J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 4023--4039

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

A Study on Choreography Strategies of Ethnic Folk Dance Movements in Southwest China Based on Graph Neural Networks

Shuai Zhao[1,✉]

[1] School of Digital Media and Performance, Sichuan Geely College, Chengdu, Sichuan, 641423, China

ABSTRACT
Ethnic folk dance in Southwest China is known for its unique regional characteristics and cultural background, and the optimization of its movement choreography strategy is especially crucial for the inheritance and development of this artistic influence. In this study, an optimized graph neural network model is used to choreograph the movements of folk dances in Southwest China. The model is equipped with multi-feature fusion, spatial modeling and temporal modeling modules, which can maximize the recognition performance of the graph neural network model. Based on the model, a framework for automatic generation of folk dance movements is designed, and the model is trained and validated using Laban-16 and Laban-48 dance movement datasets. The experimental results show that the method of this paper is well tested, and the loss value and accuracy convergence algebra of the training set and the test set are basically the same, reaching 0.25 and 96%, respectively. The lower limb motion recognition rate on Laban-16 dataset is improved by 5.21%~15.81% compared with the comparison model. Under the music of different rhythms, a variety of dance movements can be reasonably choreographed to, and the feasibility score of the model by experts is between 85 and 95, indicating that the model in this paper has practical value.

Keywords: graph neural network, multi-feature fusion, spatial modeling, temporal modeling, folk dance movement choreography

1. Introduction

As we all know, Chinese folk dance is a relatively representative form of folk art, which has an extremely deep mass base in China and is an important part of Chinese national culture. Its unique artistic and humanistic values have enriched Chinese culture. The folk dance of Southwest ethnic minorities is a typical folk art, with its distinctive dance forms and strong narrative characteristics. In the beginning, these folk dances were very loose, without the slightest generalization and systematic,

✉ Corresponding author.
E-mail address: 18280330618@163.com (S. Zhao).
Received 20 February 2024; Revised 15 May 2024; Accepted 20 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-227

scattered in different tribes, mainly in weddings and funerals, rituals and these occasions. However, with the continuous development of society and the protection of national traditional culture, these folk dances have undergone some changes and gradually developed and expanded. Therefore, it is necessary to express this extremely vivid and rich dance form with deep and accurate dance connotation through dance movement choreography to increase the artistic value and appreciation value of the dance.

Traditional dance choreography is time-consuming and labor-intensive, requiring a large amount of manual participation and repetitive labor, which seriously impedes the further development of ethnic folk dance [1]. The development of neural networks makes it possible to train models based on massive data, but the scarcity of paired data for dance and music has limited the research of intelligent choreography [2]. The diversity of dance requires a wide variety of dances and rich combinations of movements, and even for experienced motion capture companies, building a large and compliant dataset is quite a daunting task [3-4]. Nowadays, digital platforms have become an increasingly important part of people's lives, and the realization of storing, organizing classifying, retrieving, storing, and utilizing multimedia data such as dances through dance intelligent choreography and presentation systems can not only significantly promote the research and development of intelligent choreography, but also enrich people's lives [5-6].

Literature [7] explores the connection between traditional folk dance and modern choreography, showing that the rich cultural connotations of folk dance art forms provide a lot of creative inspiration for modern choreography, while the control of music and rhythm in the process of modern choreography increases the appreciation value of folk dance. Literature [8] shows that folk dance allows for stylistic changes and improvisation characteristics for the modern choreography process creates difficulties, there is an urgent need for high-quality digital technology to obtain the folk dance movement sequences and create them, in order to promote the inheritance and dissemination of folk dance. Literature [9] takes Yunnan traditional dance as the object to study the innovative creation process of folk dance, and analyzes the choreographic creation concept of traditional folk dance from the starting point of movement extraction, choreography creation and emotional expression. Literature [10] analyzes the worldwide trend of dance design and finds that the main feature of modern choreography is the amazing combination of modern dance techniques and traditional art forms, which provides theoretical support for the development of modern choreography of Chinese folk dance.

Literature [11] uses motion capture technology to accurately capture the movement details of folk dance performers, and learns and analyzes the extracted movement data through 3D convolutional neural network in machine learning, which significantly improves the accuracy, real-time and generalization performance of movement recognition. Literature [12] developed DeepDance, a cross-modal association framework based on graph convolutional neural networks, which is capable of creating one-to-many mapping relationships between music and dance movements, as a way to realize the generation of corresponding dance sequences based on music inputs, which plays an important role in the creation of improvisational dance choreography. Literature [13] emphasizes that dance choreography needs to fully consider the complex coordination between body movements and music, and proposes a choreography generation model that integrates the gated Transformer, graph neural network (GNN), and temporal convolutional network (TCN), which can not only produce creative dance movements synchronized with the music, but also capture the nuances of the art film and television. Literature [14] proposes an instruction set framework based on a data-driven learning strategy that can choreograph high-quality dance movements to a given music, by beat-aligning movement sequences with music clips to synthesize nodes of a dynamic motion graph, combined with stylistic and rhythmic evaluations, and ultimately achieving aesthetically consistent movement generation. Literature [15] proposes a music-oriented graph convolution GRU model for dance choreography, which can model the relationship between human skeleton structure and dance time sequences, and fuses the graph features by establishing path attention blocks, so that the generated dance motion graph is consistent with the music, and effectively solves the dance choreography task. Literature [16] proposes a multimodal convolutional self-encoder for the problem of prediction error accumulation in recurrent neural network choreography models, which can generate novel dance movement sequences of arbitrary length by taking 2D skeletal features and audio information as input conditions.

This paper firstly outlines the macro and micro factors affecting the choreography of folk dance movements, and specifically divides the micro factors into five categories: strength, smoothness, posture, time and rhythm, and indicates the points of attention for the use of graph neural network to choreograph dance movements. Then, through the acquisition of dance movement data and dance movement database, the automatic generation framework of dance movement based on improved graph neural network is designed. And later on, the modified graph neural network model is described in detail in terms of multi-feature fusion, spatial modeling, and temporal modeling. At the end of the article, specific experiments are conducted to evaluate the effect and feasibility of the model constructed in this paper based on the choreography of folk dance movements in Southwest China.

2. Factors affecting the choreography of folk dance movements

2.1. Choreography of Dance Movements under Macro Factors

To study the choreography of folk dance movements, we have to start from the local culture, history and living habits of the ethnic groups. Because of different living habits for folk dance unique dance movements create a superior subject matter, folk dance from the daily material elements, the use of unique movement art expression. Ethnic dance is a national characteristic, each ethnic dance movement is unique, is the embodiment of the way of life of each ethnic group, cultural development, religion and other content, so ethnic dance can not be unified form of presentation.

For example, the same Yi dance, the Yi dance of Daliang Mountain in Sichuan Province and the Yi dance of Honghe Prefecture in Yunnan Province have very different styles. From a macro point of view to analyze, although the same Yi dance, but different geographic environment, cultural background and religious beliefs affect the unique dance movements of the ethnic groups, thus forming the same type of dance, but has a different form of expression of dance movements.

2.2. Micro-factors affecting movement choreography

Ethnic dance movements usually have very strong national characteristics, dance movements in different geographic environments will be affected by it, in order to penetrate into the hearts of the people for the creation of ethnic dance, it is necessary to start from the basic constituent elements of the dance dance movements. The factors affecting the movement of folk dance are analyzed from a micro point of view in the following points.

2.2.1. Force factors. The degree of muscle exertion in dance is the expression of the element of strength. The control of the strength is the key to show the characteristics of folk dance, because each folk dance to express the spirit and emotion is different, thus requiring different strength. Controlling the strength can make the dance more national characteristics, the rhythmic beauty of different dances and the different coordination of the beauty of different national dances make the difference between the characteristics of the dance at a glance.

2.2.2. Fluidity factors. The transition links between multiple movements are a reflection of the overall fluidity of the dance. The effect of the dance and the feeling given to people should be mediated by the fluency factor. The graphic neural network movement choreography is to bring together the body parts of the dancers, according to the requirements of the dance work, and the body language of the dancers through different gestures and postures according to the needs of the dance situation. The main factor in this process of change is to grasp the fluidity between movements.

2.2.3. Attitude factors. The posture of the folk dancer demonstrates the artistic embodiment of the dancer's spatial range. The control and expression of the dance posture is an obvious feature to show the characteristic folk dance. Each folk dance is formed by each movement posture of the body, dance posture elements in folk dance is very important, because the use of graphic neural network choreography dance movements to express the spirit of the nation, the custom should be presented through the dance posture, to give the viewer a visual impact, is to watch intuitively feel the connotation of the nation.

2.2.4. Time factor. The time factor is the length of the interval between the various movements of folk dance, which expresses the speed of the movements and the continuity of the movements themselves in folk dance. There is also a big difference in the speed or rhythm of each folk dance, for example, rapid body changes in the jumping and fast rotating movements can express the enthusiasm of the nation's spontaneity, while the slow movement can reflect the body's beautiful and soothing, but also can be used to reflect the nation's introspection and the beauty of the subtle.

2.2.5. Rhythmic factors. Different ethnic groups have different choices and understandings of music, which requires dancers to ensure that they have the order and rules of movement when performing. When choreographing, it is necessary to integrate the lightness and urgency of the rhythm and the staccato, and at the same time, through the rich rhythmic changes to reflect the flavor of different ethnic dance movements, which requires the graph neural network in the dance movement choreography learning, familiar with the ethnic dance style, to make the movement coherent, smooth and harmonious, and a good grasp of the rhythm and laws of the music of the ethnic group.

When the graph neural network is utilized to create ethnic dance movements, the above five points are applied, and the dance works not only retain the ethnic dance styles, but also innovate the dance vocabulary to a certain extent, so as to form a contemporary ethnic dance works accepted by the audience today.

3. Choreography of Southwest Ethnic Folk Dance Movements Based on Graph Neural Networks

3.1. Motion Capture Data Acquisition

Motion capture [17] is a technology that quasi-tracks and measures the trajectory or attitude data of an object's movement in real three-dimensional space, and stores and 3D reconstructs the recorded motion state in a computer device.

The workflow for obtaining motion capture data using motion capture devices is as follows:

1) Preparation. Performers wear matching motion capture suits. The staff needs to debug and calibrate the hardware and software of the motion capture system, including starting the software, calibrating the home position, and setting the capture frame rate.

2) Establish the skeleton model. When ready, the performer enters the capture area and maintains the preparatory posture at the home position calibration, i.e., standing naturally with both arms raised flat and the body in a “T” shape. At this time, the software operator establishes a three-dimensional skeleton model for the performer.

3) Motion data acquisition. After the model is established, the motion capture device starts to record the movement process of the performer.

3.2. Dance movement database

There are several common storage formats for motion capture data such as BVH, ASF/AMC and HTR.In this paper, the most common BVH format is used to store motion capture data.

The BVH file contains two parts, the first part is the tree-like human skeletal structure definition area, and the second part is the motion capture data area. The first part starts with the keyword "HIERARCHY", and first defines the root node ROOT, whose initial position in 3D space is (0, 0, 0). Then, we define the name, number of channels and spatial position offset relative to the parent node for each joint point by level.

The second part starts with the keyword "MOTION" and first lists the total number of frames of the motion capture data in the segment, identified by the keyword "Frames", and then gives a fixed time interval between two frames, identified by the keyword "Frame Time". Finally, the motion data for each frame is provided in the hierarchical order of the defined zones, containing the spatial position of the root node, expressed as 3D coordinates, and the relative rotation information of the non-root nodes, expressed as Euler angles.

In order to train and validate the ethnic dance movement generation algorithm proposed in this paper, 2 ethnic dance generation datasets based on motion capture data are constructed with the following basic information:

1) Laban-16: the dance video clip of the Miao dance "Squeeze Oil Tip" in Southwest China, recorded by the author through the fieldwork method. The dance fully shows the passionate and pure character traits of the Miao A-Go-A-Mei, highlights their love for life and the yearning pursuit of love, embodies its typical and unique national character traits, is very rich in national flavor, and at the same time, reflects the most sincere and pure thoughts of the local Miao people on the love of life. The dataset, which can be used to train and validate an end-to-end continuous dance score automatic generation algorithm, contains 1,600 consecutive unsegmented human movement samples, totaling 7,121 lower limb (left and right leg) medial element movements, with a total duration of 1,119,706 frames. Each continuous motion sample randomly contained 3 to 8 elemental movements. These actions covered 8 horizontal orientations, totaling 16 labeled categories. The goal of the analysis and identification for this dataset was to map the 8 horizontal orientation symbols for the two support bars in the Laban dance score.

2) Laban-48: Dance video clips of the Miao dance "Lusheng Dance" recorded in Southwest China by the author through the fieldwork method. This dataset can be used to train and validate the end-to-end continuous dance score automatic generation algorithm, and contains 4,800 consecutive human motion samples, totaling 19,200 elemental movements of the lower limbs (left and right legs), with a total duration of 3,018,199 frames. Each continuous motion sample randomly contained 3 to 8 elemental motions. These actions covered 8 horizontal orientations and 3 vertical layers, totaling 48 labeled categories. The goal of the analysis and identification for this dataset was to map the 24 spatially distributed symbols of the two support columns in the Laban dance score.

3.3. Architecture for Automatic Dance Movement Generation

In this paper, we adopt and optimize the efficient graph neural network for human movement modeling, and focus on exploring the "end-to-end" automatic folk dance movement generation algorithm, in order to achieve the global optimization of the dance score generation system, and the high-precision and full-automatic generation of continuous human movement.

The Connection Timing Classifier CTC [18] is used to realize the automatic generation of continuous Laban dance score, and the framework of the automatic generation of folk dance movements is shown in Fig. 1. Connected Timing Classifier CTC is a multi-label prediction method widely used in sequence recognition tasks such as machine translation, speech recognition, etc. Its role in the framework of automatic generation of continuous dance movements is to decode and transcribe the results of the graph neural network's movement prediction for each frame of the input continuous movement into an optimal sequence of movement symbols.

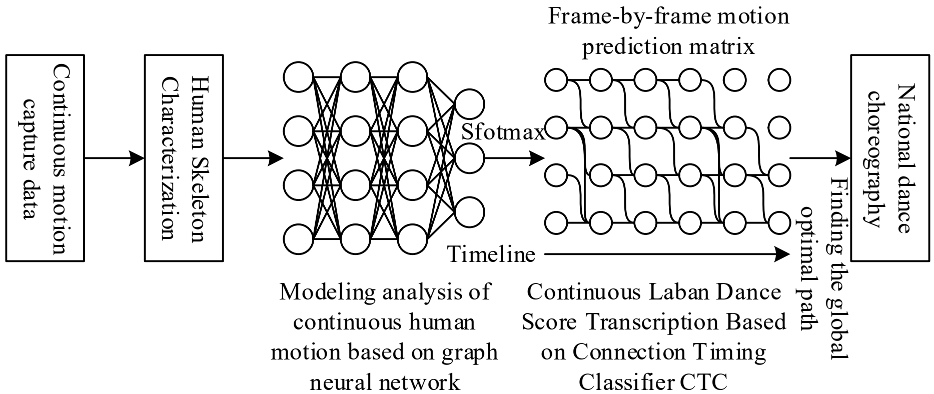

Fig. 1. National dance movement automatic generation frame

3.4. *Semantic-assisted directed graph neural network based on multi-feature fusion*

In this section, a semantically-assisted directed graph neural network model (MF-SDGNN) based on multi-feature fusion is proposed for the task of generating dance movements based on motion capture data. Figure 2 demonstrates the specific flow of the algorithm in this paper. Under the constraints of five factors, namely, strength, fluency, gesture, time and rhythm, this paper utilizes the model to test the choreographic effect of ethnic folk dance movements in Southwest China.

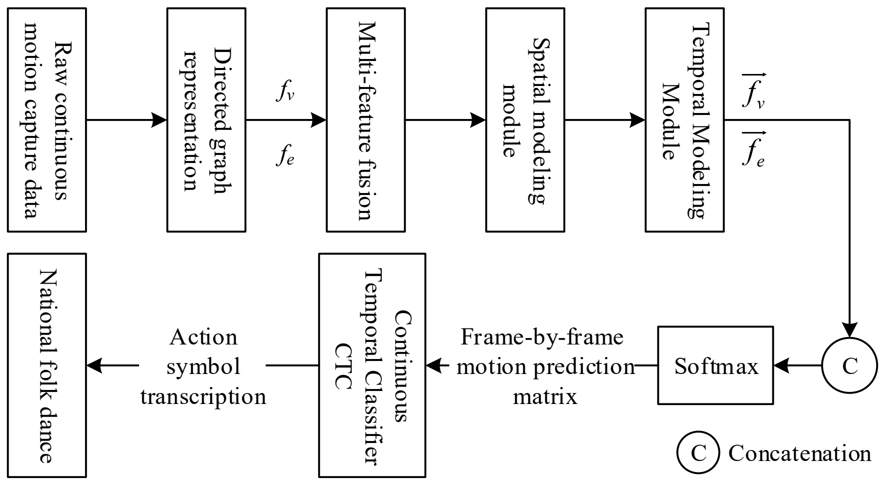

Fig. 2. The process of semantics-guided directed graph neural networks

3.4.1. Directed Graph Representation of Motion Capture Data. Equation (1) is shown:

The motion capture data records the rotational offset of each joint point of the human body relative to its parent node in terms of Euler angles (j,θ,φ), and thus its transformation relies on the Euler angles, where the three angles denote the rotational angles of the joints around the X, Y, and Z axes, respectively. Denote by $S=\{J^1,\ldots,J^T\}$ a segment of human motion of length T recorded by the BVH file, where each frame $J=\{J_i=(j,\theta,\varphi)\,|\,i=2,\ldots,N\}$ represents the pose of N a joint in the skeleton, while denoting by o_i the offset of the i th joint in each frame relative to the parent node. Assuming that the initial offset of J_i relative to the parent node at moments $t-1$, $t\le T$ is $o_i^{t-1}=(x,y,z)$, if the rotation matrix R_i is used to denote the rotation of J_i relative to the parent node at moments $t-1$ to t, the J_i new offset $o_i^t=(x',y',z')$ at moment t is shown in Eq. (1):

$$(x',y',z')=R_i\cdot(x,y,z) \tag{1}$$

A bone is the difference in coordinates between two connected joints, and if given the coordinates of two joints $p_1=(x_1,y_1,z_1)$ and $p_2=(x_2,y_2,z_2)$ that have a connecting relationship, the bone e that connects p_1 and p_2 can be calculated based on Equation (2):

$$e=p_1-p_2=(x_1,y_1,z_1)-(x_2,y_2,z_2) \tag{2}$$

In this paper, we consider that the root node is connected to itself, so the set of edges in the skeleton directed graph can eventually be represented as $E_{legs}=\{e_m\,|\,m=1,2,\ldots,N_e\}$, where N_e is the number of lower limb bones in each frame of the skeleton and has the same value as N_v.

The connectivity of vertices and directed edges in the lower extremity directed graph is described by an association matrix A of size $N_v\times N_e$. Different values of the association matrix element $A_{k,m}$ can reflect different connections of vertex v_k and directed edge e_m, where $k=1,2,\ldots,N_v$, $m=1,2,\ldots,N_e$. If vertex v_k and edge e_m are unconnected, then matrix element $A_{k,m}=0$. If vertex v_k is the target vertex of directed edge e_m, then $A_{k,m}=-1$. Eventually, each frame of the directed graph can be defined as $G=(V_{legs},E_{legs},A)$, which contains both the set of vertices of the graph, the set of directed edges, and the association matrix describing the structural information.

3.4.2. Multi-feature fusion module for directed graphs. In this section, the multi-feature fusion [19] operation is illustrated with a single-frame directed graph as an example, where the set of joints is denoted as $V_{legs} = \{v_k \mid k = 1, 2, \ldots, N_v\}$ and the set of bones is denoted as $E_{legs} = \{e_m \mid m = 1, 2, \ldots, N_e\}$. If the input of the network is a continuous sequence of directed graphs of length T, the multi-feature representation of the sequence can be obtained by repeating the following operation T times.

Based on the multi-feature fusion of joint information flow, the velocity of joint points is calculated firstly, then the joints and joint velocities are embedded into the high-dimensional space of dimension $C1$ respectively, and finally the two types of features in the high-dimensional space are fused to obtain the multi-feature representation of the joints, which is denoted as $z_k \in R^{C_1}$. The multi-feature representation of the joint information flow in each frame of the skeleton can be illustrated in Eq. (3):

$$k_v = \{z_k \mid k = 1, 2, \ldots, N_v\} \in R^{N_v \times C_1} \tag{3}$$

Multi-feature fusion based on skeletal information flow refers to the fusion of skeletal position information and skeletal velocity information in a high-dimensional space. If a bone vector $e_{t,m} = \left(x_{t,m}, y_{t,m}, z_{t,m}\right)^T \in R^3$ at the current moment $t, t \le T$ is given, the same bone vector at moment $t-1, 2 \le t \le T$ is $e_{t-1,m} = \left(x_{t-1,m}, y_{t-1,m}, z_{t-1,m}\right)^T$, and the original velocity information $v_{t,m}^e$ of bone $e_{t,m}$ can be expressed as shown in Equation (4):

$$\begin{aligned} v_{t,m}^e &= e_{t,m} - e_{t-1,m} \\ &= \left(x_{t,m} - x_{t-1,m}, y_{t,m} - y_{t-1,m}, z_{t,m} - z_{t-1,m}\right)^T \end{aligned} \tag{4}$$

Embedding the bone vectors and their velocities into a high-dimensional space of dimension C_1, and fusing the bone and its velocity features in the high-dimensional space, a multi-feature representation of the bone can finally be obtained, which is denoted as $w_m \in R^{c_1}$. As a result, the multi-feature representation of the bone information flow in each frame of the skeleton is shown in Equation (5):

$$k_e = \{w_m \mid m = 1, 2, \ldots, N_e\} \in R^{N_e \times C_1} \tag{5}$$

Finally, let $V_{legs} = k_v$, $E_{legs} = k_e$, each frame of the directed graph after the multi-feature fusion operation can be denoted as $G = \left(V_{legs}, E_{legs}, A\right)$. If the length of the continuous motion data is T, the above operation is performed on T frames of the directed graph respectively to obtain the multi-featured representation of the sequence of directed graphs, denoted as $S = \left\{G^1, G^2, \ldots, G^T\right\}$.

3.4.3. Module for spatial modeling based on dense connectivity. The architecture of the spatial modeling module is shown in Figure 3. The spatial modeling module adopts DGN as the basic unit of spatial modeling. The DGN can update the directed graph that introduces semantic information. There are two kinds of elements in a directed graph: vertices $V_{legs} = \{v_k \mid k = 1, 2, \ldots, N_v\}$ and directed edges $E_{legs} = \{e_k \mid k = 1, 2, \ldots, N_e\}$, and updating a directed graph means updating the vertices and directed edges in the graph.

For both elements, the DGN updates the element itself by aggregating the spatial information on the points and directed edges with which it has a connectivity relationship. Considering the directedness of edges, the DGN contains two aggregation functions g^- and g^+ for aggregating the spatial information on the set of vertex incoming edges e^- and the set of outgoing edges e^+, respectively. In addition, considering the different dependencies of vertices and directed edges, the DGN updates vertices and directed edges in the directed graph through two update functions h^e and h^v, respectively. If the updated versions of vertices and directed edges are denoted by $\tilde{v}_k$ and $\tilde{e}_m$, respectively, the update process of the directed graph is specified as follows in Eq:

$$\dot{e}_k^- = g^{e^-}\left(e_k^-\right) \tag{6}$$

$$\dot{e}_k^+ = g^{e^+}\left(e_k^+\right) \tag{7}$$

$$\tilde{v}_k = h^v\left(\left[v_k, \dot{e}_k^-, \dot{e}_k^+\right]\right) \tag{8}$$

$$\tilde{e}_m = h^e\left(\left[e_m, \tilde{v}_m^s, \tilde{v}_m^t\right]\right) \tag{9}$$

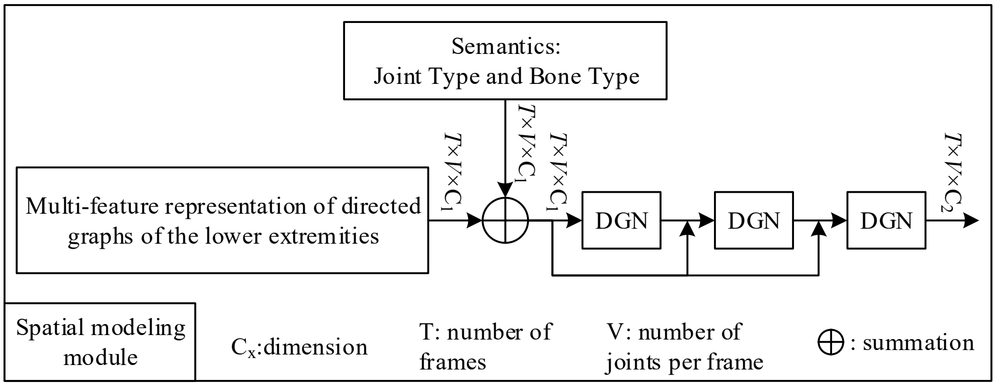

Fig. 3. The architecture of spatial modeling module

3.4.4. Residual linkage-based time modeling module. The temporal modeling module consists of a summation operation, a spatial maximum pooling layer SMP, two TCN layers, and an average pooling layer, whose basic architecture is shown in Fig. 4.

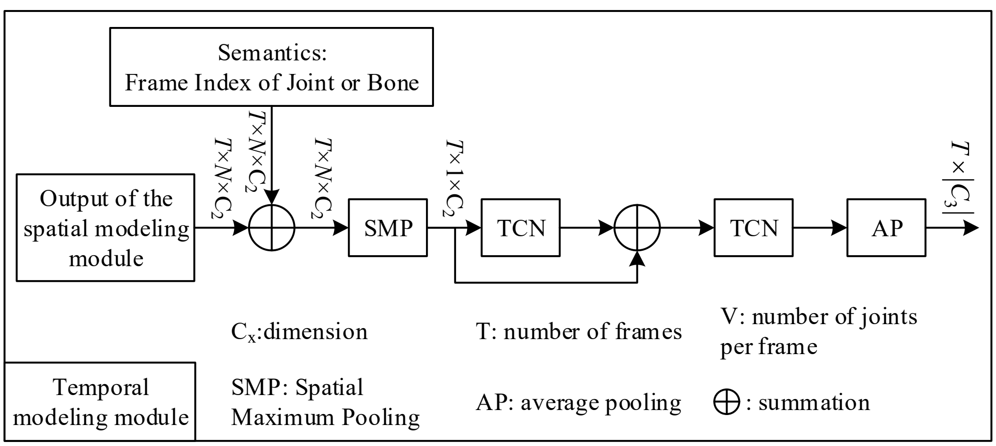

Fig. 4. The architecture of temporal modeling module

The basic unit of the temporal modeling module is the TCN, which allows temporal modeling of directed graphs.The core of the TCN is the convolution operation with convolution kernel size $k_t \times 1$. To apply it, it is only necessary to select a suitable value of k_t and use that value as its convolution kernel size along the time dimension. As a result, the TCN applies the convolution operation to the time dimension of vertices and directed edges. Finally, inspired by residual learning, jump connections from ResNet [20] are added between the two TCNs as a way to improve the training efficiency of the network.

3.4.5. Action Symbol Transcription. The frame-by-frame motion prediction matrix is generated by the module before CTC. After the temporal modeling module, the obtained joint information streams $\overrightarrow{f_v}$ and bone information streams $\overrightarrow{f_e}$ are stacked by channel dimension and integrated into a twodimensional tensor of shape $T\times 2|C_3|$. This is converted into a frame-by-frame action prediction matrix $y=\{y^t \mid t=1,2,\ldots T\}$ of size $T\times Card(L')$ using the softmax function [21], where the one-dimensional vector 6 represents the probability distribution of all dance movement types at moment t.

The CTC can define a conditional probability for any path $\pi=\{\pi_t\in L'\}_{t=1,\ldots,T}$ of length T on the frame-by-frame action prediction matrix. If $y^t_{\pi_t}$ is used to denote the probability of predicting the action at moment t as π_t, the conditional probability is defined as shown in Equation (10):

$$p(\pi \mid y)=\prod_{t=1}^{T} y^t_{\pi_t} \tag{10}$$

The training objective of the network is to minimize the negative log-likelihood of these sequence conditional probabilities. If N_X denotes the number of action samples in the training set, and $y^{(n)}$ and $l^{(n)}$ denote the frame-by-frame action prediction matrix of the nth sample and the Laban symbol sequence of the model output, the negative log likelihood of the conditional probabilities of the sequences is shown in Equation (11):

$$O=-\sum_{n=1,\ldots,N_X} \log p\left(l^{(n)} \mid y^{(n)}\right) \tag{11}$$

The most probable action in each moment is usually chosen to represent the final optimal sequence of action symbols, as shown in Equation (12):

$$l^{+}\approx B\left(\arg\max_{\pi} p(\pi \mid y)\right) \tag{12}$$

4. Organizational effectiveness test analysis

4.1. Model Training Tests

The study proposes a semantic-assisted directed graph neural network model with MF-SDGNN multi-feature fusion for Southwest folk dance movement choreography. The performance of the method is first validated. The simulation environment is Matlab (R2017b), discrete graphics card GeForceMX150, 7th generation Intel i5 processor, 2GB GDDR5, 256GB PCIe SSD, memory 8GB DDR4, DeepLeaning-Toolbox. The experiments are performed on the Laban-16 dance video movement dataset described above. The dataset is divided into training set and test set according to 3:7.

Fig. 5 shows the loss value and accuracy variation curves of the MF-SDGNN model on the selected dance action dataset. From the data in the figure, it can be seen that with the increase of the number of training rounds, the loss value of the model in this paper converges in the 30th round of the test, at which time the loss value is about 0.25. In addition, the accuracy of the MF-SDGNN model in the test increases rapidly from the beginning, and the convergence occurs in the 35th round, at which time the accuracy is as high as 96%. And the loss value and accuracy convergence between the training set and the test set are basically the same, which indicates that the model in this paper has a better training effect in the dance video action dataset.

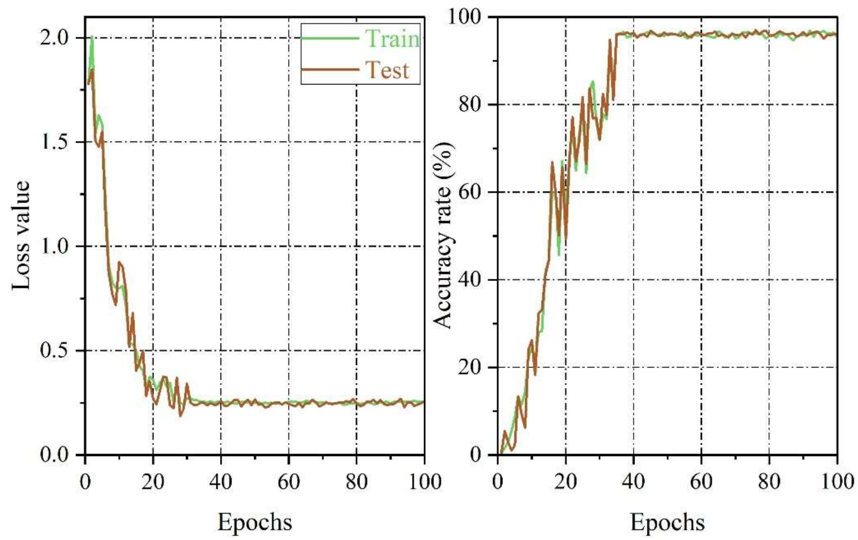

Fig. 5. MF-SDGNN model changes loss and accuracy of the dance action data

4.2. Performance evaluation tests

This section evaluates the MF-SDGNN model proposed in this paper on two datasets, Laban-16 and Laban-48. In order to correctly generate the corresponding symbols, the classes of action segments need to be learned and recognized from the data. Therefore, this problem can be seen as a specialized recognition task. The performance of the method evaluated with the average recognition rate. For the segmented training and test sets, 30% of the samples from the data for each action category are randomly selected for training and the remaining samples are tested. Due to the randomness of the algorithm, the experiment is repeated five times and the results are averaged to obtain the final results. Comparative analysis with DTW, Multi-class SVMs, Histogram-based, CRNN, and GRU-RNN, 5 groups of models are also performed to evaluate the performance of this paper's model in performing time-series modeling and recognizing dance movements to generate action symbols.

Figure 6 demonstrates the performance of the six groups of models in recognizing lower limb movements on the two datasets. From the figure, it can be known that the motion recognition rate of this paper's method is significantly higher than that of the above five methods, e.g., on the Laban-16 dataset, the recognition rates of left and right lower limb motions are 95.65% and 94.04%, respectively, which is higher than that of the DTW, Multi-class SVMs, Histogram-based, CRNN, and GRU-RNN models by 5.21% to 15.81%. This indicates that the time series dynamic modeling method based on semantic-assisted directed graph neural network model with multi-feature fusion is applicable to the analysis of motion capture data and can achieve good results.

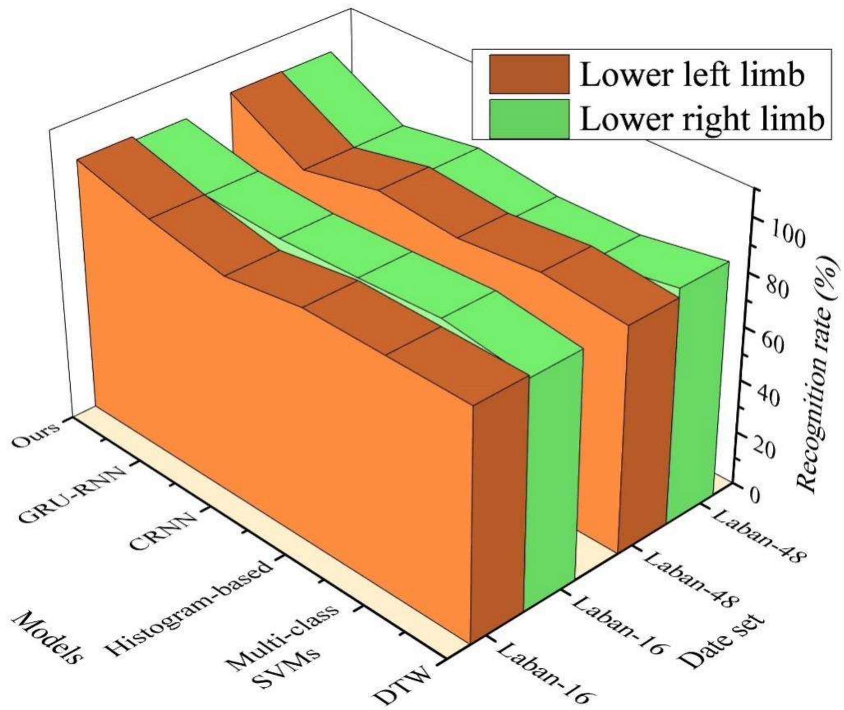

Fig. 6. The identification performance of the six group model movement

In order to further investigate the contribution of each module (i.e., multi-feature fusion module, spatial module, and temporal module) in the proposed MF-SDGNN model, we conducted an ablation study using different model configurations. The GNN model is set as the baseline method, which represents the base model without fusion of the above three modules, the MGNN represents the GNN model combined with the multi-feature fusion module, the SGNN and the DGNN are the GNN models that only fuse the spatial module as well as the temporal module, respectively, and the MF-SDGNN is the complete model proposed in this paper that fuses the above three modules. The basic GNN, MGNN, SGNN, DGNN, and MF-SDGNN are denoted as Models 1-5 in turn, and then we evaluate the impact of each module on the overall method performance through different combinations of module configurations.

The results of comparing different configuration models on the two datasets are shown in Fig. 7. From the figure, by comparing the results of configuring Model 1 with Models 2-4, we can see that the models incorporating the auxiliary modules (Models 2-4) improve the recognition rate by 9.78% to 17.46% compared to the baseline model (Model 1), which demonstrates the advantages of the proposed auxiliary modules in capturing dance movements.

In addition, by comparing Models 2-4 with Model 5, we can clearly see the performance gain from the multi-module fusion, with the recognition rate of the model used (Model 5) being above 94%. This demonstrates the effectiveness of our proposed MF-SDGNN model and that the three modules proposed in this paper can work together and improve the overall model accuracy.

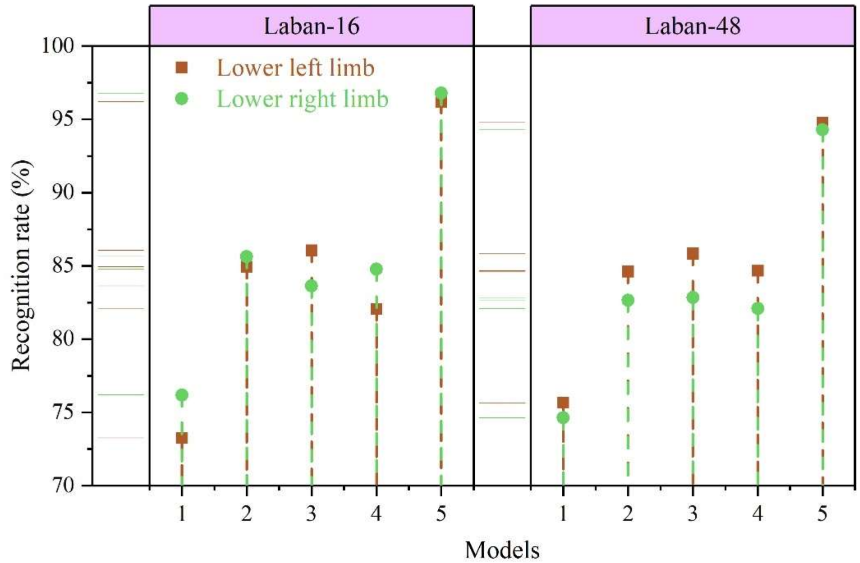

Fig. 7. Different configuration models compare results in two data sets

4.3. Choreographic features of dance movements

In this section, 10 pieces of music with different rhythms are selected and the model constructed in this paper is used for automatic generation of dance movements to realize the choreography of folk dance works. The dance movements used in each work are recorded and analyzed, and based on the data compiled and analyzed, the dance movement choreography ability of the MF-SDGNN model under different music rhythms is explored. The study only recorded the basic movements in folk dances, which are spinning, jumping, kicking, swinging, stepping, twisting, turning, shaking the shoulders, bending, fluttering the sleeves, stomping, and cross-steps, totaling 12 different basic movements.

Figure 8 demonstrates the frequency of choreographed dance movements of the MF-SDGNN model under different music tempos. From the figure, it can be seen that MF-SDGNN choreographed all the above movements under different music tempos, which indicates that the model can skillfully choreograph the same dance movements into different dance works by learning from folk dance videos, and demonstrate the regional characteristics and cultural features of Southwest China's folk dances by reasonably connecting each movement. In addition, the frequency records of movements show that the MF-SDGNN model choreographed the above basic movements 5-20 times, which indicates that the basic movements are indispensable in the choreography of sets of movements, and that good combinations can not be made without the padding of basic movements.

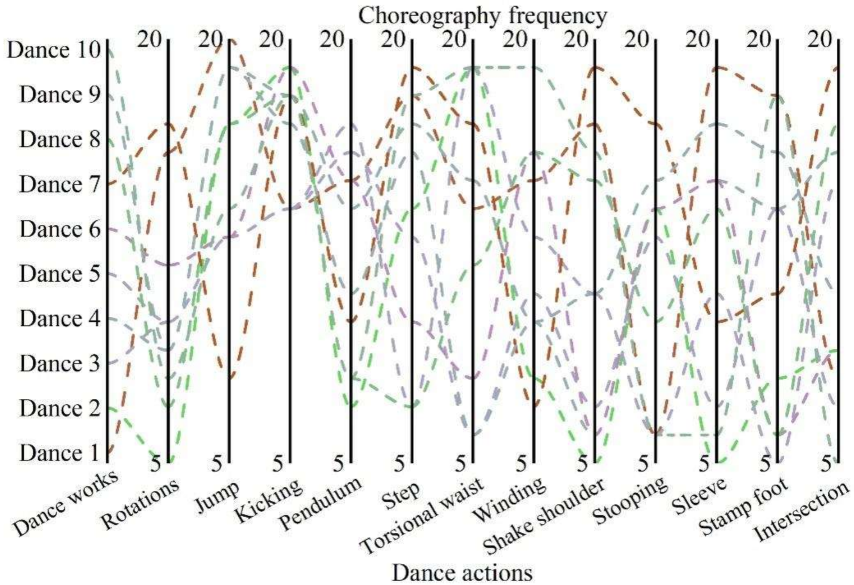

Fig. 8. The MF-SDGNN model is choreographed to dance the dance frequency

4.4. Usability testing

This section analyzes the usability of folk dance movement choreography strategies based on the MF-SDGNN model. The study invites three experts to appreciate 10 folk dance works choreographed by the above model under different music rhythms, and evaluates whether the test has feasible value by means of a questionnaire.

The questionnaire covered five micro-factors affecting movement choreography, i.e., the strength of movement choreography, the fluency of movement transitions, the gestures of folk movements, the interval transitions, and the degree of grasping the rhythm of the music. The survey adopts the expert rating system, and through the collection of statistical expert evaluation results, the feasibility level of this paper's model at each level is classified. According to the experts' suggestions, the feasibility grade division interval is shown in Table 1.

Table 1. The feasibility grade interval

Grade	Very feasible	Feasible	General	Unrecommend
Score	⩾90	70~89	60~69	<60

Fig. 9 shows the results of experts' ratings on the feasibility of action choreography of this paper's model under different levels. From the expert evaluation results, the model of this paper has high usable value. The ten dance works choreographed in this paper have achieved high ratings in the strength of movement choreography, smoothness of movement transition, gestures of ethnic movements, transformation of movement intervals, and the degree of grasping the rhythm of music, with an average rating of >85, which means that the model of this paper is feasible as a strategy for choreographing the movements of ethnic dances in Southwest China. Taking the degree of grasping the rhythm of music as an example, the average ratings of the three experts for the ten works were between 87.93 and 91.73, indicating that the dance movements choreographed by the model of this paper are able to grasp the rhythmic laws of different music, generate dance movements according to the beat of the music, and enhance the coordination and expression of emotions in folk dance performances.

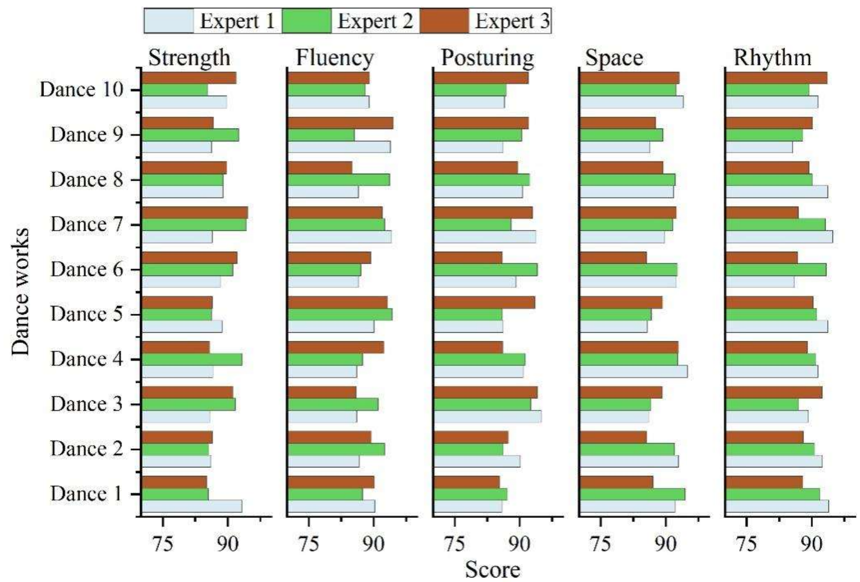

Fig. 9. The results of the feasibility score of the expert's action choreography

5. Conclusion

The article analyzes the influencing factors of folk dance movement choreography in Southwest China by taking the Yi ethnic group in Sichuan and Yunnan, China, as an example. The graph neural network model is optimized by embedding modules such as multi-feature fusion. On this basis, a framework for automatic generation of ethnic folk dance movements based on the optimized graph neural network is designed, and several experiments are carried out to explore the potential of this paper's model as an optimization strategy for dance movement choreography.

1) Experiment Summary. The MF-SDGNN model shows a good training effect as the loss value and accuracy converge to 0.25 and 96% respectively before 40 rounds of training. The model's lower limb motion recognition rate on Laban-16 and Laban-48 datasets is above 90%, which is improved to different degrees compared with the comparison model. The multi-feature fusion module, the spatial modeling module, and the temporal modeling module work together to improve the performance of the graph neural network model for dance movement recognition. The model in this paper can choreograph folk dance movements under different music beats, and the feasibility score is higher than 85, which can be used as a strategy for choreographing folk dance movements in Southwest China.

2) Insufficiency and Prospect. Intelligent choreography has insufficient data, which is a challenge for the training and accuracy of graph neural networks. Without enough data, the algorithm may not be able to learn enough patterns and features, which affects the quality and diversity of folk dance movement choreography. With the development of augmented reality and virtual reality technologies, graph neural networks can enable a more immersive choreography and learning experience on these platforms.

References

[1] Zhanna, R. (2020, September). Digitization of Intangible Cultural Heritage as a method to save and actualize it on the example of Kazakh folk dance. In Proceedings of the 6th International Conference on Engineering & MIS 2020 (pp. 1-7).

[2] Kico, I., Grammalidis, N., Christidis, Y., & Liarokapis, F. (2018). Digitization and visualization of folk dances in cultural heritage: A review. Inventions, 3(4), 72.

[3] Doulamis, A. D., Voulodimos, A., Doulamis, N. D., Soile, S., & Lampropoulos, A. (2017, February).

Transforming Intangible Folkloric Performing Arts into Tangible Choreographic Digital Objects: The Terpsichore Approach. In VISIGRAPP (5: VISAPP) (pp. 451-460).

[4] Grammalidis, N., Kico, I., & Liarokapis, F. (2020). Analysis of Human Motion based on AI technologies: Applications for safeguarding folk dance performances. In Strategic Innovative Marketing and Tourism: 8th ICSIMAT, Northern Aegean, Greece, 2019 (pp. 321-329). Springer International Publishing.

[5] Li, R., Zhao, M., Zhang, X., & Li, X. (2021, September). Dance to Music: Generative Choreography with Music using Mixture Density Networks. In 2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 348-353). IEEE.

[6] Zhang, A. (2023). Optimization Simulation of Match between Technical Actions and Music of National Dance Based on Deep Learning. Mobile Information Systems, 2023(1), 1784848.

[7] Doulov, U. (2024, September). THE INFLUENCE OF TRADITIONAL FOLK DANCES ON CONTEMPORARY CHOREOGRAPHY. In INTERNATIONAL CONFERENCE ON INTERDISCIPLINARY SCIENCE (Vol. 1, No. 9, pp. 7-10).

[8] Aristidou, A., Stavrakis, E., & Chrysanthou, Y. (2014, October). Motion Analysis for Folk Dance Evaluation. In GCH (pp. 55-64).

[9] Yang, Q., & Wu, L. (2021, January). Opinions on the Choreography Creation of Yunnan Traditional Folk Dance. In The 6th International Conference on Arts, Design and Contemporary Education (ICADCE 2020) (pp. 77-81). Atlantis Press.

[10] Bazela, D., Pavliuk, T., Babych, O., Batieieva, N., & Kyzim, P. (2022). Contemporary global trends in choreographic art development at the beginning of the 21st century. SPORT TK-Revista EuroAmericana de Ciencias del Deporte, 60-60.

[11] Li, M. (2024). Ethnic dance movement recognition based on motion capture sensor and machine learning. International Journal of Information and Communication Technology, 25(8), 81-96.

[12] Sun, G., Wong, Y., Cheng, Z., Kankanhalli, M. S., Geng, W., & Li, X. (2020). Deepdance: music-to-dance motion choreography with adversarial learning. IEEE Transactions on Multimedia, 23, 497-509.

[13] Beevi, S. Z., & Malini, M. (2024). GATED TRANSFORMER WITH GRAPH NEURAL NETWORKS AND TEMPORAL CONVOLUTIONAL NETWORKS FOR DANCE CHOREOGRAPHY. JOURNAL OF BASIC SCIENCE AND ENGINEERING, 21(1), 334-346.

[14] Au, H. Y., Chen, J., Jiang, J., & Guo, Y. (2022, October). Choreograph: Music-conditioned automatic dance choreography over a style and tempo consistent dynamic graph. In Proceedings of the 30th ACM International Conference on Multimedia (pp. 3917-3925).

[15] Hou, Y., Yao, H., & Li, H. (2023, July). Graph Convolutional GRU for Music-Oriented Dance Choreography Generation. In 2023 IEEE International Conference on Multimedia and Expo (ICME) (pp. 2111-2116). IEEE.

[16] Kritsis, K., Gkiokas, A., Pikrakis, A., & Katsouros, V. (2022). Danceconv: Dance motion generation with convolutional networks. IEEE Access, 10, 44982-45000.

[17] Ariane P. Lallès,Geoffroy Moucheboeuf,Emilie Doat,Hélène Pillet & Xavier Bonnet. (2025). IMU Calibration Effect on Lower Limbs Kinematics Against Optical Motion Capture in Post-Stroke Gait. IRBM(1),100873-100873.

[18] LiruHua,XinyiMa,ChihangZhao,BailingZhang,ZijunSu & YuhangWu. (2024). Recognition of vehicle license plates in highway scenes with deep fusion network and connectionist temporal classification. IET Image Processing(13),4066-4080.

[19] He Zhao,Jingjing Wang,Jiali Guo,Xin Hui,Yunling Wang,Dongyu Cai & Haijun Yan. (2024). Detecting Water

Stress in Winter Wheat Based on Multifeature Fusion from UAV Remote Sensing and Stacking Ensemble Learning Method. Remote Sensing(21),4100-4100.

[20] Haoyuan Sun,Xiaoping Gu,Yubo Zhang,Fulin Sun,Sheng Zhang,Duojin Wang & Hongliu Yu. (2025). An enhanced ResNet deep learning method for multimodal signal-based locomotion intention recognition. Biomedical Signal Processing and Control107254-107254.

[21] Jing Luo,Wanqing Zhao & Jinye Peng. (2025). Causal softmax for out-of-distribution generalization. Digital Signal Processing(PB),104861-104861.