ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

The Protection, Inheritance and Development of Folk Dance in the Environment of Modern Technology

Simeng Gao[1,2], *Peera Phanlukthao*[2] *and Kai Guo*[3,*]

[1] Henan Postdoctoral Innovation Practice Base, Luoyang Vocational College of Science and Technology, Luoyang, Henan, 471000, China
[2] The School of Fine and Applied Arts, Mahasarakham University, Mahasarakham, 30100, Thailand
[3] School of Economics and Management, Luoyang Vocational College of Science and Technology, Luoyang, Henan, 471000, China

SUMMARY: *Traditional folk dance represents a significant form of intangible cultural legacy in China, The use of modern technology to identify and capture folk dance movements facilitates their understanding and inheritance. This paper collects folk dance motion data via Kinect to build a typical dance movement dataset, and adopts an attention-enhanced spatio-temporal graph convolutional network to implicitly learn skeletal sequence features for dance motion recognition. Furthermore, it proposes a sensor-based auxiliary training method. By constructing a 3D human skeleton joint model, the system reconstructs learners' movements to assess positional accuracy, thereby enabling assisted dance training and supporting heritage preservation.The test results show that compared with other motion recognition systems, this system has the smallest error between the motion recognition angle and joint positioning and the Kinect standard value, and the recognition accuracy rate is as high as 99.1%. A thorough examination validates that the suggested sensor - based supplementary training approach allows for the precise acquisition and automated identification of folk dance motions, which attests to its efficacy.*

KEYWORDS: *attention mechanism; Spatial-temporal graph convolution network; action recognition; motion capture sensor; folk dance*

1 Introduction

In the past few years, as the creative adaptation and innovative advancement of outstanding traditional Chinese culture have been vigorously promoted, remarkable ethnic dances have made the shift from the traditional style to the modern one. Contemporary technology has presented a brand - new chance for the creative adaptation and innovative progress of ethnic dance culture [1]. Novel technologies expedite the combination with dance, facilitating the conversion of ethnic dance resources from being "digitized" to becoming "assets", and from being confined to the "cultural sphere" to reaching the "social sphere"[2, 3]. However, the rapid development and evolution of new technologies have also brought many new problems and dilemmas to the industry and communication ecology of folk dance. The excessive reliance on digital technology by the technology-first instrumental pragmatism, the one-sided pursuit of efficiency and convenience, and the resulting fragmentation, superficiality, and transient

*18604043426@163.com
https://doi.org/10.65102/is2026572
Published: 30 April, 2026

characteristics have led to the digital development of folk dance becoming a lifeless cultural specimen detached from the ecology [4]. Based on this, how digitalization empowers the new scene of "use for protection" of folk dance culture, how to realize the symbiosis between reality and reality of multilayered narratives in the cultural space, and how to bridge the development and dissemination of the heritage of folk dance is an urgent problem to be solved.

Since the 1990s, folk dance has undergone digital conservation and research on a global scale, amassing a substantial body of literature. This article delves into its conservation, inheritance, and advancement in the context of modern technology, specifically focusing on motion capture, 3D, VR, AR, and other immersive technologies. Motion capture, initially utilized by Disney in the 1970s, is now extensively employed in 3D animation production. This advanced technology employs contemporary technological devices to monitor the location of every movement of an individual in motion within a three - dimensional environment. It then documents the location of each joint segment in the format of numerical coordinates[5-7]. This data is formed by recording the human body movements through sensing devices and animating the human body movements to a computer. Reference [8] centers on the utilization of motion capture technology in dance visualization and methodically delineates the fundamental principle of this technology, based on the conversion of three - dimensional animation data, it formulates a comprehensive technical roadmap spanning from dynamic capture to the creation of virtual scenes. Through model building, costume crafting, and scene integration, it achieves the digital representation and intricate visual manifestation of dance art. Moreover, it offers technological backing for the inheritance, teaching, and innovative performances of dance.It provides technical support for dance inheritance, teaching and innovative performance. Literature [9] on the basis of Kinect motion capture technology, constructed a three-dimensional motion data acquisition system suitable for modern dance, through the integration of body movement and facial expression capture, realized the dance action data and animation model of the efficient drive; the study shows that this technology can be employed not only for the digital safeguarding and transmission of dance art but also can yield positive social and economic gains in various sectors such as animation, film - and - TV production, and virtual reality. Research indicates that it has the capacity to preserve and pass on dance art digitally while also generating favorable social and economic outcomes in the areas of animating, creating films and TV shows, and the realm of virtual reality. Reference [10] put forward an approach for the intelligent auxiliary training of stage performances. This approach is tailored to the distinctive cultural traits of ethnic minority dances. It employs human motion capture technology, extracts human body regions in the video via human posture estimation, and integrates 3D - SIFT and optical flow characteristics for multi - dimensional feature extraction.Reference [11] delves into the fundamental principles and distinctive characteristics of optical motion capture technology. It assesses the benefits of this technology in the digital production of dance, including the ability to record with high precision, restore dynamic details, and conduct digital editing. Additionally, it devises a workflow for digitizing traditional dance based on motion capture. The feasibility and effectiveness of this workflow are then validated through practical examples.

If one intends to utilize computer motion capture technology to gather, document, and transmit three - dimensional data of ethnic dances, it is necessary to create three - dimensional character models of the dancers using modeling software. Additionally, the clothing designs with ethnic features of the dancers need to be simulated. This is done to more realistically replicate the details of the dance movements that are characteristic of ethnic minorities[12]. Subsequent research delves into the practical worth of 3D technology for the preservation of dance heritage. Reference [13] evaluates its use in safeguarding ethnic minority dances in Southwest China. Case studies on Tibetan dances illustrate that 3D technology can efficiently

document and examine the nuances of dance. Moreover, it enhances the sense of immersion and interactivity in cultural inheritance. This technology presents novel approaches for the digital conservation, instruction, promotion, and cross - generational spread of ethnic minority dances. A research in literature [14] introduces a 3D character modeling approach through multi - view stereo reconstruction. When integrated with motion capture technology, this method propels the 3D model to create dance animations. As a result, it facilitates the low - cost, user - friendly, and high - quality digital conservation and distribution of ethnic dances. Reference [15] delves into the geographical distinctiveness of Chinese folk dance. It amalgamates 3D digital technology with human gesture recognition technology, putting forward a folk dance learning approach that incorporates small - sample learning. Through human body detection and tracking technology, the dance movement data is gathered. This data is then incorporated into the AAM model for 3D digital modeling. Moreover, streaming sorting is employed to preserve the characteristic details of dance movements. Reference [16] contends that the swift advancement of 3D technology offers novel methods for the conservation, renovation, and presentation of both tangible and intangible cultural heritage, which can not only record architectural and cultural relics, but also effectively preserve living cultural heritage, such as performing arts and handicrafts, and provide key support for its inheritance in the context of globalization and population mobility.

The immersive technology mainly based on virtual reality and augmented reality can record folk dance performances, costumes, props, and stages in a 360-degree panorama so that users can walk freely and explore, and obtain a sense of space and presence beyond the traditional video. A study [17] delves into the multi - sensory, immersive, interactive, real - time, and conceptual characteristics of virtual reality (VR) technology. It also examines how this technology can play a role in strengthening the dissemination of Manchu dance culture, as well as expanding its protection, inheritance, and innovation. Another piece of literature [18] approaches the use of VR and digital dance image technology (DDIT) from cultural, ethical, educational, and technological viewpoints. It discusses how these technologies can be used for the preservation, display, and spread of ethnic dances. Moreover, it constructs a platform that combines tradition and innovation to promote the sustainable development of dance art. A research [19] looks into the influence of VR on the teaching of Shanxi folk dance. The findings indicate that it enhances skill training and cultural immersion. The study also recommends integrating VR into the curriculum to modernize heritage education and draw in younger inheritors. Literature [20] developed a mobile augmented reality application prototype to complete dance digitization by recording professional folk dancers' performances and generating selectable virtual character models simultaneously.

This research paper commences by examining the traits of ethnic dances. It hand - picks 20 representative dance movements and employs the Kinect device to gather skeletal data, which is utilized to build a dataset for dance movements. Subsequently, it takes the Spatio - Temporal Graph Convolutional Network (ST - GCN) as the baseline network. Additionally, an adaptive module is incorporated to fine - tune the graph topology and forge connections between joints at a greater distance. In the meantime, an attention mechanism is incorporated to enable the network to concentrate on crucial features and enhance recognition efficiency. Comparative experiments confirm the efficacy of the enhanced model. Moreover, a self - service dance training approach relying on motion recognition and capture sensors is put forward. This approach constructs a motion database and a 2D human skeletal joint model from video frames,convert the two-dimensional model to a three-dimensional model using the matrix BVH parsing algorithm, select the sensors to perform the motion capture, filter and normalize the results, and based on the captured motion values, incorporate them into the human body movement model and compare them with the movement database by calculating the point - to

- point distances against the standard actions stored in the database. Ultimately, the motion capture sensor put forward in this paper is employed to carry out systematic test experiments on the folk dance training approach.

2 Digital preservation of folk dance in a modern technological environment

2.1 Folk Dance Typical Movements Collection and Data Set Construction

2.1.1 Study of Typical Movements in Folk Dance

1) Characteristics of folk dance movements

In this research paper, five distinct types of ethnic dances have been chosen from a wide array of ethnic dance forms. These include the Dai dance, Tibetan dance, Viennese dance, Mongolian dance, and Hmong dance. These five ethnic dances are frequently utilized in ethnic dance instruction. Their movement features and styles have been summarized, and the movement characteristics of these five ethnic dances are as follows: (1) Dai dance: Among the numerous ethnic groups in China, the Dai people are an ethnic group boasting a rich and long - standing history and cultural heritage.

(1) Dai Ethnic Dance: In China, which is home to numerous ethnic minorities, the Dai people are an ethnic group boasting a rich historical background and a well - established cultural tradition. Their dance, as a significant means of cultural heritage, has been handed down through the ages. This has led to the formation of a distinct national cultural style and artistic features unique to the Dai ethnic group.

(2) Tibetan Dance: A Cultural Art Form of the Tibetan CommunityTibetan dance is a long - standing traditional dance style practiced by the Tibetan ethnic group. Shaped by the unique geographical characteristics of the plateau region, the power and dynamism of Tibetan dance are predominantly manifested in the lower part of the body. Dancers achieve an aesthetic appeal by precisely and rhythmically flexing and extending each joint.In Tibetan dance, the body's center of gravity is positioned forward. At the same time, the dancers let their arms hang down naturally by their sides, creating a distinct and captivating visual effect.

(3) Uyghur Dance: Uyghur Dance is a typical folk performing art of the Uyghur people after different stages of development. The Uyghur dance features an erect and curvaceous form. Movements are distinctively displayed in various parts of the body, namely the head, neck, shoulders, chest, waist, and feet.

(4) Mongolian Dance: Mongolian Dance is characterized by a large range of movements and a fast rhythm.

(5) Miao Dance: In this research paper, the focus is on the Jinji Dance, a long - standing folk dance hailing from Qiandongnan, Guizhou. The golden pheasant, which serves as the totem of the Miao ethnic group, is a symbol through which the Miao people convey their gratitude and reverence for their ancestors via this dance.

2) Selection of typical movements of ethnic dances

After editing the collected typical movements of the five types of folk dances, we made movement example videos for the preparation of the subsequent movement acquisition experiments. The 20 clips of typical movements were screened, with 4 movements for each category of folk dance.

2.1.2 Typical movement dataset construction for folk dance

1) Data preprocessing

During the experiment, the absent joints in the initial skeleton data were replenished, and the coordinates of the missing joints were substituted with 0. For the missing frames of the movement sequence in the experiment, the missing skeleton data of a certain frame was filled with the skeleton data of the previous frame.The Kinect device records sequences of the skeletal movements in traditional folk dances at a rate of 30 frames per second. Considering that the movements in these dances are usually slow and drawn - out, we sample one frame out of every five. This approach helps to eliminate redundant data.

2) Data set structure

The data set comprises 1,000 skeletal specimens that encompass five ethnic dance forms, featuring a total of 20 movement classifications. Each of these classifications consists of 30 specimens. The skeletal data are saved in TXT files, where each file documents a full - fledged movement carried out by an individual subject, which stands for one specimen. The text data is recorded as the three-dimensional coordinate data of 20 human joints, and the rows in the data matrix of the txt file are the x, y, and z coordinates of the joints, of which the size range of the x and y coordinates is (1,-1), and the size range of the z coordinates is (0.5,5). Every 20 lines in the file is a complete human skeleton, i.e., one frame of the action. The number of rows for each action sample data is a multiple of 20. The labeling of the action is reflected in a specific file naming method, the file is named in the format of a00_s00_e00, where a represents the action serial number, s represents the character body, e represents the number of times the action is executed, and a01_s01_e01, for example, represents the action data of the 1st person executing the 1st action for the 1st time.

2.2 Skeletal Behavior Based Folk Dance Movement Recognition Model

2.2.1 Convolutional Networks for Spatio-Temporal Maps

1) Spatio-temporal map of human skeleton

Skeletal behavior identification entails the extraction of human skeletal data and its spatial - temporal characteristics. Usually, skeletal sequences are depicted by 2D or 3D joint coordinates throughout multiple frames. To build a spatial - temporal map for a sequence with N joints and T frames, two steps are required. First, physically connected joints within each frame are joined. Second, the same joints are linked across successive frames. The resultant map is presented in Fig. 1.

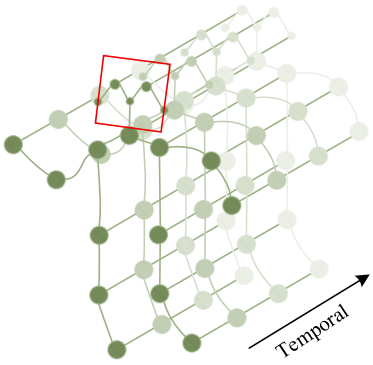


Figure 1: Spatiotemporal diagram of the human skeleton

The whole spatio - temporal graph is symbolized as , where represents the collection of all the juncture points of the human body. indicates the linkage between the juncture points in the skeletal diagram. It is composed of and , with signifying the physical linkages of the human skeleton within the same frame. Meanwhile, pertains to the connections of the same joints

across consecutive frames.

2) Spatial graph convolution

The conventional convolution operation is proficient in dealing with data that has a grid - like structure. In this context, both the natural image and its corresponding feature map can be regarded as a two - dimensional grid of pixels. The output of the convolution is presented in equation (1), where the sampling function and the weight function are and respectively.

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

To carry out convolution on the graph, the definitions of and must be re - established. In a two-dimensional image, the sampling function samples the adjacent pixels of the central position x to extract features. Likewise, within the graph, the sampling function is delineated within the neighborhood set of the node. Moreover, the criteria for the neighborhood set can be self - defined. Typically, when the shortest distance from a particular node in the graph to the root node along any given path is smaller than a specific value, it can be considered that this node is in the neighbor set of v_{ti}, and in ST-GCN, this value is 1.

Given that the graph pertains to non - Euclidean data, the quantity of neighboring nodes fluctuates. Meanwhile, the number of parameters in the convolution remains constant. To establish a correspondence between these two aspects, ST - GCN devises a mapping function. This function partitions the set of neighbors into a fixed number of K subsets.

ST - GCN puts forward three partitioning approaches: single - label partitioning, distance - based partitioning, and spatial structure - based partitioning.

The result of graph convolution within ST - GCN when employing the single label partitioning approach is presented as follows:

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

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

In the case of distance partitioning and spatial structure partitioning strategies, as the collection of neighboring elements is split into several subsets, the adjacency matrix is broken down into multiple matrices. Consequently, Equation (2) is transformed into:

$$f_{out} = \sum_j \Lambda_j^{-\frac{1}{2}} A_j \Lambda_j^{-\frac{1}{2}} f_{in} W_j \tag{3}$$

In the realm of spatial graph convolution, ST - GCN makes use of a straightforward attention mechanism. A weight matrix that can be learned is utilized to indicate the disparity in the significance of edges during graph convolution. In Equation (3), is substituted with. Here, represents the element - wise multiplication of the matrix and the corresponding elements of. The elements of are all initially set to 1. As a result, Equation (3) is transformed into:

$$f_{out} = \sum_j W_j (f_{in} \Lambda_j^{-\frac{1}{2}} A_j \Lambda_j^{-\frac{1}{2}}) \otimes M_j \tag{4}$$

2.2.2 Adaptive graph convolution module

In this research paper, the graph convolutional network undergoes enhancement, and the ultimate adaptive graph convolutional layer is presented as follows:

$$f_{out} = \sum_j W_j f_{in} \left(\alpha A_j + B_j + C_j \right) \tag{5}$$

To begin with, the structure here bears resemblance to that in Equation (4). Equation (4) features an adjacency matrix that depicts the physical connectivity associations among human joints.

An adjacency matrix that can be learned is similar to the function of another element in that it is data - driven. It is acquired through network learning and can concurrently indicate both the existence or non - existence of connections between two nodes and the intensity of these connections. The disparity lies in the method of fusion with another component. In this case, two elements are added together. As a result, some zero elements in a certain matrix can be transformed into non - zero elements after the calculation.

In this research paper, a matrix is incorporated. This matrix has the ability to construct a distinct graph for each individual sample. Moreover, it can signify whether there are links between different joints along with the intensity of those links.The detailed implementation process involves the following steps. First, the input is mapped to a lower - dimensional space. This is achieved through dimensionality reduction using the and functions. Subsequently, the feature similarity between each pair of nodes is calculated via multiplication. Finally, this similarity matrix is normalized by the activation function to yield the ultimate adjacency matrix.This entire process can be represented as follows:

$$C_j = \tanh\left(f_{in}^T W_{\theta j}^T W_{\phi j} f_{in} \right) \tag{6}$$

2.2.3 Attention module

(1) Channel Attention

The channel attention module is primarily composed of two components: compression and excitation. Initially, the output from the spatial map convolution serves as the input for the compression process. The corresponding expression is presented as follows:

$$z = \frac{1}{T \times N} \sum_{i=1}^{T} \sum_{j=1}^{N} u_c(i, j) \tag{7}$$

Subsequently, the excitation phase ensues. This phase modifies the feature map using the subsequent formula:

$$s = \sigma\left(W_2 \delta\left(W_1 z \right) \right) \tag{8}$$

(2) Spatial attention

The formulation for the spatial attention module is presented as follows:

$$s = \sigma\left(w_s(AvgPool(f_{in}))\right) \tag{9}$$

The module input is denoted as [input]. Meanwhile, [pooling operation] represents the global average pooling carried out in the time dimension. Once the pooling process is finished, the information within the time dimension gets condensed, and the dimension of the feature map turns into [dimension]. Eventually, the spatial attention weight [weight] is acquired.

(3) Temporal attention

The configuration of the temporal attention module is identical to that of the spatial attention module, and its formulation is presented as follows:

$$s = \sigma\left(w_t(AvgPool(f_{in}))\right) \tag{10}$$

The distinction lies in the fact that there is a global average pooling operation here, which condenses the information within the spatial dimension. The ultimate outcome will be a temporal attention weight with a specific dimension.

2.2.4 Network basic units and structures

The fundamental unit of the network is composed of a GCN module and a TCN module that share the same structure. The spatial and temporal graph convolutional layers are responsible for extracting the spatial and temporal characteristics of skeletal sequences, respectively. The batch normalization layer helps to alleviate the problem of gradient vanishing. Meanwhile, the ReLU activation function boosts the model's ability to express complex patterns through nonlinearity. The adaptive module is integrated into the spatial graph convolutional layer.

The entire framework of the network amounts to a combination of these fundamental units, making up a total of nine layers. At the start, the quantity of output channels is 64, and it gets multiplied by two every three layers. The Batch Normalization (BN) layer standardizes the input data. During the process of training the network, the dropout technique is employed. This involves eliminating specific nodes from the network as a means to avoid overfitting. The features undergo pooling through global averaging and are then input into a softmax classifier to yield the final classification outcomes. Figure 2 depicts the overall architecture of the network.

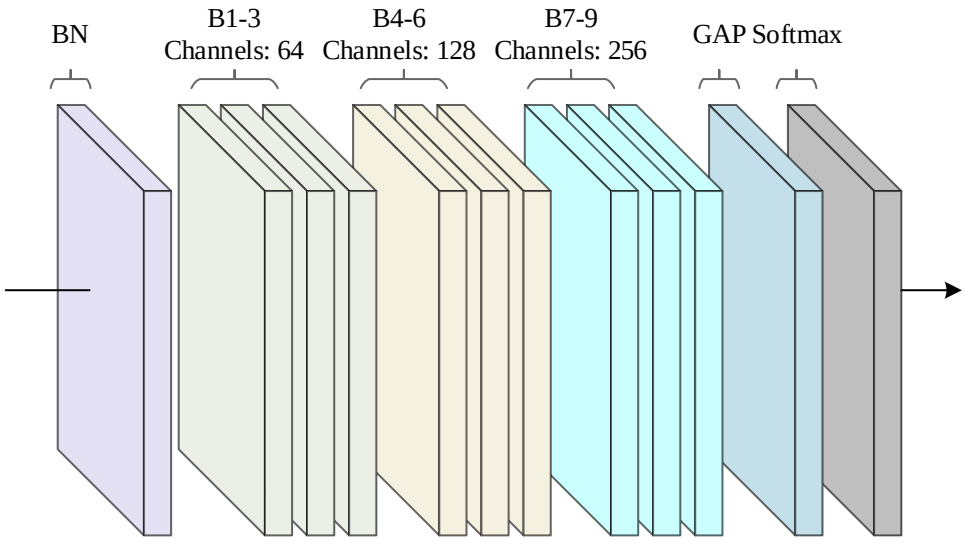


Figure 2: Overall network structure

2.3 Comparative analysis of experiments

In this paper, two datasets are used for experimental validation, NTU RGB+D dataset and NDT dataset of typical movements of folk dances constructed in the previous paper.NTURGB+D datasetThe NTURGB+D dataset is the largest and the most commonly employed movement identification data set at present, which contains 56,758 movement clips of 60 movement categories, and is completed by 20 volunteers. Three cameras positioned at the same elevation but different perspectives record every action. The dataset includes depth map series, RGB video recordings, infrared video footage, and 3D skeletal sequences for each action specimen. In the skeletal sequence, each video of the action has 25 joints. Moreover, the original documentation of the dataset suggests two benchmarks.

1) Cross - subject (X - Sub): This refers to a situation where the experimental subjects in the training set and the validation set are distinct. The training set consists of 40598 videos, while the validation set comprises 16160 videos.

2) Cross - View (X - View): This refers to the situation where the same action is filmed from various perspectives. In this benchmark, the training dataset consists of 38,526 videos, which were recorded by cameras 2 and 3. Meanwhile, the validation dataset comprises 18,232 videos, captured by camera 1.

In this paper, we follow this division and output the action recognition rate on the validation set for both the benchmarks.

2.3.1 Effects of attentional mechanisms on action recognition

On the basis of the basic ST-GCN, the attention mechanism is introduced, different weights are assigned to human joints, and the dynamic adjustment can more effectively capture the degree of influence of each joint on the action in human movement. Table 1 presents a comparison of the recognition rates associated with various strategies. When the attention mechanism is incorporated into ST - GCN, it attains recognition rates of 89.25% (X - Sub) and 96.46% (X - View) on the NTURGB+D dataset. By effectively capturing spatio - temporal dynamics, this approach surpasses the baseline by 11.52% and 6.25% respectively. On the folk dance dataset, the recognition rate achieved is 41.23%. This represents an improvement of 10.81% compared to the baseline. However, it is significantly lower than the rates obtained on the NTURGB+D dataset.

Table 1: Comparison of recognition rates of different improvement strategies

Dataset	Basic ST-GCN	Fusion attention mechanism
NTU RGB+D(X-Sub)	77.73	89.25
NTU RGB + D(X-View)	90.21	96.46
A dataset of typical movements in ethnic dances	30.42	41.23

2.3.2 Extended Hierarchical Temporal Convolutional Networks for Action Recognition

In the fundamental Spatio-Temporal Graph Convolutional Network (ST - GCN), when extracting features from the time dimension of an action sequence, only a single layer of standard two - dimensional convolution is employed. As a result, the long - term correlations within the actions are not extracted. By integrating ST - GCN with an augmented hierarchical temporal convolutional network that incorporates residual mechanisms, both long - term and short - term action dependencies can be effectively captured. As presented in Table 2, this refined model attains an accuracy of 88.13% (X - Sub) and 97.55% (X - View) on the NTURGB+D dataset. This performance outstrips the baseline by 10.88% and 6.19%

respectively. When applied to the folk dance dataset, it achieves a score of 41.89%, marking a 10.87% improvement. Although this approach enhances performance, the improvement is less significant compared to the use of the attention mechanism alone. This implies that joint correlation plays a more crucial role in action recognition.

Table 2: Comparison of recognition rates of different improvement strategies

Dataset	Basic ST-GCN	Fusion extended hierarchical temporal convolutional network
NTU RGB+D(X-Sub)	77.25	88.13
NTU RGB+D(X-View)	91.36	97.55
A dataset of typical movements in ethnic dances	31.02	41.89

2.3.3 Results and comparison of improved models

The outcomes of the experiment suggest that both joint dynamic correlations and temporal dependencies have an impact on recognition performance. Consequently, to extract these features concurrently, this paper incorporates the attention mechanism and the extended hierarchical temporal convolutional network into ST - GCN. Table 3 presents the corresponding recognition rates. The presented table indicates that the suggested model substantially outperforms the baseline Spatio-Temporal Graph Convolutional Network (ST-GCN). This superiority is achieved by concurrently capturing both joint spatio-temporal correlations and temporal dependencies. On the NTU RGB+D dataset, the model attains an accuracy of 91.91% for the X-Sub category and 97.43% for the X-View category. This represents an improvement of 9.62% and 5.3% respectively when compared to the baseline.The higher accuracy in the X-View category can be attributed to the collection of data from multiple angles. This approach enables better extraction of motion features. In contrast, the relatively lower accuracy in the X-Sub category suggests that subject differences have a more significant impact on the results.On the folk dance dataset, the model achieves an accuracy of 38.34%, which is a 5.29% improvement over the baseline.

Table 3: Comparison of strategies for improved strategy

Dataset	Basic ST-GCN	AAST-GCN
NTU RGB+D(X-Sub)	82.29	91.91
NTU RGB+D(X-View)	92.13	97.43
A dataset of typical movements in ethnic dances	33.05	38.34

The recognition rates for the NTU RGB+D (X - Sub, X - View) and the folk dance dataset are depicted in Figures 3 to 5. These figures display the accuracy over the course of iterations under various strategies. During the training process, the accuracy experiences significant fluctuations. In the initial 150 epochs, it rises at a rapid pace. Subsequently, its growth decelerates and eventually stabilizes after 300 epochs. The figures clearly illustrate that the suggested model surpasses the baseline ST - GCN and other enhanced approaches.

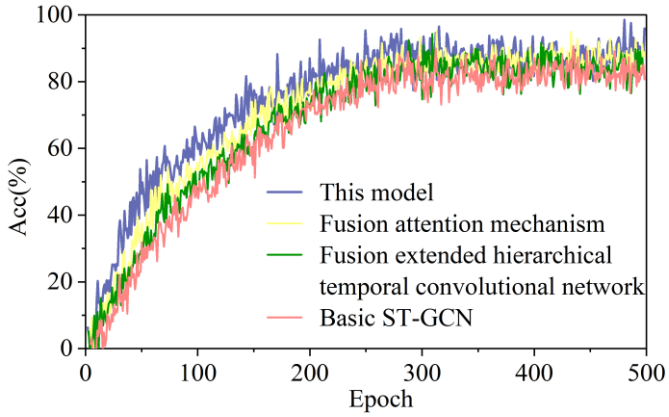


Figure 3: Action recognition rate of NTU RGB+D(X-Sub)

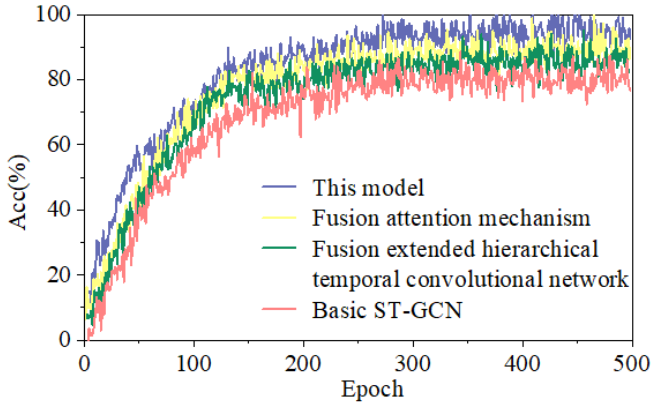


Figure 4: The action recognition rate of NTU-RGB+D(X-View)

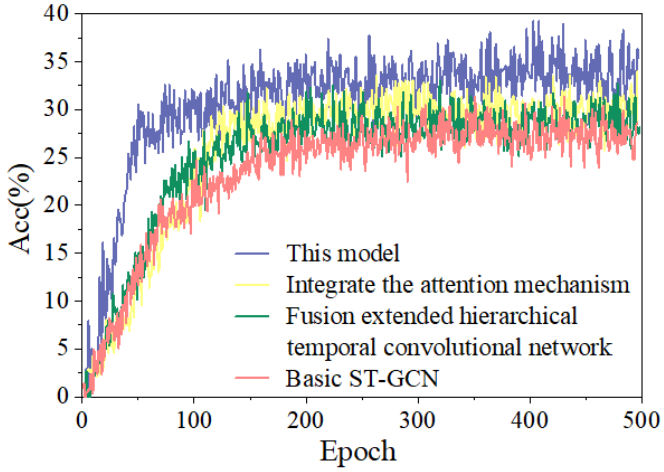


Figure 5: The motion recognition rate of the national dance typical action data set

2.3.4 Comparison of the improved model with other models

To assess the performance, the suggested model is contrasted with other cutting - edge approaches on the two datasets. For the sake of clarity, the results are rounded to one decimal place. As depicted in Fig. 6, on the NTU RGB + D dataset, comparisons are carried out against Clips - CNN + MTLN, DPRL + GCNN, SR - TSL, HCN, AS - GCN, and 2S - AGCN.Under

the X - Sub setting, the proposed model outperforms these methods by 11.1%, 5.1%, 5.1%, 4.1%, 2.7%, and 1.7% respectively. In the X - View setting, it exceeds the performance of Clips - CNN + MTLN, DPRL + GCNN, SR - TSL, HCN, and AS - GCN by 10.8%, 3.8%, 3.2%, 4.2%, and 1.1% respectively.

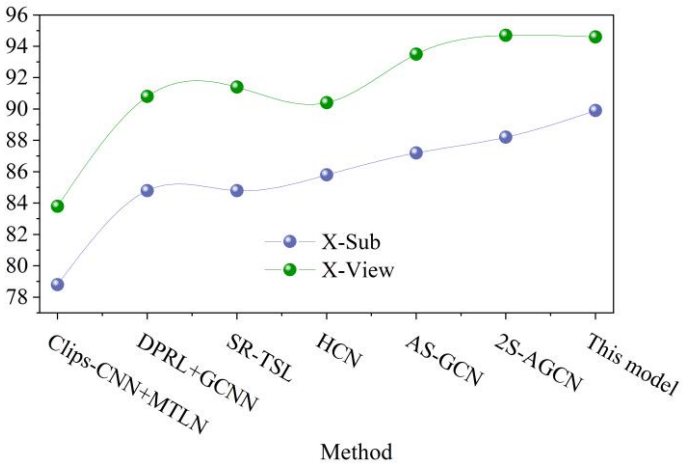


Figure 6: Comparison with other Models on NTU RGB+D dataset

The model presented in this research is juxtaposed with Deep LSTM, Temporal Cony, AS - GCN, and 2S - AGCN respectively using the characteristic movement dataset of folk dance. The comparison of this model with the other models on the characteristic movement dataset of folk dance is depicted in Figure 7. As can be observed from the figure, on this particular dataset, the model in this paper shows improvements of 22.9%, 16.7%, 2%, and 0.5% when compared to Deep LSTM, Temporal Cony, AS - GCN, and 2S - AGCN respectively. In summary, the experimental outcomes have demonstrated that the model proposed in this paper exhibits a significant enhancement in performance compared to the other models.

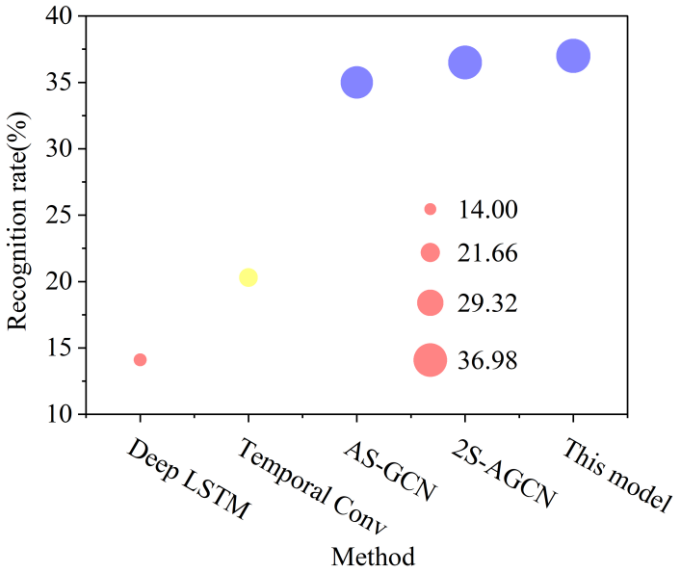


Figure 7: Comparison with other Models on the national dance typical action data set

3 Development of movement-recognition-based assisted training for folk dance

3.1 Motion Capture Sensor Based Assisted Training for Folk Dance

3.1.1 Establishment of a human movement database

The creation of the human motion database consists of several steps, labeling the user's action information by the user's actions and the frame values in the video, and encoding the action data in the database, as well as matching specific encoding methods and dynamically adjusting the results. Within this research paper, the database of human actions is symbolized by , , where stands for the action, indicates the overall quantity of frames included in the action, and the collection of frames of action - related data is presented as follows:

$$X = \left(F_1, F_2, \cdots, F_N\right)^T \tag{11}$$

In this context, denotes the initial frame, signifies the data of the -th frame, and stands for the filter compensation value at the moment of recording. Within the database, an average action is delineated using the subsequent formula:

$$\phi = \frac{1}{N}\sum_{i=1}^{N} F_i \tag{12}$$

Then the established action matrix C is defined as follows:

$$C = \frac{1}{N}\sum_{i=1}^{N} d_i d_i^T \tag{13}$$

Let be defined as , which represents the disparity between an action and the mean action. The eigenvectors within the action matrix can be directly obtained by setting , and the eigenvalues are . Subsequently, the proportion of information stored in the database is as follows:

$$\delta = \frac{\sum_{i=1}^{i=p} \lambda_i}{\sum_{i=1}^{i=N} \lambda_i} \tag{14}$$

Here, represents the mean value of the quantity of frames. The value of the information stored in the human motion database is derived from these frames.

3.1.2 Modeling human movement

The human action model established in this paper is represented by 21 joint points as well as 20 bones, and the human skeleton model is shown in Fig. 8.

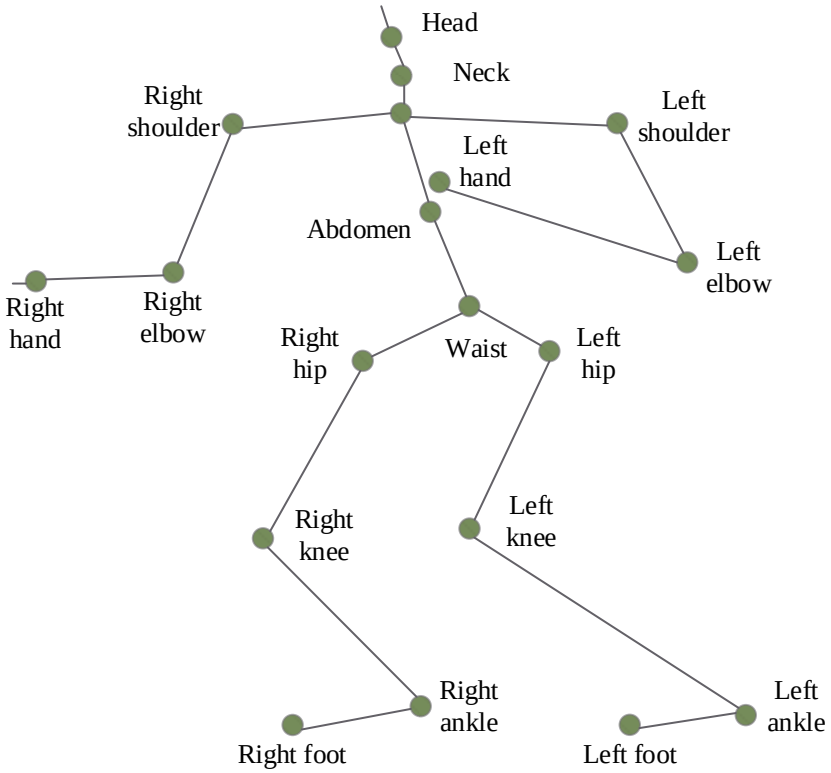


Figure 8: Human skeleton model

In the figure, the long bones are modeled as rigid structures, through key connections in this model as rigid bodies, representing both deformation and invariance of the bones. In this paper, the length of the bones obtained from the human skeleton model is calculated. Considering the different sizes of the video characters saved in the human motion database, the joints are solved to calculate the length of each bone and the skeleton model is scaled. According to the skeleton model, the correct position of the bones as well as the motion state is converted into an animation using the matrix BVH parsing algorithm with the following equations:

$$v' = Mv \tag{15}$$

where v' represents the transformed points in the skeletal model. v represents the original points in the skeletal model. M represents the transformation matrix. While a single bone can be transformed in motion using chi-square coordinates, the transformation remains in the transformation matrix, yielding $M = TRS$, with S representing the size of the bone fed back into the model.The matrix serves as the translation matrix, while the other matrix stands for the rotation matrix. The local transformation of a bone depicts its orientation within the local coordinate system. To obtain the overall transformation matrix of the bone, it is essential to multiply the local transformation by the bone's global transformation. Additionally, the local transformation within the bone needs to be multiplied by that of the parent bone. After obtaining the locations of the key joint points, the 3D model is built from the behavioral data.

3.1.3 Learner Motion Capture

Motion capture sensor choose ADXL345 acceleration sensor, the sensor can measure the dynamic acceleration caused by movement and impact as well as static acceleration, simultaneously, the sensor features a high level of integration and offers two distinct communication methods.The angular velocity sensor opts for the L3G4200D gyroscopic sensor.

This sensor has a sensitivity of 8.25 millidegrees per second per digit within the range of plus or minus 250 degrees per second and can fulfill the requirements of dance motion capture. The schematic illustration of the L3G4200D is presented in Figure 9.

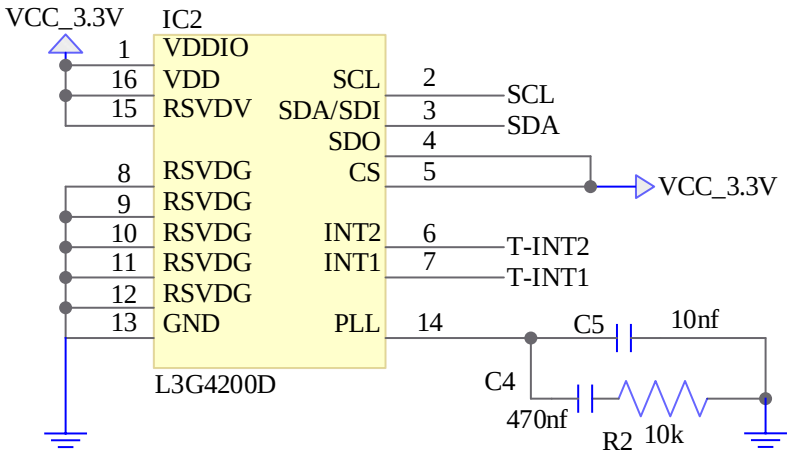


Figure 9: Schematic diagram of L3G4200D

An HMC5883L magnetic resistance sensor was added to the conventional motion sensor to detect the magnetic field errors that occur during motion capture and to correct the deviations. Considering that in the process of capturing the movements of learners, it is necessary to ensure the synchronization of the data output on each sensor in terms of time, the data from the sensors is processed using FIR filtering. The sampling rate in the filtering process is 75Hz, the order is 74, the normalized passband frequency of the filtering is 0.151, the normalized stopband frequency is 0.316, the normalized cutoff frequency is 0.347, and the ripple of the normalized passband is 0.0335dB.

Once the filtering procedure is finished, the pose will be computed using the complementary filtering algorithm as follows:

$$pos = a*(pos + gyro*dt) + (1-a)*(acc\_mag\_pos) \tag{16}$$

Here, denotes the action computed by the sensor via the vector observation approach. stands for the integration interval. The is the update interval of the filtering outcome using the angular velocity integration technique, and the value of is determined as follows:

$$a = \frac{\tau}{\tau + dt} \tag{17}$$

where τ represents the time constant. Logical judgment of the action can adjust τ to improve the complementary filtering effect.

3.1.4 Action Reproduction and Action Error Correction

Once the learner's actions have been recorded, a measure of distance is employed to assess the level of resemblance to those stored in the action database. The two - dimensional representation of the action is depicted in Figure 10.

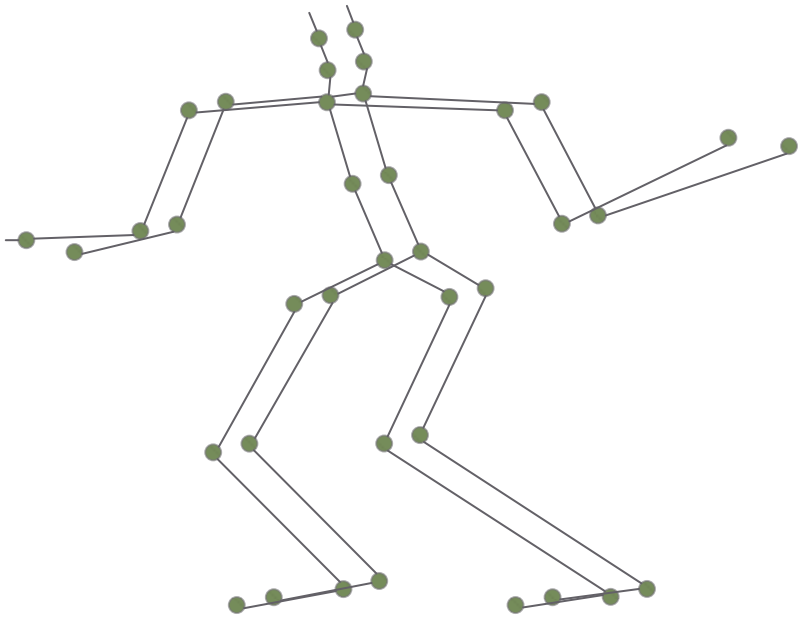


Figure 10: Action reproduction diagram under the model

In the course of an action, every point exerts a distinct level of impact on that action. In the process of modeling, the end - points tend to have a more significant influence. Consequently, when carrying out calculations, it is necessary to take into account the variance of change. Simultaneously, the Euclidean distance between the points should also be factored in. When conducting an action comparison, it is crucial to select an appropriate reference point to guarantee the accuracy of the comparison. In this research paper, the hip joint of the performer is selected as the reference point. The distance of the resulting joint points is calculated as follows:

$$D_i = \sqrt{\sum_{j=1}^{N} \left(\Omega_i^j - \Omega_\tau^j \right)^2 \Big/ \sigma_j} \tag{18}$$

$$\sigma_j = \sum_{i=1}^{N} \left(M - j_i \right)^2 \Big/ N \tag{19}$$

Let σ represent the standard deviation of each joint within the entire action. Denote p as the position of the joint within the action, x as the coordinate of the joint at frame t in the gathered data, and y as the target position of the joint at frame t. Once the positional disparity between movements has been computed, a threshold is established. In the event that this disparity surpasses the set threshold, the system notifies learners of incorrect postures. In this way, it enables assisted training for folk dance.

3.2 Experimental Evaluation of Folk Dance Motion Capture

3.2.1 Functional module testing

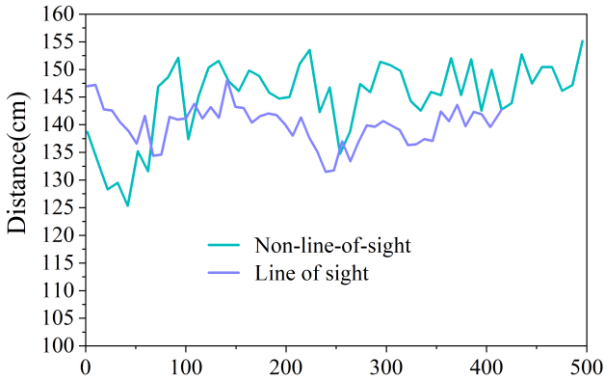
1) Motion Capture Sensor Module Test

To determine whether the motion capture sensor module of the method presented in this paper can efficiently gather dance movement data in real - time, an indoor test was conducted. Motion capture sensors were affixed to 16 joints of 8 test subjects. The test subjects were then instructed to execute 8 different dance movements, including leaning, inclining, punching, rotating, another instance of rotating, kicking, squatting, and leaping. Each movement was

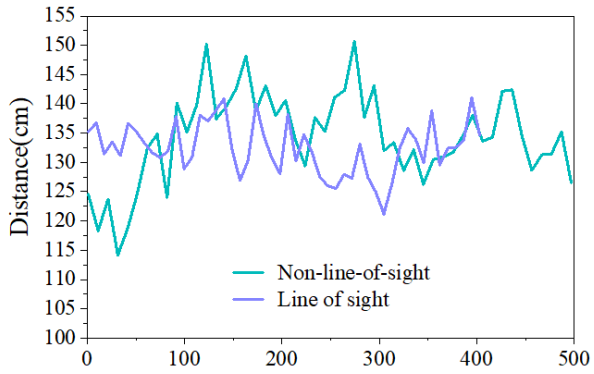
repeated 4 times and each movement was completed within 10s. Then use the information collection equipment connected to the computer through the USB interface, set the baud rate to 113600, read the motion capture sensor data.

2) Human displacement localization module test

To evaluate the performance of the human displacement localization module in the method presented in this paper, the motion capture sensor and base station A and computer connection, and base station B and C were placed on the top of the base station A and the right side, the label and the motion capture sensor were placed in the center of the square field coordinates, human limbs and waist, and let the testers into the center of the square field coordinates. The results of human displacement localization are shown in Fig. Figure 11a and 11b present a comparison of the data along the X - axis and Y - axis, respectively. The tester moved around the sensor at a steady pace for a full circle, and the position data was documented over ten repeated trials. Subsequently, the mean and standard deviation were computed to derive the positioning outcomes. The data indicates that under line - of - sight circumstances, there is stable displacement and positioning. When it comes to non - line - of - sight conditions, the values on the X - axis vary between 125 and 155, while those on the Y - axis range from 120 to 150. The fluctuations have small amplitudes and mostly overlap with the results obtained under line - of - sight situations. This validates the outstanding positioning performance of the proposed human displacement localization module.



(a) Comparison of X-axis data



(b) Comparison of Y-axis data

Figure 11: Human body displacement positioning results

3.2.2 Experimental data

To precisely identify the dance motions, the OrdroAC5 device was employed to capture the dance movements in real - time. Meanwhile, the Kinect system was utilized to store the data related to these dance movements. In total, 15 distinct dance movements were obtained. These movements were then classified into 6 decomposed motions. Additionally, the Kinect was used to gather joint - point data, which served as the standard set of movements.

3.2.3 Experimental results

1) Motion Capture Experiment and Analysis

To validate the impact of the suggested dance - assisted training approach on ethnic dance motion capture, the joint errors are contrasted between the inertial sensor based on the RBF neural network and the least squares MEMS inertial sensor. Using Kinect data as the benchmark, the coordinate errors of the three methods are examined, as depicted in Figure 12. The head coordinate error of the proposed method is 1.08, which is just 0.05 different from the standard value of 1.13. In contrast, the other two methods have errors of 1.46 and 0.68 respectively. This indicates that the proposed method attains a higher level of capture precision and aligns more closely with real - life motion.

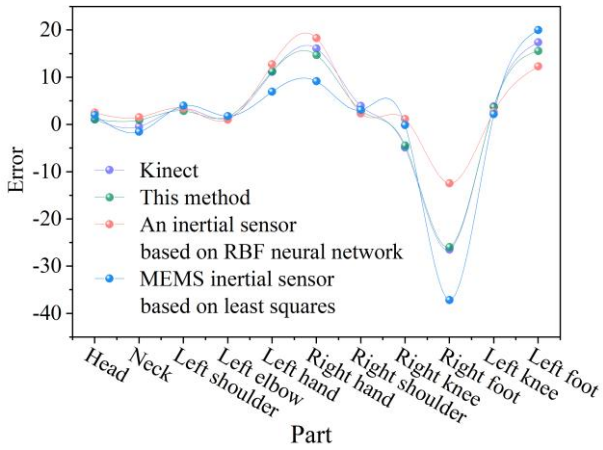


Figure 12: Comparison of error coordinates in dance movements

To further illustrate the preeminence of the suggested approach, motion paths are graphically presented for a comparison with conventional techniques and the Kinect system, as depicted in Figure 13. The outcomes indicate that the path captured by the suggested approach is more similar to that of the Kinect than those obtained from the RBF neural network - based inertial sensor and the least squares MEMS inertial sensor. This validates its ability to precisely capture dance movements.

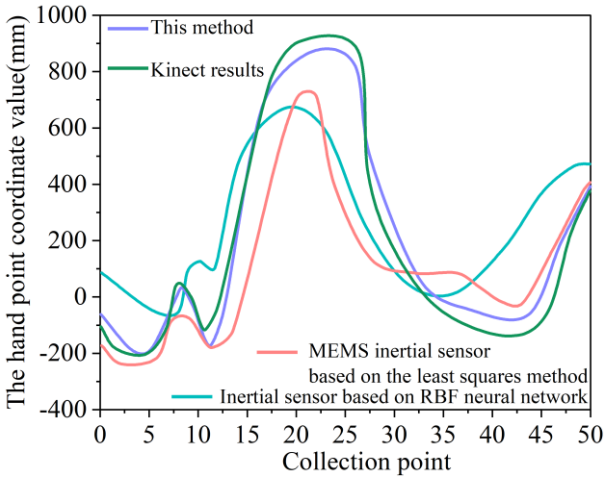


Figure 13: Dance movement trajectory detection results

To conduct a more in - depth analysis of the joint angle capture performance, a series of experiments were carried out, and the outcomes are presented in Table 4. When compared to conventional approaches, the suggested method captures the angles of the left wrist - elbow - shoulder, neck - right shoulder - elbow, and waist - left knee - ankle at values of 170, 150.2, and 157.8 respectively. These values deviate from the Kinect standard by only 0.5, 1, and 1.2. On the contrary, the other two methods exhibit errors that are approximately 20% greater. This clearly demonstrates that the proposed method attains a higher level of accuracy in motion capture and recognition.

Table 4: Comparison of the Angle of dance action

	Kinect	This method	An inertial sensor based on RBF neural network	MEMS inertial sensor based on least squares
Left wrist-left elbow-left shoulder	169.5	170	225.8	145.6
Right wrist-right elbow-right shoulder	113.1	112.4	160.1	190.1
Neck-right shoulder-right elbow	151.2	150.2	187.9	139.9
Neck-left shoulder -left elbow	169.1	170	182	189.1
Waist-left knee-left ankle	159	157.8	169.1	192.2
Waist-right knee-right ankle	99.7	101.6	120.9	134.5

A thorough examination of the aforementioned charts reveals that the application of this approach can precisely detect and recognize the position of a dancer's motions. By conducting comparisons and making matches with the standard movements stored in the database, it becomes possible to carry out training and rectification of dance movements. This, in turn, can assist dance trainers in enhancing their self - training capabilities.

2) Experiment and analysis of movement recognition

To conduct a more comprehensive verification of the automatic recognition performance, the system undergoes testing on six distinct dance motions: stretching, chest stretching, body rotation, jumping, full - body movement, and the final pose. Table 5 presents comparative experiments between this system and the Kinect - based system. The outcomes demonstrate that the proposed system attains a maximum accuracy of 99.1% and sustains an accuracy of over 93% for all six motions. In contrast, the Kinect - based system reaches a peak accuracy of

only 89%. This validates the superior dance recognition ability of the proposed system.

Table 5: Action recognition result

	Action	Recognition rate	Rejection rate	False recognition rate
This system	Stretching	93.5	6.5	0
	Enlargement	94.8	2.6	2.6
	Bulk transport	99.1	0.5	0.4
	Jumping motion	97.8	1.9	0.3
	General motion	93	7	0
	Collating motion	94.6	2.8	2.6
Motion recognition system based on Kinect	Stretching	87.3	12.7	0
	Enlargement	86.1	3.4	10.5
	Bulk transport	88.5	7.4	4.1
	Jumping motion	87	2.4	10.6
	General motion	89	9.4	1.6
	Collating motion	84.1	6	9.9

Following a statistical comparison and analysis of the recognition accuracy of the two systems, the action recognition outcomes of the two systems are presented in Table 6. The data in the table indicates that, when contrasted with Kinect's action recognition system, the recognition precision of the method proposed in this paper is notably higher. Moreover, it can attain high - quality recognition results across multiple dance movements.

Table 6: The action recognition results of the two systems

	Organizing exercise	Full-body exercise	Jump	Body rotation	Chest expansion	Stretch
This system	100	100	100	83.04	73.39	100
Recognition system based on Kinect	92.50	80.46	87.98	79.18	67.37	94.85

4 Conclusions and Strategies

4.1 Conclusion

This article addresses the issue of folk dance preservation and inheritance and development, and proposes a more complete set of folk dance movements from movement recognition to movement capture to assist dance training. The article draws the following conclusions:

1) The skeletal behavior recognition method based on adaptive and attentional mechanisms is experimentally validated on two publicly available action datasets, NTU RGB+D and Ethnic Dance Typical Action Dataset, and compared with other existing models, and in cross view (X-View), this paper's method is better than Clips-CNN+MTLN, DPRL+GCNN, SR-TSL, HCN, AS-GCN by 10.8%, 3.8%, 3.2%, 4.2%, and 1.1%, respectively. The experiments confirmed that the enhancement approach presented in this paper outperforms other models.

2) The outcomes of the tests on ethnic dance motion capture technology indicate that the approach presented in this paper boasts a high level of accuracy in automatically recognizing dance movements. It is capable of achieving the automatic identification of various dance motions, including swooping, tilting, and rushing. Compared with the standard SVM

recognition model and KNN algorithm, the algorithm in this paper has higher recognition accuracy, with an average accuracy of 90.93%, which is obviously effective and superior.

4.2 Strategies

4.2.1 Transmission and application at the technical level

Combined with the demand preferences and usage habits of the current new media audience, use VR panoramic cameras and professional audio equipment to produce Vlog and Podcasting works centered on folk dance inheritors, and put them on the self-media side, so that the inheritance, teaching, and application of folk dances can be more effectively disseminated and exchanged. Modern science and technology and means can also be used to produce folk dance-related short videos, animation, H5, VR, mobile live broadcasting and game works, so that folk dance can enter today's all-scene era, and through cross-media narratives, can be recognized, accepted and loved by more young audiences.

4.2.2 Transmission and application at the media level

With the influence of the integrated media platform, we will launch the column or small program of “Learning Ethnic Dance with Non-Genetic Inheritors”, focusing on the promotion of the animal dance part of the ethnic dance, which is both interesting and experiential, and can also gain the attention and participation of new media users.

4.2.3 Transmission and application at the social level

In addition, with the help of media promotion and public opinion influence, combining the effectiveness of media communication with social influence, the media can be utilized to show the cultural and artistic value of folk dance, and then boost the development of cultural tourism industry. Today's tourists attach more importance to the cultural heritage and connotation of tourist destinations, and folk dance is precisely a kind of figurative and visual way to show the Dongba culture, which can be introduced into all kinds of stage performances after scientific refinement and artistic processing, realizing the artistic presentation and modern expression of traditional culture.

References

[1] Wan, D., & Yang, S. (2025). Digital environments for preserving Chinese national dance culture. Research in Dance Education, 1-16.

[2] Pathrapoowanun, N., Kwangmuang, P., Siritaratiwat, A., & Lan, N. T. (2025). Enhancing Creative Learning in Thai Classical Dance Education: An Integration of Digital Preservation Framework with Ubiquitous Learning Environment. Digital Applications in Archaeology and Cultural Heritage, e00460.

[3] Chaturvedi, A. (2025, February). Preserving the Legacy of Rabindra Nritya: Global Efforts in Preservation, Protection, and Digitization of Music and Dance. In International Conference On Innovative Computing And Communication (pp. 299-314). Singapore: Springer Nature Singapore.

[4] Gao, S., Phanlukthao, P., & Guo, K. (2025). Exploring the Preservation, Inheritance, and Digital Development of Ethnic Dance in a Modern Technological Environment.

International Journal for Housing Science & Its Applications, 47(1).

[5] Qianwen, L. (2024). Application of motion capture technology based on wearable motion sensor devices in dance body motion recognition. Measurement: Sensors, 32, 101055.

[6] Strutt, D. (2022). Motion capture and the digital dance aesthetic: Using inertial sensor motion tracking for devising and producing contemporary dance performance. In Dance data, cognition, and multimodal communication (pp. 131-147). Routledge.

[7] Tao, R. (2024). Virtual Simulation of Dance by Integrating VR Technology and Motion Capture Technology. Informatica, 48(15).

[8] Sun, K. (2022). Research on dance motion capture technology for visualization requirements. Scientific programming, 2022(1), 2062791.

[9] Yao, R. (2020). Three-dimensional Digitizing of Modern Dance Based on Kinect Motion Capture System. Computer-Aided Design & Applications, 17.

[10] Shi, Y. (2022). Stage performance characteristics of minority dance based on human motion recognition. Mobile Information Systems, 2022(1), 1940218.

[11] Wang, M., & Yu, R. (2022). Digital production and realization for traditional dance movements based on Motion Capture Technology. Front. Soc. Sci. Technol, 4(11), 13-18.

[12] Xie, L. (2022). Dance Performance in New Rural Areas Based on 3D Image Reconstruction Technology. Computational Intelligence and Neuroscience, 2022(1), 7122053.

[13] He, Y. (2025). 3D technologies for preserving China's Intangible cultural heritage: a case study of dance culture of Southwest ethnic minorities. Research in Dance Education, 1-17.

[14] Zhang, D. (2022). Analysis of the Style Characteristics of National Dance Based on 3D Reconstruction. Computational Intelligence and Neuroscience, 2022(1), 2419175.

[15] Zhang, N. (2022). 3D Digital Model of Folk Dance Based on Few-Shot Learning and Gesture Recognition. Computational Intelligence and Neuroscience, 2022(1), 3682261.

[16] Skublewska-Paszkowska, M., Milosz, M., Powroznik, P., & Lukasik, E. (2022). 3D technologies for intangible cultural heritage preservation—literature review for selected databases. Heritage Science, 10(1), 3.

[17] Zhang, C. (2024). Research on the Technology of Virtual Reality Empowering Manchu Dance Cultural Communication. Academic Journal of Humanities & Social Sciences, 7(2), 254-259.

[18] Qu, Y. (2023). Innovative Research on the Application of Digital Dance Imaging Technology in Dance Presentation. Frontiers in Art Research, 5(15).

[19] Zhang, B., Sukirman, S. N., & Shek, S. Z. B. S. (2025). The Application of Virtual Reality in Folk Dance Teaching in Shanxi. Metaverse Basic and Applied Research, 4, 2.

[20] Kico, I., & Liarokapis, F. (2020). Investigating the learning process of folk dances using mobile augmented reality. Applied Sciences, 10(2), 599.