A Study on the Recognition of Typical Movement Characteristics of Ethic Folk Dances Based on Movement Data

Ying Wang
School of Dance, Northwest Normal University, Lanzhou, Gansu 730070, China
Corresponding address: No. 967, Anning East Road, Anning District, Lanzhou City, Gansu 730070, China
Email: wc23470@163.com

Keywords: folk dance, typical movement, motion data, feature recognition, skeleton joint

Received: November 8, 2023

Ethnic folk dances possess significant cultural value and require documentation and preservation. This article begins by recognizing the distinctive movement characteristics found in ethnic folk dances. It then collects skeletal motion data of the human body while executing typical movements in ethnic folk dances using Kinect V2. Two primary features, namely angle and relative distance, were extracted. Deep learning was combined with the attention mechanism to design a three-layer BiLSTM-attention method. Experiments were conducted using the typical movement feature set of ethnic folk dance and the MSR-Action3D dataset. It was found that the three-layer BiLSTM method exhibited superior performance when compared to other configurations of BiLSTM layer. Additionally, the results derived from the BiLSTM model surpassed those achieved with RNN or LSTM models. Furthermore, the inclusion of the attention layer led to a noteworthy 0.0234 increase in the ACC value compared to models without it. The processed features demonstrated enhanced performance compared to the raw skeletal motion data. ACC values exceeding 0.95 were achieved for the recognition of typical movement features in various types of ethic folk dances. Notably, the ACC value of the three-layer BiLSTM method for the MSR-Action3D dataset was 0.9767, which was superior to the other methods. These outcomes validate the robustness of the methodology presented in this paper for recognizing typical movement features in folk dance and suggest its potential for practical applications.

Povzetek: Raziskava predstavlja analizo značilnih gibalnih vzorcev ljudskih plesov s pomočjo tri nivojske arhitekture z globokim učenjem.

1 Introduction

Dance, as a performing art, has evolved significantly in response to societal developments. Its styles have become increasingly diverse, encompassing traditional folk dance, modern dance, jazz dance, and more. As a result, it has gained popularity among a growing number of enthusiasts [1]. Ethnic folk dance, in particular, emerges from the unique cultural and regional influences of each nation, encapsulating rich national culture and spirit. Consequently, the study of ethnic folk dance plays a pivotal role in preserving and spreading cultural heritage. Typically, each ethnic folk dance has its own unique movements and gestures. By recognizing these characteristic movement features, it is possible to classify and document various ethnic folk dances. However, recognizing the distinctive features of ethnic folk dances poses great difficulty due to the complex variations in their movements. Fortunately, with the continuous advancement of emerging technologies like sensors and computers, an increasing number of methods have been applied to the field of movement recognition [2].

2 Related works

According to Table 1, currently in the field of action recognition, most research focuses on human daily activities, with some studies also exploring gesture movements. However, there is relatively little research on dance movements. Furthermore, when it comes to discussing the application of deep learning methods in movement recognition, there is still potential for further improving recognition effectiveness. Additionally, existing features mostly come from videos and sensors, with limited consideration given to skeletal motion data. Therefore, this article focuses on the recognition of typical movement characteristics in ethnic folk dance. Based on Kinect device to acquire skeletal motion data, a method using long short-term memory (LSTM) neural network was designed. Through experimental analysis, the effectiveness of this method has been proven, providing a new approach for recording and inheriting ethnic folk dance and offering theoretical support for further application of Kinect device in movement recognition.

Table 1: A summary table of related works

Literature	Feature	Recognition method	Results
Gao et al. [3]	Integration of deep video and RGB trichromatic features	A global coding algorithm and a convolutional neural	An average classification accuracy rate of 85.79%

		network	
Barkoky et al. [4]	Complex network-based features extracted from RGB-D data	Meta-paths in complex networks	Good recognition performance on both MSR-Action Pairs and MSR Daily Activity3D
Li et al. [5]	M-mode ultrasound	A support vector machines (SVM) and a back-propagation neural network (BPNN)	The average classification accuracy was 98.83% ± 1.03% for SVM and 98.70% ± 0.99% for BPNN.
Athavale et al. [6]	Daily human activity signals recorded by cell phone accelerometers	VGG16-SVM	An accuracy of 79.55% and an F-value of 71.63%

3 Skeleton motion data

In the realm of movement recognition, commonly utilized motion data include RGB data, and optical flow data [7]. Skeleton data comprises the 3D coordinates of skeletal joints and offers several advantages, such as smaller dimensions, ease of acquisition, and good stability. This type of data can be directly obtained through depth sensors like Kinect [8], which circumvent issues associated with traditional camera-based motion data collection, such as susceptibility to light and background interference. Consequently, skeleton data has gained increasing prominence in the field of movement recognition. This paper predominantly utilizes the Kinect sensor as the device for acquiring the motion data to facilitate the capture of skeletal motion data in the context of folk dance.

The performance of Kinect V2 has been improved in all aspects compared to Kinect V1. The comparison of the two devices is shown in Table 2. It can be found that the detection range of Kinect V2 is larger and the number of joints detected is more. As a result, it has better performance in extracting skeleton motion data. This study used the Kinect V2 as the acquisition device.

Table 2: Comparison of two generations of Kinect devices

Equipment parameters	Kinect V1	Kinect V2
Number of people	6	6
Number of detected joints	20 joints/person	25 joints/person
Effective detection range	0.8-4.0 m	0.5-4.5 m
Horizontal angle detection range	57°	70°
Vertical angle detection range	43°	60°
USB interface	2.0	3.0
APP	Single	Multiple

Kinect achieves data acquisition and processing through the Kinect SDK. The Kinect SDK comprising various tools and software libraries, serving as the foundation for human-computer interaction development. The process of extracting skeletal motion data is as follows. The Kinect SDK analyzes a single depth image, classifies different body parts using a random decision forest to determine whether each pixel belongs to a skeletal joint, and then updates the three-dimensional coordinates in real time. This algorithm operates at a speed of approximately 5 milliseconds per frame, providing robust real-time performance and effectively meeting practical requirements.

The 25 joints extracted by the Kinect V2 are shown in Figure 1.

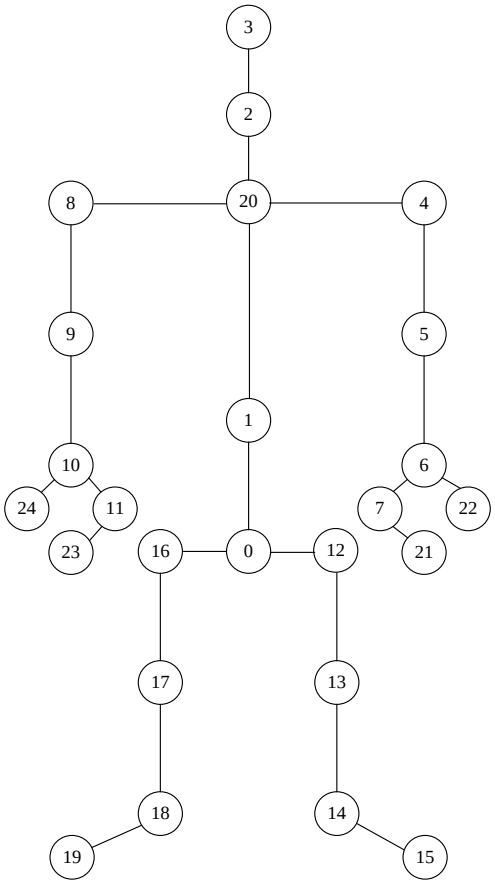

Figure 1: Kinect V2 skeletal joint nodes.

The 25 joint nodes shown in Figure 1 contain a detailed division of hand and foot joint nodes, but in the actual application process, too many joint nodes may instead lead to an increase in the amount of computation and noise;

therefore, considering the actual movement of the human body in the process of accomplishing the folk dance, this paper simplified the skeletal joint nodes. Finally, the 15 joint nodes shown in Table 3 were used for computation.

Table 3: Simplified 15 skeletal joint nodes

Serial number	Name
0	Spine base
3	Head
4	Shoulder left
5	Elbow left
6	Wrist lest
8	Shoulder right
9	Elbow right
10	Wrist right
12	Hip left
13	Knee left
14	Ankle left
16	Hip right
17	Knee right
18	Ankle right
20	Spine shoulder

The Kinect coordinate system is written as (x_k, y_k, z_k), and the human body coordinate system is written as (x_h, y_h, z_h). It is assumed that the mapping of the angle between the shoulder line and X_k on the $X_k O Z_k$ plane is α, then the skeletal joint node is converted from the Kinect coordinate system to the human body coordinate system. The equation is:

$$\begin{bmatrix} x_h \\ y_h \\ z_h \end{bmatrix} = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}.$$

The processed skeletal joint nodes' coordinates have certain limitations if directly used as the input for subsequent movement feature recognition methods, which would require a significant amount of computation and increase computational complexity. Therefore, the coordinate data need to be processed again to extract more representative features. In feature extraction, it is important to consider that the features should not vary significantly due to differences in human skeletal structure, while still effectively reflecting changes in human movement. During the execution of ethnic folk dances, various joints of the body exhibit different angles, and these angles also differ for different dance movements. Additionally, there are certain patterns in the variations of relative distances. This paper extracted the following two types of features in terms of angle and distance considerations.

(1) Angles. During the movement process, the skeletal joint nodes will have different sizes of angles. As shown in Figure 2, taking the angle formed by joint nodes 4, 5, and 6 as an example, it consists of two vectors, i.e., the vector formed by joint nodes 5 and 4, and the vector formed by joint nodes 5 and 6. The former is defined as r_i, and the latter is defined as r_j. The formula for θ can be written as:

$$\theta_n = arccos \frac{r_{i1} \times r_{j1} + r_{i2} \times r_{j2} + r_{i3} \times r_{j3}}{\sqrt{r_{i1}^2 + r_{i2}^2 + r_{i3}^2} + \sqrt{r_{j1}^2 + r_{j2}^2 + r_{j3}^2}}.$$

By this method, the 15 angles in Figure 2 can be calculated.

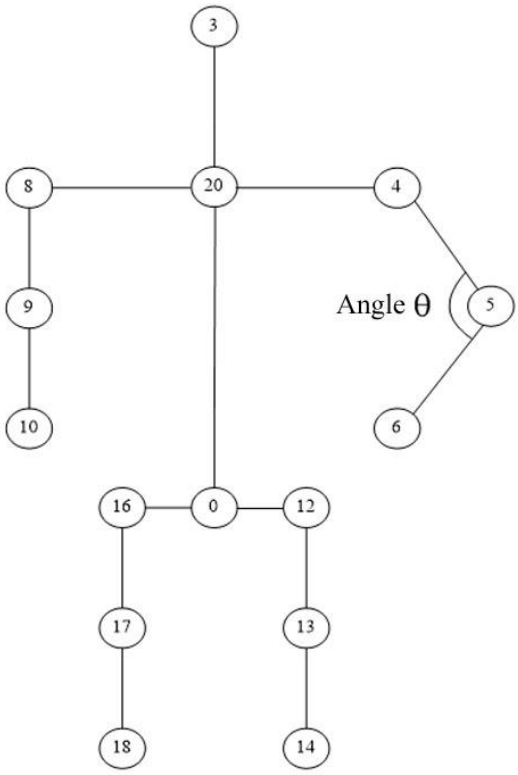

Figure 2: Schematic diagram of joint angles.

(2) Relative distance. During the movement process, the spatial position of the skeletal joint nodes will also change. For example, when performing various different folk dances, the hands, feet, and human spine have different relative distances; therefore, this paper considers joint node 0, i.e., the spine base, as the center point. The calculation formula is:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}.$$

According to the above equation, the relative distance between the remaining 15 joints and the center point can be calculated.

To avoid the influence of individual differences on the results, the calculated relative distances were normalized:

$$D_{ij}' = \frac{D_{ij}}{d}$$

where d is the distance between joint node 20 and joint node 0 in Figure 2.

4 Motion feature recognition method

The recurrent neural network (RNN) has proven to be highly effective in addressing temporal problems [9]. However, RNNs suffer from the problems of vanishing and exploding gradients [10], as well as limitations in handling long input sequences and preserving long-term dependencies. In contrast, LSTM is a type of RNN that can learn long-term dependencies [11]. They not only alleviate the issues of vanishing and exploding gradients but also effectively filter temporal information through gate structures, allowing for flexible retention of important information. Given that skeletal movement data inherently contains abundant temporal information, this paper leveraged the LSTM to design a movement feature recognition method.

LSTM realizes the processing of information through three gates, and its application to the recognition of movement features of folk dance can obtain better results. It is assumed that the input of the current moment is x_t, the output of the previous moment is h_{t-1}, and LSTM measures the degree of updating the input information through input gate i_t:

$$i_t = \sigma(U_i \cdot [h_{t-1}, x_t] + b_i).$$

Forget gate f_t is used to control the forgetting degree of historical information, which can be written as:

$$f_t = \sigma\big(U_f \cdot [h_{t-1}, x_t] + b_f\big).$$

At time t, the update formula of a memory unit can be written as:

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(U_c \cdot [h_{t-1}, x_t] + b_c).$$

Output gate o_t is used to control the output quantity of output value h_t in the LSTM, which can be written as:

$$o_t = \sigma(U_o \cdot [h_{t-1}, x_t] + b_o).$$

Finally, output gate o_t and unit state c_t jointly determine the output of LST. It is expressed by the following equation:

$$h_t = o_t \tanh(c_t),$$

where U and b denote the weight and bias term of each layer.

LSTM has a limitation in that it cannot encode information from the backward to forward. The human skeletal motion data is very complex. Unidirectional LSTM can only process the motion data in one direction, while BiLSTM can simultaneously capture all information from both the forward and backward directions. In order to extract more features from the data, this paper adopted BiLSTM, which learns features through a forward LSTM and a backward LSTM. It is expressed as:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(h_{t-1}, x_t, c_{t-1}),$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(h_{t+1}, x_t, c_{t+1}).$$

Then, the final hidden layer output of BiLSTM is:

$$h_t = \overrightarrow{h_t} + \overleftarrow{h_t}.$$

In addition, in order to avoid the inadequacy of one-layer BiLSTM in data feature extraction, this paper proposed a multi-layer BiLSTM and combined it with the attention mechanism [12], so as to further improve the performance of movement feature recognition. The established three-layer BiLSTM-attention structure is shown in Figure 3.

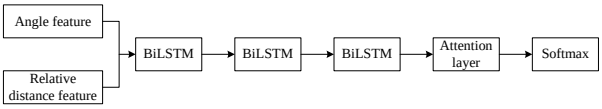

Figure 3: Three-layer BiLSTM-attention structure.

As in Figure 3, the features are inputted into the BiLSTM for learning, and then the output of the previous BiLSTM layer is used as the input of the next BiLSTM layer to further mine the deeper associations between features and strengthen the feature learning ability. Then, output h_t of the three-layer BiLSTM is sent to the attention layer, and weights are assigned according to the importance of features. The corresponding equations are:

$$h_t' = \sum_{t=1} a_t\, h_t,$$
$$a_t = softmax(score_t),$$
$$score_t = f(Query, h_t),$$

where a_t is the weight corresponding to each feature vector, which is calculated by $score_t$, and f is a scoring function. The relational degree between h_t and current object $Query$ is calculated to realize the assignment of attention. Then, h_t' output from the attention layer undergoes softmax classification to obtain the final results for ethic folk dance movement feature recognition:

$$\hat{y} = softmax(h_t').$$

Using the cross entropy as a loss function, the model is trained with the goal of minimizing the cross entropy:

$$L = -\sum_{i=1}^{C} y_i \log \hat{y}_i,$$

where C is the movement category.

5 Results and analysis

5.1 Experimental setup

The experiments were carried out on a Windows 10 computer, and the algorithm was implemented using the Keras framework. In the three-layer BiLSTM-attention model, the hidden layer comprised 512 neurons. Dropout was used and set as 0.3 to mitigate overfitting. The input nodes were set as 30 dimensions, i.e., there were 30 input nodes. The learning rate was set at 0.001, and the total number of iterations was fixed at 1,000.

The experimental data comprised two distinct parts. The first part consisted of a typical movement feature set derived from ethnic folk dances, which were collected in the laboratory from 50 dance majors. These dances included five different styles: Uighur, Mongolian, Han, Tibetan, and Dai dances. For each dance style, a representative movement was selected, as depicted in Figure 4. The description of different types of ethnic folk dances is as follows.

(1) Uighur dance: The basic characteristics of Uyghur dance in terms of posture are upright head, chest out, and straight waist. It achieves graceful movements through continuous slight tremors or variations in knee movements, as well as decorative actions such as neck movement and wrist flipping.

(2) Mongolian dance: Mongolian dance is robust, graceful, and bold, characterized by continuous undulations and a combination of softness and strength. It relies on the coordination of various parts of the body to meet rhythmic requirements.

(3) Han dance: The movements of Han ethnic dance are graceful and fluid, with unique hand gestures that showcase the dancers' elegance and strength through body rotations, twists, and bends. The footwork is light and agile, emphasizing precision and rhythm.

(4) Tibetan dance: Tibetan dances often involve standing or half-squatting postures, with graceful movements such as jumps and spins. The footwork is agile and varied, allowing the dancers to showcase their flexibility through bending the waist, arching the back, and loosening the hips.

(5) Dai dance: Dai ethnic dance showcases the flexibility of the lower legs, characterized by graceful knee movements. It forms a basic rhythm through the bending of arms and various joints in the body, resulting in elegant and intricate hand and foot gestures.

Each student refrained from strenuous exercise for 24 hours prior to data collection. The students were then guided by experimental staff to perform the specified movements within the effective range of the Kinect device. Each movement was executed ten times, resulting in the collection of 500 samples per movement and a total of 2,500 samples. If there was any missing action sequence frame, the skeletal data from the previous frame was used for filling. Furthermore, in order to reduce information redundancy in the original data, only one frame was retained for every five frames. To maintain uniformity in the number of frames across samples, the last frame was duplicated to match samples with fewer frames than the maximum frame count. The second part of the experimental data was a widely used dataset in the field of movement recognition known as MSR-Action3D [13]. This dataset included 20 distinct movements performed by ten individuals. As each movement was executed 2-3 times, there were 557 samples of movement. To ensure reliable results, both datasets underwent experimentation utilizing a five-fold cross-validation.

Figure 4: Typical movement characteristics of ethnic folk dances (from left to right: Uighur dance: double hat support; Mongolian dance: soft arms; Han dance: push the fan horizontally with both hands diagonally downward; Tibetan dance: peacock reflecting its image by the water's edge; Dai dance: small jumps by stomping and bending legs).

The performance of the algorithm was assessed using the accuracy of movement feature recognition, i.e., the ratio of correctly recognized samples to the total samples, which is expressed as:

$$ACC = \frac{TP}{TP+FN},$$

where TP stands for the number of samples that are actually positive and also identified as positive, and FN stands for the number of samples that are actually positive but are recognized as negative.

5.2 Analysis of results

Firstly, the performance of the three-layer BiLSTM-attention method was analyzed using the feature set of folk dance movements to compare the effect of different BiLSTM layers on the performance. The results are presented in Figure 5.

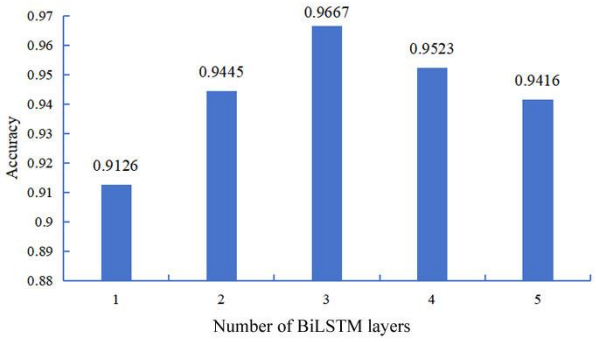

Figure 5: Effect of number of BiLSTM layers on recognition performance.

From Figure 5, it can be found that the performance of the algorithm was extremely poor when only one layer of BiLSTM was used, and the ACC value for recognizing folk dance movement features was 0.9126, which indicated that one layer of BiLSTM did not adequately learn the features. Then, as the number of BiLSTM layers increased, the ACC value of the algorithm also increased. It reached a maximum of 0.9667 when there were three BiLSTM layers, which was 0.0541 higher compared with that of one layer. In the case where the number of BiLSTM layers continued to increase, the complexity of the model

also increased, which brought a burden to the recognition of ethnic folk dance movement features, resulting in a decrease in the ACC value. Specifically, when there were five BiLSTM layers, the ACC value was 0.9416, which was 0.0251 lower compared to three layers. The results in Figure 4 showed that the best results could be obtained when using a three-layer BiLSTM.

Then, the effect of replacing BiLSTM with RNN, LSTM, and removing the attention layer on the performance was compared, and the results are displayed in Table 4.

Table 4: Effect of BiLSTM layer and attention layer on performance.

	ACC value
Three-layer RNN-attention	0.8962
Three-layer LSTM-attention	0.9189
Three-layer BiLSTM	0.9433
Three-layer BiLSTM-attention	0.9667

From Table 4, when replacing BiLSTM with RNN, the ACC value of the algorithm became 0.8962, which was reduced by 0.0705 compared to the three-layer BiLSTM-attention method. When replacing BiLSTM with LSTM, the ACC value of the algorithm became 0.9189, which was reduced by 0.0478. This suggested that BiLSTM outperformed LSTM in the learning of folk dance movement features. The comparison of the results obtained in the presence and absence of the attention layer demonstrated that the ACC value of the three-layer BiLSTM method without adding the attention layer was 0.9433, which was reduced by 0.0234 compared with adding the attention layer. The result revealed the importance of the attention layer.

In the feature processing, this paper chose the angle and relative distance features extracted from the preprocessed skeleton motion data. The effects of different choices of features on the recognition results were compared, and the results are presented in Table 5.

Table 5: Effect of feature selection on performance.

	ACC value
Original data	0.8527
Angle features	0.9216
Relative distance features	0.9423
Angle features + relative distance features	0.9667

From Table 5, the ACC value obtained by inputting the original data into the three-layer BiLSTM-attention method for recognition was 0.8527. Under this condition, the training data was too large, which made training difficult. Additionally, the features extracted by the algorithm from the original skeleton movement data were insufficient to effectively distinguish between different categories of ethnic folk dances. As a result, the ACC value was also low. When using angle features or relative distance features individually, the ACC values were 0.9216 and 0.9423, respectively, which were improved by 0.0689 and 0.0896 compared to using the original features. This suggested that the relative distance features contributed more to the recognition of ethnic folk dance movement features and contained more information. Finally, the ACC value of the algorithm in the case of using angle features + relative distance feature was 0.9667, which was improved by 0.114 compared to the case of using the original features. This result proved the reliability for the processing of skeleton movement data.

The recognition performance of the three-layer BiLSTM-attention method for different categories of ethnic folk dances in the five-fold cross-validation is shown in Table 6.

Table 6: Recognition performance for different categories of folk dances.

	Uighur dance	Mongolian dance	Han dance	Tibetan dance	Dai dance
1	0.9732	0.9824	0.9661	0.9552	0.9577
2	0.9821	0.9644	0.9634	0.9502	0.9662
3	0.9882	0.9712	0.9516	0.9531	0.9656
4	0.9884	0.9778	0.9717	0.9507	0.9703
5	0.9841	0.9712	0.9532	0.9513	0.9582
Average	0.9832	0.9734	0.9612	0.9521	0.9636

From Table 6, it can be found that the results of the algorithm obtained in the five experiments were relatively stable, and the differences in ACC values were small. Then, from the comparison of different categories, the algorithm exhibited the best performance in recognizing Uighur dance, reaching an ACC value of 0.9832, followed by Mongolian dance with an ACC value of 0.9734, and it achieved the lowest ACC value (0.9521) when recognizing Tibetan dance. These results demonstrated that the three-layer BiLSTM-attention method achieved ACC values above 0.95 for recognizing various categories of ethnic folk dance movement features.

Finally, the method proposed in this paper was compared with other methods in the literature using the MSR-Action3D dataset:

(1) differential RNN: an LSTM incorporating differential gating scheme [14];

(2) linear SVM [15];

(3) spatio-temporal LSTM (ST-LSTM) [16].

The comparison of the results is presented in Figure 6.

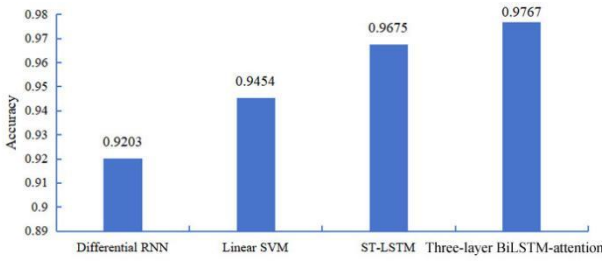

Figure 6: Comparison with other methods.

The ACC values of these methods for the MSR-Action3D dataset were above 0.9 in Figure 6. In comparison, the differential RNN method had the lowest ACC value (0.9203), while the three-layer BiLSTM-attention method obtained the highest ACC value, which was 0.0564 higher than the differential RNN method, 0.0313 higher than the linear SVM method, and 0.0092 higher than the ST-LSTM method. These results verified the recognition performance of the three-layer BiLSTM-attention method.

6 Discussion

Movement recognition is an important research direction in computer vision, aiding computers in identifying human or object movements from data. It has wide applications in fields such as video surveillance, human-computer interaction, and virtual reality. With the advancement of technology, deep learning methods have made significant progress in movement recognition, with models like RNN being widely used for feature extraction and identification of movements. The performance of movement recognition methods depends on the diversity of training data, while data collected based on images or videos is likely to be affected by changes in shooting angles and positions, which can impact the performance of movement recognition. Additionally, the collaborative motion of multiple parts also increases the complexity of movement recognition. Currently, most movement recognition methods are based on daily human activities such as walking and gestures. However, ethnic folk dances involve complex and diverse motion variations, making it difficult for traditional movement recognition approaches to achieve satisfactory performance. Therefore, this study combined skeleton motion data with deep learning methods to investigate the approach for recognizing the motion characteristics of ethnic folk dances.

The original skeletal data is too complex and contains a lot of redundant information, which hinders the performance and efficiency of movement feature recognition. Based on the characteristics of ethnic folk dance movements, this paper proposed using joint angles and relative distances as features. To improve the recognition performance of LSTM, this paper designed a three-layer BiLSTM-attention method and conducted experiments on two datasets. Firstly, from the recognition results of ethnic folk dances, both the feature selection method and action recognition method designed in this paper demonstrate excellent performance. They are capable of capturing valuable features for action recognition more effectively from skeletal motion data. The utilization of BiLSTM structure and the addition of an attention layer contribute to improving the accuracy of dance movement recognition. The recognition accuracy for different types of ethnic folk dances was consistently above 0.95. Then, the method designed in this paper was compared with other current recognition methods on the MSR-Action3D dataset, further demonstrating the effectiveness of this approach in movement recognition.

The research in this article provides a novel and reliable method for recognizing the movement characteristics of ethnic folk dances, which contributes to the preservation and inheritance of ethnic folk dance culture. It is of great significance for documenting and protecting some endangered ethnic folk dances that are on the verge of being lost. Additionally, this research further enriches the field of movement recognition, promoting the development of deep learning and movement recognition.

7 Conclusion

This paper primarily focuses on the movement feature recognition method for folk dance based on skeleton movement data. A three-layer BiLSTM-attention method was developed. Through experimental analysis, it was observed that the algorithm achieved its optimal performance in recognizing movement features of folk dance when employing three BiLSTM layers. Removing the attention layer resulted in a decrease of 0.0234 in the ACC value. The algorithm demonstrated the highest ACC value of 0.9667 when utilizing both the angle and relative distance features. Furthermore, the ACC values for recognizing various types of folk dance movement features consistently exceeded 0.95. Finally, the proposed method outperformed other techniques when the MSR-Action3D dataset was used, further confirming its reliability for movement recognition. The three-layer BiLSTM-attention method holds promise for broader application and implementation in practical contexts. However, there are still some limitations in this study. For instance, the experimental dataset was limited and did not cover a wider range of ethnic folk dance movements. Therefore, in future research, the author will expand the scope and quantity of data collection to further validate the proposed method's reliability in recognizing characteristics of ethnic folk dance movements. Additionally, the author will explore the possibility of applying this method to other fields such as sports movement recognition and investigate methods to improve recognition performance by comparing and analyzing the effects of various deep learning approaches on movement recognition.

References

[1] Schupp K (2018). Dance Competition Culture and Commercial Dance: Intertwined Aesthetics, Values, and Practices. *Journal of Dance Education*, 19, pp. 1-10. https://doi.org/10.1080/15290824.2018.1437622.

[2] Li L (2021). Mirror motion recognition method about upper limb rehabilitation robot based on sEMG. *Journal of Computational Methods in Sciences and Engineering*, 21, pp. 1021-1029. https://doi.org/10.3233/JCM-204812.

[3] Gao P, Zhao D, Chen X (2020). Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework. *IET Image Processing*, 14, pp. 1257-1264. https://doi.org/10.1049/iet-ipr.2019.0588.

[4] Barkoky A, Charkari N M (2022). Complex Network-based features extraction in RGB-D human action recognition. *Journal of Visual Communication & Image Representation*, 82, pp. 1-9. https://doi.org/10.1016/j.jvcir.2021.103371.

[5] Li J, Zhu K, Pan L (2022). Wrist and finger motion recognition via M-mode ultrasound signal: A feasibility study. *Biomedical Signal Processing and Control*, 71, pp. 1-11. https://doi.org/10.1016/j.bspc.2021.103112.

[6] Athavale V, Kumar D, Gupta S (2021). Human Action Recognition Using CNN-SVM Model. *Advances in Science and Technology*, 105, pp. 282-290. https://doi.org/10.4028/www.scientific.net/AST.105.282.

[7] Ji Y, Yang Y, Shen F, Shen H, Zheng WS (2020). Arbitrary-view Human Action Recognition: A Varying-view RGB-D Action Dataset. *IEEE Transactions on Circuits and Systems for Video Technology*, 31, pp. 289-300. https://doi.org/10.1109/TCSVT.2020.2975845.

[8] Li G, Li C (2020). Learning skeleton information for human action analysis using Kinect. *Signal Processing Image Communication*, 84, pp. 1-5. https://doi.org/10.1016/j.image.2020.115814.

[9] Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L, Brunner D (2018). Reinforcement Learning in a large scale photonic Recurrent Neural Network. *Optica*, 5, pp. 1-5. https://doi.org/10.1364/OPTICA.5.000756.

[10] Huang Y, Bai C, Li H, Zhang J, Chen S (2020). Image Captioning Based on Conditional Generative Adversarial Nets. *Journal of Computer-Aided Design & Computer Graphics*, 32, pp. 911-918. https://doi.org/10.3724/SP.J.1089.2020.18003.

[11] Kumar J, Goomer R, Singh AK (2018). Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters. *Procedia Computer Science*, 125, pp. 676-682. https://doi.org/10.1016/j.procs.2017.12.087.

[12] Zang Y, Yu Z, Xu K, Chen M, Yang S, Chen H (2023). Fiber communication receiver models based on the multi-head attention mechanism. *Chinese Optics Letters*, 21, pp. 1-6. https://doi.org/10.3788/COL202321.030602.

[13] Gharahdaghi A, Razzazi F, Amini A (2021). A non-linear mapping representing human action recognition under missing modality problem in video data. *Measurement*, 186, pp. 1-10. https://doi.org/10.1016/j.measurement.2021.110123.

[14] Veeriah V, Zhuang N, Qi GJ (2015). Differential Recurrent Neural Networks for Action Recognition. *2015 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Santiago, Chile, pp. 4041-4049, https://doi.org/ 10.1109/ICCV.2015.460.

[15] Vemulapalli R, Arrate F, Chellappa R (2014). Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, OH, USA, pp. 588-595, https://doi.org/ 10.1109/CVPR.2014.82.

[16] Liu J, Shahroudy A, Xu D, Kot AC, Wang G (2017). Skeleton-Based Action Recognition Using Spatio-Temporal LSTM Network with Trust Gates. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 40, pp. 3007-3021. https://doi.org/10.1109/TPAMI.2017.2771306.