wrestling competition stipulate that in the classical and freestyle wrestling competitions, if the score is evenly divided within the specified time, the score will not be divided. In the event of a victory or defeat, the referee will draw lots to decide which side will attack first, while the other side can only choose to defend, so there is a lot of random luck in the result of the game, because the side that takes the initiative to attack has a lot of dominance and initiative power. But at the same time, it can be seen that sometimes the two sides of the wrestling match are indeed equal in strength, especially the athletes from both sides of the Olympic gold medal final.

4. Conclusion

This paper put forward the architecture and design ideas of the wrestling competition analysis system, and elaborated the system analysis, design and implementation methods from the aspects of system requirements, system framework construction, specific functional module design, database construction, etc., and developed a wrestling match video analysis and retrieval system for the latest wrestling competition rules. According to the research content of this paper and the implementation plan of the system, the main work of this paper is as follows: 1. It researched and proposed a deep learning-based video key action classification algorithm for wrestling competitions, using the fine-tuning method to achieve three wrestling actions (standing, lifting and hugging, kneeling support) classification, the experimental results showed the effectiveness of the method. 2. It completed the system requirement analysis and overall architecture design, divided the system development into two subsystems, the wrestling match video analysis subsystem and the wrestling match information online retrieval subsystem. 3. It designed and implemented the video analysis subsystem of wrestling match. And based on .NetFramework architecture it achieved the entire system. The real-time data transmission between the two systems was accomplished through the http protocol. 4. It designed and implemented the online retrieval subsystem of wrestling match information based on B/S architecture. Based on the Web application

frameworks—Spring and Hibernate, and in accordance with the standard process of software information system development, the architecture design of the server system and the design of specific functional modules were completed. JAVA was used to develop the system, and showed the main interface completed by the system development.

Reference

- [1] Wang W, Hou ZG and Cheng L, “Toward Patients' Motion Intention Recognition: Dynamics Modeling and Identification of iLeg—An LLRR Under Motion Constraints,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, pp. 980-992, 2016. [Article\(CrossRef Link\)](#)
- [2] Laraba S, Brahimi M and Tilmanne J, “3D skeleton-based action recognition by representing motion capture sequences as 2D-RGB images,” *Computer Animation and Virtual Worlds*, vol. 28, pp. e1782.1-e1782.11, 2017. [Article\(CrossRef Link\)](#)
- [3] Kim YK, Yoon YS and Oh TG, “Real-time VR Strategy Chess Game using Motion Recognition,” *Journal of Digital Contents Society*, vol. 18, pp.1-7, 2017. [Article\(CrossRef Link\)](#)
- [4] Wu Q, Shao J and Wu X, “Upper Limb Motion Recognition Based on LLE-ELM Method of sEMG,” *International Journal of Pattern Recognition & Artificial Intelligence*, vol. 31, pp. 1750018-1750024, 2017. [Article\(CrossRef Link\)](#)
- [5] He J, Chen S and Guo Z, “A comparative study of motion recognition methods for efficacy assessment of upper limb function,” *International Journal of Adaptive Control and Signal Processing*, vol. 33, pp.1248-1256, 2019. [Article\(CrossRef Link\)](#)
- [6] H Kristjánisdóttir, KR Jóhannsdóttir and Pic M, “Psychological characteristics in women football players: Skills, mental toughness, and anxiety,” *Scandinavian Journal of Psychology*, vol. 60, pp.609-615, 2019. [Article\(CrossRef Link\)](#)
- [7] X Wu, D Wu and X Pan, “Analysis of Psychological Characteristics and Influencing Factors of COVID-19 Patients,” *Advanced Journal of Nursing*, vol. 1, pp.12-23, 2020.
- [8] Ding H, He Q and Lei Z, “Motion intent recognition of individual fingers based on mechanomyogram,” *Pattern Recognition Letters*, vol. 88, pp.41-48, 2017. [Article\(CrossRef Link\)](#)
- [9] Anam K, Jumaily AA and Maali Y, “Index finger motion recognition using self-advise support vector machine,” *International Journal on Smart Sensing & Intelligent Systems*, vol. 7, pp.644-657, 2014. [Article\(CrossRef Link\)](#)
- [10] Sakalys P, Savulioniene L and Savulionis D, “Research of robotic systems control methods using motion recognition tool, machine learning and skeletalization algorithm,” in *Proc. of the International Scientific Conference*, vol. 5, pp.448 -458, 2021. [Article\(CrossRef Link\)](#)
- [11] Li L, “Mirror motion recognition method about upper limb rehabilitation robot based on sEMG,” *Journal of Computational Methods in Sciences and Engineering*, vol. 21, pp. 1021-1029, 2021. [Article\(CrossRef Link\)](#)
- [12] Luo W and Ning B, “High-Dynamic Dance Motion Recognition Method Based on Video Visual Analysis,” *Scientific Programming*, vol. 2022, pp.1-9, 2022. [Article\(CrossRef Link\)](#)
- [13] Kim T, J Kim and Koo B, “Effects of Sampling Rate and Window Length on Motion Recognition Using sEMG Armband Module,” *International Journal of Precision Engineering and Manufacturing*, vol. 22, pp.1401-1411, 2021. [Article\(CrossRef Link\)](#)
- [14] Wu H, “Design of Embedded Dance Teaching Control System Based on FPGA and Motion Recognition Processing,” *Microprocessors and Microsystems*, vol. 83, pp.103990 - 103997, 2021. [Article\(CrossRef Link\)](#)
- [15] Zhou B, Peng S and Liu X, “Two-Layer Motion Semantic Recognition by Fusing the Restricted Boltzmann Machine Based Generative Model and Discriminative Model,” *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/Journal of Computer-Aided Design and Computer Graphics*, vol. 29, pp. 689-698, 2017.

- [16] Shin D, “Changes of performance of Motion-recognition game based DTT on the Cognitive and Motor Skills of Student with Disabilities in Special Class,” *Journal of special education & rehabilitation science*, vol. 59, pp. 115- 135, 2020. [Article\(CrossRef Link\)](#)
- [17] Tateno S, Liu H and Ou J, “Development of Sign Language Motion Recognition System for Hearing-Impaired People Using Electromyography Signal,” *Sensors*, vol. 20, pp.5807-5832, 2020. [Article\(CrossRef Link\)](#)
- [18] Dabwan BA and Jadhav ME, “A review of sign language and hand motion recognition techniques,” *International Journal of Advanced Science and Technology*, vol. 29, pp.4621-4635, 2020. [Article\(CrossRef Link\)](#)
- [19] Lee JY and Kwon JS, “Application of motion recognition technology for interactive implementation in space,” *Journal of Digital Contents Society*, vol. 21, pp.1171-1179, 2020. [Article\(CrossRef Link\)](#)
- [20] Han K, “Motion Recognition Algorithm in VR Video Based on Dual Feature Fusion and Adaptive Promotion,” *IEEE Access*, vol. 8, pp.201134-201146, 2020. [Article\(CrossRef Link\)](#)



Lianhong Yuan was born in Hangzhou, Zhejiang Province, China in 1965. He received his M.S. degree in Software engineering from School of Software Engineering, Dalian University of Technology in 2015. Since 1985, he was a Research Assistant with the Institute of Internet of Things Technology, Hangzhou Polytechnic. His research interests include wireless sensor network and computer technology.



Yuan Yuan received her B.S. degree from Zhijiang college of Zhejiang university of technology in June 2020. Her major research are in the field of electrical automation and embedded development.



Jun Liu is a researcher and vice-president of Hangzhou Automation Research Institute. Since 1989, he has been engaged in automated software and basics resources of artificial intelligence and technology platform research.