

## *Retraction*

# **Retracted: Study about Football Action Recognition Method Based on Deep Learning and Improved Dynamic Time Warping Algorithm**

### **Mobile Information Systems**

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

Copyright © 2023 Mobile Information Systems. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

### **References**

- [1] J. Xing and H. Li, "Study about Football Action Recognition Method Based on Deep Learning and Improved Dynamic Time Warping Algorithm," *Mobile Information Systems*, vol. 2022, Article ID 3861620, 10 pages, 2022.

## Research Article

# Study about Football Action Recognition Method Based on Deep Learning and Improved Dynamic Time Warping Algorithm

Jiarui Xing<sup>1</sup> and Huilin Li <sup>2</sup>

<sup>1</sup>Graduate School, Jose Rizal University, Mandaluyong, 10048 Manila, Philippines

<sup>2</sup>College of Physical Education, Xuchang University, Xuchang, 461000 Henan, China

Correspondence should be addressed to Huilin Li; [xcxingahongyou@xcu.edu.cn](mailto:xcxingahongyou@xcu.edu.cn)

Received 7 January 2022; Revised 19 February 2022; Accepted 4 March 2022; Published 4 April 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Jiarui Xing and Huilin Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artificial intelligence recognition of human actions has been used in various fields. This article is based on deep learning and improved dynamic time regularization algorithms to study football action postures. This paper proposes a hierarchical recurrent network for understanding team sports activities in image and position sequences. In the hierarchical model, this article integrates the proposed multiple human-centered features on the time series based on LSTM output. In order to realize this scheme, the holding state is introduced as one of the external controllable states in LSTM, and the hierarchical LSTM is extended to include the integration mechanism. Test outcomes demonstrate those adequacies of the recommended framework, which includes progressive LSTM human-centred benefits. In this study, the improvement of the reference model in the two-stream LSTM-based method is shown. Specifically, by combining human-centered features and meta-information (e.g., location data) into the postfusion framework proposed in the article, the article also proves that the action categories have increased, and the observations enhanced the robustness of fluctuations in the number of football players. The experimental data shows that 67.89% of the postures of football players through this algorithm can be recognized by the improved dynamic time warping algorithm.

## 1. Introduction

At the beginning of the 21st century, artificial intelligence is constantly being used in various fields, and computer vision is also an important part of it. It can assist and capture related video sequences. Human action recognition is the current research boom in algorithms, and it can be widely used in many areas of life such as video surveillance, entertainment, human-computer interaction, and video portrait search. Combining deep learning with recognition of human gestures is now an important breakthrough point.

Aiming at the current video surveillance and judgment in football matches, it only stays at manual judgment. This article judges and analyzes football actions based on deep learning and improved dynamic time warping algorithm. Although the artificial accuracy rate is relatively high at present, the artificial judgment and energy are limited after all. It is impossible to observe every action of the entire football

field in detail, not to mention that people will experience mental and visual fatigue. Compared with the continuous operation of computers and monitoring equipment, algorithms can be used to make reasonable analysis and judgments on the actions of football in the video, which greatly reduces manual consumption.

In recent years, deep learning has been rapidly developed in major technology companies and has become a hot topic of research. Trabelsi et al. suggested a novel sort of hand vein pattern identification method for human body recognition in the article. Hand vein characteristics can be regarded more dependable in the field of biometrics than other biometrics (such as palm prints and fingerprints) since the veins are situated in the volume, making the feature more durable to test circumstances. To extract hand vein patterns, this work proposes a rotation-invariant texture descriptor dubbed Circle Difference and Statistical Direction Mode (CDS DP). Its histogram may be thought of as an attribute

vector. The statistical direction information of the ship is contained in the CSDSP, which is a weighted circular discrepancy [1]. Lu et al. spoke on the effectiveness of the presently prevalent Bunch transformation mobile CNNs and provided some principles for building mobile CNNs. This research provides completely linked convolutional neural networks (CNN based on these rules (FSC-Nets)) [2]. J.D. Freitas obtained information and detailed insights from all of these information sources which is a big issue for customized medicine and then the next healthcare deployment. Deep learning refers to a type of neural network-based machine learning techniques that allow for successful modeling, representation, and learning from highly complex and diverse data sets [3].

The innovation of this paper is that (1) through the research on the deep basic neural network in human behavior recognition, and further specific to the judgment of football movements and the collection of posture templates in videos, on the basis of deep learning, relevant research and recognition of actions are carried out. (2) In the theory of thinking, the above simulation affirms the analysis and judgment of behavior and action recognition through deep learning. (3) The time series template of the video is combined with the convolutional neural network to realize the analysis and judgment of different angles of football actions. And through the recognition method of human skeleton dynamic time regularization, by comparing the similarity between a single action sequence and a template group, it can identify the complex actions of people playing football.

## 2. Analysis of Football Action Recognition Method

**2.1. Extreme Learning Machine Model.** Assume  $\{x_i, t_i\}, i = 1, \dots, n$  is the knowledge set gathered during training,  $n$  is the total number of generated samples,  $x_i \in R_d$  is indeed the recovered concentration and attributes, which is the dense trajectory feature, while refers to the number of replicated samples.  $t_i \in R_d$  is the actual action collection class, based on deep learning characteristics of motion information, where  $q$  is the total amount of class computations for all training set. When a hidden layer's activation function is defined at  $g(x)$ , then hidden layer usually has 1 neuron.

$$\min \|M\alpha - T\|^2 + \|\alpha\|. \quad (1)$$

$i = 1, \dots, n, j = 1, \dots, L, H$  is the output vector associated to the center of the hermetic barrier, when  $H_{ij}$  is really the output of the  $j$  constrained layer component, and  $X_i$  indicates a constricted layer node in the equation  $N = \{NI\} = \{G(a_j, b_j, x_j)\}, i = 1, \dots, n, j = 1, \dots, L, H. \alpha = [\alpha_1, \alpha_2, \dots, \alpha_L]^T, M(x_i) = [m_1(x_i), m_2(x_i), \dots, m_L(x_i)], T = [t_1, t_2, \dots, t_n]^T$ .

The following formula can be used to solve the formula (1), as per the library [4].

$$\alpha = N + T. \quad (2)$$

$M +$  seems to be the generalised invert of state matrix in the expression. The original learning machine idea was to cope with the flaws of a simple impermeable element's feedback neural network, but later on, a substantial number of work-related individuals expanded the idea of excessive learning to issues that were not network neural, confirming the limit. The learning machine's relevant requirements are less stringent than those of the vector and least squares mechanisms [5], as is the case with the unsupervised algorithm in this paper. Overfitting means that the prediction of the model is very accurate on the training data, and the prediction is very poor on the unknown data. Overfitting is mainly due to abnormal points in the training data, which deviate significantly from the normal position. Multiple small convolution kernels are used to replace large convolution kernels, and the  $n \times n$  convolution kernel is split into  $1 \times n$  and  $n \times 1$  convolution kernels. This method can reduce parameters, speed up the operation, and reduce the phenomenon of overfitting.

The extreme learning machine's basic developed for solving issue is denoted by the following formula.

$$\min \text{LPELM} = \frac{1}{2} \|\alpha\|^2 + C \frac{1}{2} \sum_{i=1}^n \|\gamma_i\|^2. \quad (3)$$

The constraints are

$$h(x_i)\alpha = t_i^T - \gamma_i^T, i = 1, \dots, n. \quad (4)$$

In the equation  $i = [\gamma_{i,1}, \gamma_{i,2}, \dots, \gamma_{i,q}], i = [i, 1, 2, \dots, q]^T$  is the regularization variable, and  $c$  is the deviation of something like the action gathering matrix of both the instance  $x_i$  of the  $q$  output layers. The optimization issue encountered may be translated into the following equation using Kadoyili's conditions [6].

$$\alpha = H^T \varphi, \varphi_i = C\gamma_i, h(x_i)\alpha - t_i^T + \gamma_i^T = 0, i = 1, \dots, n, \quad (5)$$

where the likelihood ratio matrix is denoted by  $T$ . The output data mass is computed using the formula below.

$$\alpha = H^T \left( \frac{I}{C} + HH^T \right)^{-1} T \quad (6)$$

As a result, the extreme learning machine's output function may be characterized as the following formula.

$$f(x_j) = h(x_j)H^T \left( \frac{I}{C} + HH^T \right)^{-1} T, j = 1, \dots, n. \quad (7)$$

Because the learning algorithm technique and indeed the support-vector principle are already so similar, you may transform shown above knn into an unsupervised algorithm and

then confine the circumstances under which it can work. The condition of limitation output may be transformed into the continuity formula using Merece's theorem.

$$f(x_j) = h(x_j)\alpha = [K(x_j, x_j) \cdots K(x_j, x_n)]^T \left( \frac{T}{C} + K \right)^{-1} T. \quad (8)$$

$j = 1, \dots, n$  in the formula is the application which may categorize and quantify the video card just after forwarded streamlined processing.

Deep learning is supported by a large volume of data and a substantial majority of difficult calculations. To optimize the training impact of neural networks, huge number of iterations and millions of parameter adjustments are required for each training process. The Amax XG-48201G network engine is chosen as the machine learning software platform for this subject since it must analyze a significant quantities of information and build and test the multilayer perceptron. Amax XG-48201G network engine can support up to 8 GPU cards, providing powerful performance comparable to 100 CPU server nodes, paving the way for the most complex application research in cancer research, climate model building, energy, artificial intelligence, and so on. Table 1 depicts the platform's exact configuration:

**2.2. Dynamic Time Warping.** As an important part of operation research, dynamic programming (DP) is a mathematical calculation method to find the optimal solution and achieve the goal of optimization. The dynamic programming algorithm is similar to the divide and conquer method. Its basic idea is to decompose the problem to be solved into several subproblems. First solve the subproblems and then get the solution of the original problem from the solutions of these subproblems. In the 1950s, the famous American physicist discovered the world-famous optimization method (principle of optimal PQ) when analyzing the optimization problem of the optimal solution (multistep decision process, MDP). A computationally intensive and complex problem is divided into many small branches. These branches are necessarily much simpler than the original problem, thus solving the problem and the optimal solution—dynamic programming. Therefore, dynamic time warping is a way of thinking about optimization problems, which corresponds to the output of standard templates and related functions. Analyze the planning function obtained when the two variables are compared. The original founder of the dynamic time warping algorithm was AKOE. He proposed it in the 1980s to solve the problem of people's classification of time in a large number of sequences in the process of language recognition. Algorithm concept: The basis of measuring similarity is to use a dynamic program to calculate the distance between time series of different lengths and to solve the problem that the Euclidean distance cannot calculate due to different lengths. Similar with traditional time series, Sakoe provides various route optimization schemes and distance calculation strategies.

TABLE 1: Documentation about the deep learning computer platform's configuration.

Part name	Selection type
CPU	E5-2650 V4
GPU	NVIDIA Tesla V100×8
RAM	DDR4 16G×2
Hard disk	SSD 480G×2
System	Ubuntu16.04
BIOS version	2.0a
CUDA version	9.0
Deep learning framework	Caffe/Tensorflow/Pytorch

Dynamic time planning is mainly used to match text data. Speech processing and image shape recognition are given two time series  $Q$  and  $C$  of length, and the distance matrix  $D$  between them is

$$Q = q_1, q_2, \dots, q_n, \quad (9)$$

$$C = c_1, c_2, \dots, c_n, \quad (10)$$

$$D = \begin{matrix} d(q_1, c_1) & d(q_1, c_2) & \dots & d(q_1, c_n) \\ d(q_2, c_1) & d(q_2, c_2) & \dots & d(q_2, c_n) \\ \dots & \dots & \dots & \dots \\ d(q_n, c_1) & d(q_n, c_2) & \dots & d(q_n, c_n) \end{matrix} \quad (11)$$

The curve's route is specified as the route of curve and  $W$  is described as a collection of uninterrupted matrix elements with distinct connections between data series at various lengths of time series analysis  $D$ .

$$W = \{w_1, w_2, \dots, w_n\}. \quad (12)$$

The following requirements are satisfied by the curved path:

(1) Bounded

$$\text{Max}(m, n) \leq k \leq m + n - 1. \quad (13)$$

(2) Defined boundaries

$$w_1 = D(1, 1), w_k = D(n, m) \quad (14)$$

(3) Continuity

$$w_1 = D(i, j), w_{k-1} = D(O, P), I - O \leq 1, J - P \leq 1. \quad (15)$$

## (4) Monotonicity

It is forbidden to occur at the very same period.

$$I - O \leq 1, J - P \leq 1. \quad (16)$$

Determine the incremental calculation of the best path using the following specification.

**2.3. Optical Flow Method.** When the human eye observes the movement of the target in the three-dimensional world. The continuous trending stream of the shape of the moving target creates a set of constantly changing images on the retina. These constant changes pass through the human retina. Similar to the type of optical flow, it is the same, so it is called optical flow in terms of target detection. Optical flow refers to the movement of the grayscale pattern surface in the picture. On the object plane, it is really the double translation of the four-vector field of the target vehicle. The motion detection does not only provide its mobility details of the small object to convey the change and placement trend in the video. It also shows the movement trend so that the goal movement may be determined. The luminous flux may be thought of as the velocity flow field created by the mobility of gray pixels on the picture plane, given that now the visual environment is uninterrupted and navigable in spacetime. Set the exposures of a mainly intended  $(x, y)$  in the picture at period  $tI(x, y, t)$  in the coordinate dispersion in the image because then  $u(x, y)$  and  $v(x, y)$  may be treated as sites  $(x, y)$  once the motion deconstruction parameters inside the  $x$  and  $y$  axes are decomposed. The location  $(x, y)$  is shifted here to photo location  $(x + dx, y + dy)$  in a very brief time differential  $dt$ , and  $dx = udt$ ,  $dy = vdt$ , predicated on the notion of equal luminosity, that seems to be, the adjacent dots within every frame and along specifically requires curved has the same gray value. This graph depicts the link between the optically fluid velocity and the spatial gray of the monochrome in the image. Thus,  $u$  and  $v$  are unknown in the image feature constraint equation. So there is only one restraint equation, the emission intensity cannot be computed in a unique way. To establish  $u$  and  $v$ , further restrictions must be applied. New techniques of determining light output will emerge when different limitations are introduced. The differential image feature technique, often known as the divergence approach, is the most well known. To obtain the path to the destination from every pixel in the image, the majority of them rely on the gradient value of the grayscale scale. The Horn-Schunck code [7] and Lucas-Kanada algorithm [8] are the two most prevalent gradient-based optical flow algorithms. A limitation in the Horn-Schunck algorithm is predicated on the premise that perhaps versus the overall flow, normalizing restriction is implemented under constant intensity circumstances. It is considered that the change in luminous flux throughout the whole picture is regular, i.e., that the moving target's motion vector is regular. Because neighboring pixels change at the same pace, this approach determines the field's dense luminous flux. Therefore, the spatial velocity change rate of the local area is zero. The basic idea of the stress condition proposed

by Horn is that the luminous flux should be as smooth as possible to minimize the smoothing limitation constraint. The Lucas-Kanada code was researched in the early 1980s. The algorithm has a very important and fixed condition that is to keep the collected motion vectors unchanged in a confined space [9, 10], and then, all the resulting vector is weighted by the least square method. The L-K algorithm assumes that the uniformity of pixel motion in a certain local area replaces the smoothing term of the speed. The sparse streamer field needs to be limited to three assumptions:

- (1) The appearance of moving target pixels with constant brightness between images remains unchanged. For grayscale images, it is assumed that the pixel brightness remains constant throughout the tracking period
- (2) Time continues the movement of the image is quite slow compared with time. In actual use, this means that the proportion of time relative to the motion in the image is small enough. And the target movement between adjacent frames is relatively small
- (3) Regional coherence motion of adjacent points on the same surface in the same scene is similar, and the projection of these points on the image is also in the adjacent area. LK's optical flow approach, on the other hand, necessitates more use of than edges. Confirm that the vector is reversed in the  $n \times n$  window above to assure also that image sequence limiting formula has a solution. As a result, the LK luminous flux computation technique must incorporate both angle and border data in real time in order to produce a sparsely light transmitted vector [11, 12].

### 3. Experimental Data Collection

**3.1. Two-Layer Core Extreme Learning Machine.** This article divides the extreme radioactive teaching engine project into two main categories. The first layer, as depicted in Figure 1, combines the deep learning function's core with the manual function's core. The estimated value of characteristic core 3 is the outcome of the first layer. The classifier is trained in the second layer. All predicted scores are mapped towards the ultimate running category in this way.

- (1) Fusion of characteristic cores

Different combinations of main attributes can combine attributes of different video dimensions. Therefore, this article combines the core of the manual function and the core of the deep learning function and uses L2 general to calculate the linear kernel. L2 regularization increases the sum of squares of the ownership weight  $W$  parameters, forcing all  $w$  to approach zero but not zero as much as possible. Each fitting point of the regularization function needs to be considered, because there is a large fluctuation in the final fitting point of the regularization function. The linear kernel is the desired linear model. The polynomial kernel is similar, but

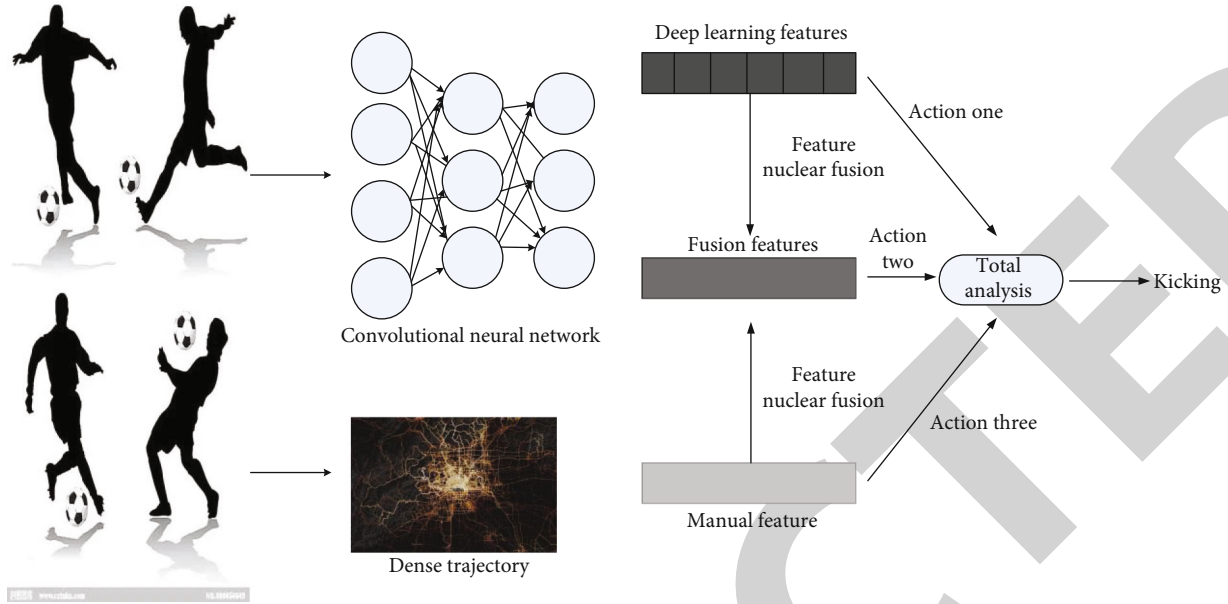


FIGURE 1: The framework of this article nuclear extreme learning machine.

the boundary is some defined but arbitrary order. By determining the characteristic provider core matrix's overall average. Calculate the core of the image pairs. After the functional core is merged, three axes are given: deep learning functional core. The manual function axis then uses the kernel limit machine to calculate the expected number of executions of the kernel function. [13, 14]

- (2) Deep learning characteristics that are dependent on accelerometers

To recognize human actions, this research provides a visual description of motion data. It highlights the significance of movement data in various time domains. Improve the video material's low-frequency action discrimination. Figure 2 illustrates the general design of this module.

- (3) Motion representation scheme for motion recognition

The synchronization model can restore the entire motion sequence of the image. As a result, a temporal model is used in this article's changes in physical. The temporal framework is designed by counting the graded frequencies of the film measurements. And use the difference between the video images to calculate the motion data between the images. This can be seen in Figure 3 which calculates visual movement energy (MEI), and  $\mu_2$  calculates visual image history (MHI). MEI assigns identical weight to all accelerometer in the time-frequency domain since  $\mu_1$  is just staying in contact, and  $\mu_2$  is just a nonlinear cumulative function. As a result, MHI inserts the most crucial one towards the end of the video section.  $\mu_3$  seems to be the least significant to the very last short video because this is a straight iterative function.  $\mu_4$  assigns the highest priority to the sequence, placing video in the centre of the temporal domain. Ulti-

mately, procedures  $\mu_2$ ,  $\mu_3$ , and  $\mu_4$  focus on the temporal domain's starting, end, and middle areas, accordingly.

3.2. *Human Motion Information Capture.* Evaluate how to recognize the proposed action. The currently popular action data set MSRA action3D was provided by the Microsoft Corporation in the United States. It was selected to evaluate the performance in the test [15, 16]. The average template representation of the action and the BOW representation are combined by the method of augmented feature multicore learning, and the contribution of the two is adjusted by introducing the learning weight; through the above two improvements, the accuracy of action recognition is improved. The capture of this system is shown in Figure 4:

MSRA action3D is a data set publicly provided by the Microsoft Corporation of the United States. The data set contains bone information and data and consists of 20 action sequences of 10 individuals and 2 or 3 iterations each. The resolution of the depth data is  $320 \times 240$ . The 20 action data and experimental classification are shown in Table 2:

Provide a fair comparison with existing methods. Therefore, the same experimental settings as in the article were applied. In each set of actions, a set of action 1 (A S1), a set of action 2 (A S2), and a set of action 3 (A S3), randomly select half of the data as the action model. The remaining half will be used as the output sample information of the test.

3.3. *Catch the Displacement Optical Flow Network.* As the network's input, the single-mode optical circulation system overlays two video frame pictures. There are nine convolution layers in the overall network. The fully convolutional system automatically selects the image feature information from different pictures using the special implementation

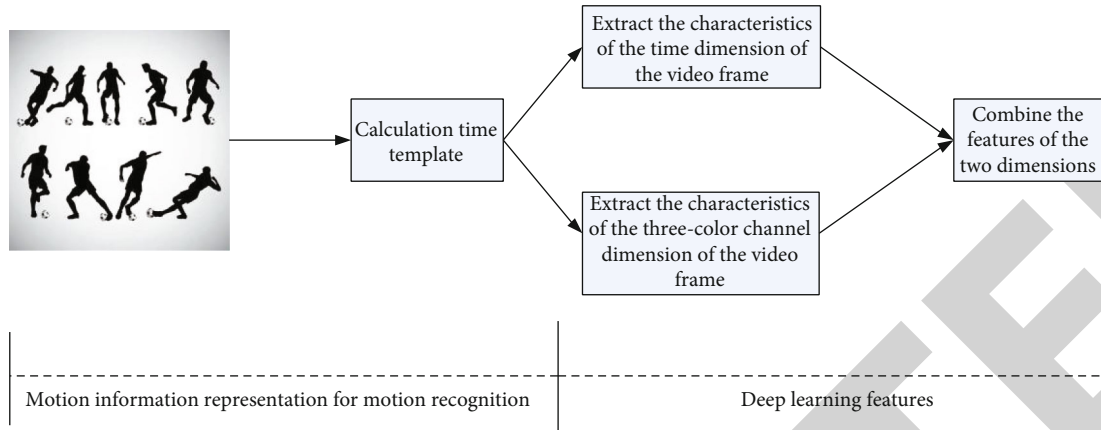


FIGURE 2: Deep learning process based on motion information

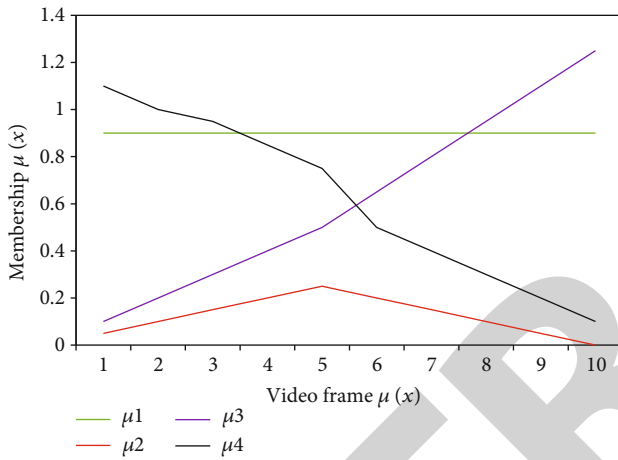


FIGURE 3: Graphs of 4 fuzzy membership functions.

approach of superimposing the two data frames together. The overall network structure diagram is shown in Figure 5: The specific network structure is arranged as follows:

- (1) The input of the network is two superimposed video frame images. The video frame images need to be converted into a database file before being input to the network. At the same time, the image size is reset to a three-channel color image with a uniform size of  $320 \times 256$  pixels
- (2) A convoluted layer “conv1” plus a pooled plane “pool1” makes up the first convolution operation. The convolution operation of “conv1” inside the convolutional is  $7 \times 7$  pixels in size, and each layer contains 6 pixels. The convolution kernel’s step length is one conventional unit, and the boundary compensation is one. As a result, the convolutional feature size that may be estimated is  $320 \times 256$ ; the “pool1” layer utilizes the greatest downsampling layer, while the “pool2” layer uses the smallest down perfect choice, and the window size of 22 is set to the setting style of the pooling layer, the moving displacement distance of the entire window is two, the

overall setting of the pooling layer is the same, and the image size after down sampling is  $160 \times 128$ , which is used as the input of the next layer

- (3) A convnet “conv2” and a pool layer “pool2” make up the second convolution layer. The characteristic map is created by calculating 64 kernels of  $5 \times 5$  length in each layer. The picture following the pooling layer has a dimension of  $140 \times 168$  pixels and a total of  $68 \times 84$  pixels
- (4) The second cnn model is made up of two convolution layers (conv3 and conv3 1) as well as a pooling layer (pool3). The convolution operation in these two convolutions has a size of  $4 \times 4$ . The “conv3” processing element has 64 convolution kernels, whereas the “conv3 1” convolution layer has double conv3 convolution operation, for a total of 128. The default step size for the convolution kernel is 1 image, and the edge correction is also 1 pixel. The feature map’s width may be determined to be  $40 \times 32$  pixels
- (5) The fourth convolutional layer has two convolution layers (conv4 and conv4 1) as well as a pooling layer (pool4). The deep network comprised in the two activation functions has a surface area of  $4 \times 4$ . The “conv3” processing element has 128 convolution kernels, while the “conv3 1” convolution layer has 256 convolution kernels, which is twice as many as the “conv3” convolution layer. The default step size for kernel size is one pixel. The convolutional feature size may be computed as  $20 \times 16$  by accounting for 1 pixel
- (6) The fifth convolution layer contains two convolution layers “conv5” and “conv5\_1” and a pooling layer “pool5”. The convolution kernels of the two convolution layers are both  $3 \times 3$ ; “conv3” volume, both the product layer and the “conv3\_1” convolution layer, contains 512 convolution kernels. By assuming that the step size of the convolution kernel is 1, the size of the feature map can be calculated to be  $8 \times 10$

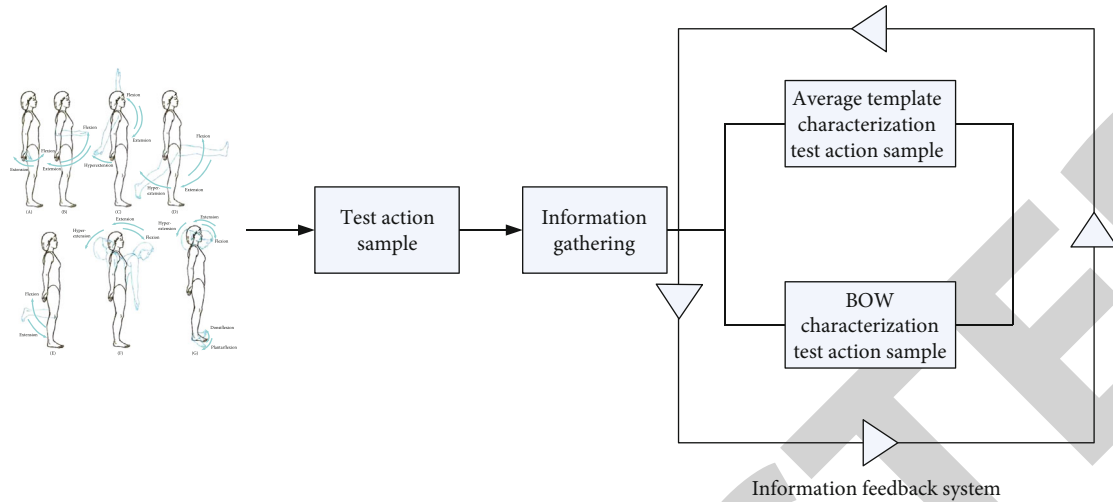


FIGURE 4: Information capture feedback.

TABLE 2: 20 action information and experimental classification.

AS1	AS2	AS3
Horizontal	High arm wave	Arm wave
Arm wave	High throw	Draw tick
Hammer	Side kick	Draw circle
Forward punch	Jogging	Two hand waving
High throw	Tennis swing	Side boxing
Hand clap	High arm wave	Hammer
Bend	Forward punch	Pick up
Tennis serving	Forward punch	Throw

Select the data set intercepted from the animation and compare it with the real optical flow to evaluate the effect of the output optical flow image. As shown in Figure 6, it is a comparison between the real optical flow and the optical flow image calculated by the single-input optical flow network. It can be seen that the optical flow image calculated by the single-input optical flow network can roughly estimate the movement trend of the subject of the moving target, but the outline of the moving target is relatively fuzzy, ignoring many details of the movement, and it is not reflected in the movement of small displacement. Obviously, even the target information with small displacements is lost.

The University of Central Florida created UCF101, an action game dataframe for action categorization tasks [17, 18]. This data collection is made up of 15,412 video clips from YouTube that include a wide range of activity types, including footage of women interacting with items, footage of people interacting with other people, videos of people who play stringed instruments, and different sports videos. During the data set’s creation, the video’s format was standardized. Table 3 shows the UCF101 data set’s individual video characteristics as well as the data set’s scale:

There are less video data featuring deviant behaviors in the UCF101 data set’s action video categories, which include

more frequent scenarios in life including sports scenarios and playing various music instruments. CASIA behavior data set was added to the experiment procedure in order to accomplish the building of the aberrant behavior data set. CASIA’s applied behavior database comprises a total of 1,446 surveillance video from three improperly calibrated fixed cameras taken at the same time. These data are dispersed across the external world in horizontal, tilt, and aerial angles, giving observational evidence for consumer behaviour research. The information is separated into two categories: single-person behavior and multiperson interaction. Walking, sprinting, twisting, climbing, squatting, dying, having trouble, and wrecking a car are all examples of human activity. A total of 16 persons are involved in each act. Heavy balls, passing, pursuing the ball, keeping the ball, confronting, meeting, and passing the ball are all examples of behavioral interventions. Table 4 shows the number of persons who saw the movie either once or twice, the video characteristics, and the size of the CASIA systematic database data set:

The dual-stream system is designed here on UCF101 information source with rich action classifications to evaluate the efficiency of the profound ResNet in the recognition job of football actions. The network is transferred to once the deep double, developed over years, model is built. The work of sports action recognition and categorization has been performed in the CASIA behavior database [19].

#### 4. Data Analysis

4.1. *Experimental Analysis of UCF101 Data Set.* The UCF101 collection of data may be found on the Internet. The training data set has a high level of complexity. All of the films have clear and cluttered backdrops, and UCF101 features a record of 16,420 video clips in 102 action subcategories. For the sample, this article employs three partitions techniques from the default testing phase. For each score pattern, 7 of the 26 video sequences will be selected as the test sequence, and the other 19 videos will be used as the training sequence. Those test will be performed as stated by the default three-segment



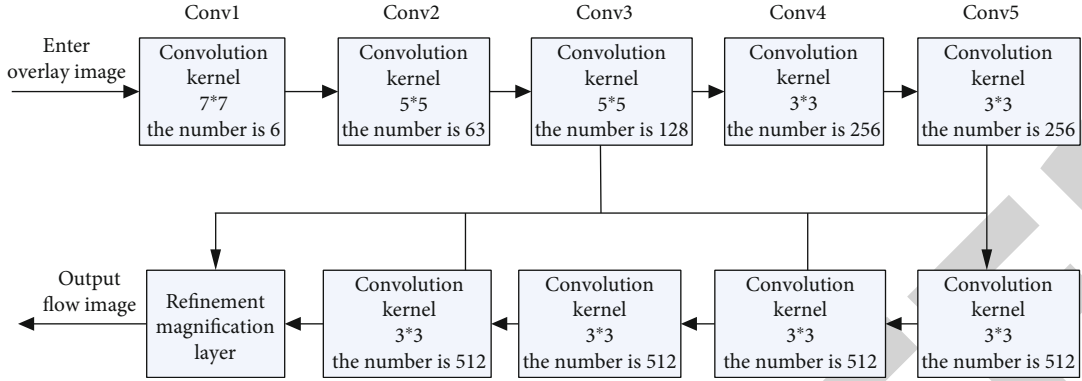


FIGURE 5: The overall network structure diagram of a single-input optical flow network.

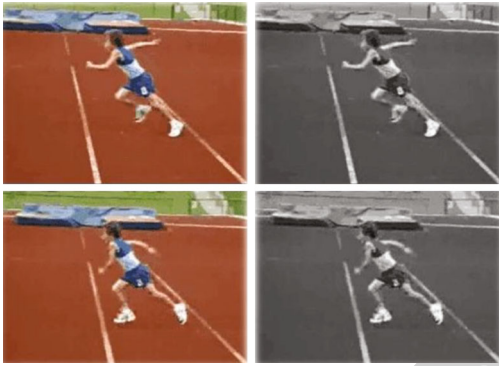


FIGURE 6: Comparison of real optical flow and calculated optical flow.

TABLE 3: Specific parameters of UCF101 data set.

Project	Parameter
Action category	80
Number of videos	13310
Average video duration	8.21
Total time length	1400 min
Maximum duration	72.5 s
Longest	2.06 s
Frame rate	30fps
Resolution	320*240

TABLE 4: Specific parameters of CASIA behavior analysis database.

Project	Parameter
Action category	16
Number of videos	1336
Video duration	5 s—28 s
Frame rate	25fps
Resolution	320*240

model of the UCF101 information set, and the normal of the three information sets will be taken concerning illustration those last test aftereffect. In place to assess the distinguish-

ment execution from claiming this algorithm on the UCF101 information set, look at this calculation for other activity distinguishment algorithms, namely, convolution neural system (CNN-) based activity distinguishment algorithm [20], dependent upon those calculations for upgrading thick trajectories (IDT), distinguishment dependent upon profound taking in (C3D), and human activity distinguishment calculation in view of neural system association (CNNT), distinguishment movement algorithm in light of spacetime web-domain and vector system (SVM).

4.2. *The Impact of the Sampling Frequency of the Measurement System on the Mapping.* In order to analyze the influence of the sampling frequency of the measurement system on the mapping, the sampling frequency of the laser tracker was set to 25 Hz, 50 Hz, and 80 Hz, and two sets of trajectory data were obtained. 20 pairs of data points, 40 pairs of data points, and 68 pairs of data points are substituted into ISO algorithm, DTW algorithm, and CDTW algorithm to analyze the influence of sampling frequency on the three mapping methods. The experimental results are shown in Table 3. The ISO algorithm is used for mapping between trajectories. As the sampling frequency increases, the distance between the sampling points becomes smaller, the trajectory tangent becomes steep, and the trajectory points with mapping errors increase. It can be seen from Table 3 that when the sampling frequency is increased from 25 Hz to 80 Hz, the trajectory accuracy decreases from 1.160 mm to 1.532 mm, and the standard deviation increases from 0.304 mm to 0.535 mm. The increase in sampling frequency leads to a decrease in the accuracy of the trajectory after ISO algorithm mapping, and an increase in the overall error fluctuation range. The DTW algorithm is a point-to-point mapping. As the sampling frequency increases, the density of trajectory points increases, and one-to-one matching between trajectory points can be better achieved. From Table 3, it can be seen that the DTW mapping method is just the opposite of the ISO mapping method. When the sampling frequency is increased from 25 Hz to 80 Hz, the trajectory accuracy increases from 0.678 mm to 0.508 mm, and the standard deviation decreases from 0.242 mm to 0.159. mm. The increase of the sampling frequency leads to the improvement of the accuracy of the trajectory after

the DTW algorithm is mapped, and the overall error fluctuation range is reduced.

**4.3. Analysis of Characteristics of Spatial Flow Network.** The GoogLeNet neural network fine-tuned by the pretraining model is used to convolute and aggregate the appearance images and corresponding optical flow features in the video in a given time window layer by layer. Then, the semantic feature sequence of spatiotemporal flow with high-level significant structure is obtained by using long-term and short-term memory multilayer recursive network cross-perception. After obtaining the characteristics of the time flow network and the space flow network of the deep residual network with a dual-stream structure, the temporal characteristics are merged. Through the analysis of the fusion method in Chapter 4, the weighted average fusion method is used to fuse the probability vectors of multiple video frames into the probability vector of the entire video, and the weight coefficient of the spatial stream network is taken as 1/4, and the weight coefficient of the spatial stream network is taken as 2/5, the recognition accuracy of the two networks is merged, and the final recognition classification result is 88.94%. In order to prove that the dual-stream structure model and the deep residual network are useful and applicable in action recognition, the final recognition result is compared with the accuracy based on the UCF101 behavior data set in other documents and deepened to the 101-layer dual-stream deep residual compared with other methods, the behavior recognition effect of the UCF101 data set on the network is improved, and because the optical flow features extracted by the combined optical flow network are more accurate, the behavior characterization information is extracted.

## 5. Conclusion

Through experiments, it can be found that when recognizing some small actions, a large amount of data in the background of the video are required to analyze and judge them. The first step of the dual-core extreme learning machine uses the method of trajectory density and learning characteristics to determine the action of kicking football. Perform analysis, you can fully characterize the action. The second step combines the time sequence template of the video with the convolution neural network to achieve the analysis and judgment of different angles of football action. And through the recognition method of the dynamic time regularization of the human skeleton, the system recognizes the complex actions of a person when playing football by comparing the similarity between a single action sequence and a template group. The shortcomings of this article are mainly that the algorithm for identifying football actions is too complicated or not simple and efficient enough, and there is no standard for rapid response, so the algorithm still needs to be optimized and strengthened. Prospects for follow-up research: First of all, in the area of data collection, I hope to get as much data as possible. When collecting data, the scene cannot be too single, and multiple sites need to be replaced to achieve the universality of the data. In addition,

different professional personnel are required to calibrate and collect the template sequence in the video, so that when the data is compared later, a single template can more appropriately correspond to the actions in the video template.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

There are no potential competing interests in our paper.

## Authors' Contributions

All authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

## References

- [1] R. B. Trabelsi, A. D. Masmoudi, and D. S. Masmoudi, "Hand vein recognition system with circular difference and statistical directional patterns based on an artificial neural network," *Multimedia Tools and Applications*, vol. 75, no. 2, pp. 687–707, 2016.
- [2] Y. Lu, G. Lu, J. Li, Z. Zhang, and Y. Xu, "Fully shared convolution neural networks," *Neural Computing and Applications*, vol. 23, no. 4, pp. 1–14, 2021.
- [3] X. Zheng, X. Xiao, and Y. Wang, "Harmonic impedance measurement based on an improved binary regression algorithm and dynamic time warping distance," *International Journal of Electrical Power & Energy Systems*, vol. 130, no. 1, p. 106907, 2021.
- [4] E. Eslami, Y. Choi, Y. Lops, A. Sayeed, and A. K. Salman, "Using wavelet transform and dynamic time warping to identify the limitations of the CNN model as an air quality forecasting system," *Geoscientific Model Development*, vol. 13, no. 12, pp. 6237–6251, 2020.
- [5] Z. Shang, X. Liu, W. Li, G. Maosheng, and Y. Yan, "A rolling bearing fault diagnosis method based on fastDTW and an AGBDBN," *Insight-Non-Destructive Testing and Condition Monitoring*, vol. 62, no. 8, pp. 457–463, 2020.
- [6] Z. Yi, L. Qi, E. Chen, Y. Ge, and J. L. Zhao, "Exploiting multi-channels deep convolutional neural networks for multivariate time series classification," *Frontiers of Computer Science*, vol. 10, no. 1, pp. 96–112, 2016.
- [7] K. Zhang, Z. Liu, and L. Zheng, "Short-term prediction of passenger demand in multi-zone level: temporal convolutional neural network with multi-task learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1480–1490, 2020.
- [8] G. Amirthalingam, "Multi-biometric authentication using deep learning classifier for securing of healthcare data," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 4, pp. 1340–1347, 2019.
- [9] K. Simonyan and A. Zisserman, "Two-stream convolutional net-works for action recognition in videos," *Advances in*

- Neural Information Processing Systems*, vol. 1, no. 4, pp. 568–576, 2016.
- [10] Y. Jiang, Y. Qi, W. K. Wang et al., “EventDTW: an improved dynamic time warping algorithm for aligning biomedical signals of nonuniform sampling frequencies,” *Sensors*, vol. 20, no. 9, p. 2700, 2020.
- [11] S. Li and Y. Lei, “Quantitative assessment of stroke patients based on dynamic time warping algorithm,” *Sheng wu yi xue gong cheng xue za zhi=Journal of biomedical engineering=Shengwu yixue gongchengxue zazhi*, vol. 35, no. 1, pp. 139–144, 2018.
- [12] S. Chen, S. Jin, X. Y. Li, and W. Yang, “Nonstretching normal-moveout correction using a dynamic time warping algorithm,” *Geophysics*, vol. 83, no. 1, pp. V27–V37, 2018.
- [13] X. Wang, F. Yu, W. Pedrycz, and L. Yu, “Clustering of interval-valued time series of unequal length based on improved dynamic time warping,” *Expert Systems with Applications*, vol. 125, pp. 293–304, 2019.
- [14] W. Li, Z. Luo, and X. Xi, “Movement trajectory recognition of sign language based on optimized dynamic time warping,” *Electronics*, vol. 9, no. 9, p. 1400, 2020.
- [15] L. Livingston, P. Deepika, and M. Benisha, “An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition,” *International Journal of Engineering Trends and Technology*, vol. 35, no. 11, pp. 506–510, 2016.
- [16] R. Varatharajan and G. Manogaran, “Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm,” *Cluster Computing*, vol. 21, no. 6, pp. 1–10, 2018.
- [17] H. L. Li, Y. Liang, and S. C. Wang, “Review on dynamic time warping in time series data mining,” *Control and Decision*, vol. 33, no. 8, pp. 1345–1353, 2018.
- [18] Y. Li, J. Sun, and R. Li, “Human action recognition based on dynamic time warping and movement trajectory,” *International Journal of Simulation: Systems*, vol. 17, no. 46, pp. 33.1–33.5, 2016.
- [19] G. Ekim, A. Atasoy, and N. Kizler, “A new approach for eye-blink to speech conversion by dynamic time warping,” *Traitement du Signal*, vol. 38, no. 2, pp. 369–377, 2021.
- [20] V. M. Kumar and D. Thipesh, “Robot arm performing writing through speech recognition using dynamic time warping algorithm,” *International Journal of Engineering, Transactions B:Applications*, vol. 30, no. 8, pp. 1238–1245, 2017.