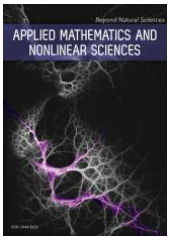
sciendo Applied Mathematics and Nonlinear Sciences, 10(1) (2025) 1-20



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on Sports Dance Training and Teaching in Modern Colleges and Universities Combined with Deep Learning

Jie Jiao[1,†]

1. Luoyang Institute of Science and Technology, Luoyang, Henan, 471000, China.

Submission Info

Communicated by Z. Sabir
Received October 13, 2024
Accepted February 10, 2025
Available online March 21, 2025

Abstract

Deep learning technology, one of the key technologies in the field of artificial intelligence, has shown a strong potential for application in many fields. The purpose of this paper is to explore the application of deep learning technology in the training and teaching of sports dance in modern colleges and universities, with a view to improving training efficiency and teaching quality. The OpenPose algorithm is used to realize the posture estimation of sports dance trainers, and the sports dance movement recognition model based on TAR-DL is constructed, and then the sports dance movement evaluation method based on the improved DTW algorithm is proposed.The sports dance movement recognition rate of the TAR-DL model is as high as 99.72%, which is significantly better than that of other 3D recognition methods. Meanwhile, the recognition rate of this paper's method for the six basic sports dance movements is between 95% and 99%, which is better than the recognition effect. Compared with the DTW algorithm, the Improved-DTW algorithm improves the accuracy by 3.14%, while reducing the time consumed by 0.37ms, which proves the effectiveness of the algorithm improvement strategy designed in this paper. In addition, the evaluation results based on the improved DTW proposed in this paper are closer to those of the professional sport dancer teacher, which fully proves the superiority and effectiveness of the Improved-DTW algorithm in the sport dance movement recognition task, and it can be used in the movement evaluation task of sport dance training and teaching in colleges and universities.

Keywords: OpenPose algorithm; TAR-DL method; DTW algorithm; Deep learning; Sport dance.
AMS 2010 codes: 68T01

†Corresponding author.
Email address: cleanj@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2025-0602

1 Introduction

With the improvement of people's living standard, more and more people begin to pay attention to health problems, and sports have become a popular way of fitness. In colleges and universities, sports dance is a set of sports, art, physical beauty in one and as a beautiful art form and an important sports program, by a wide range of attention and support [1-4]. Sports dance has a low threshold, rich in forms of expression, and has become a popular sport in colleges and universities, and the quality of dance teaching should also be mentioned to a new height [5-6]. However, at present, there is a serious disconnection between classroom teaching and extracurricular training in college sports dance, which directly affects the improvement of sports dance performance and is not conducive to the long-term development of sports dance specialty. In order to improve students' competitive level and physical and mental health, the combination of teaching and training of college sports dance is increasingly important [7-10].

Training content and method are two important aspects of specialized physical training in college sports dance. In terms of training content, it is necessary to develop a scientific and reasonable training program according to different training objectives and individual differences of students, focusing on details and skill training [11-13]. In terms of training methods, diversified training methods can be used, such as single training, combination training and cycle training, and at the same time, competitive training and application of technical means can also be increased to improve the training effect and interest [14-16]. Sports dance special physical training should focus on comprehensiveness, which means that the training program needs to cover all aspects of the athlete's body, including endurance, explosive force, flexibility, balance and so on. In training, it is necessary to set up corresponding training programs for different abilities in order to comprehensively improve the physical quality of athletes [17-20].

In this paper, the OpenPose algorithm, the TAR-DL method, and the improved DTW algorithm are comprehensively applied to successfully assess the accurate movement of sports dance training movements and provide teaching aids for dance teaching. Firstly, the key points of human skeleton in the sports dance movement images are extracted based on the OpenPose algorithm, and the estimation of human posture is realized by matching the key points for the sports dance trainers in the images and aggregating different types of key points into human bodies. Then, the TAR-DL method, which embeds the (3D time + 3D channel + 3D space) attention module with BCEF loss function in SlowFast network, is proposed to realize the accurate recognition of sports dance movements. Then, an improvement strategy for the DTW algorithm is proposed, and the sports dance movement evaluation is realized based on the improved DTW algorithm. Finally, the model is applied and evaluated.

2 Deep Learning-based Training Model for College Dance Sport

In order to realize the research on sports dance training and teaching in modern colleges and universities, this paper firstly extracts the key point information of human skeleton by using OpenPose algorithm, then constructs a sports dance movement recognition model based on TAR-DL method, and finally applies DTW algorithm to evaluate the normality of the recognized dance movements.

2.1 Human Pose Estimation Based on OpenPose Algorithm

Since the sequence of human skeletal keypoints can effectively express the human action posture information in videos, more and more computer vision and machine learning tasks are realized based

on human posture estimation algorithms, including image recognition, action video analysis, etc. OpenPose algorithm is the world's first real-time 2D human posture estimation open-source library developed based on convolutional neural networks and supervised learning, and its good results in both single and multi-person human pose detection applications [21].

The OpenPose algorithm uses a bottom-up idea to implement a human pose estimator, which first predicts all the keypoints in the image as candidate points, and then matches the keypoints for each person in the image, aggregating different types of keypoints into a human body. Therefore, no matter how many people are in the image, the pose of all the people can be estimated with only one inference, thus achieving near-real-time processing performance. Meanwhile. The OpenPose algorithm also has excellent performance in estimating poses of blurred or overlapping person objects.

2.1.1 Overall process and network structure of OpenPose

The Open Pose algorithm adopts a two-branch multi-stage convolutional neural network structure, and the network structure and overall processing flow are shown in Fig. 1. First, the original RGB image of size $w \times h$ is fed as input to the first 10 layers in the VGG-19 network for initialization and fine-tuning to generate a set of input feature maps F, which are fed to two convolutional neural network branches in the first stage [22]. Among them, the top branch is responsible for predicting the set of confidence maps $S^1 = \rho^1(F)$ for all skeletal keypoints in the image, and the bottom branch is responsible for reasoning about the affinity vector field (PAF) $S^1 = \varnothing^1(F)$ between skeletal keypoints. Where ρ^1 and $\varnothing^1$ denote the two convolutional neural network branches in the first stage, and in each subsequent stage, the prediction results of the two network branches in the previous stage are fused with the original feature maps, so as to iteratively refine the confidence maps and PAFs, and to produce more accurate prediction results. The iterative formulas are shown in Eqs. (1) and (2):

$$S^t = \rho^t\left(F, S^{t-1}, L^{t-1}\right), t \geq 2 \tag{1}$$

$$L^t = \phi^t\left(F, S^{t-1}, L^{t-1}\right), t \geq 2 \tag{2}$$

where ρ^1 and $\varnothing^1$ are the two convolutional neural network branches in stage t.

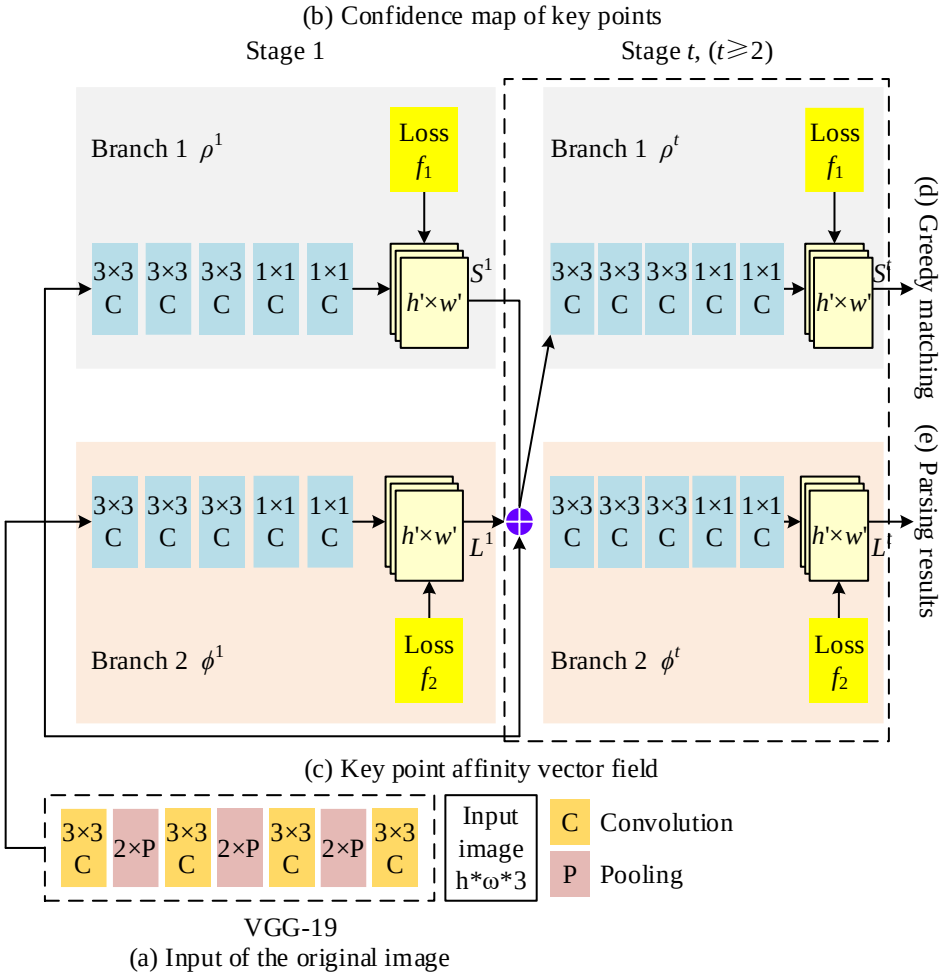


Figure 1. Two-branch multi-stage convolutional neural network

Corresponding loss functions are applied at the end of each stage of the two network branches to guide the network iterations in generating the confidence map ensemble and PAF, where a standard L_2 loss function is used. At the same time, the problem in the dataset caused by the fact that some of the data do not fully label all the human targets is addressed by weighting the loss function in space. The loss functions of the two convolutional neural network branches for a particular stage t are shown in Eq. (3) and Eq. (4), respectively:

$$\sum_{j=1}^{J}\sum_{d} W(p)\cdot\left\|S_j^t(p)-S_j^*(p)\right\|_2^2 \tag{3}$$

$$\sum_{c=1}^{c}\sum_{p} W(p)\cdot\left\|L_c'(p)-L_c^*(p)\right\|_2^2 \tag{4}$$

Where, S_j^* is the confidence map corresponding to the key points of the human skeleton. L_c^* is the PAF of the human skeleton keypoints. W is a binary mask to avoid the error penalty during the training process. $W(p)=0$ denotes that position p is not labeled in the image, but the network still predicts the confidence map and PAF of the corresponding skeleton keypoints. In this case, the error penalty can be avoided by using the binary mask W. Meanwhile, in order to solve the problem of

disappearing gradients, the OpenPose network periodically replenishes the gradients at various stages in between. Equation (5) shows the overall loss function of OpenPose network:

$$f = \sum_{t=1}^{T} \left(f_S^t + f_L^t \right) \tag{5}$$

where f_S^t denotes the loss function corresponding to the human skeletal keypoint confidence map prediction network branch, and f_L^t denotes the loss function corresponding to the human skeletal keypoint affinity vector field inference network branch.

2.1.2 Human critical point confidence and partial affinity vector fields

During the OpenPose model training process, a 2D confidence map S^* of the corresponding human skeletal keypoints is generated based on the labels, which is used to represent the confidence level of the skeletal keypoints appearing at each pixel location in the image. Each pixel in the confidence map corresponds to a confidence value, the size of which is determined by the distance between the pixel and the label location: the closer the distance, the higher the value, and the overall Gaussian distribution. Let $x_{j,k} \in R^2$ be the true labeled position of the j th skeletal key point corresponding to the k rd human body in the image, then for position $p \in R^2$ the value in the confidence map $S_{j,k}^*$ is shown in Equation (6):

$$S_{j,k}^* = \exp\left(-\frac{\left\| p - x_{j,k} \right\|_2^2}{\sigma^2} \right) \tag{6}$$

where σ serves to limit the spread of the confidence map peaks. When there are multiple human bodies in the graph, each skeletal critical point of each human body matches a confidence map, so the individual confidence maps of each human body must be aggregated together by a maxima-taking operation to obtain a single confidence level for calculating the loss S_j^*. The aggregation formula is shown in Equation (7):

$$S_j^*(p) = \max_k S_{j,k}^*(p) \tag{7}$$

Partial Affinity Vector Field PAF for Human Skeletal Keypoints PAF is a two-dimensional vector field that encodes the orientation of a human limb from one skeletal keypoint to another, with each type of human limb having a corresponding partial affinity vector field connecting the two associated body part maps. The greatest advantage of this feature representation is that it preserves both positional and directional information about the limb support region.

The single limb keypoint partial affinity vector field is schematically shown in Fig. 2, and let the left elbow bone keypoint j_1 of person k in the image be at position $x_{j1,k}$ in the image, and the left wrist keypoint j_2 be at position $x_{j2,k}$ in the image. if the pixel point p is located in the region corresponding to the left arm limb, the corresponding partial affinity $L_{c,k}^*(p)$ at p behaves as a unit vector pointing from j_1 to j_2, while for other pixel points, $L_{c,k}^*(p)$ behaves as a zero vector. Therefore, the partial affinity vector field $L_{c,k}^*$ at pixel point p can be defined as:

$$L_{c,k}^{*}(p)=\begin{cases} v & physically \\ 0 & Other \end{cases} \tag{8}$$

$$v=\frac{x_{j_2,k}-x_{j_1,k}}{\left\|x_{j_2,k}-x_{j_j,k}\right\|_2} \tag{9}$$

where v denotes the unit vector in the direction of the limb. The set of points on the limb satisfies the following three conditions:

$$0\le v\cdot\left(p-x_{j_1,k}\right)\le l_{c,k} \tag{10}$$

$$v_{\perp}\cdot\left(p-x_{j_1,k}\right)\le \sigma_l \tag{11}$$

$$l_{c,k}=\left\|x_{j_2,k}-x_{j_1,k}\right\|_2 \tag{12}$$

where $v_{\perp}$ is a vector perpendicular to v, σ_l is the width of the limb characterized in terms of pixel distance, and $l_{c,k}$ is the limb length. The PAF of a pixel point in the image is the average of the PAFs of all corresponding skeletal keypoints located at that point, as shown in Equation (13):

$$L_c^{*}(p)=\frac{1}{n_c(p)}\sum_k L_{c,k}^{*}(p) \tag{13}$$

where $n_c(p)$ denotes the number of non-zero vectors on pixel point p in the image.

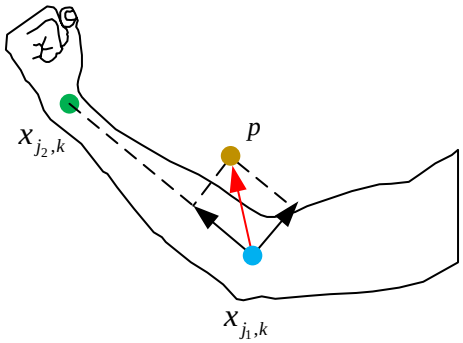


Figure 2. Schematic diagram of affinity vector field of key points of a single limb

2.2 Sports dance movement recognition model based on TAR-DL

In this paper, we propose the TAR-DL method, which embeds the (3D temporal + 3D channel + 3D spatial) attention module with the BCEF loss function in the SlowFast network to improve the average accuracy of the model for sports dance movement recognition.

2.2.1 TAR-DL network structure

The structure of the TAR-DL network in this paper is schematically shown in Fig. 3. First, consecutive multiple video frames with temporal information are input into the TAR-DL network, and the spatial

and motion information in the video are captured using the Slow Fast network, after which the generated spatial and motion information are horizontally connected to generate a feature map with five dimensions of data information [23]. These five dimensions are the number of video frames, the number of channels, the time dimension, the width and the height of a single input model, i.e., $C \times T \times W \times H$. Compared to the image-generated Feature map, the video-generated Feature map has an additional time dimension T. Then, the SlowFast-generated Feature map $F \grave{o} R^{C \times T \times H \times W}$ is used as the input Feature map to the (3D time + 3D channel $+3D$ spatial) attention Convolution module, respectively, from the time, channel and space of the three different perspectives of the information in $F \in R^{C \times T \times H \times W}$, so as to realize the use of the attention mechanism to pay attention to the "when", "what", "where" in the video. Finally, a multi-label classifier is used to realize the classification prediction, and the difference between the prediction result and the real value is calculated using BCEFLoss.

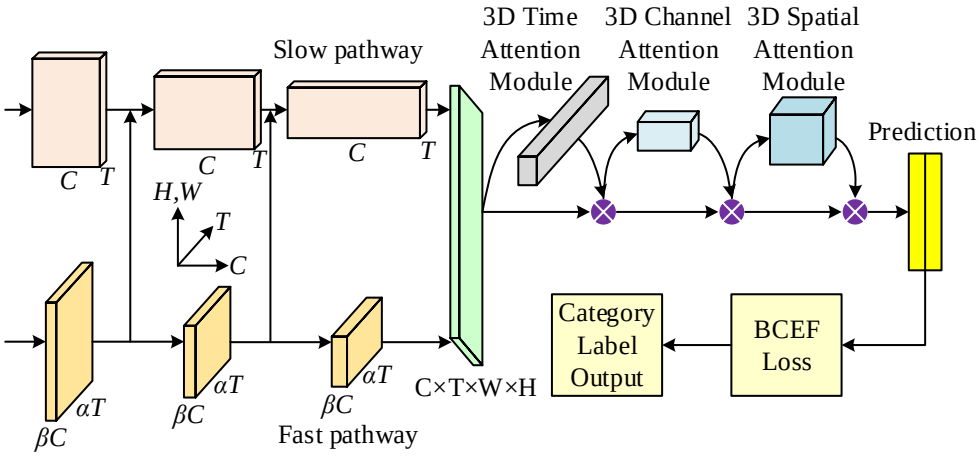


Figure 3. Network structure of the TAR-DL

(3D time + 3D channel + 3D space) The computation of the attention convolution module is shown in Eqs. (14) to (16):

$$F' = M_t(F) \otimes F \tag{14}$$

$$F'' = M_c(F') \otimes F' \tag{15}$$

$$F''' = M_s(F'') \otimes F'' \tag{16}$$

Where $M_t \in R^{1 \times T \times 1 \times 1}$ is the 3D temporal module attention computation function, $M_c \in R^{C \times 1 \times 1 \times 1}$ is the 3D channel module attention computation function, $M_s \in R^{1 \times 1 \times H \times W}$ is the 3D spatial module attention computation function, and $\otimes$ is the element-by-element multiplication. From the calculation process, it can be seen that after three times of element-by-element multiplication, the attention value generated by each attention module will be passed to the next layer, so the final output attention graph F''' contains temporal attention feature information, channel attention feature information, and spatial attention feature information.

In the (3D time + 3D channel + 3D space) attention convolution module, Global Maximum Pooling (GMP) and Global Average Pooling (GAP) are used to focus on the information of time, channel and space, and then Squeeze and Excitation operations are performed to obtain the corresponding weight information.

2.2.2 Three-dimensional attention and the BCEF loss function

1) Attention Modules

A common practice to improve the average recognition accuracy of existing network models is to embed an attention module, and the performance of many networks is improved by embedding an attention module. Thus, in this paper, a lightweight attention module, CBAM, is improved and embedded into the existing SlowFast network to improve the average recognition accuracy of network models [24].

CBAM consists of two sub-modules: channel attention CAM and SAM, CAM performs channel Attention on Feature map, i.e., retains channel information, compresses spatial information, and focuses on the "what" in the network. SAM performs spatial Attention on Feature map, i.e., retains spatial information, compresses channel information, and focuses on "where" in the network.

$F \in R^{C\times H\times W}$ is the Feature map of the input CBAM module, where the channel attention map is denoted by M_c and the spatial attention map is denoted by M_s. Then the calculation of the channel attention and spatial attention is shown in Eqs. (17) to (18):

$$F' = M_c(F) \otimes F \tag{17}$$

$$F'' = M_s\left(F'\right) \otimes F' \tag{18}$$

Eqs. $M_c \in R^{C\times 1\times 1}$, $M_s \in R^{1\times H\times W}$.

CBAM only focuses on both channel and spatial information, while ignoring the temporal information. For this reason, this paper proposes (3D time + 3D channel + 3D space) attention convolution module to improve the average recognition accuracy of the existing network SlowFast while preserving the temporal information between consecutive video frames.

2) 3D temporal attention part

The 3D temporal attention part uses the temporal information of the Feature map obtained from SlowFast network to generate a 3D temporal attention map, which focuses on the "when" information in the input video. The specific process is as follows:

First, the 3D Temporal Attention module uses 3DGMP with 3DGAP to aggregate the channel information and spatial information in the input Featuremap to generate two different feature maps: $F_{avg}^{t} \in R^{1\times T\times 1\times 1}$, $F_{\max}^{t} \in R^{1\times T\times 1\times 1}$. The 3D temporal attention module then feeds it into the shared network to generate $M_t \in R^{1\times T\times 1\times 1}$. In this case, the shared network consists of two parts, the multilayer perceptron (MLP) and the hidden layer (HL), and the activation size in the HL is set to $\square^{1\times T/r\times 1\times 1}$. The $r = 8$ in the expression, which represents the squeeze rate, is summed up element by element after Squeeze and Excitation of the shared network. Finally, the 3D temporal attention module multiplies the Feature map obtained by element-by-element summing with the input Feature map to generate a temporal attention map with the same size as the input Feature map.

The formulas of 3D temporal attention module are shown in Eqs. (19) to (20):

$$M_t(F) = \sigma MLP\left(3DA(F)\right) + MLP\left(3DM(F)\right) \tag{19}$$

$$M_t(F) = \sigma\left(W_1\left(W_0\left(F_{avg}^t\right)\right) + W_1\left(W_0\left(F_{\max}^t\right)\right)\right) \tag{20}$$

Where, σ denotes the sigmoid function, F denotes the input Feature map, $W_0 \in R^{C/r \times C}$, $W_1 \in R^{C \times C/r}$.

3) 3D channel attention part

The 3D channel attention module uses the feature map obtained from SlowFast network to generate the 3D channel attention map, which focuses on the "what" information in the input video. The specific process is as follows:

First the 3D Channel Attention module uses 3DGMP with 3DGAP to make the network model focus on temporal and spatial information to generate two different 3-dimensional Featuremaps: $F_{avg}^c \in R^{C\times1\times1\times1}$, $F_{\max}^c \in R^{C\times1\times1\times1}$. Then the 3D Channel Attention module feeds the two 3-dimensional Featuremaps into the shared network to generate the Temporal Attention Map $M_c \in R^{C\times1\times1\times1}$, with the activation size set to $R^{C/r\times1\times1\times1}$, and in the experiments set to $r = 16$. After the shared network's Squeeze and Excitation, the elements are summed up element by element. Finally, the 3D Channel Attention module multiplies the Feature map obtained by element-by-element summing with the input Feature map to generate a channel attention map with the same size as the input Feature map.

The formulas of the 3D channel attention module are shown in Eqs. (21) to (22):

$$M_c(F) = \sigma\left(MLP\left(3DA(F)\right) + MLP\left(3DM(F)\right)\right) \tag{21}$$

$$M_c(F) = \sigma\left(W_1\left(W_0\left(F_{avg}^c\right)\right) + W_1\left(W_0\left(F_{\max}^c\right)\right)\right) \tag{22}$$

4) 3D Spatial Attention

The 3D spatial attention part uses the feature map obtained from SlowFast network to generate the 3D spatial attention figure $M_s \in R^{1\times1\times H\times W}$, which focuses on the "where" information in the input video.

The computation of the 3D spatial attention module is shown in Eqs. (23) to (24):

$$M_s(F) = \sigma\left(f^{3\times3\times3}([3DA(F); 3DM(F)])\right) \tag{23}$$

$$M_s(F) = \sigma\left(f^{3\times3\times3}\left[F_{avg}^s; F_{\max}^s\right]\right) \tag{24}$$

where $f^{3\times3\times3}$ denotes a convolution operation with a convolution kernel of size $3\times3\times3$.

5) Loss function

In the multi-label classification action recognition task with spatio-temporal information, the loss function used by the conventional deep learning network model is the loss function of binary categorization crossover (BCE Loss).The formula for BCE Loss is as follows:

$$L_{Bce} = -y\log y' - (1-y)\log(1-y') \tag{25}$$

where y' is the predicted category probability of the model and y is the true category sample value. However, BCE Loss does not take into account the impact of the long-tail effect on the model training, while Focal Loss can increase the loss weight of the tail data categories during the training process and attenuate the impact of the long-tail effect [25].The formula of Focal Loss is as follows:

$$L_F = \begin{cases} -(1-y')^{\gamma}\log y', y=1 \\ -y'^{\gamma}\log(1-y'), y=0 \end{cases} \tag{26}$$

Compared with BCE Loss, Focal Loss increases the weight value of $(1-y')^{\gamma}$, i.e., if the probability of the correct category predicted by the network model increases, the value of $(1-y')^{\gamma}$ decreases, so the loss weight of the correctly categorized samples decreases, which leads to an increase in the weight of the incorrectly categorized, i.e., the sample category with a small amount of data.

In order to increase the loss weight of the sample categories of the tail class data and reduce the influence of the long-tail effect on the model training, this paper fuses the BCE Loss with the Focal Loss and proposes the BCEF Loss to improve the average identification accuracy of the model.The formula of the BCEF Loss is shown in equations (27) to (29):

$$A = -\alpha\log(1-y')^{\gamma}\log(y')y \tag{27}$$

$$B = (1-\alpha)(y')^{\gamma}\log(1-y')(1-y) \tag{28}$$

$$L_{BCEF} = (A-B) \tag{29}$$

α is the weight factor, which takes a value in the range of $\alpha \in [0,1]$ and is used to reduce the weight of the positive or negative category samples with the aim of solving the problem of imbalance in the positive and negative category samples. γ is the adjustable focus parameter, $(1-y')^{\gamma}$ is the weight value that changes with γ, y' is the predicted category probability of the model, and y is the true category sample label.

The core idea of this paper is to improve the overall average correct recognition rate of the model by adjusting the focus parameter γ to reduce the loss weight of the head class samples and increase the loss weight of the sparse samples in the tail class. When γ takes the value of 0, BCEFLoss is equivalent to the loss function of binary crossover. When γ takes a gradually increasing value, the overall impact of $(1-y')^{\gamma}$ on the model will also gradually increase, and the experimental results show that the TAR-DL method proposed in this paper has the best performance when $\gamma = 5$.

2.3 DTW-based assessment method for sport dance movements

In order to further assess the completion quality of prescribed movements in sport dance, this paper develops a movement assessment strategy. The training set was selected to integrate the prescribed movement template, and at the same time, the qualified range of movement quality was determined, collectively known as the range of movement similarity index, through which the range of similarity index was used to determine the grade to which the prescribed movements belonged to, so as to achieve the purpose of movement quality assessment. The data in the test set are used as the actual action sequences, and the optimized DTW algorithm is used as the evaluation model, which calculates the action similarity index by comparing the actual action sequences with the sequence of prescribed action templates, and then determines the action quality grade in the action evaluation grade table, and then evaluates and provides targeted guidance to the actual actions according to the grade [26].

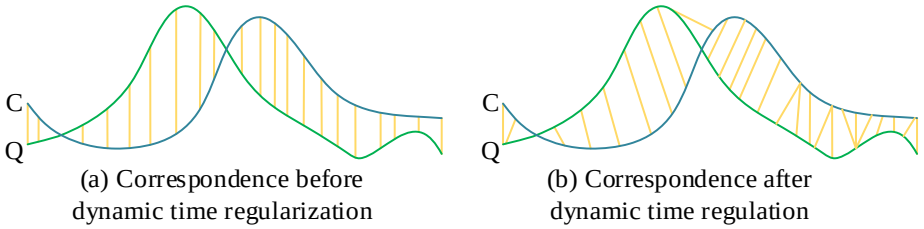
2.3.1 Action similarity and rating assessment

The DTW algorithm can be seen as employing a dynamic programming strategy to search for the minimum path of two time series of unequal lengths. By aligning the two sequences, the algorithm is able to measure the similarity between them more precisely. Similarly, DTW has a strong potential for application in sports dance movement quality assessment. It can match the realistic movement sequences with the template movement sequences, calculate the similarity between them, and provide a scientific basis for assessing movement normality and precision.

The main idea is: Let there be two different time sequences of the same sport dance movement, sequence C and sequence Q, with time lengths of M and N, respectively:

$$\begin{aligned} C &= C_1, C_2, \ldots, C_i, \ldots C_M \\ Q &= Q_1, Q_2, \ldots, Q_j, \ldots Q_N \end{aligned} \tag{30}$$

The correspondence of sequence points between Sequence C and Sequence Q in each frame before and after dynamic time planning is shown in Fig. 4.



(a) Correspondence before dynamic time regularization

(b) Correspondence after dynamic time regulation

Figure 4. Mapping between sequential points before and after dynamic time planning

The process of utilizing the DTW algorithm for sports dance movement evaluation is as follows:

First, in order to align sequence C and sequence Q, a 2D matrix R of $M*N$ needs to be constructed. $R(m,n)$ denotes the Euclidean distance when the m th point of sequence C is aligned with the n th point of sequence Q, i.e., $R(m,n)$ denotes the Euclidean distance value $d(m,n)$ when C_m is aligned with Q_n.

Let $D(C,Q)$ be the total distance of a path, i.e., the cumulative sum of the distance values $d(1,1)$ from C_1 and Q_1, through the distance values $d(i,j)$ from $C_i(1\le i\le M)$ and $Q_j(1\le j\le N)$, as shown in Eq. (31):

$$D(C,Q)=\sum d(i,j),1\le i\le M,1\le j\le N \tag{31}$$

There exists a shortest path in $D(C,Q)$ for the global path optimal solution, and this shortest path needs to satisfy the following three conditions:

1) Boundary conditions: the first and last points of the two sequences must be one-to-one correspondence, to ensure the completeness and coherence of the sequence alignment, and to avoid the missing sequence fragments.

2) Monotonicity: due to the specificity of the individual, each person performs the action at a different speed, but the action sequence must be constant, so the action sequence advances with time. Therefore, if $R_{\alpha-1}=(a,b)$ and $R_\alpha=(a^\varsigma,b^\varsigma)$ in the two-dimensional grid, it must satisfy $a^\varsigma-a\ge 0, b^\varsigma-b\ge 0$, to ensure that the optimal path must be monotonically increasing with the time sequence.

3) Continuity: The next point of the current point can only be a point directly adjacent to it. That is, the next point of the current point (i,j) must be one of $(i+1,j),(i,j+1),(i+1,j+1)$, to avoid the occurrence of many-to-one or one-to-many non-continuous cases in the matching process, in order to ensure that every coordinate in the sequences C and Q appears in the shortest path.

The above conditions show that if the coordinate of the current point is (m,n) then the next point must be one of the coordinates of $(m+1,n),(m,n+1),(m+1,n+1)$. The final formula of the DTW algorithm is obtained as shown in equation (32):

$$D(C,Q)=d(M,N)+\min\begin{cases}D(M-1,N)\\ D(M-1,N-1)\\ D(M,N-1)\end{cases} \tag{32}$$

In this way, the DTW distances of sequences C and Q can be obtained $D(C,Q)$. In order to more intuitively assess the execution of actions, this paper introduces the action similarity index as a measure of the quality of action completion. This index is obtained based on the DTW distance normalization process, which transforms the unbounded DTW distance into a similarity index that lies in the range of $(0,1]$ by means of the estimated upper and lower bounds. The estimated upper bound refers to the maximum possible distance between sequences C and Q, while the lower bound represents a relatively small distance value. This normalization makes the estimation results more understandable and comparable, as described in the following normalization procedure:

$$\psi(C,Q)=1-\frac{D^\Delta(C,Q)-D_\varsigma(C,Q)}{D^\varepsilon(C,Q)-D_\varsigma(C,Q)} \tag{33}$$

$$D_{\zeta}(C,Q)=\max\begin{cases}\left|First(C)-First(Q)\right|\\ \left|Last(C)-Last(Q)\right|\\ \left|\max(C)-\max(Q)\right|\\ \left|\min(C)-\min(Q)\right|\end{cases} \tag{34}$$

$$D^{\varepsilon}(C,Q)=\max(M,N)\times\max\begin{cases}\left|\max(C)-\min(Q)\right|\\ \left|\min(C)-\max(Q)\right|\end{cases} \tag{35}$$

where $\Psi(C,Q)$ is the similarity index of sequences C and Q, $D^{\Delta}(C,Q)$ is the DTW distance of sequences C and Q, $D_{\zeta}(C,Q)$ is the estimated lower bound, $D^{\varepsilon}(C,Q)$ is for the estimated upper bound, and $D^{\varepsilon}(C,Q)$ takes the maximum length in the two sequences multiplied by the diagonal path length as the maximum possible value of the DTW distance. The range of $\Psi(C,Q)$ is limited to $(0,1]$, and when $D^{\Delta}(C,Q)=D_{\zeta}(C,Q)$, the similarity index $\Psi(C,Q)$ reaches a maximum value of 1, indicating that the two sequences are perfectly matched. As the dissimilarity of the two sequences increases, $\Psi(C,Q)$ will continue to decrease, converging to 0 but not equaling 0.

0.6 is selected as the critical value, when $0.6\le\Psi(C,Q)\le 1$, it can be seen that the similarity between sequence C and sequence Q is higher, and the completion of the action is better, which is rated as grade A. On the contrary, the similarity between the two sequences is lower, and the completion of the action is unsatisfactory. If the completion of the movement is unsatisfactory, the bone point of $0<\Psi_i<0.6$ will be searched out and rated as grade B. It is considered that the bone point of Δ_i does not meet the requirements of the angle of the bone point in the prescribed movement template, and this result will be fed back to the user to provide targeted guidance.

2.3.2 Improved DTW algorithm

For the task of sports dance movement evaluation, the DTW algorithm does show high accuracy. However, at the same time, its complexity should not be neglected.The DTW algorithm calculates the Euclidean distance, which leads to both time and space complexity of $O(M\times N)$ when dealing with large-scale datasets. Especially in the case of $M=N$, the complexity is as high as $O(N^2)$. The computational burden of the DTW algorithm grows geometrically as the time-series data continues to grow, which is a great challenge for evaluating sports dance movements in real time. At the same time, due to the restriction of the DTW algorithm on the global path direction, i.e., the slope is set within 0.5~2. Although the problem of too large or too small path slopes is prevented to a certain extent, the computation of the algorithm is also aggravated at the same time.

In order to improve the performance of the DTW algorithm in the task of sports dance movement evaluation. The following improvement strategies are adopted for the DTW algorithm:

1) Relax the global path restriction: expand the slope range from 0.5~2 to 0.2~3, which reduces the constraints of the algorithm in finding the optimal path, and thus reduces the amount of computation. Meanwhile, due to the continuity and fluidity of sports dance movements, relaxing the slope restriction can better capture the similarity between movements and improve the accuracy of evaluation.

2) Coarse-graining search for shortest path. First, the full-resolution matrix R is coarsely granularized by setting the four adjacent cells to be merged into a new cell, which can significantly reduce the number of cells in matrix R, thus reducing the complexity of the algorithm. Then, the shortest path is found in the reduced resolution matrix R. Due to the smaller size of the matrix after coarse-graining, the search for the shortest path will be much faster. Finally, the searched shortest path is mapped back to the original full-resolution matrix R to get the shortest path in the full-resolution matrix. This method retains the main features of the original data and reduces the computational complexity.

3 Model application and analysis

3.1 Comparison of sports dance movement recognition algorithms

In order to verify the effectiveness of this paper's method for sports dance basic movement recognition, the TAR-DL network model was trained using the training set of sports dance basic movement dataset, and the accuracy of the model was tested using the test set. Compared to the action recognition method based on video data, the recognition method based on skeletal coordinate information has higher recognition efficiency. Therefore, in the comparison experiments, this paper first compares different methods based on 3D action recognition on the NTU RGB-D dataset, and the comparison results of different methods on the NTU RGB-D dataset are shown in Table 1.

The results show that 3D-based action recognition with richer joint relationships helps to capture more useful patterns, and additional motion prediction and complementation based on skeleton features in the NTU RGB-D dataset improves the recognition efficiency. To deal with noise and occlusion in 3D skeleton data, ST-LSTM introduces a gating mechanism in LSTM to learn the reliability of sequential input data and adjusts its effect on updating long term contextual information stored in memory cells accordingly, with a recognition rate of 77.7%. And the recognition efficiency of ST-GCN reaches 88.3%, which is a combination of the GCN model as well as the TCN model, which is a dynamic skeleton model of spatio-temporal dual streams, and the spatio-temporal dual streams based recognition method is configured with three-dimensional convolutional filters, and the accuracy of this method is better than the network structures such as ST-LSTM, TSRJI, Clips+CNN+MTLN, and so on. Where the hierarchical structure of the GCN model and the data in the action recognition task are diverse, the topology of the graph is heuristically set and fixed to all model layers and input data for processing the data with different rules of the graph structure.The recognition accuracy of MS-AAGCN is 96.2%. The action recognition method that uses TEM (Time Extension Module) in addition to MS-AAGCN has a recognition accuracy of 96.5%. And the TAR-DL method used in this paper, embedding (3D time + 3D channel + 3D space) attention module with BCEF loss function in SlowFast network, the recognition rate of this method is up to 99.72%, which is obviously better than other sports dance movement recognition methods.

Table 1. Comparison of 3D-based recognition methods in NTU RGB-D

Methods	CV/%	CS/%
ST-LSTM	78.94	70.31
TSRJI	81.41	74.25
Clips+CNN+MTLN	85.67	80.46
ST-GCN	89.44	82.17
DPRL	90.75	84.38
SGN	94.38	87.52
2S-AGCN	96.24	89.43
2S-NLGCN	96.24	89.43
MS-AAGCN	97.16	90.27
Sym-GNN	97.58	90.49
MS-AAGCN+TEM	97.69	91.84
TAR-DL	99.72	97.41

Given that sports dances are categorized into 10 dance categories, this paper targets the basic movements of Tango among them to conduct recognition experiments. In order to accurately observe the TAR-DL classification results of the self-constructed sports dance video dataset, this paper uses a confusion matrix to evaluate the performance of TAR-DL. The resulting confusion matrix for sports dance movement recognition is shown in Fig. 5, where it is known to assume that the diagonal elements are equal to the percentage of real numbers.

The confusion matrix shows that TAR-DL can effectively solve the problems of shape change and bone noise in large-scale data. Sport Dance's pas de deux format causes occlusion problems because of the interaction between dancers, and the consistency of each dancer's movements can have an impact on the experiment's results. As can be seen from Figure 5, the recognition rate of the six basic movements of "Progressive Side Step", "Closed Promenade", "Lock Turn", "R.F&L.F Lock Turn", "Progressive Link" and "Walk" is more than 90%, indicating that the TAR-DL model in this paper has a good recognition effect on the movement of dance sport and is feasible. Specifically, "Progressive Side Step", "Closed Promenade", and "Progressive Link" are confused, with 1% of "Progressive Side Steps" being mistaken for "Closed Promenade" and 1% being mistaken for "Progressive Link". In "Closed Promenade", 1% were mistaken for "Progressive Side Step" and 2% were mistaken for "Progressive Link". In "Progressive Link", 1% were mistaken for "Progressive Side Step" and 2% were mistaken for "Closed Promenade". At the same time, "Lock Turn" is also confused with "R.F&L.F Lock Turn", and 4% of "Lock Turns" are mistaken for "R.F&L.F Lock Turns". The reason for the confusion is mainly because in the dance sport, the extraction of each dancer's movements has the problem of different active and passive interactions, the lack of information about the relationship between skeleton joints, and the similarity between some movements, so it is difficult to capture and distinguish.

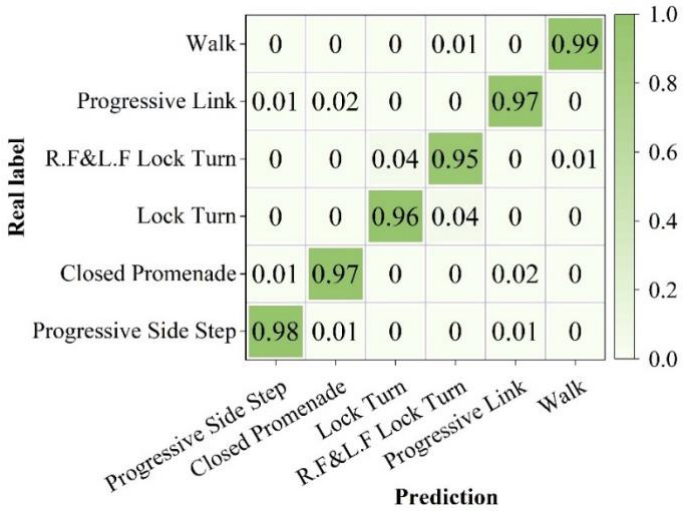
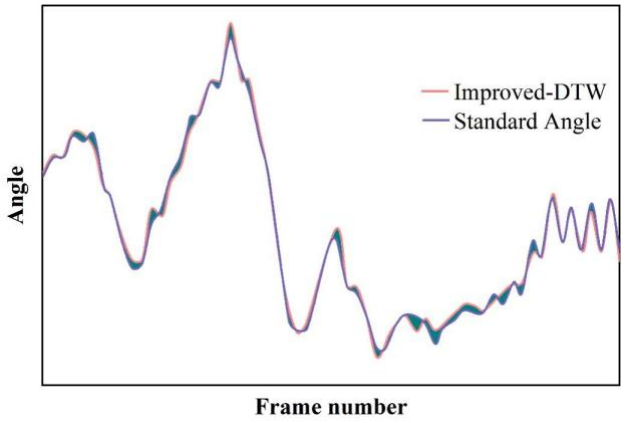


Figure 5. Confusion matrix of ballroom dancing basic movement recognition

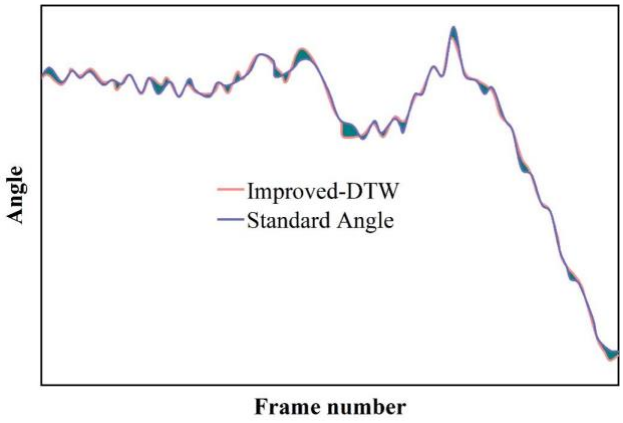
3.2 Performance evaluation of improved DTW algorithm

In order to verify the effectiveness of the DTW algorithm improvement strategy designed in this paper, a sports dance athlete competition video in a standard dataset is randomly selected as the prediction data, and the distance of the three basic movements of "Progressive Side Step", "Closed Promenade" and "Progressive Link" is analyzed by the Improved-DTW algorithm proposed in this paper. As shown in Figure 6, (a)~(c) are the comparison of the three dance movements of "Progressive Side Step", "Closed Promenade" and "Progressive Link", respectively. The shaded portion depicts the distinction between the standard sequence and the matched sequence of the Improved-DTW algorithm.

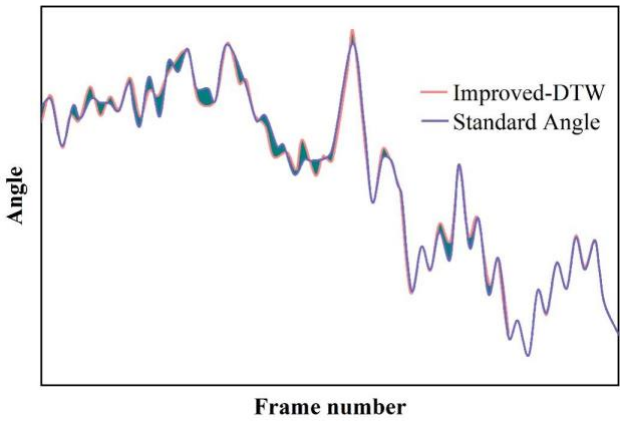
From Figure 6, it can be clearly observed that the two trajectories are closely adjacent to each other during each movement implying that the athletes' foot movement trajectories in this process are relatively standard and highly compatible with the ideal dance movements. However, if a significant distance between the matching line and the standard line is found during the movement process, it proves that there are deficiencies in the athletes' sports dance movement postures, which fail to fully meet the standard requirements of the movement. Additionally, the data collected during this experiment revealed the dynamic changes in the athlete's front and back swing during every move.



(a) Progressive Side Step



(b) Closed Promenade



(c) Progressive Link

Figure 6. Comparison of foot angle change of athletes with standard sequence

In order to verify the performance of Improved-DTW algorithm, DTW algorithm, FastDTW algorithm and HMM algorithm are used in this section of experiments for experimental comparison respectively, and the recognition results of each algorithm are shown in Table 2.

Through the comparative analysis, it can be clearly seen that the Improved-DTW algorithm proposed in this paper significantly outperforms the other compared algorithms in terms of accuracy. In the comparison results with the DTW algorithm, due to the Improved-DTW adopts the improvement strategy of relaxing the global path restriction and coarse-graining the search for the shortest path, the algorithm outperforms the DTW algorithm in terms of time-consumption and accuracy, with an improvement of 3.14% in terms of accuracy, and a reduction of 0.37ms in terms of time-consumption, which proves the effectiveness of the algorithmic improvement strategy designed in this paper. The results of experimental comparisons fully demonstrate the superiority and effectiveness of the Improved-DTW algorithm in the task of sports dance movement recognition.

Table 2. Experimental comparison

Algorithms	Accuracy rate/%	Time consuming/ms
DTW	91.25	2.04
FastDTW	86.73	1.74
HMM	90.68	2.18
Improved-DTW	94.39	1.67

3.3 Results and Analysis of Sports Dance Movement Assessment

In this paper, each dancer's posture during practicing the same piece of sports dance is selected, and the similarity scores between the sports dance posture and the standard posture of different dancers are calculated separately. When performing the calculation, the weights of all body parts are set to 1, and the weights of each category of scores are also set to 1. The sports similarity scores calculated using this paper's sports dance movement assessment method based on the improved DTW algorithm are shown in Table 3. Where A1~A7 denote the numbers of different dancers, A1 and A2 are professional dancers, and A3~A7 are amateur dancers.

As seen in Table 3, the scoring results of professional dancers A1 and A2's sport dance postures were 90.0 and 90.3 respectively, which were much higher than those of other dancers.The dance scores of

amateur dancers, such as A3~A7, were lower than those of the professional dancers, with scores all lower than 72. This reflects the validity of the method of this paper to some extent.

Table 3. Ballroom dancing movement similarity score using the method in this paper

Dancer	S_{pos}	S_{vel}	S_{angle}	S_{dist}	S_{total}
A1	91.9	93.0	90.6	84.6	90.0
A2	89.8	90.8	91.0	89.7	90.3
A3	70.1	72.2	69.2	72.1	70.9
A4	79.5	80.9	63.7	55.2	69.8
A5	77.9	81.4	65.9	60.8	71.5
A6	65.8	58.9	61.1	60.5	61.6
A7	56.0	60.3	49.7	56.8	55.7

In order to further prove the feasibility of this paper's method, the results obtained from this paper's method are compared with the scoring results of professional sports dance teachers, and the comparison of scoring results is shown in Table 4. As can be seen from Table 4, the results of the sports dance movement evaluation method based on the improved DTW algorithm described in this paper are relatively close to the scoring results of professional sports dance teachers. It shows that the method of this paper can be applied to the movement scoring task of sports dance training and teaching in modern colleges and universities.

Table 4. Comparison of scoring results

Dancer	Upper-body fluidity	Lower-body fluidity	Musical timing	Body balance	Choreography	S_{total}
A1	4	4	5	5	4	90.0
A2	4	5	5	5	5	90.3
A3	4	3	5	3	4	70.9
A4	4	5	3	2	5	69.8
A5	4	3	4	4	4	71.5
A6	4	2	4	2	3	61.6
A7	2	3	3	2	3	55.7

In addition, this paper also compares the scoring results without and with movement sequence alignment as shown in Table 5, which shows that the scoring effect is closer to the scoring results of professional sport dance teachers after using movement sequence alignment.

Table 5. Evaluation results with and without alignment processing

Dancer	The score of sports dance teachers after the normalization process	Align the processed S_{total}	Do not use aligned S_{total}
A1	93	90.0	67.4
A2	93	90.3	38.5
A3	82	70.9	60.2
A4	78	69.8	44.8
A5	82	71.5	52.9
A6	58	61.6	30.6
A7	50	55.7	41.5

4 Conclusion

Based on OpenPose algorithm, TAR-DL method and improved DTW algorithm, this paper explores the application of deep learning technology in the training and teaching of sports dance in modern colleges and universities, and provides data references for dance teaching through accurate recognition of sports dance movements.

The 3D-based motion recognition has richer joint relationships, which helps to capture more useful patterns, and the recognition effect is better than that of traditional 2D recognition methods. The TAR-DL method used in this paper, embedding (3D time + 3D channel + 3D space) attention module and BCEF loss function in the SlowFast network, has a recognition rate of 99.72% for dance sports training movements, which is significantly better than other 3D recognition methods. At the same time, in the TAR-DL classification experiment of sports dance movements, the recognition rate of six basic sports dance movements of "Progressive Side Step", "Closed Promenade", "Lock Turn", "R.F&L.F Lock Turn", "Progressive Link" and "Walk" reached more than 90%. The results show that the TAR-DL model in this paper has a good recognition effect on sports dance movements and is feasible.

When the Improved-DTW algorithm analyzed in this paper is applied to recognize sports dance movements, the trajectory of the athlete's feet during each movement is relatively consistent with the ideal dance movement. And comparing with the DTW algorithm, the Improved-DTW algorithm is significantly improved in terms of time consumption and accuracy, and its time consumption is reduced by 0.37ms while the accuracy is increased by 3.14%, which proves the effectiveness of the algorithmic improvement strategy designed in this paper. In addition, the Improved-DTW algorithm is used in the actual dancer movement evaluation task, and the evaluation results obtained are extremely close to those of professional sport dance teachers, thus strongly proving the superiority of the Improved-DTW algorithm in the sport dance movement recognition task.

References

[1] Wang, A., & Wang, C. (2021). Research on the application of sport dance in colleges and universities in the healthy development of sports. Revista Brasileira de Medicina do Esporte, 27, 464-467.

[2] Tang, J. (2024). The Influence of Sports Modern Dance on the Psychological Health of College Students. Journal of Sport Psychology/Revista de Psicología del Deporte, 33(1).

[3] Zhang, L., Zhao, S., Weng, W., Lin, Q., Song, M., Wu, S., & Zheng, H. (2021). Frequent sports dance may serve as a protective factor for depression among college students: a real-world data analysis in China. Psychology research and behavior management, 405-422.

[4] Chen, L., & Hu, F. (2017, April). Study on the Aesthetic Education Value of College Sports Dance and Its Realization Approach. In 7th International Conference on Education, Management, Information and Mechanical Engineering (EMIM 2017) (pp. 1611-1614). Atlantis Press.

[5] Wang, M. H. (2023). Influence of sport dance courses on female college students. Revista Brasileira de Medicina do Esporte, 29, e2022_0783.

[6] Zheng, C., & Ji, H. (2021). Analysis of the intervention effect and self-satisfaction of sports dance exercise on the psychological stress of college students. Work, 69(2), 637-649.

[7] Tang, H., & Guan, L. (2022). DANCE SPORTS INFLUENCE FEMALE UNIVERSITY STUDENTS'PHYSICAL HEALTH IN ETHNIC UNIVERSITIES. Revista Brasileira de Medicina do Esporte, 29(spe1), e2022_0182.

[8] Zhang, M. (2017, July). Effect Analysis of the Sports Dance Teaching on the Healthy Personality of College Students. In 2017 3rd International Conference on Economics, Social Science, Arts, Education and Management Engineering (ESSAEME 2017) (pp. 41-44). Atlantis Press.

[9] Li, Q. (2023). The Promoting Effect of Sports Dance Art Learning on College Students' Mental Health. Revista de Psicología del Deporte (Journal of Sport Psychology), 32(3), 428-436.

[10] Guo, S., Yang, X., Farizan, N. H., & Samsudin, S. (2024). The analysis of teaching quality evaluation for the college sports dance by convolutional neural network model and deep learning. Heliyon, 10(16).

[11] Yang, P. (2019, May). Practical Analysis of Sports Dance Teaching in Colleges and Universities. In 2019 4th International Conference on Social Sciences and Economic Development (ICSSED 2019) (pp. 201-204). Atlantis Press.

[12] Ge, L. (2022). RESEARCH ON THE TEACHING PRACTICE OF COLLEGE SPORTS DANCE FROM THE PERSPECTIVE OF BEHAVIORAL PSYCHOLOGY. Psychiatria Danubina, 34(suppl 4), 91-91.

[13] Guan, Y. (2014, November). Research on Sports Dancing to Healthy Development of College Students' EQ. In 2014 International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC-14) (pp. 1475-1478). Atlantis Press.

[14] Yang, F., Wu, G., & Shan, H. (2022). Real-Time Monitoring of College Sports Dance Competition Scenes Using Deep Learning Algorithms. Wireless Communications and Mobile Computing, 2022(1), 1723740.

[15] Yi, L. I. U. (2018). Impact of sport dancing on the dynamics characteristic of foot movement of college students. Revista de Pielarie Incaltaminte, 18(2), 109.

[16] Yuan, F. (2023). The Influence of Psychological Ability on the Performance of Sports Dance. Revista de Psicología del Deporte (Journal of Sport Psychology), 32(1), 352-360.

[17] Chen, Y., & Li, X. (2021, January). Research on the application of flipped classroom model in college sports dance teaching. In 2021 International Conference on Information Technology and Contemporary Sports (TCS) (pp. 508-511). IEEE.

[18] Wang, Y., Charupheng, M., & Srisawat, P. (2024). The The Influence of Sports Dance on Self-esteem Among College Students. International Journal of Sociologies and Anthropologies Science Reviews, 4(5), 487-492.

[19] Hong, Y., & Nor, N. B. M. (2023). Research on the Application of Micro-courses in Colleges and Universities Sports Dance Teaching. Frontiers in Educational Research, 6(21), 71-75.

[20] Zhang, X., & Li, Z. (2022). Investigation and analysis of the status quo of sports dance based on mobile communication. Mobile Information Systems, 2022(1), 7240810.

[21] Zhang Siqi, Jin Jie, Wang Chaofang, Dong Wenlong & Fan Bin. (2022). Quality Evaluation Algorithm for Chest Compressions Based on OpenPose Model. Applied Sciences(10),4847-4847.

[22] Wang Kaihang & Li Kexin. (2023). Research on road damage recognition and classification based on improved VGG-19.Mathematical Models in Engineering(4),115-129.

[23] Xuegang Wu, Jiawei Zhu & Liu Yang.(2024).Faster-slow network fused with enhanced fine-grained features for action recognition. Journal of Visual Communication and Image Representation104328-104328.

[24] Imen Labiadh, Larbi Boubchir & Hassene Seddik. (2024). Optimization of 2D and 3D facial recognition through the fusion of CBAM AlexNet and ResNeXt models. The Visual Computer(prepublish),1-16.

[25] Weijia Xiang, Yunru Wu, Cheng Peng, Kaicheng Cai, Hongbing Ren & Yuming Peng.(2024).A Fault Diagnosis Method for Electric Check Valve Based on ResNet-ELM with Adaptive Focal Loss. Electronics(17),3426-3426.

[26] MengJuan Han. (2024). Systematic financial risk detection based on DTW dynamic algorithm and sensor network. Measurement: Sensors101257-101257.