International Journal for Housing Science and Its Applications
Publish September 9, 2025. Volume 47, Issue 1 Pages 181-192

https://doi.org/10.70517/ijhsa47115

Exploring the Preservation, Inheritance, and Digital Development of Ethnic Dance in a Modern Technological Environment

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

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

Corresponding authors: (e-mail: 18604043426@163.com).

Abstract This paper addresses the issue of preserving and digitizing ethnic dance movements by analyzing the characteristics of ethnic dance movements, Kinect-based motion capture technology, and skeleton-based motion recognition methods. It proposes a deep learning-based method for recognizing typical ethnic dance movements. Using Kinect sensors to collect data on typical ethnic dance movements, a dataset of typical ethnic dance movements was constructed, and 3D CNNs were used for recognition. Finally, strategies for the protection, inheritance, and digital development of ethnic dances are proposed, and effective pathways for the digital preservation of minority ethnic dances are explored. The results indicate that both the detection of two-dimensional joints and the extraction of three-dimensional joint information can to a certain extent meet the requirements for three-dimensional human motion reconstruction. Additionally, the results of the motion capture system setup and three-dimensional human motion reconstruction are also satisfactory, with experimental errors around 3%. Compared to traditional methods, the motion capture joints and angles under the proposed method are closer to the Kinect standard values, and the motion capture trajectories have the smallest error compared to the Kinect method. Furthermore, the recognition accuracy rate of the proposed method remains above 95%, with a maximum accuracy rate of 99.76%, demonstrating that the proposed method has certain feasibility and application prospects in the preservation and inheritance of ethnic dances.

Index Terms Kinect, 3D CNN, ethnic dance, digital development, preservation and inheritance

I. Introduction

China is a unified multi-ethnic nation, and each ethnic group has distinct differences in terms of living environments and cultural customs. It is precisely these differences that have given rise to the rich and diverse ethnic dances of the Chinese nation [1]-[3]. Ethnic dance is not only an art form of physical expression but also a manifestation of ethnic culture and spirit, serving as a carrier for the inheritance of traditional Chinese culture. It plays a crucial role in preserving cultural heritage and conveying emotions [4]-[6]. From a modern aesthetic perspective, ethnic dance is not only a form of dance art but also a precious intangible cultural heritage, a brilliant crystallization of Chinese traditional culture [7], [8]. The dance of each ethnic group contains profound cultural beauty, reflecting the pursuit of "beauty" by various ethnic groups [9]. With the development of the times and changes in people's aesthetic preferences, the inheritance and development of ethnic dance are facing numerous challenges and crises [10]. Therefore, it is necessary to identify the specific obstacles to the inheritance of ethnic dance in the contemporary era and use this as a guide to implement effective inheritance pathways. This will lay a solid foundation for the sustainable inheritance and modern development of ethnic dance, injecting inexhaustible internal momentum for its brilliance in the fields of dance and art [11]-[14].

In the digital age, numerous scholars have focused their attention on cross-cultural dissemination strategies and innovative practices in ethnic dance, aiming to explore pathways for the protection and transmission of ethnic dance in the context of the new era. For example, Li, L. and Kang, K. developed a classification framework for audiovisual short video content and investigated the relationship between minority groups' creation of cultural content and audience viewing interests, providing digital protection and transmission channels for minority cultures, including ethnic dance [15]. Meanwhile, Heyang, T. and Martin, R. demonstrated that social media platforms can create spaces and venues for dance education and learning, noting that the educational experiences and dance culture provided by social media platforms can effectively enhance dance teaching practices in higher education environments [16]. Their analysis clarified that emerging social media platforms are important platforms for the display and dissemination of ethnic dance, enabling outstanding ethnic dance works to be rapidly disseminated while also allowing learners to access a vast amount of ethnic dance information. For example, Zhao, Y. introduced virtual reality environment technology for teaching ethnic minority dances, which provides students with immersive and interactive learning experiences, helping them complete ethnic dance learning tasks more efficiently [17]. Hajdin, M. et al. explored the application of three-dimensional virtual reality teaching methods in ethnic dance education, finding that this method better stimulates learners' potential in dance learning compared to video-based dance instruction [18]. Kico, I. and Liarokapis, F. developed a mobile augmented reality prototype application to assist in learning ethnic dance. By digitizing ethnic dance and presenting it synchronously through virtual avatars, this application enhances learners' ability to master ethnic dance movements [19]. It can be observed that the introduction of virtual reality and augmented reality adds rich digital information to real ethnic dance performances, enabling users to subtly acquire ethnic dance knowledge through immersive experiences and modern visual presentations. Additionally, Kishore, P. V. V. et al. proposed a method for classifying Indian classical dance movements based on convolutional neural networks, by collecting gesture and posture data from dance videos to achieve accurate recognition of complex dance movements [20]. This demonstrates that motion capture technology provides a scientifically precise method for the digital preservation and inheritance of ethnic dances, and precise digital data of ethnic dances also brings innovation to dance education.

This paper focuses on the detection and tracking methods for human targets in Kinect depth images, including an investigation into the principles of Kinect, discussions on methods for acquiring and processing depth information, and an explanation of the method for identifying skeletal joint nodes of human targets using depth information. To capture motion information across multiple consecutive frames, this study investigates action recognition methods based on skeletal information from both spatial and temporal dimensions. A three-dimensional convolutional neural network (3D CNN) model is adjusted for the recognition of typical movements in ethnic dances, with an exploration of the model's recognition performance. Finally, considering practical circumstances, this study discusses current issues related to the teaching and preservation of Chinese ethnic dances, providing technical support for the protection, inheritance, and digital development of ethnic dance movements.

II. Deep learning-based ethnic dance movement recognition technology

II. A. Kinect-based dance motion detection and tracking technology

II. A. 1) Kinect Depth Information Measurement

In order to use Kinect[21] to measure depth, it is necessary to first calibrate the Kinect's RGB lens and depth sensor. Kinect collects human depth values to calculate the actual distance between each part of the human body and the Kinect sensor, thereby obtaining the world coordinates of each joint point of the human body. Let the human depth value obtained by Kinect be d_r. The actual distance between Kinect and the user can be calculated using the following formula:

$$d = K \cdot \tan(H \cdot d_r + L) - 0 \tag{1}$$

where $K = 12.55cm$, $H = 4 \times 10^{-4} rad$, $L = 1.5rad$, and $O = 4.0cm$.

The user's depth coordinates can be obtained from the actual distance as (x_d, y_d, z_d). Let the user coordinates obtained by Kinect be (x_k, y_k, z_k). The corresponding relationship should be:

$$\begin{cases} x_k = (x_d - \frac{w}{2}) \cdot (z_k + D') \cdot F \cdot (\frac{w}{h}) \\ y_k = (y_d - \frac{h}{2}) \cdot (z_k + D') \cdot F \\ z_k = d \end{cases} \tag{2}$$

Among them, $D' = -10.5$, $F = 0.001$, w, and h are the length and width of the Kinect color camera resolution, respectively, while $w = 1950$ and $h = 1020$ are the height and width, respectively.

II. A. 2) Model-based human detection

Kinect detects human targets using a model-based approach. It acquires depth images, performs noise reduction and smoothing, then performs edge detection and shape detection. Using the edge information and relevant depth changes in the image, it locates human targets using a second-order head detection process. The human detection process is shown in Figure 1.

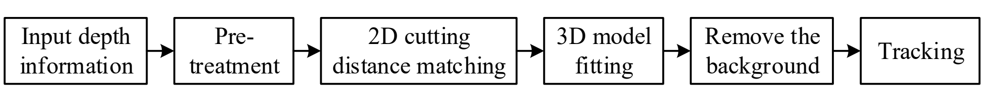


Figure 1: Flowchart of Human body detection

(1) Preprocessing

In actual use, even in continuous spaces, the depth images obtained by Kinect will always generate some noise caused by pixel shifts. The effects of this noise can be eliminated by restoring the depth values. This requires the use of the nearest neighbor interpolation algorithm, which assumes that all missing pixels have empty values and then fills them with the depth values of their neighboring pixels.

(2) Two-dimensional cutting distance matching

First, the Canny edge detection algorithm is used to extract the contours of the depth image. Then, the areas where human targets are likely to appear are determined by detecting contour information. Finally, the cutting distance matching algorithm is used to detect the human body.

(3) 3D model fitting

A cubic regression function is constructed for the height and depth values of the human head. The cubic function is obtained:

$$-y = p_1 \cdot x^3 + p_2 \cdot x^2 + p_3 \cdot x + p_4 \tag{3}$$

Among them $p_1 = -1.4 \times 10^{-9}$, $p_2 = 1.85 \times 10^{-5}$, $p_3 = -0.09$, $p_4 = 190.00$.

When identifying a human body, the first step is to locate the head. After obtaining the depth information of the human body target, use formula (3) to calculate the standard height of the head in the depth image, and then determine the radius R of the search range according to the following formula

$$R = 1.35h / 2 \tag{4}$$

Normalize the depth in the circular region CR with a radius of R_t using the following formula:

$$d\_n(a,b) = d(a,b) - \min(d(a,b)) \quad a,b \in CR \tag{5}$$

where $d(a,b)$ represents the depth value of pixel point (a,b), and $d\_n(a,b)$ represents the normalized depth value. Then, the mean square error between the 3D model and the human head region $T(a,b)$ is:

$$Er = \sum_{a,b \in CR} | d_n(a,b) - T(a,b) |^2 \tag{6}$$

(4) Contour extraction

The region growing algorithm is used to generate all regions that need to be segmented. The generated target regions are the regions of the human body. Then, methods such as Canny edge detection are used to extract the contours of the regions to obtain the overall contour of the human body. That is:

$$S(x,y) = | d(x) - d(y) | \tag{7}$$

Among these, $d(\cdot)$ represents the depth value of a pixel or region. During the region growing process, whenever a new pixel is added to the region, the average depth value of all pixels in the region must be calculated using the following formula, and this value is then used as the new depth value for the region. That is:

$$d(R) = \frac{1}{N} \sum_{i \in R} (d(i)) \tag{8}$$

(5) Tracking

Assuming that the changes in the coordinates and velocity of the target human body during movement are smooth, tracking of the human body can be achieved by locating the center point of the target area of the human body. First, the three-dimensional coordinates of this point are obtained from the depth image, and then the velocity of the human body is calculated based on the changes in the coordinates of the center point between adjacent frames in the video. The energy value of the changes in the position and velocity of the target human body is calculated using the following equation:

$$E = (p - p_0)^2 + (v - v_0)^2 \tag{9}$$

Among these, p represents the coordinates of the center point of the human target in the current frame, p_0 represents the coordinates of the center point of the human target in the previous frame, and v and v_0 represent the velocities of the target human in the current frame and the previous frame, respectively. By calculating all possible target humans, the one with the smallest energy value E is the result.

II. A. 3) Human Joint Point Recognition Based on Kinect

The human skeleton joint model is shown in Figure 2. The principle of human joint recognition is to divide each frame of depth images obtained into several human body parts that are suspected to contain the skeleton joints of interest, and then reproject the divided parts into the world coordinate system to form a prediction of the 3D position of each skeleton joint. The human skeleton joint recognition process can be divided into three steps:

(1) Separate the human target from the background. Use the distance between the object and the camera within the depth sensor's field of view to determine the approximate location of the human target. Then, use the region growth algorithm to identify the actual area of the human target. Finally, employ edge detection algorithms to extract the human contour.

(2) Use a feature matching algorithm based on contour features to identify and locate the head, limbs, and torso of the human contour.

(3) Complete the identification of human joint points, and infer the coordinates of all 25 joint points based on the identified positions of the various body parts.

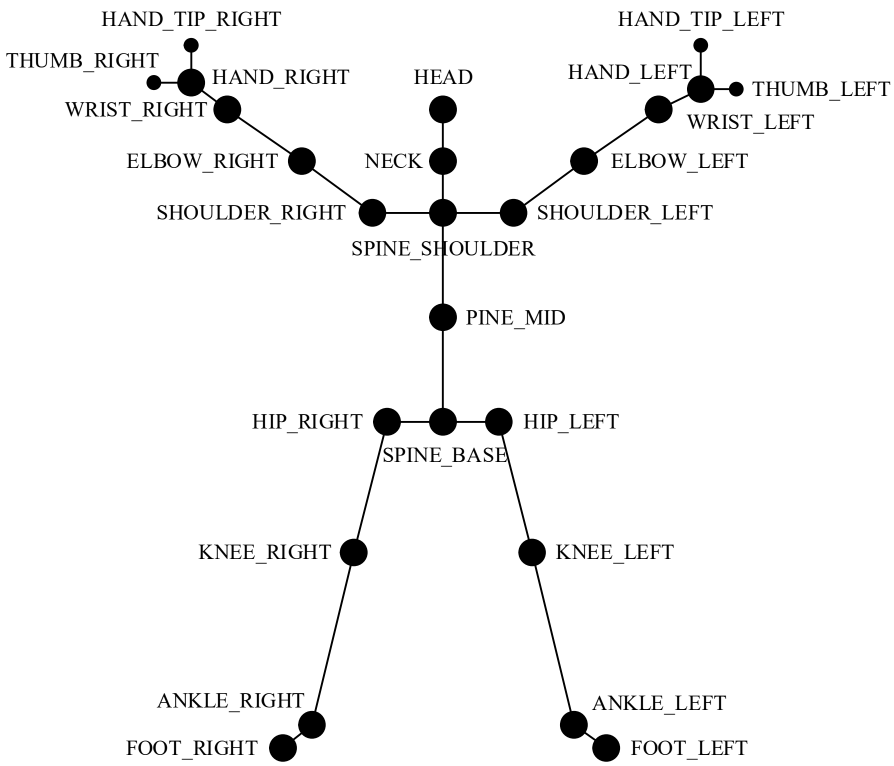


Figure 2: Model of human skeletal joint points

The density estimates for human body parts are defined as follows:

$$f_c(\hat{x}) \propto \sum_{i=1}^{N} w_{ic} \exp(-\| \frac{\hat{x}-\hat{x}_i}{b_c} \|^2) \tag{10}$$

Among them, $\hat{x}$ represents the world coordinates, N represents the number of image pixels, w_{ic} represents the pixel weights, $\hat{x}_t$ represents the reprojection of pixel point x_i in the world coordinates based on a specific depth $d_I(\chi_i)$, and b_c represents the calculated width of each part of the human body.

The definition of pixel weight w_{ic} takes into account both the probability of which body part it belongs to and the surface area of that part in the real world system. That is:

$$w_{ic} = P(c \mid I, x_i) \cdot d_I(x_i)^2 \tag{11}$$

Based on the different structural characteristics of various parts of the human body, the posterior probability $P(c \mid I, x_i)$ was obtained by accumulating data from a collection of partial parts. This ensured that the density estimate was stable and greatly improved the accuracy of joint position prediction.

II. B. Typical Movement Recognition in Ethnic Dance Based on Deep Learning

II. B. 1) Classic action recognition methods based on 3D CNNs

Since the emergence of convolutional neural networks (CNNs), there has been a surge in research on action recognition using

CNNs in the field of computer vision. However, most of these methods have been applied to 2D static image recognition and classification problems. Previous researchers proposed an asymmetric 3D-CNN method for action recognition tasks, which aims to avoid training two separate networks on RGB and optical flow fields, thereby improving computational efficiency. Skeleton recognition methods based on 3D CNNs can more efficiently extract spatio-temporal features from human skeleton sequences, are more robust to noise in skeleton sequences, and exhibit better generalization performance. Action recognition methods based on 3D CNNs have achieved satisfactory recognition results. This paper builds upon the 3D CNN framework to design a 3D CNNs model for ethnic dance action recognition.

II. B. 2) Ethnic dance motion recognition method based on 3D CNNs

(1) Network Overview

This paper employs 3D CNNs to identify skeletal information of typical movements in ethnic dances. Skeletal information is single-channel data, resulting in lower computational requirements and improved model recognition performance. The 3D deep convolutional neural network used in this paper comprises four convolutional layers, two downsampling layers, two fully connected layers, and one Softmax classification layer. The downsampling layers utilize Max-pooling with a kernel size of 3×3×3 and a stride of 1.

(2) Detailed structure of the model

To capture motion information in multiple consecutive frames, features are calculated from spatial and temporal dimensions. The value of the unit with position coordinates (x, y, z) in the j feature map of the i layer is:

$$V_{ij}^{xyz} = f(b_{ij} + \sum_{r}\sum_{l=0}^{l_i-1}\sum_{m=0}^{m_i-1}\sum_{n=0}^{n_i-1}\omega_{ijr}^{lmn} v_{(i-1)r}^{(x+l)(y+m)(z+n)}) \tag{12}$$

Among them, the time dimension of the 3D convolution kernel is n_i, the position is (l,m,n), and the weight value of the convolution kernel connected to the r feature map is ω_{ijr}^{lmn}.

The ReLU function [22] is the most commonly used activation function in deep learning models. This function can make the model parameters sparse, thereby reducing overfitting. In addition, it can also reduce the model's computational load. The ReLU activation function is defined as:

$$f(x) = \max(0,x) = \begin{cases} 0,(x \le 0) \\ x,(x > 0) \end{cases} \tag{13}$$

The calculation of max pooling in the model is shown below:

$$V_{x,y,z} = \max_{0\le i\le s_1, 0\le j\le s_2, 0\le k\le s_3}(\mu_{x\times s+i, y\times t+j, z\times r+k}) \tag{14}$$

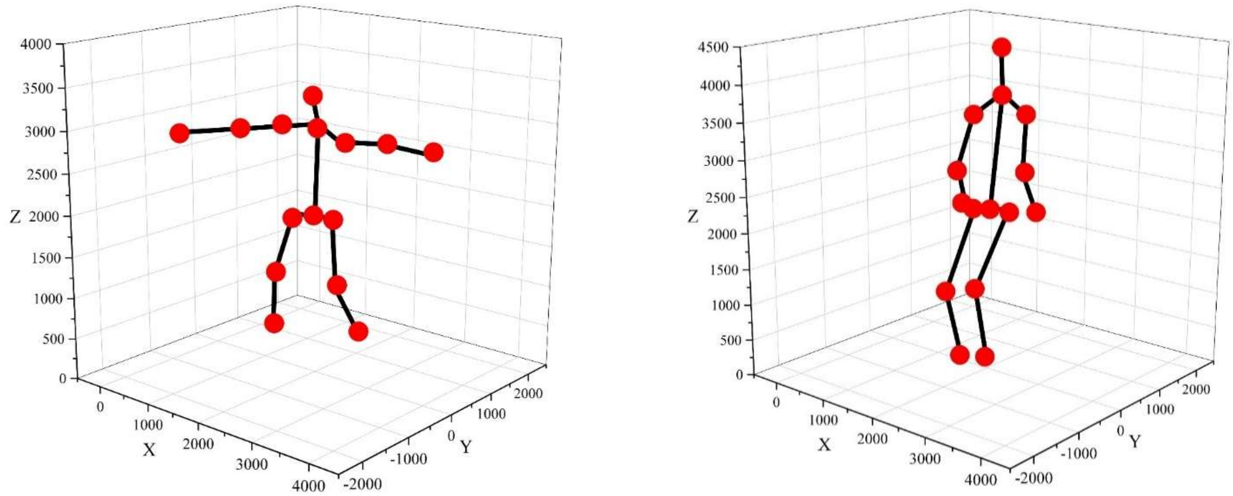
Among these, μ represents the three-dimensional input vector, V represents the output after pooling, and s,t,r represents the sampling stride in the direction. The Softmax function is commonly used in the final layer of classification tasks. It maps an n-dimensional vector x to a probability distribution, such that the probability of the correct category approaches 1, the probabilities of other categories approach 0, and the sum of the probabilities of all categories equals 1.

III. Exploring the Protection, Inheritance, and Digital Development of Ethnic Dance

III. A. Analysis of the effectiveness of ethnic dance movement recognition

III. A. 1) Experimental Data and Time Test Results

Experimental data: In this paper, we first used Kinect technology to detect and track ethnic dance movements from a large three-dimensional dataset, resulting in a three-dimensional skeleton reconstruction model. For the deep learning-based ethnic dance motion capture method proposed in this paper, the average processing time for the entire algorithm workflow is 82 ms per frame. The average time for two-dimensional human pose detection and joint point identification is approximately 21 ms per frame, the time for calculating three-dimensional joint coordinates is approximately 26 ms per frame, and the time for three-dimensional reconstruction is approximately 28 ms per frame. The overall processing speed can reach approximately 16 frames per second. After the system is set up, a single stereo vision ZED camera can be used to achieve 3D reconstruction of ethnic dance motion poses in real-world scenes, enabling motion capture. In this study, the captured image resolution was 1080×1080. The camera was connected to the PC via an interface, with the ZED camera responsible for image acquisition and transmission to the PC via the interface, while the PC handled data processing and result display. Two sets of experimental photos were randomly selected from the publicly available datasets AIST** and NTU RGB120, defined as Group 1 and Group 2, respectively. The three-dimensional skeleton reconstruction results required for this experiment were then constructed using the method described in this paper. The three-dimensional skeleton reconstruction results are shown in Figure 3, where (a) and (b) represent the three-dimensional skeleton reconstruction results for Group 1 and Group 2, respectively.



(a) Group 1 (b) Group 2

Figure 3 Results of 3d skeleton reconstruction of human body

III. A. 2) Data Error Analysis and Action Evaluation

The discussion of experimental data and error analysis in this paper primarily focuses on two aspects: depth information error analysis and motion evaluation using similarity metrics. This paper first conducts a depth value error analysis, with the primary objective of the binocular stereoscopic vision method being to obtain the spatial three-dimensional coordinates of joint points. Two sets of three-dimensional joint node coordinates obtained from randomly selected actions were used. The target numbers correspond to the labels of the human body model. The depth values obtained using the parallax ranging method were compared with the actual measured values, with the shooting distance ranging from 1m to 3m. The results of the depth information error analysis are shown in Table 1. The results indicate that the error values between the system-calculated values and the actual measured values in the two sets of images are extremely small, ranging from 0% to 0.27%. This demonstrates that the depth information error obtained during experiments conducted at approximately 2 meters does not exceed 0.3%, which falls within the error range specified for model accuracy validation in this paper. It also confirms that the method proposed in this paper possesses good effectiveness and feasibility.

Table 1: Results of deep information error analysis

Human body joints	The depth values of the joints acquired by the first group of movements			The depth values of the joints acquired by the second group of movements		
	System calculated value(mm)	Actual measured values(mm)	Error value(%)	System calculated value(mm)	Actual measured values(mm)	Error value(%)
Head	1763.38	1761	0.15	2476.2	2477	0.03
Left upper arm	1780.36	1781	0.06	2435.4	2432	0.13
Left forearm	2213.08	2213	0.00	2419.43	2420	0.03
Right upper arm	2224.25	2221	0.16	2260.52	2264	0.17
Right forearm	1777.99	1779	0.05	2229.01	2229	0.01
Left thigh	1782.35	1784	0.09	2277.28	2276	0.07
Left leg	2235.13	2234	0.03	2433.59	2437	0.13
Right thigh	2215.34	2217	0.07	2416.0	2416	0.00
Right leg	2079.07	2083	0.18	2420.87	2417	0.17
The left arm is a special plane	2235.8	2235	0.05	2737.38	2733	0.15
The right arm is very flat	2249.79	2253	0.14	2783.88	2776	0.27
The left leg is very flat	2188.42	2189	0.01	2736.71	2742	0.18
The right leg is very flat	2256.55	2262	0.24	2355.18	2353	0.08
Chest in a frontal view	2243.92	2243	0.03	2346.05	2344	0.08
Hip feature plane	2276.67	2276	0.03	2375.39	2378	0.09

Next, we validated the depth information at different distances. The error analysis results of the experimental results at different distances are shown in Figure 4. Error validation was performed at distances ranging from 1 m to 10 m, and the maximum error and average error of the depth values obtained at different distances were analyzed and calculated, with the unit being mm. As the distance increased, the number of experiments also increased to obtain a more accurate error analysis.

It can be concluded that as distance increases, the accuracy of locating 3D keypoints in space decreases. The primary reasons are twofold: first, at longer distances, the detection of human 2D posture and keypoint identification are affected; second, the error in stereo calibration parameters also influences the results.

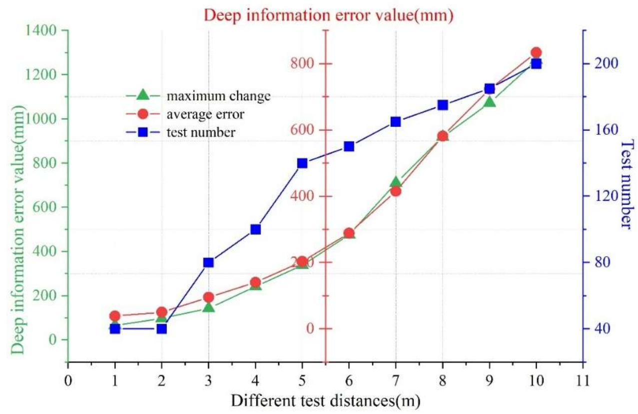


Figure 4: Analysis of experimental results and errors at different distances

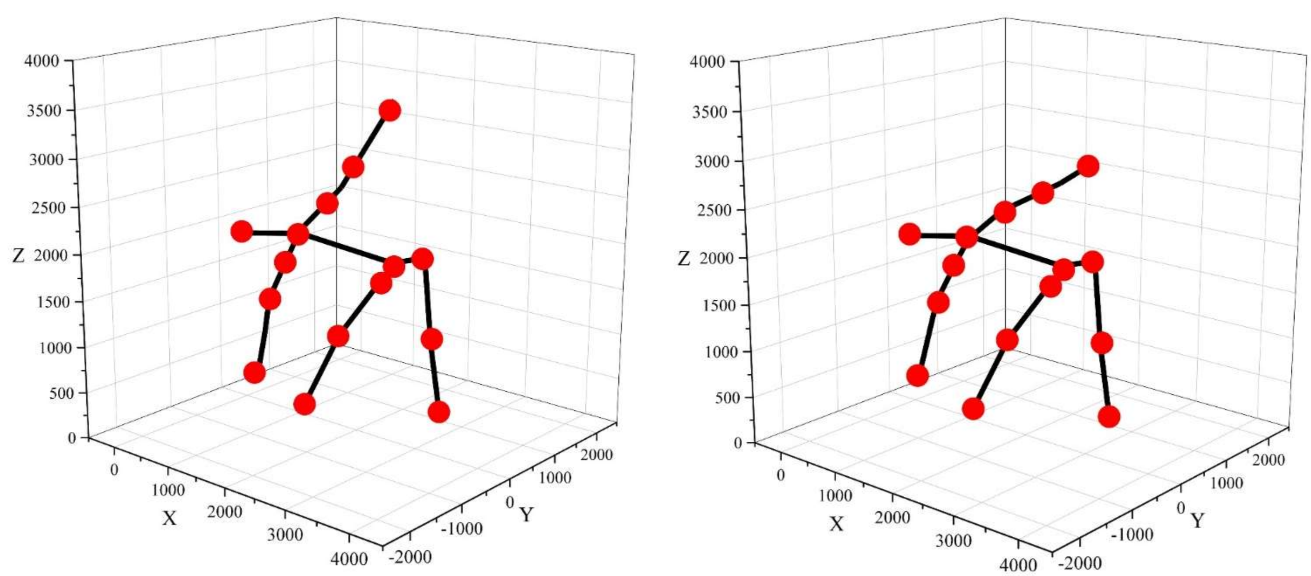
III. A. 3) Action similarity analysis

This paper uses cosine similarity to evaluate the similarity of movements. Compared with the Euclidean distance method, cosine similarity is more accurate in comparing the differences between the directions of two vectors. The formula for calculating and comparing the feature vectors of movements in different parts based on cosine similarity is as follows:

$$similarty_i = \frac{V_i \bullet V_i^{'}}{|V_i| \times |V_i^{'}|} \tag{15}$$

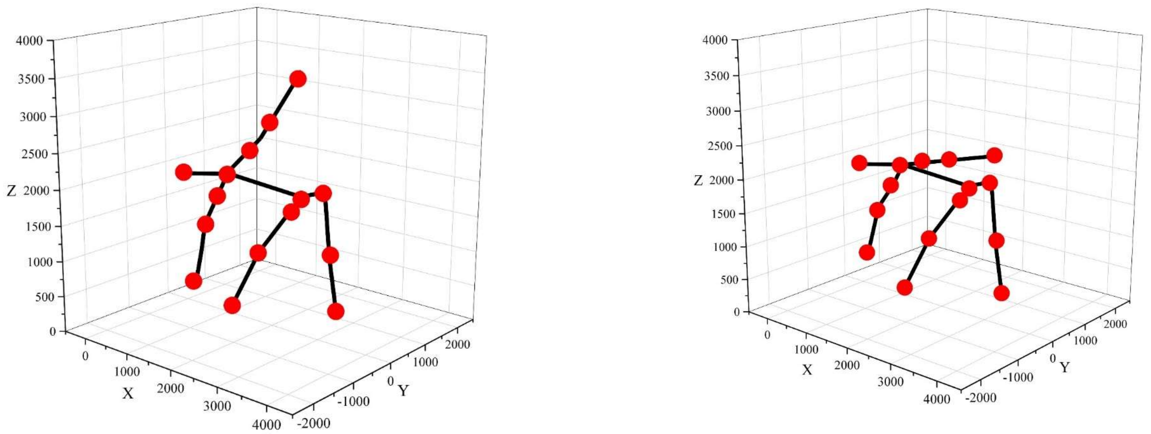
In the formula, v_i represents the i joint feature vector of the standard action, $v_i^{'}$ represents the i joint feature vector of the action to be tested, and $similarity_i$ represents the similarity, with a value ranging from $[-1,1]$. When comparing the overall similarity of the actions, the results of the similarity comparisons obtained for each part need to be added together.

To validate the effectiveness of this motion evaluation method, two sets of yoga motions were selected for comparison. The test diagrams for the first and second sets of motions are shown in Figures 5 and 6, where (a) represents the standard motion, and (b) to (c) represent the test dance motions 1, 2, and 3, respectively.



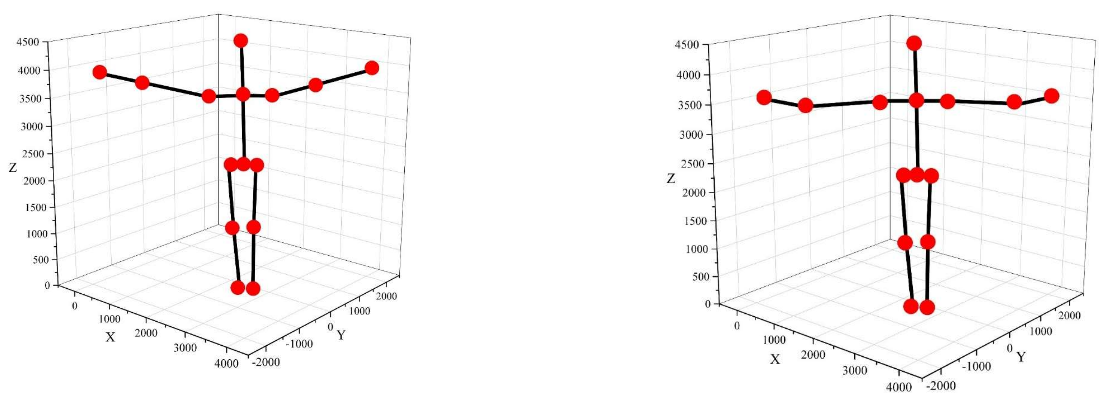
(a)Standard moves

(b)Action 1 is to be tested

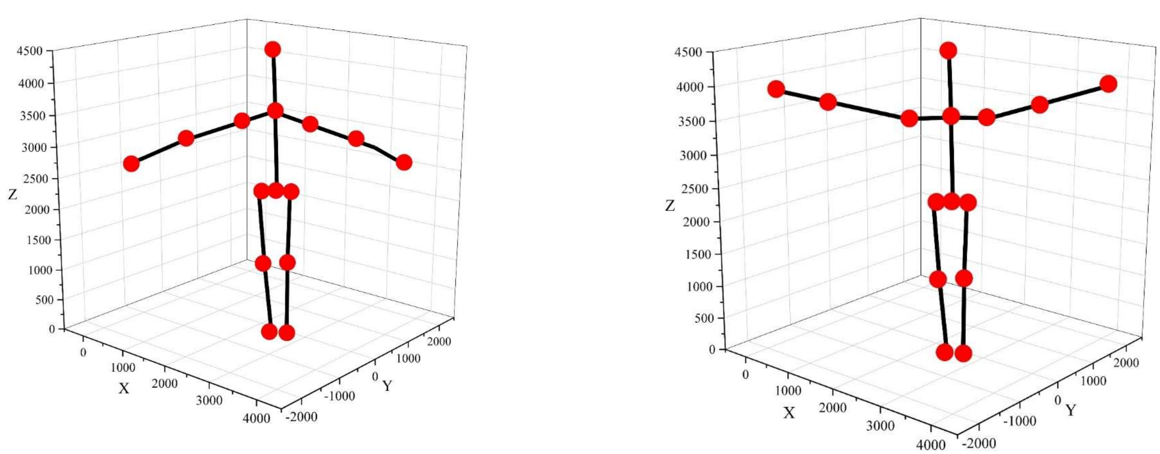


(c)Action 2 is to be tested (d)Action 3 is to be tested

Figure 5: Schematic diagram of test for first set of movement evaluation



(a)Standard moves (b)Action 1 is to be tested



(c)Action 2 is to be tested (d)Action 3 is to be tested

Figure 6: The test diagram of the second group of action evaluation

The cosine similarity measure method described above was used to calculate the similarity between standard actions and test actions. The results of the similarity assessment for ethnic dance actions are shown in Table 2. The larger the calculated value, the higher the degree of similarity, indicating a better match between the test action and the standard action. The action with the highest total value is considered the most standard action in the evaluation of this group of actions. The evaluation results of the two groups of actions using the method proposed in this paper indicate that the most similar actions are the tested action 2 and the tested action 3, which are consistent with the results observed by the naked eye. This demonstrates the feasibility of the action evaluation method proposed in this paper and also validates the reasonableness of the selected feature vectors, which can be used for the evaluation of ethnic dance actions.

Table 2: Results of similarity assessment of ethnic dance movements

Action name	Action 1	Action 2
Action 1 is to be tested	13.4119	13.2927
Action 2 is to be tested	14.0944	11.7832
Action 3 is to be tested	11.2503	14.5601
Most like action	Action 2 is to be tested	Action 3 is to be tested

III. B. Case Study of Motion Recognition Based on Ethnic Dance

The experiment will use Python programming for ethnic dance motion recognition. The hardware environment includes a Hacker DDR4 2500 GB memory module, a WD10EZEX 1T hard drive, and a GTX graphics card with a capacity of 2.5 GB. To accurately recognize dance movements, OrdroAC5 is used for on-site recording of dance movements, and Kinect is utilized to save the dance movement data, resulting in 15 dance movements. These movements are divided into 6 decomposed movements, and the Kinect-collected joint point data is used as the standard movements.

III. B. 1) Motion capture experiments and analysis

To validate the effectiveness of the proposed method in capturing ethnic dance movements, the experiment compares the joint position errors of the proposed method with those of inertial sensors based on Radial Basis Functions (RBF) neural networks and least squares (MEMS) inertial sensors. The coordinate errors of the three methods in capturing movement errors are compared and analyzed with the standard values collected by Kinect. The results of the coordinate error comparison for dance movements are shown in Table 3. It can be clearly seen that the coordinate error between the movements captured by the proposed method and the Kinect standard values is the smallest across all body parts. Taking the head as an example, the error between the proposed method and the Kinect standard value is only 0.02, while the errors for the other two methods are as high as 1.54 and 0.79, respectively, far exceeding the error of the proposed method. Therefore, it can be concluded that the method proposed in this paper achieves higher accuracy in dance motion capture and is closer to the actual effect.

Table 3: Coordinate comparison results of dance movement errors

Position	Kinect	3D CNNs	RBF	MEMS
Head	1.17	1.15	2.71	1.96
Left upper arm	−0.48	−0.41	−0.71	−2.83
Left forearm	3.14	2.95	4.01	3.21
Right upper arm	1.37	1.32	1.02	−0.22
Right forearm	10.76	11.03	13.02	12.55
Left thigh	15.85	15.93	13.69	14.54
Left leg	4.27	4.51	7.38	3
Right thigh	−5.07	−4.82	−6.62	−7.58
Right leg	−26.54	−27.14	−30.23	−20.02
The left arm is a special plane	4.01	3.79	2.23	3.84
The right arm is very flat	16.89	16.56	16.09	14.92
The left leg is very flat	1.17	1.12	2.01	1.96
The right leg is very flat	−0.48	−0.41	−0.71	−2.83
Chest in a frontal view	3.14	2.95	4.01	3.21
Hip feature plane	1.37	1.32	1.02	−0.22

To more clearly demonstrate the effectiveness of the method proposed in this paper, the experiment will plot the motion trajectories of different methods and the Kinect method, thereby verifying the superiority of this system. The results of dance motion trajectory detection are shown in Figure 7. As can be seen from the motion trajectories, compared to traditional methods based on RBF neural networks and MEMS inertial sensors using the least squares method, the proposed method produces motion trajectories that are closer to those of the Kinect method for dance trainees, indicating that the proposed method can accurately capture dance movements with precision.

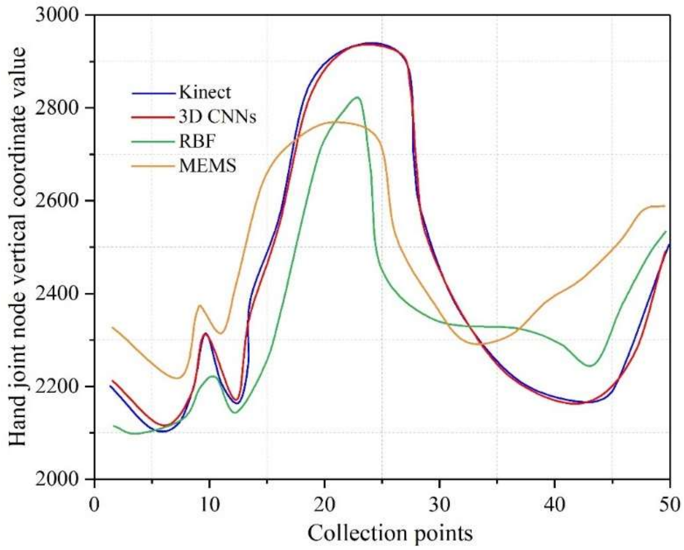


Figure 7: Dance movement trajectory detection results

III. B. 2) Action Recognition Experiment and Analysis

To further validate the automatic recognition performance of this system for dance movements, the experiment tested the recognition accuracy of the system for six types of dance movements: stretching, chest expansion, body rotation, jumping, full-body movements, and warm-up/cool-down exercises. The system was compared with a Kinect-based motion recognition system. The results of dance movement recognition under both systems are shown in Table 4. The results show that, across the six different dance movements, the recognition accuracy of this system reaches up to 99.76% compared to the Kinect-based motion recognition system, with all results maintaining at 95% or above. In contrast, the Kinect-based system achieves a maximum recognition accuracy of 88.34%, with results generally around 86%. Compared to the Kinect-based motion recognition system, our system demonstrates significantly higher recognition accuracy and achieves superior results across multiple dance movements. This indicates that the method proposed in this paper is more effective for recognizing ethnic dance movements.

Table 4: Dance motion recognition results under the two systems

Method	Movement	Discrimination (%)	Reject rate(%)	Misrecognition rate(%)
3D CNNs	Stretching exercise	98.76	1.24	0
	Thoracic expansion exercise	96.23	3.57	0.2
	Kinetic movement	97.78	2.22	0
	Jumper motion	99.76	0.24	0
	Move your whole body	98.92	1.08	0
	Get organized	95.34	4.18	0.48
Kinect	Stretching exercise	87.7	11.22	1.08
	Thoracic expansion exercise	85.65	12.07	2.28
	Kinetic movement	84.8	7.43	7.77
	Jumper motion	87.69	6.06	6.25
	Move your whole body	83.95	5.87	10.18
	Get organized	88.34	5.8	5.86

III. C. New Explorations in the Inheritance and Digital Development of Ethnic Dance Culture

III. C. 1) Cultural Heritage of Ethnic Dance in the Context of Modern Technology

(1) Ethnic dance should inherit tradition.

Ethnic dance culture is very important to China and is a valuable treasure left behind from China's long history. Therefore, ethnic dance should inherit traditional dance culture. For example, some specific ethnic meanings in dance are very important in today's ethnic dance. Only by including all of these elements in the process of inheritance can the culture of ethnic dance be passed on.

(2) Ethnic dance should be developed and innovated

For ethnic dance, there are many important elements that constitute its essence. These elements should be preserved, while those with little practical significance should be removed. New elements should be added to innovate ethnic dance and infuse it with new vitality; Additionally, ethnic dance can leverage modern communication methods for teaching purposes. For instance, the 3D CNNs method proposed in this paper can be used to identify and detect ethnic dance movements. Subsequently, under the backdrop of modern technology, ethnic dance can be disseminated through the internet, enabling more people to understand it and learn it through online videos.

(3) Recognizing the importance of teaching

Ethnic dance is a cultural tradition with a long history and educational significance. Therefore, in order to ensure the effective transmission of ethnic dance, it is essential to recognize the importance of teaching. Only by utilizing digital technology to extract and preserve the content of ethnic dance can its teaching be effectively disseminated, enabling more people to understand and develop a passion for ethnic dance. Through the study of ethnic dance, one can also gain insights into its historical context and the symbolic meanings embedded in its movements.

III. C. 2) Effective Pathways for the Digital Preservation of Ethnic Dance

(1) Expand the scope and frequency of digital resource collection for ethnic dance

Conduct thorough data collection. Research various applicable spaces from different angles and perspectives to enhance the efficiency of digital resource collection and enrich the ethnic dance database. Mobilize the subjective initiative of all parties involved in image collection. Increase public participation in the collection of digital resources for ethnic dance and leverage the strong information collection capabilities of all parties involved. Especially in regions with unique ethnic dances, the initiative of grassroots organizations and individuals should be leveraged. Utilize digital devices such as Kinect depth measurement technology, 3D CNNs methods, multimedia cameras, and stereoscopic laser scanners to conduct on-site filming, scanning, detection, identification, and organization of ethnic dance-related materials. These materials should then be entered into the digital database. Through relevant online platforms, this provides a solid foundation for professionals to conduct specialized research, dance choreography, and other related work.

(2) Creating a database of ethnic dances

In order to create a database of ethnic dances, valuable data should be collected from various perspectives, such as historical culture and dance forms. This includes data on dancers, dance movements, costumes and props, dance music, and other information, which will ensure the effective collection and use of ethnic dances.

(3) Utilizing digital technology to innovate the presentation of ethnic dance

In light of the current state of ethnic dance preservation, we should adopt a forward-looking perspective to promote the preservation and inheritance of ethnic dance in a modern technological environment. We should focus on the cultural space of ethnic dance and promote the organic integration of digital technologies such as Kinect and 3D CNNs with the preservation and inheritance of ethnic dance. While emphasizing the authentic reproduction of ethnic dance, we should also innovate its presentation forms.

(4) Intensify the promotion of digital cultural and creative products related to ethnic dance

By deeply extracting representative elements and connotations of ethnic dance, digital technology can be utilized to design audio souvenirs, smart collectibles, and other cultural and creative products that are popular among consumers. Additionally, QR codes can be incorporated into these products, allowing people to scan them to learn about ethnic dance, and appreciate the unique culture and humanistic charm of minority regions. These digital cultural and creative products not only do not require material costs but are also not constrained by production and processing techniques. They can be presented and disseminated using mobile devices such as smartphones and tablets as output devices. Furthermore, the 3D CNNs method described in this paper can be used to identify ethnic dance movements, and digital materials can be utilized to create short videos, micro-documentaries, and other digital products aligned with public demand, helping people gain a deeper understanding of ethnic dance and promoting its protection and inheritance.

IV. Conclusion

This paper proposes a deep learning-based method for identifying typical movements in ethnic dances, building upon Kinect dance motion detection and tracking technology, to assist in the preservation, inheritance, and digital development of ethnic dances.

(1) The method calculates human 3D joint coordinates and 3D reconstruction at speeds of 26 ms/frame and 28 ms/frame, respectively, demonstrating overall high speed and efficient model application.

(2) The error between the system-calculated values and actual measured values for the two sets of images (0%–0.27%) falls within the error range validated for the model's accuracy. However, as the recognition distance increases, the spatial localization accuracy of 3D joint points decreases, and the number of experiments required to obtain more precise results increases.

(3) The test results for the most similar actions under the proposed method are consistent with those observed by the naked

eye, demonstrating the method's good feasibility. Additionally, the proposed method can accurately capture and recognize the dancer's action positions, maintaining an identification accuracy rate above 95%, clearly indicating its applicability for evaluating ethnic dance movements.

References

[1] Lin, J. C., & Jackson, L. (2022). Just singing and dancing: Official representations of ethnic minority cultures in China. International Journal of Multicultural Education, 24(3), 94-117.

[2] Li, F. (2020). Study on the characteristics of dance art of minorities in West Hunan. Frontiers in Art Research, 2(6), 44-50.

[3] Li, R. (2025). Etiquette and dance—an analysis of the cultural phenomenon of the etiquette and custom dance of the ethnic minorities in Southwest China. Mediterranean Archaeology and Archaeometry, 25(2), 525-529.

[4] He, X. (2025). Meditative Water Drum Rhythm in De'ang Dance. Arts, Culture and Language, 1(3).

[5] Guo, Z., & Yang, J. (2016, January). The Folk Custom Projection of Bloodline Heritage of Ethnic Group—Inspection and Analysis of Gejia People's Music and Dance. In International Conference on Education, Management, Computer and Society (pp. 521-523). Atlantis Press.

[6] Yi, L., & Ling, L. Y. (2025). Embodied Heritage and Cultural Memory: A Comparative Study of Oroqen, Evenki, and Daur Dance Traditions in Inner Mongolia. Journal of Integrated Social Sciences and Humanities, 1-9.

[7] Tang, X. (2023). Research on the Value Construction of Lingnan Dance Culture in the New Era. J. Sociol. Ethnol, 5(11), 96-102.

[8] Blanchier, R. (2017). Mongolian Dance in the Face of Globalisation. Mongolian Responses to Globalisation, 149-174.

[9] Chu, G. (2025). Cultural Identity and Community Resilience: The Role of Ethnic Dance in Maintaining Rural Traditions and History. Mediterranean Archaeology and Archaeometry, 25(3).

[10] Zhang, S., & Wu, C. (2023). Revitalizing endangered traditions: Innovative approaches to safeguarding Yunnan's ethnic minority music as intangible cultural heritage. Herança, 6(1), 101-128.

[11] Gratsiouni, D., Koutsouba, M., Venetsanou, F., & Tyrovola, V. (2016). Learning and digital environment of dance–the case of Greek traditional dance in YouTube. European Journal of Open, Distance and E-learning, 19(2), 80-95.

[12] Harrington, H. (2020). Consumer dance identity: the intersection between competition dance, televised dance shows and social media. Research in Dance Education, 21(2), 169-187.

[13] Zhang, Y., He, X., Wang, J., Bai, X., & Ma, M. (2024, June). Exploration and research on the digital protection methods of ethnic dance. In International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024) (Vol. 13180, pp. 1718-1725). SPIE.

[14] Zhang, C. (2023). Exploration of Ethnic Dance Education and the Development Pathways for Dance Culture Inheritance. Frontiers in Art Research, 5(17), 67-71.

[15] Li, L., & Kang, K. (2021). Exploring the relationships between cultural content and viewers' watching interest: A study of Tiktok videos produced by Chinese ethnic minority groups. In 18th International Conference on e-Business. SCITEPRESS-Science and Technology Publications.

[16] Heyang, T., & Martin, R. (2024). Teaching through TikTok: A duoethnographic exploration of pedagogical approaches using TikTok in higher dance education in China and Norway during a global pandemic. Research in Dance Education, 25(3), 343-358.

[17] Zhao, Y. (2022). Teaching traditional Yao dance in the digital environment: Forms of managing subcultural forms of cultural capital in the practice of local creative industries. Technology in Society, 69, 101943.

[18] Hajdin, M., Kico, I., Dolezal, M., Chmelik, J., Doulamis, A., & Liarokapis, F. (2018, September). Digitization and visualization of movements of slovak folk dances. In International Conference on Interactive Collaborative Learning (pp. 245-256). Cham: Springer International Publishing.

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

[20] Kishore, P. V. V., Kumar, K. V. V., Kiran Kumar, E., Sastry, A. S. C. S., Teja Kiran, M., Anil Kumar, D., & Prasad, M. V. D. (2018). Indian classical dance action identification and classification with convolutional neural networks. Advances in Multimedia, 2018(1), 5141402.

[21] E. Protopapadakis,A. Grammatikopoulou,A. Doulamis & N. Grammalidis. (2017). FOLK DANCE PATTERN RECOGNITION OVER DEPTH IMAGES ACQUIRED VIA KINECT SENSOR. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,XLII-2/W3,587-593.

[22] Nawaf Waqas,Muhammad Islam,Muhammad Yahya,Shabana Habib,Mohammed Aloraini & Sheroz Khan. (2025). Med-ReLU: A Parameter-Free Hybrid Activation Function for Deep Artificial Neural Network Used in Medical Image Segmentation. Computers, Materials & Continua,84(2),3029-3051.