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

Combinatorial Press

Optimization Algorithm Based Design of Tap Dance Movement Training Path in Dance Professional Construction

Lin Fan[1], Luyao Gong[2,] ✉

[1] *Popular Music Academy, Sichuan Conservatory of Music, Chengdu, Sichuan, 610500, China*
[2] *Dance Academy, Sichuan Conservatory of Music, Chengdu, Sichuan, 610500, China*

ABSTRACT

Key frame extraction is an important research content for human motion capture data analysis and processing, for this reason, a key frame extraction method for motion capture data based on quantum particle swarm optimization algorithm is proposed, which can either extract a definite number of key frame sequences or extract key frame sequences according to the objective function. In this paper, the spatio-temporal graph convolutional network is selected as the benchmark network for tap dance action recognition, and the dance action recognition is realized by combining adaptive and attention mechanisms. The comprehensive index of tap dance is introduced and used as a constraint, and the golden section algorithm is used to optimize the training path of the dance action to obtain an ergonomic training path. The experimental results of this paper show that the key frame extraction method of motion capture data based on quantum particle swarm optimization algorithm meets the need of real-time compression of motion capture data. By constructing the validation dataset, the accuracy improvement of AAST-GAN algorithm and the effect of gesture extraction are compared and verified, and the recognition accuracy reaches more than 86%, which is a good recognition accuracy for each tap dance action. The dance movement training path proposed in this paper ensures the effectiveness and comfort of tap dance movements.

Keywords: quantum particle swarm, dance action recognition, path planning, spatio-temporal graph convolution

1. Introduction

In this fast-paced era, people increasingly need a way to relax and release stress. Dance, as an elegant and energetic art form, is gradually becoming a fitness method chosen by people. And tap dance, as a unique dance form, is loved by many people [1-4]. Tap dance can be performed in a variety of forms, such as solo performance, duo performance, team performance and so on. In social occasions, tap

✉ Corresponding author.
E-mail address: gongluyao1115@gmail.com (L. Gong).
Received 05 August 2024; Revised 25 October 2024; Accepted 15 February 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-199

dance is often used as one of the programs for entertainment and social activities, bringing joy and relaxation to people [5-8]. In formal stage performances, tap dance can be combined with other dance forms, such as ballet and modern dance, to form a unique performance style. In addition, tap dance can be combined with other art forms such as music and drama to create colorful art works [9-12].

Dancing while walking tap dance is a dance form that combines walking and tap dance, which combines the unique rhythmic sense of tap dance and the ease and naturalness of walking, so that people can feel the power of nature and the charm of dance at the same time [13-16]. Through walking while dancing tap dance, people can exercise their bodies and release pressure while enjoying the happiness brought by dance. The movement training of tap dance mainly includes the sense of rhythm, foot strength and body coordination, and dancers need to master various steps and rhythm changes through repeated training and practice [17-20]. In addition, tap dance has high requirements for dance shoes, and professional tap shoes can enhance the contact between the feet and the ground, so that dancers can better perform their skills [21-23]. Training tap dance requires patience and perseverance, and it also needs to be combined with other dance elements, such as body posture and dance movements [24-25].

Tap dance, as a traditional dance, is known for its unique rhythms and graceful steps, and is popular for its upbeat tempo, flexibility, and enthusiastic performance style. Pallares, M. et al. assessed the effects of untrained components and retention, stability, endurance, and application by initiating tap dance training with multiple novices, revealing that a precise teaching framework is useful for people who want to improve their motor performance training and discusses its limitations and directions for development [26]. Lewis,L aims to develop an appreciation for tap as a performing art and outlines the history of tap, major works, artists, and the physical and mental benefits of this dance [27]. Goldberg,T.L. states that teaching and learning tap effectively enriches the experience of both teachers and students, realizing the importance of integration of music theory, composition, etc., with the history of tap dance and advocates for a curriculum with multiple facets that is effectively implemented in K-16 programs [28]. Polascik, B. A. et al. examined the peak vertical force, average vertical foot speed, and maximum/minimum ankle angle produced by tap dancers of different levels performing the steps of the toe galley, heel galley, and other steps, indicating that performing these steps forces are associated with joint injury risk [29]. Hartley,D. provides an in-depth overview of the development of tap dance and points out that tap's rhythm and musicality, the language of learning tap, and the steps from basic steps to advanced technique guide beginning tap dancers and extend the technique of more experienced dancers as the perfect companion [30]. Silao,N. examined dance educators and tap dance experts breaks down the improvisational portion of teaching for students, reveals dance as an interdisciplinary art form, and offers suggestions for in-depth learning and providing creative lesson plans [31].

In this paper, a key frame extraction algorithm for motion capture data based on quantum particle swarm optimization algorithm is proposed to encode the human motion frame sequences with ordered integers, and then select the key frame sequences with optimal reconstruction error by using the global optimization ability and fast search ability of quantum particle swarm optimization algorithm. On the basis of ST-GCN, the adaptive module is embedded in the spatio-temporal map convolutional layer, and combined with the attention module, which further improves the recognition accuracy on the public dataset, and optimizes the model's recognition performance for tap dance movements. Finally, on this basis, the design is based on the ergonomics and tap dance comprehensive index constraints of human lower limb feature output function, using the golden section method to optimize the training path of dance movements to improve the training efficiency.

2. Recognition of tap dance movement training based on optimization algorithm

2.1. Tap Dance Movement Campaign Data Analysis

2.1.1. Motion Data Representation. The more commonly used motion capture data formats are ASF/AMC, BVH, HTR, etc. In this paper, the file format of ASF/AMC is used.The ASF file describes the skeleton model of the motion capture data as shown in Fig. 1, which defines the basic pose of the skeleton as well as the information of skeleton joints and bone segments.

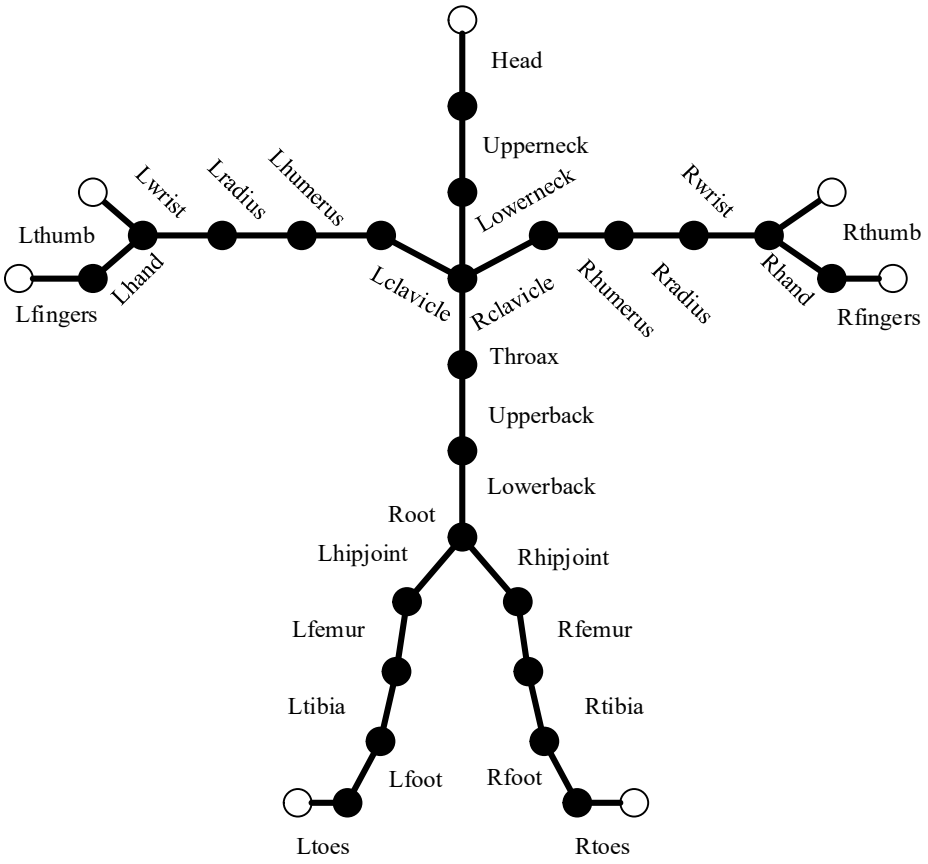


Fig. 1. ASF / AMC skeletal model

The motion capture data in the AMC file is represented as:

$$M = \{f_1, f_2, \cdots, f_{n-1}, f_n\} \tag{1}$$

where: $f_i(i = 1, 2, \cdots, n)$ denotes the motion data of each frame, i is the temporal position of the frame in the original motion sequence. And the human body motion data obtained at each sampling point is expressed as:

$$f_i = \{p(i), q_1(i), q_2(i), \cdots, q_m(i)\} \tag{2}$$

where: $p(i)$ is the coordinate of the root joint in frame i, i.e., the translation between frames; $q_j(i)$ is the rotation of bone segment j in frame i; and m denotes the number of bones per frame. Since the reconstruction interpolation used in this paper is a spherical linear interpolation of quaternions, and the original format of the motion capture data is Euler angles, it is first necessary to convert the Euler angles to the quaternion format.

2.1.2. Motion interpolation reconstruction. Motion interpolation reconstruction refers to reconstructing the original motion by means of keyframe interpolation. The simplest motion reconstruction method is linear interpolation, but the reconstruction effect of linear interpolation is poor. Considering that the goodness of the interpolation algorithm determines the result of key frame extraction and the effect of reconstruction, this paper adopts the Quaternion Spherical Linear Interpolation (Slerp) algorithm, which has a better interpolation effect, and this algorithm is able to generate smoother reconstruction frames according to the key frames, which improves the reconstruction efficiency.

If $f(t_1)$ and $f(t_2)$ are the motion data of two neighboring keyframes, $p(t_1)$ is the root coordinate position of $f(t_1)$, and $p(t_2)$ is the root coordinate position of $f(t_2)$, the interpolation of the root coordinates adopts linear interpolation in this paper:

$$p(t)=\frac{t_2-t}{t_2-t_1}p(t_1)+\frac{t-t_1}{t_2-t_1}p(t_2) \tag{3}$$

where: $t_1<t<t_2$, t are the frame numbers between moments t_1 and t_2.

If $q_i(t_1)=[w_1,x_1,y_1,z_1], q_i(t_2)=[w_2,x_2,y_2,z_2]$ is two unit quaternions and θ is the angle between them:

$$\theta=\arccos(w_1w_2+x_1x_2+y_1y_2+z_1z_2) \tag{4}$$

Then the spherical linear interpolation $q_i(t)$ between them is:

$$q_i(t)=slerp(q_i(t_1),q_i(t_2),t)=\frac{\sin(1-t)\theta}{\sin\theta}q_i(t_1)+\frac{\sin t\theta}{\sin\theta}q_i(t_2) \tag{5}$$

where: 0 < t < 1.

2.1.3. Calculation of reconstruction error. The reconstruction error of the algorithm in this paper is measured by the average inter-frame distance between the original motion sequence and the reconstructed motion sequence. For an original motion segment om containing n frame of data, the first and last frames of om are used as the key frames, and Eqs. (1)(2) are used to interpolate and reconstruct to obtain the reconstructed motion segment rm. In calculating the reconstruction error, it is necessary to take into account not only the positional error between the original motion and the reconstructed motion, but also the difference in the rates between them. Let om,rm be the original motion sequence and reconstructed motion sequence respectively, whose sequence lengths are both n. In this paper, the reconstruction motion sequence error is defined as:

$$D(om,rm)=D_p(om,rm)+uD_v(om,rm) \tag{6}$$

where: $D_p(om,rm)$ describes the positional error of the human body posture; $D_v(om,rm)$ indicates the difference of the joint rate; u is the proportion of the adjusted rate difference, according to the experiments in this paper can be seen, the difference of the rate of the proportion is very small, so it is set to 1. That is:

$$d(om^i,rm^i)=\sum_{j=1}^{m}wj\,\|\,om_j^i-rm_j^i\,\|^2 \tag{7}$$

$$D_p(om,rm)=\frac{1}{n}\sum_{i=1}^{n}d(om^i,rm^i) \tag{8}$$

$$D_v(om,rm)=\frac{1}{n}\sum_{i=1}^{n-1}|\,d(om^{i+1},om^i)-d(m^{i+1},m^i)\,| \tag{9}$$

where: $d(om^i, rm^i)$ denotes the Euclidean distance between frame i of the original motion and frame i of the reconstructed motion; w_j is the weighting value of joint j in the error calculation, which represents the importance of joint j; and om_j^i, rm_j^i is the coordinate position of the jth joint in frame i of the original motion and reconstructed motion, respectively.

2.2. Key frame extraction based on quantum particle swarm optimization algorithm

2.2.1. Coding and population initialization. In this paper, we use integer encoding, the length of the encoding is the number of target keyframes, and the value of the encoding is an integer between 1 and the total number of frames. When the population is initialized, a set of specified number of mutually different ordered integer sequences are randomly generated, and they are taken as an individual of the population, and the population size is set to 50 in this paper.

2.2.2. Definition of the fitness function. When keyframes are extracted, it is desired that the number of extracted keyframes should be as small as possible and the quality of reconstructed frames should be as high as possible. However, if the number of key frames is too small, the reconstruction error will be larger, so the two important factors for the optimal solution of extracting key frames are the reconstruction error of the motion sequence and the number of extracted key frames, i.e., minimizing the reconstruction error and optimizing the compression rate are taken as the objectives. So the defined fitness function is:

$$f(kfnum) = \frac{1}{wE(kfnum) + (1-w)R(kfnum)} \tag{10}$$

where: $kfnum$ is the number of keyframes; $w(0 \le w \le 1)$ is the user-defined weights. In the trade-off between the reconstruction error and the number of keyframes, the fewer keyframes are extracted, the greater the reconstruction error, so the fitness function consists of a weighted union of these two terms.

$$R(g) = \frac{kfnum}{totalnum} \tag{11}$$

where: $R(g)$ is the compression rate; *totalnum* is the total number of frames of the motion sequence, i.e., the ratio of the number of key frames extracted from this motion sequence to the total number of frames of the motion sequence.

$$E(kfnum) = \frac{E_{kfnum}}{E_{1,end}} \tag{12}$$

where: E_{kfnum} is the minimum reconstruction error value obtained when the number of keyframes is $kfnum$, and $E_{1,end}$ is the reconstruction error value when only the first and last frames constitute the reconstruction error value.

2.2.3. Updating the position of a particle swarm. The particle positions are updated according to Eqs. (13)~(16).

$$p = a \times Pbest(i) + (1-a) \times Gbest \tag{13}$$

$$mbest = \frac{1}{M} \times \sum_{i=1}^{M} Pbest(i) \tag{14}$$

$$b = 1.0 - \text{gen}/\max \text{ gen} \times 0.5 \tag{15}$$

$$X_{(i+1)} = p \pm b \times |mbest - X(i)| \times \ln(1/u) \tag{16}$$

where: a, u is a random number between 0 and 1; Pbest is the optimal solution found by an individual particle, i.e., the individual extreme value; Gbest is the optimal value found by the whole population, i.e., the global extreme value; M is the number of particles; mbest is the average of the individual extreme values of the quantum particle swarm; b is the contraction expansion coefficient, which is linearly decreasing during the convergence process of the QPSO; gen is the current number of evolved generations; maxgen is the set maximum number of evolutionary generations; $X(i)$ is the particle population of the i th generation.

2.2.4. Keyframe extraction based on number of settings. Given the sequence of frames and the number of keyframes, return the sequence of keyframes with the minimum reconstruction error.

keyframe=QPSO(frame, m) quantum particle swarm algorithm.

Keyframe extraction is performed with the number of keyframes determined, and optimization is performed with the objective of optimizing the reconstruction error to obtain the keyframe sequence with the minimum reconstruction error. Error is the reconstruction error value of the computed motion sequence. Particle swarm individuals are ordered integer sequences that are not identical to each other from 1 to the total number of frames. When the position of the particle swarm is updated, the position of individual particles will be out of the range from 1 to the total number of frames. In this case, their positions need to be corrected by taking 1 when the position value is less than 1, and taking the total number of frames when the position value is greater than the total number of frames. In the update the average extreme value mbest and the contraction expansion coefficient b will both appear as floating point numbers, which will result in the final key frame sequence will appear as floating point numbers, for this reason it is necessary to round up the floating point numbers to ensure that the algorithm is encoded in integer numbers.

2.3. AAST-GCN based tap dance training action recognition

2.3.1. Convolutional Networks for Spatio-Temporal Maps. The action recognition algorithm in this paper is realized on the basis of spatio-temporal graph convolution network ST-GCN for improvement, and this section will introduce the construction method of the spatio-temporal graph of the human skeleton involved in the algorithm, as well as the knowledge related to the more important spatial graph convolution in the network structure.

1) Spatio-temporal map of human skeleton. ST-GCN first solves the above problem by utilizing the human skeleton spatio-temporal map to represent the skeletal sequence. Constructing a human skeleton spatio-temporal map for a skeletal sequence with N joints and T frames is done in two steps. First, according to the physical connection of the human body structure, the joints belonging to the two ends of the same skeleton in the same frame are connected with edges, and then the same joints in consecutive frames are connected, thus completing the construction of the skeleton spatio-temporal graph.

2) Spatial graph convolution. Traditional convolution operations specialize in handling data with a grid-like topology, where both the natural image and its feature map can be viewed as a two-dimensional grid of pixels. The convolution output is shown in equation (17), where p and $_w$ are the sampling function and weight function, respectively:

$$f_{out}(x) = \sum_{h=1}^{K} \sum_{w=1}^{K} f_{in}(p(x,h,w)) \cdot w(h,w) \tag{17}$$

The output of graph convolution in ST-GCN under single-label partitioning strategy is shown below:

$$f_{out} = \Lambda^{-\frac{1}{2}}(A+I)\Lambda^{-\frac{1}{2}} f_{in} W \tag{18}$$

where A is the adjacency matrix representing the physical connectivity relationship between the joints of the human body and I is the unit matrix. $\Lambda^{ii} = \sum_j (A^{ij} + I^{ij})$, $A^{-\frac{1}{2}}(A+I)A^{\frac{1}{2}}$ denotes the normalization of the adjacency matrix, f_{in} is the input feature map, and W is the weight matrix.

For the distance partitioning and spatial structure partitioning strategies, the adjacency matrix is decomposed into multiple matrices because the set of neighbors is partitioned into multiple subsets, $A+I = \sum_j A_j$, and so Eq. (18) transforms into:

$$f_{out} = \sum_j \Lambda_j^{-\frac{1}{2}} A_j \Lambda_j^{-\frac{1}{2}} f_{in} W_j \tag{19}$$

In spatial graph convolution, ST-GCN employs a simple attention mechanism where a learnable weight matrix M is used to reflect the difference in the importance of edges in the graph convolution, replacing A_j in Eq. (19) with $A_j \otimes M$, $\otimes$ denotes the multiplication of the corresponding elements of matrices A_j and M, and the elements of M are initialized to 1. Eq. (19) transforms to:

$$f_{out} = \sum_j W_j (f_{in} \Lambda_j^{-\frac{1}{2}} A_j \Lambda_j^{-\frac{1}{2}}) \otimes M_j \tag{20}$$

2.3.2. Adaptive graph convolution module. ST-GCN uses a learnable weight matrix M to realize the attention mechanism, and corrects the degree of dependency between joints through continuous learning, so as to better model the relationship between joints, but this attention mechanism is not flexible enough, M is directly multiplied with the neighbor matrix according to its elements, and if there exists an element of 0 in the neighbor matrix A, then no matter what the value is of the corresponding element of M, the final result will be 0, i.e., M will create a connection other than the original physical connection, which cannot solve the “applause” problem mentioned above. If there is an element of 0 in the neighbor matrix A, then no matter what the value of the corresponding element of M is, the final result will be 0, i.e., it will create a connection other than the original physical connection, which will not solve the “applause” problem mentioned above.

In this paper, the graph convolutional network is improved, and the final adaptive graph convolutional layer is as follows:

$$f_{out} = \sum_j W_j f_{in} (\alpha A_j + B_j + C_j) \tag{21}$$

2.3.3. Attention module:

1) Channel Attention. The input feature dimension of graph convolutional network is $C \times T \times N$, which represents the number of channels, the number of frames, and the number of joints per frame, respectively. During the convolution process, in order to minimize the loss of information, multiple convolution kernels are often required, which will lead to an increase in the number of channels of the network, and the information contained in each channel has different degrees of importance, which is expressed by calculating the weight for the signals on each channel separately, and the larger the weight indicates that the signals on the channel are the more important, the more attention is needed [32]. Based on this, the channel attention module is introduced in this paper.

The channel attention module mainly consists of two parts: compression and excitation. Firstly, the output $f_{out} \in \mathbb{R}^{C \times T \times N}$ of the spatial map convolution is taken as the input for the compression operation, and the expression is as follows:

$$z = \frac{1}{T \times N} \sum_{i=1}^{T} \sum_{j=1}^{N} u_c(i, j) \tag{22}$$

Specifically the compression operation corresponds to a global average pooling operation that takes a C × T × N feature map and levitates it into a C × 1 × 1 feature map, an operation that serves to levitate global spatial and temporal information into the channel descriptors.

Next is the excitation part, which transforms the feature map z with the following expression:

$$s = \sigma(W_2 \delta(W_1 z)) \tag{23}$$

The incentive operation specifically contains two fully connected layers, with $w_1 \in R^{C \times \frac{c}{r}}, w_2 \in R^{\frac{c}{r} \times c}$ representing both fully connected layers.

2) Spatial attention. In this paper, a nonlocal mechanism is introduced in constructing the adjacency matrix of the graph, which initially implements spatial attention, and calculating the relationship between nodes can be understood as reweighting the matrix, and the more critical nodes will have larger weights, thus solving the problem of which nodes should be focused on in the same frame. And in this section, the quality of spatial features extracted by the network is further improved by introducing the spatial attention module.

The expression of the spatial attention module is as follows:

$$s = \sigma(w_s(AvgPool(f_{in}))) \tag{24}$$

where $f_{in} \in \mathbb{R}^{C \times T \times N}$ is the module input, AvgPool represents global average pooling in the time dimension, after pooling is completed the information in the time dimension is compressed and the feature map dimension becomes C × 1 × N. Because the number of joints is relatively small, there is no need for dimensionality reduction operations as in the case of the Channel Attention Module, w_s is a one-dimensional convolutional operation, and the dimensionality of the convolutional operation weights is 1 × C × K, and σ represents the Sigmoid activation function. The final spatial attention weight $s \in \mathbb{R}^{1 \times 1 \times N}$ is obtained.

3) Temporal Attention. The structure of the temporal attention module is the same as that of the spatial attention module, and the expression is as follows:

$$s = \sigma(w_t(AvgPool(f_{in}))) \tag{25}$$

2.3.4. Network basic units and structures. The basic unit of the network is shown in Fig. 2, which mainly contains a GCN module and a TCN module, the structure of the two modules is the same, in which the spatial map convolutional layer and the temporal map convolutional layer are used to extract the spatial and temporal features of the skeleton sequences, respectively, the batchnormal layer can alleviate the problem of gradient vanishing by executing the batch normalization operation, and the Relu activation function can enhance the model expression ability by adding the nonlinear factors to enhance the expressive power of the model. The adaptive module is located in the spatial map convolution layer.

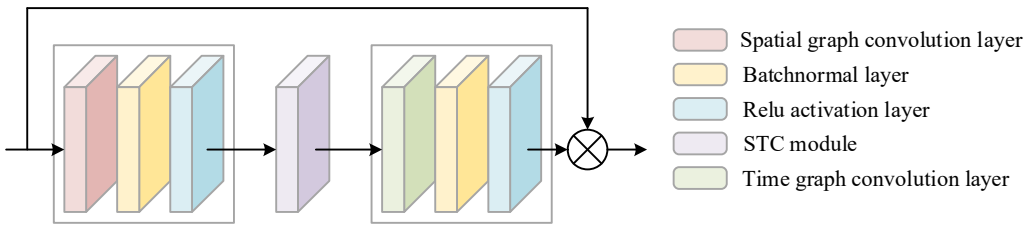


Fig. 2. Network basic unit

The overall structure of the network is a superposition of these basic units in nine layers. The number of output channels is 64 at the beginning and then the number of output channels is doubled every three layers.The BN layer normalizes the input data. While training the network, dropout method is used to remove certain nodes from the network as a way to prevent overfitting. The features are pooled by global average and fed into a softmax classifier to obtain the final classification results.

3. Tap dance movement training recognition results

3.1. Tap Dance Movement Keyframe Extraction

In this paper, Matlab2006a is used to implement the key frame extraction method for motion capture data used in the paper. The motion data used for the experiments were obtained from CMU and HDM05 motion databases with a sampling frequency of 120 frames/s. All the experiments were carried out on an IntelP42. 8GHz PC with 1GB RAM. In the experiments, 10 different types of motions such as CLIMB and JUMP were selected for testing.

3.1.1. Compression rate and maximum reconstruction error for pre-selected keytons. The compression ratios and maximum reconstruction errors of the preselected keyframes of the motion clips in the Hdm05 database are shown in Figures 3 and 4. From Fig. 3, it can be seen that the preselected keyframes of the vast majority of the motions are 60% or less of the original number of motion chastity, and only a small number of motions, such as run, rotate arms, and skier, have a larger ratio of preselected keyframes. The vertical coordinate compression ratio in Fig. 3 is calculated as the ratio of the number of preselected key frames to the number of original motion frames. From Fig. 4, it can be seen that the maximum reconstruction error of the preselected keyframes for most of the motions is less than 11 cm (110 mm), and most of them are concentrated in less than 5.5 cm (55 mm). The pre-selected keyframes strategy extracts the “boundary” poses as candidate keyframes on the one hand, and reduces the computation amount for selecting the keyframes that finally meet the requirements on the other hand.

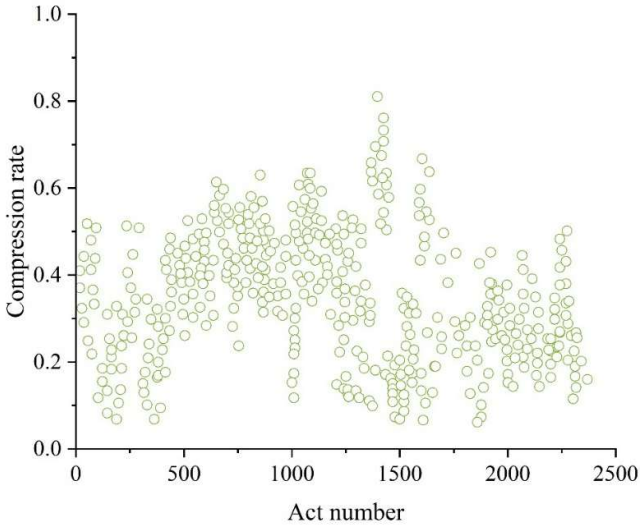


Fig. 3. Compression rate for each motion segment

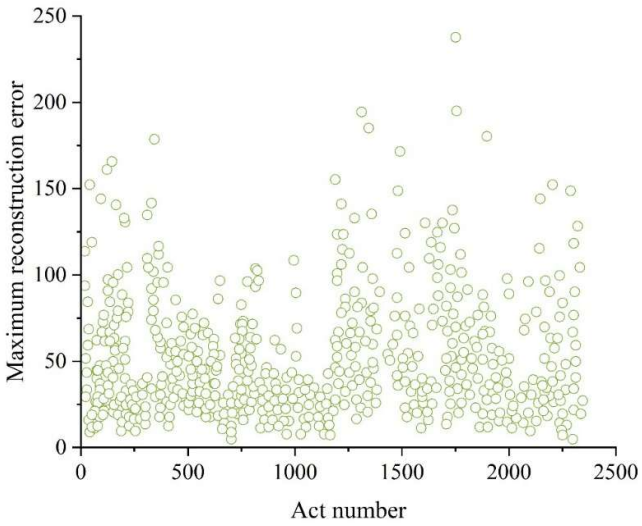


Fig. 4. Maximum reconstruction error for critical tons

3.1.2. Reconstruction Error Based Keyframe Extraction. Given the maximum reconstruction error thresholds of 11-105 (mm), respectively, the results of compression ratios, average errors, and processing rates are analyzed for the final keyframes of these 10 types of motions at a step size of 11 mm. Figure 5 shows the variation of the compression ratio distribution. When the maximum reconstruction error of each joint is required to be no more than 11mm, the compression ratios of these 10 types of motions vary from 4.7 times to 17.9 times, which is related to the type of motion, and the compression ratios of all types of motions will be increased with the relaxation of the requirement of the maximum reconstruction error, the siting, climbing, and lie down floor motions are slow to change so they have a higher compression ratio, the The cartwheel, jump and run movements are intense and have low compression ratios. It is generally believed that the human eye has difficulty in detecting the maximum reconstruction error of less than 2cm, so the optimal compression ratio can be obtained by setting the maximum reconstruction error at 20mm.

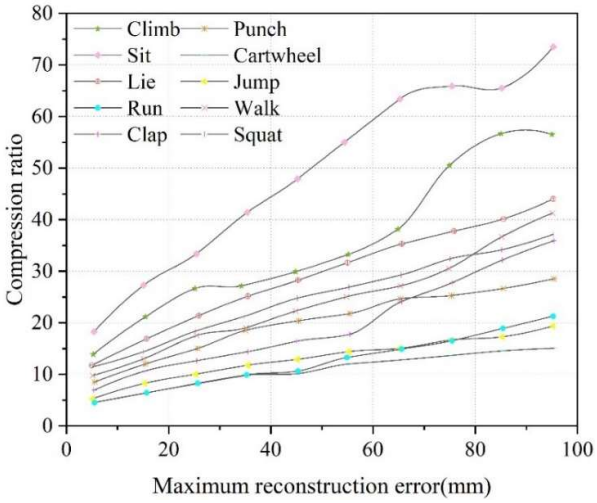


Fig. 5. Schematic of the compression ratio of maximum reconstruction error

3.1.3. Key frame extraction based on quantum particle swarm optimization algorithm. Table 1 shows the results of comparing the method of this paper with the latest method among the traditional methods of keyframe extraction. The method uses the average reconstruction error as the evaluation criterion for frame reduction and obtains the optimal number of keyframes (columns) based on the reconstruction error curve of the reduced frames. In this paper, the same number of keyframes are extracted by using the keystroke extraction method based on two-way decimation and splitting strategy. From the table, we can see that the average error of this method is close to that of the traditional method, and the maximum error is slightly lower than that of the traditional method, but the traditional method extracts the optimal keycalls without considering the reconstruction effect, and the maximum reconstruction error is more different. Since the traditional method eliminates each frame one by one, and then calculates the average reconstruction error of different numbers of key frames to obtain the "optimal" number of key tilts, the processing rate is slow, and the processing rate of Quantum Particle Swarm Optimization algorithm is much higher than that of traditional method, which is suitable for real-time compression and key frame extraction.

Table 1. Comparison of the performance of the key-dun extraction algorithms

Action	Length	Initial key frame	Optimal key frame	Maximum reconstruction error (mm)		Mean reconstruction error (mm)		Execution time (s)	
				Traditional method	This article	Traditional method	This article	Traditional method	This article
Climb	459	253	21	39.55	25.18	6.49	7.81	18.86	1.86
Jump	829	515	36	217.47	153.11	22.72	30.17	53.37	2.61
Punch	965	752	55	62.71	43.73	6.64	8.65	71.02	3.15
Run	919	434	37	137.49	170.28	26.62	30.49	64.29	2.69
Sitt	1789	1268	66	40.18	22.51	5.45	6.69	228.87	4.95
Walk	658	386	27	64.47	61.09	9.36	10.74	34.35	2.23
Cartwheel	545	173	27	177.86	157.65	20.01	26.32	25.17	1.75
Cleap	182	75	19	25.19	14.72	3.63	4.71	5.53	1.12
Lie	904	164	37	67.88	49.41	8.29	10.03	62.39	2.34
Squat	436	115	26	57.52	31.73	8.72	10.1	18.25	1.52

3.2. Recognition effects of tap dance movement training

In this paper, we take tap dance action poses as an example, and a total of six actions need to be recognized: touch step, stomp step, stutter step, forward wipe step, backward wipe step, and walk step, and a total of 300 video sequences are captured, which contain 1532 action sequences of multiple human targets at different viewpoints and different distances. The parameter information of network training is shown in Table 2. The accuracy and loss of the training process are shown in Fig. 6 and Fig. 7.

Table 2. Information on the training parameters

Parameter name	Content
Action category	6
Basic learning rate	0.001(The 20-0f 80th is 1000%, and 130 is reduced by 10%)
Batch size	32
epoch	150
Dropout	0.5

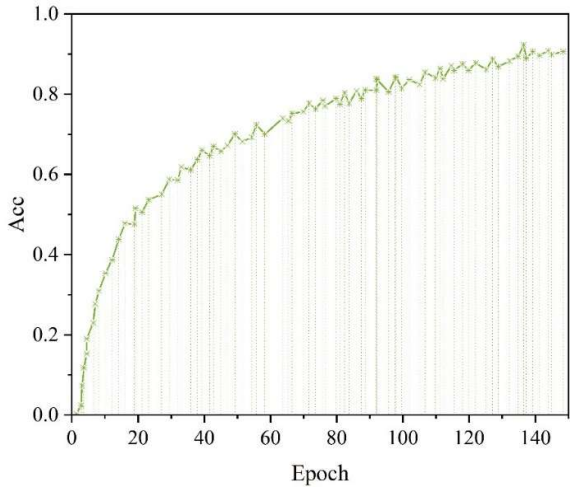


Fig. 6. Training accuracy diagram

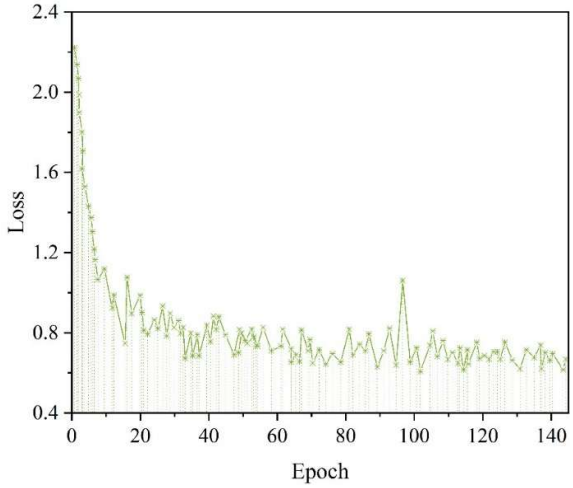


Fig. 7. Plot of the training loss

In order to verify the precision of the AAST-CCN action recognition method in the task of dance action recognition, this paper chooses to take recall, precision and F1 as the evaluation indexes, and

test the trained AAST-GCN model on the test dataset, and the specific precision data are shown in Table 3.

As can be seen from the table, each action has a good recognition precision, especially the two simpler actions of touch step and stomp step, the recognition precision reaches more than 86%. The other four dance poses are more complex, so the accuracy is lower than the first two. The horizontal water shooting action has body occlusion in some viewpoints, resulting in the lack of gesture information, so its recognition accuracy is significantly lower than the other actions. In summary, this paper uses spatio-temporal graph convolutional network for tap dance action recognition, which has high accuracy and can well recognize the target's action based on the dancer's temporal pose information.

Table 3. Verification of action recognition accuracy

Action category	Precision	Rcall	F1
Touch the step	89.57%	84.77%	86.97%
Stamp	88.31%	82.29%	84.30%
A step	84.28%	81.41%	82.81%
Before the brush step	82.18%	79.21%	80.68%
After the brush step	78.39%	71.87%	74.63%
walk with the ball	82.77%	76.51%	79.54%

4. Optimization of tap dance movement path based on comprehensive indexes

4.1. Path planning based on the output of human lower limb motion characteristics

4.1.1. Limb Coordination Tap Dance Lower Limb Motion Characterization Output Based on Limb Coordination. In order to realize the recognition of dance movement poses, training path planning should be performed based on the human motion feature outputs, and the accuracy of the selected human motion feature outputs reflecting the human movements of the subjects determines the effectiveness of the path planning. The selected human motion feature output should be able to be expressed as a function of time and can describe the motion state of the human body. The regular human lower limb motion feature outputs are the joint angles of the hip, knee, and ankle joints, and for the subject's leg lifting action under the tap dance action state in this paper, the sagittal plane knee joint angle is selected as the human motion feature output to describe the rehabilitation action performed by the subject as shown in Figure 8.

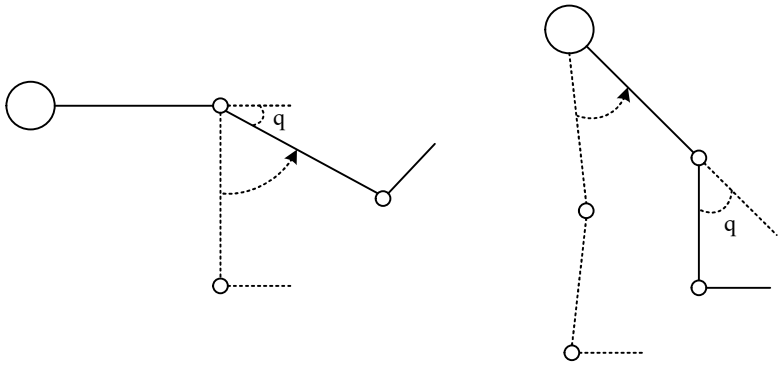


Fig. 8. Posture output of human lower limbs

4.1.2. Normalized output function based on the output of motion features. After accumulating the knee joint data of the leg lifting action in the tap dance action state of the subjects, it was homogenized

to prevent data mutation in a single data in the test, and the processed data was used for action training path planning. Fourier fitting was performed on the collected output information of human motion characteristics of the subjects, i.e., the knee joint angle, which was expressed as a continuous function with respect to time, in order to seek the regularity of the lower limb tap dance action, and the normalized output function obtained from the fitting was as follows:

$$y(t) = a_0 + a_1 \cos(\omega t) + b_0 \sin(\omega t) \tag{26}$$

where $y(t)$ denotes the output value of the human knee angle, t denotes time, and Equation (26) is the normalized output function of the human lower limb output with respect to time.

4.2. Optimization of tap dance movement paths with integrated metrics

4.2.1. Tap dance composite indicators. Fitts' law is a model that describes the movement characteristics, movement time, movement range and movement accuracy during human movement. In this section, the Fitts' theorem and the human joint angle activity threshold are utilized to design a comprehensive rehabilitation index for human uni-joint trajectory planning of the lower extremity in the context of the lower extremity unilateral leg, as shown in (27):

$$\begin{cases} J_d = J_1 + J_2 \\ J_1 = \lambda_1 \log_2(q_\omega / 2y(t)) \\ J_2 = \lambda_2 \cos((100 - y(t))/10) \end{cases} \tag{27}$$

4.2.2. The golden section algorithm. Selecting the joint angle within the joint activity threshold that makes the comprehensive index of tap dance take the extreme value needs to be calculated by a suitable algorithm for searching the optimal value, and the golden section algorithm is used in this experiment [33]. The golden section algorithm, also known as the 0.618 method, makes the search interval containing the extreme value points shrink continuously by comparing the function values of the trial points in order to realize the search for the extreme value.

The golden section algorithm needs to determine the search interval $[a_0, b_0]$ and the allowable error $h > 0$, and calculate the initial test points p_0 and q_0, which are selected as shown in (28):

$$\begin{cases} p_0 = a_0 + 0.382(b_0 - a_0) \\ q_0 = a_0 + 0.618(b_0 - a_0) \end{cases} \tag{28}$$

where the function values of initial test points p_0 and q_0 are $f(p_0)$ and $f(q_0)$ and set $i := 0$. The specific steps of the golden section algorithm are as follows:

Step 1: Judge the size relationship between $f(p_i)$ and $f(q_i)$, if $f(p_i) \le f(q_i)$ then proceed to the second step, otherwise proceed to the third step.

Step 2: Calculate the left test point, if $|q_i - a_i| \le h$ then stop calculating and output p_i, otherwise use formula (29), calculate the value of $f(p_{i+1})$, update the value of i to $i := i + 1$, turn to step one.

$$\begin{cases} a_{i+1} := a_i \\ b_{i+1} := q_i \\ f(q_{i+1}) := f(p_i) \\ q_{i+1} := p_i \\ p_{i+1} := a_{i+1} + 0.382(b_{i+1} - a_{i+1}) \end{cases} \tag{29}$$

Step 3: Calculate the right trial point, if $(b_i - p_i) \le h$ then stop the calculation and output q_i, otherwise use equation (30), calculate the $f(q_{i+1})$ value and update the i value to $i := i + 1$, turn to step one.

$$\begin{cases} a_{i+1} \coloneqq p_i \\ b_{i+1} \coloneqq b_i \\ f(p_{i+1}) \coloneqq f(q_i) \\ p_{i+1} \coloneqq q_i \\ q_{i+1} \coloneqq a_{i+1} + 0.618(b_{i+1} - a_{i+1}) \end{cases} \tag{30}$$

In this experiment, the confidence interval of the tap dance action pose training path is used as the search interval of the golden sectioning algorithm, and the comprehensive index of tap dance is used as the objective function of the search, and the golden sectioning algorithm continuously searches for the joint angle that makes the comprehensive index of tap dance take the extreme value within the confidence interval, which is the output of the optimized human dance action features based on the comprehensive index of tap dance.

4.2.3. Optimized normalized output function. The original normalized output function is optimized based on the tap dance composite index through the golden section algorithm, the optimized joint angle values are used as the new human lower limb outputs, and the new human kinematic feature outputs are fitted to the optimized normalized output function through Fourier fitting, as shown in (31):

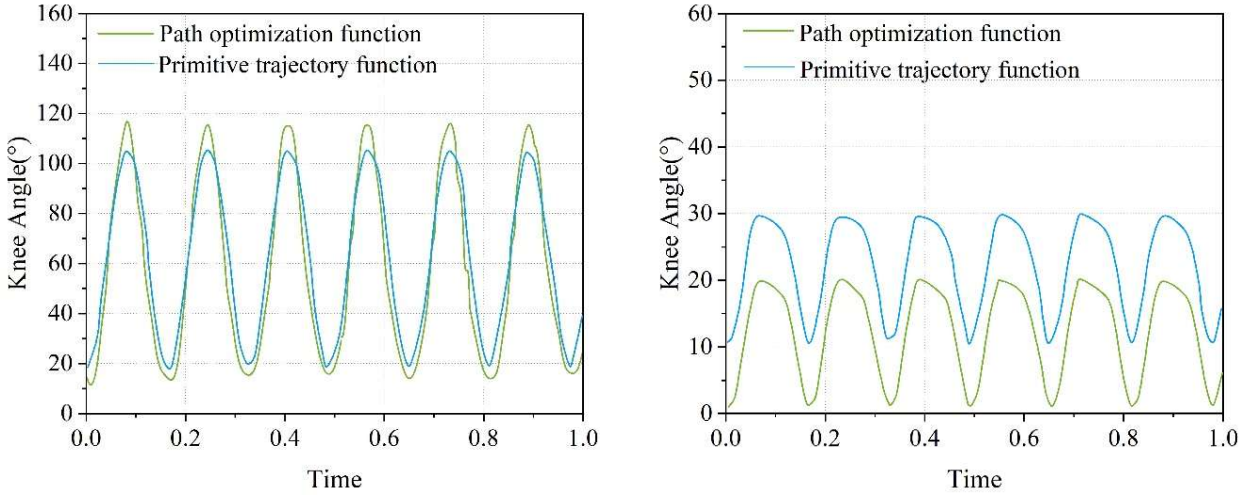
$$y(t) = a_0 + a_1 \cos(\omega t) + b_0 \sin(\omega t) \tag{31}$$

where $y(t)$ is the optimized human knee angle output value, t denotes time, and Equation (31) is the normalized output function of the optimized human lower limb output with respect to time.

4.3. Simulation Experiments

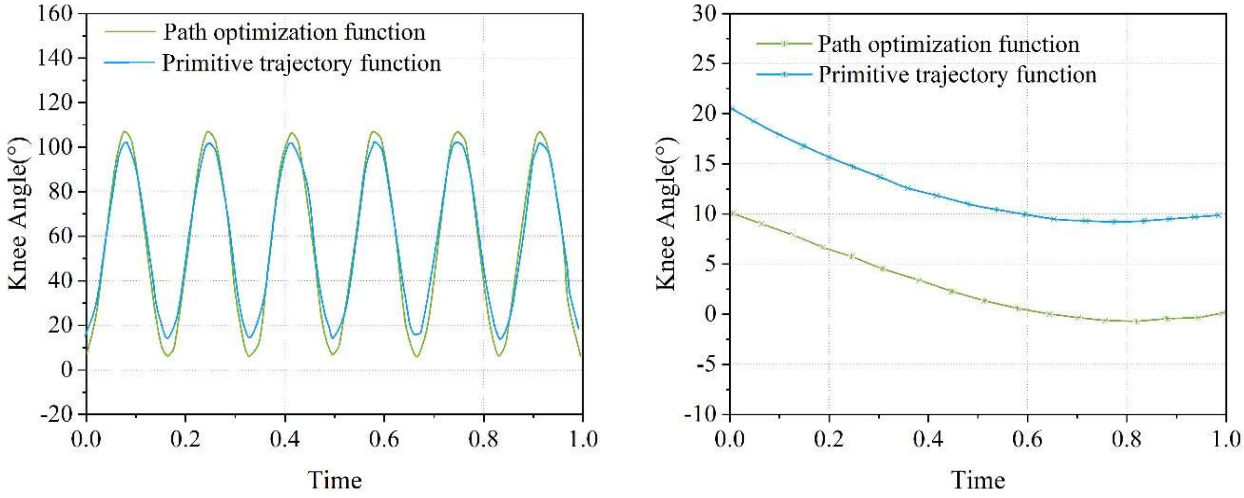
In order to make the subject dance movement can be based on the current tap dance movement training difficulty of the lower limb training optimal path, improve the training efficiency, to ensure the effectiveness and comfort of the lower limb training, based on the human body movement characteristics output function and tap dance comprehensive index using the golden section algorithm model of the path optimization of the three tap dance movement training task.

As the training proceeds, the muscular endurance and control of the subject's lower limbs have been improved to a certain extent, and it is necessary to increase the training difficulty for the patient's damaged muscle groups to carry out sufficient stimulating training, to set the complexity weighting coefficient λ_1 to 0.8 and the comfort weighting coefficient λ_2 to 0.2, for example, the weighting coefficients are substituted into the time-varying nonlinear constrained optimization problem to solve the optimization of the path of the lower limb dance movement obtained after the solution is as shown in Fig. 9 ((a) and (b) for flexion of knee joint and extension of leg joint respectively) and Figure 10 ((a) and (b) for knee joint abduction and leg joint adduction respectively) are shown, when the subject is performing the knee flexion-extension movement, the position of the knee joint in the initial state is adjusted, and the joint range of motion of knee flexion and extension is increased based on the constraints of the comprehensive index of tap dance under the premise of guaranteeing the safety of the subject's dancing movement, which makes the affected side of the knee joint muscles such as the biceps brachii receive stimulating training that is more adapted to the patient's current state, improving the actual training efficiency. The experimental results show that through the training trajectory, the subjects can obtain better effectiveness and comfort during the dance training process.



(a)Knee joint in flexion-extension (b)Leg joint flexion-extension

Fig. 9. Knee flexion-extension exercises



(a)Knee joint abduction-adduction (b)Leg joint abduction-adduction

Fig. 10. Knee abduction – adduction movement

5. Conclusion

In this paper, we propose a key frame extraction method for motion capture data based on quantum particle swarm optimization algorithm, and then use the spatio-temporal graphical convolutional network AAST-GCN based on adaptive and attentional mechanisms to recognize tap dance movements. The original normalized output function is obtained by Fourier fitting of the human motion feature output, and based on ergonomics, a comprehensive index of tap dance is introduced to optimize the training path of dance movements. Finally, the golden segmentation algorithm is used to find the knee joint angle that makes the tap dance comprehensive index take the extreme value within the confidence interval of the human motion feature output, and the Fourier fitting is performed on the finding result to obtain the optimized normalized output function. The results show that the key frame extraction method for motion capture data based on the quantum particle swarm optimization algorithm can efficiently propose key frames that meet the requirements from the original motion

sequence, which better meets the needs of real-time compression of motion capture data. The accuracy of recognizing tap dance movements using the AAST-GCN method reaches more than 86%. Using the golden segmentation of the lower limb output data constrained by the comprehensive index of tap dance, we obtain the optimal path for lower limb movement training that can independently adjust the training difficulty according to the current training stage of the subject, which improves the training efficiency and ensures the effectiveness and comfort of the training.

References

[1] Hanna, J. L. (2017). Dancing to resist, reduce, and escape stress. The oxford handbook of dance and wellbeing, 99.

[2] Gurusathya, C. (2019). Dance as a catalyst for stress busting. Central European Journal of Sport Sciences and Medicine, 26, 15-29.

[3] Leng, P. G., Mak, F. M., Saearani, M. F. T., & Sampurno, M. B. T. (2025). The Effectiveness of Improvisational Dance in Decreasing Stress Levels among Tertiary Students in Malaysia. Arteterapia, 20, e97816.

[4] Rocha, P., McClelland, J., Sparrow, T., & Morris, M. E. (2017). The biomechanics and motor control of tap dancing. Journal of Dance Medicine & Science, 21(3), 123-129.

[5] Wang, Q., & Zhao, Y. (2021). Effects of a modified tap dance program on ankle function and postural control in older adults: a randomized controlled trial. International Journal of Environmental Research and Public Health, 18(12), 6379.

[6] Zhao, Y., Cai, K., Wang, Q., Hu, Y., Wei, L., & Gao, H. (2021). Effect of tap dance on plantar pressure, postural stability and lower body function in older patients at risk of diabetic foot: a randomized controlled trial. BMJ Open Diabetes Research and Care, 9(1), e001909.

[7] Murgia, M., Prpic, V., O, J., McCullagh, P., Santoro, I., Galmonte, A., & Agostini, T. (2017). Modality and perceptual-motor experience influence the detection of temporal deviations in tap dance sequences. Frontiers in Psychology, 8, 1340.

[8] Schroeder, J. (2021). Tap Dance and Cultural Memory: Shuffling with My Dancestors. Cultural Memory and Popular Dance: Dancing to Remember, Dancing to Forget, 25-38.

[9] Michiels Hernandez, B. L., Ozmun, M., & Keeton, G. (2013). Healthy and Creative Tap Dance: Teaching a Lifetime Physical Activity. Journal of Physical Education, Recreation & Dance, 84(4), 29-40.

[10] Purvis, D. (2017). Meter and Musicality in Tap Class. Journal of Dance Education, 17(4), 158-161.

[11] Wagoner, R. (2017). Rhythm in shoes: student perceptions of the integration of tap dance into choral music. Voice and Speech Review, 11(3), 279-295.

[12] Wilson, L., & Moffett, A. T. (2017). Building bridges for dance through arts-based research. Research in Dance Education, 18(2), 135-149.

[13] Morrison, M. (2024). Tapping the Margins: Feminist Research in Tap Dance History. Dance Chronicle, 1-26.

[14] Hernández, O. (2021). Continuation and the Break: Notes on Black Lives and Tap Dance. Brooklyn Rail.

[15] Crawford-Shepherd, S. (2022). Independent beats to global feet: the evolution of English tap jams. Theatre, Dance and Performance Training, 13(1), 22-33.

[16] Strâmtu, A. (2019). THE FINALITIES OF TEACHING-LEARNING TAP-DANCE TO ACTORS. Educația Plus, 22(1), 43-50.

[17] Jin, X., Wang, B., Lv, Y., Lu, Y., Chen, J., & Zhou, C. (2019). Does dance training influence beat sensorimotor synchronization? Differences in finger-tapping sensorimotor synchronization between competitive ballroom dancers and nondancers. Experimental Brain Research, 237, 743-753.

[18] Kendall Lynch, P. T. (2022). Hip Pointer in a Professional Tap Dancer: A Case Report. Orthopaedic Physical Therapy Practice, 34(2), 111-118.

[19] Heins, N., Pomp, J., Kluger, D. S., Vinbrüx, S., Trempler, I., Kohler, A., ... & Schubotz, R. I. (2021). Surmising synchrony of sound and sight: Factors explaining variance of audiovisual integration in hurdling, tap dancing and drumming. Plos one, 16(7), e0253130.

[20] Heins, N., Pomp, J., Kluger, D. S., Trempler, I., Zentgraf, K., Raab, M., & Schubotz, R. I. (2020). Incidental or intentional? Different brain responses to one's own action sounds in hurdling vs. tap dancing. Frontiers in neuroscience, 14, 483.

[21] Willis, C. M. (2023). Black Tap Dance and Its Women Pioneers. McFarland.

[22] Thomas, S. (2017). Articles Educated Feet: Tap Dancing and Embodied Feminist Pedagogies at a Small Liberal Arts College. Feminist Teacher, 27(2-3), 196-210.

[23] Hill, C. V. (2021). Brotherhood in Rhythm: The Jazz Tap Dancing of the Nicholas Brothers. Oxford University Press.

[24] Ormonde, J. (2017). Tap dancing at a glance. Read Books Ltd.

[25] Giguere, M. (2023). Beginning modern dance. Human Kinetics.

[26] Pallares, M., Newsome, K. B., & Ghezzi, P. M. (2021). Precision teaching and tap dance instruction. Behavior Analysis in Practice, 14, 745-762.

[27] Lewis, L. (2023). Beginning Tap Dance. Human Kinetics.

[28] Goldberg, T. L. (2018). Release, relax, ready: A rhythm-first, holistic approach to teaching tap dance. Dance Education in Practice, 4(3), 17-24.

[29] Polascik, B. A., Jiang, Y., & Schmitt, D. (2024). Step type is associated with loading and ankle motion in tap dance. Plos one, 19(5), e0303070.

[30] Hartley, D. (2018). The Essential Guide to Tap Dance. The Crowood Press.

[31] Silao, N. (2021). RIFFS: A Theory of Instruction and Reference for Step Progression in Tap Dance. Dance Education in Practice, 7(3), 6-17.

[32] Su Yang,Wang Jun,Wu Peng,Wu Chengke,Yue Aobo & Shou Wenchi. (2025). Automatic Completion of Underground Utility Topologies Using Graph Convolutional Networks. Journal of Computing in Civil Engineering(1),

[33] Fatma Zohra Zoubiri,Rachida Rihani & Fatiha Bentahar. (2020). Golden section algorithm to optimise the chemical pretreatment of agro-industrial waste for sugars extraction. Fuel117028-117028.