Hindawi
Mobile Information Systems
Volume 2023, Article ID 1784848, 10 pages
https://doi.org/10.1155/2023/1784848

*Research Article*

# Optimization Simulation of Match between Technical Actions and Music of National Dance Based on Deep Learning

**Aimin Zhang**

*Department of Music, School of Special Education, Beijing Union University, Beijing 100075, China*

Correspondence should be addressed to Aimin Zhang; 1928104141@siit.edu.cn

Received 24 March 2022; Revised 28 April 2022; Accepted 30 April 2022; Published 18 May 2023

Academic Editor: Liping Zhang



In the match between technical movements and music of folk dance, the most important thing is to extract features effectively. DL algorithm is one of the most efficient methods to extract video features at present. In this study, the DL method is applied to the matching optimization of technical movements and music in folk dance. Using DL to train the corresponding relationship between the technical movements and music of national dance, the given dance movements and corresponding movements are adapted to the musical beat points. To better reflect the degree of correlation between music and movement changes, the change rate of feature value is used instead of feature value itself in correlation calculation. The matching degree between this method and genetic theory method and spatial skeleton timing diagram method is compared. The experiment shows that the matching method of technical movements and music of national dance optimized by DL can achieve 95.78% accuracy, and the matching synchronization of technical movements and music of national dance can reach 96.17%. Therefore, the method proposed in this study can fully reflect the synchronization of music and movement changes, and the optimized movement matching method matches the national dance technical movements—music matching quality is better. This study expands a new perspective for the research of dance and music matching technology. It has certain practical and theoretical significance.

## 1. Introduction

As a multiethnic country in China, each ethnic minority has a long history of traditional culture, while folk dance of ethnic minorities is the representative of traditional culture of ethnic minorities [1]. With the development of modern society, the living space of ethnic minority dance cultural heritage is increasingly impacted by industrial civilization and economic globalization [2]. Minority dance is an important part of intangible cultural heritage. However, due to the change in cultural and ecological environment, some excellent Chinese folk performing arts are on the verge of loss and extinction, and it is urgent to give scientific and effective protection and inheritance through different means [3]. The most common expression of music is to match music with dance [4]. The most common form is the dance action matching technology in the choreography of dance music according to the change in music [5]. Dance with music is a kind of advanced art that combines auditory and visual enjoyment. Because people are sensitive to changes in music and human movements, they can detect some subtle changes. For example, the same music rhythm can express different emotions, and the same movements can show different dance styles. The synchronous occurrence of these subtle changes is of great significance for enhancing the understanding of music and dance.

For the analysis of specific targets in dance movements, the traditional manual analysis methods can no longer meet the existing needs, so the intelligent processing of video data has become a key issue. In modern music choreography, people's matching effect of music and dance requires strong synchronization between music changes and dance movements, and they must have a deep understanding of music and a strong ability to grasp it. At present, music-dance movement matching technology cannot determine the movement rhythm of dance with the change in music. In this case, deep learning (DL) [6–8] came into being. DL comes from the study of artificial neural network (NN) in machine learning [9–11]. DL is a computational model composed of multiple processing layers to learn data representation methods with multiple levels of abstraction. The foundation of deep neural network (DNN) [12, 13] is a multilayer perceptron, which usually consists of input layer, output layer, and multiple hidden layers. In the process of DL recognition, it is not necessary to artificially select features as input data for recognition, but to learn by automatically selecting features of objects represented by the model itself, which is an end-to-end recognition method. The training of human behavior recognition model is completed by determining the parameters of the model through autonomous learning. The difference between artificial choreography and music-dance action matching technology is that real dance is based on the interaction between the body's energy, sense of balance, and environment, which is called internal cause-oriented. The music-dance action matching technology is based on the results, mainly considering the spatiality and rhythm of the dance, which is called exogenous orientation. Based on DL technology, this study makes an in-depth study on the optimization simulation of the match between technical movements and music of folk dance. Its innovations are as follows:

(1) In this study, the DL method is applied to the matching optimization of dance and music, which opens up a new perspective for studying the matching technology of dance and music. To ensure the cohesion of adjacent actions, this study analyzes the connectivity of adjacent actions according to the action distance. Then, intensity matching is carried out on the input audio and all possible resultant movements, and the best matching dance movements are screened out.

(2) In this study, to improve the accuracy, the mapping of music-action segments is restricted, which greatly improves the accuracy of searching action segments. At the same time, to measure the effect of action cohesion, a fluency function is proposed. The matching degree between this method and genetic theory method and spatial skeleton timing diagram method is compared. Experiments show that the method proposed in this study can fully reflect the synchronization of music and movement changes, and the optimized movement matching method matches the national dance technical movement and music matching with better quality.

The main goal of this study was to improve the compatibility of technical movements and folk dance music. The following is a list of the sections: introduction is the first section. The research content, research background, and research significance of this study are all presented in this section. The research innovation and organizational structure of this article are briefly introduced in this study. The second section is about work that is related to the first. This section summarizes the domestic and international research literature on the paper's research issue, as well as the current state of research on the music and dance feature matching model. The third section briefly covers the relevant foundations and theories of DL, such as the basic structure of NN and the principle of network parameter design. The method for extracting dance-music features is also provided. Then, to address the drawbacks of previous methods, a DL-based optimization strategy for matching technical movements and folk dance music is provided, along with a detailed implementation methodology. The experimental section is in the fourth section. The performance of the approach described in this research was investigated in this section, which included a significant number of tests and analyses on the "DL-based matching optimization of technical movements and music of national dance." The summary and prospects are presented in the fifth section. This section highlights and describes the paper's research findings and accomplishments. It also points out the paper's flaws and areas for improvement, as well as discusses the study direction of music-dance matching in future work.

## 2. Related Work

In recent years, motion capture technology has been widely used to get realistic dance movements, gradually replacing the original manual dance generation. Dance and music matching technology has become a research hotspot in the field of music understanding and dance synthesis. Zhang *J* et al. proposed an optimization method based on genetic theory for dance technique movement and music matching. This method first determines the music style, obtains the underlying style features of music and dance, and performs correlation analysis on the underlying features of music and dance to remove redundant feature pairs. Then, the genetic theory is used to achieve the optimal selection to satisfy the corresponding relationship between the music matching accuracy and the matching efficiency, and the pose re-quantification technology is used to align the directions of the action joints corresponding to the matched dance moves [14]. The corresponding relationship formed by the matching method proposed by An FP et al. can effectively represent the synchronization of changes in music and dance movements, but there is a problem that the calculation process is cumbersome and time-consuming [15]. Zhou S et al. proposed a method based on musical similarity to generate dance motion sequences, which avoids motion feature extraction and motion segmentation compared with other methods [16]. Wei H et al. proposed a machine learning-based optimization method for matching dance technical movements and music in music choreography, but there is a problem of low matching quality between music and dance moves [17]. Ijjina EP et al. proposed a new dance movement articulation algorithm. This method first generates multiple interpolated action nodes between two action clips and then uses the degree of action correlation to select the most natural nodes to combine to generate action clips [18]. Akula A et al. proposed an optimization method for dance technique movement and music matching based on greedy theory [19]. Saha S et al. proposed a dance action recognition algorithm based on a spatial skeleton timing diagram. The algorithm first uses PAFs to extract skeleton nodes for dancers in dance videos and then generates skeleton sequences. Finally, it is combined with LSTM for action recognition [20]. Parsons A et al. outlined the role of motion capture technology in the digital development of minority folk dances and discussed the process and direction of digital development of folk dances based on motion capture technology [21]. He Y et al. proposed a method to synthesize robot dances based on the rhythmic and emotional features of music and actions [22]. Dawar N et al. proposed to automatically synthesize dances by applying the Gaussian processing to human-generated dance movements. The algorithm utilizes a large number of human-created actions and corresponding music on the Internet and extracts the nonlinear mapping relationship between the music beat and the coordinates of each joint of the action through a Gaussian model. The clustered Gaussian models correspond to various types of dances. This method solves the problem of insufficient amount of original action data and monotonous style and improves the diversity of generative dances [3].

Although domestic and foreign literature has done a lot of research on the matching technology of dance movements and music, there is little research on the matching technology of dance movements and music of national dance. Based on the previous research results on the matching between dance movements and music, this study puts forward a DL-based optimization method for the matching between dance movements and music. Aiming at the specific style of music and dance, this study uses DL to train the correspondence between music and movements, so as to adapt the given dance movements and corresponding movements to the music beat points. To better reflect the degree of correlation between music and movement changes, the change rate of feature value is used instead of feature value itself in correlation calculation. Then, music-dance action matching is carried out, and the dance action segments that meet the music rules are retrieved from the corresponding matching database. Finally, the appropriate dance moves are selected to complete the arrangement of all dance moves. The experimental results show that the rhythm and intensity of synthetic dance are basically the same as that of music, and the matching effect of technical movements of national dance and music is better.

## 3. Methodology

*3.1. DL.* The model of DL is NN with multiple hidden layers. In the hidden layer of DL, by means of feature combination, the original input can be transformed into shallow and middle-level features layer by layer, and finally, the high-level features can be obtained to achieve the task goal. DNN processes the required features through the hidden layer and completes the mapping from input to output through multiple neurons. Its feature learning ability is excellent, and it can characterize data more essentially, which is beneficial to classification [23]. As a machine learning method of training sample distribution by multilayer NN fitting, DL alleviates the local optimum problem of traditional NN algorithm when training multilayer NN, and its training process does not depend on sample label information. The DL layer-by-layer training method solves the gradient diffusion problem easily caused using the BP algorithm when there are too many layers in the network, and the local optimum problem is caused using gradient descent.

The number of parameters in the learning of a DL network model grows in proportion to the model's complexity. The amount of data necessary for network training is also considerably increased to properly train the parameters in the model and avoid insufficient network learning [24]. The parameter calculation in one layer of the model is the product of the filter size, the feature map size, and the feature map depth, whereas the pool layer does not increase the parameter. The full-connection layer comes last, and it can synthesize the features of the preceding layers to produce the network's classifier. In a network, the full-connection layer is the layer containing the most parameters. The product of the number of input channels and the number of output channels is used to calculate the whole connection layer's parameters. The parameter quantity should be doubled by 2 due to forward and backward propagation in training. In comparison with DLNN, it emphasizes network structure depth, clarifies the importance of learning features, improves sample feature representation and classification and prediction accuracy using layer-by-layer feature transformation, and better extracts and characterizes the intrinsic information in data by learning features from big data [25]. Between convolution layers, neurons map data using a nonlinear function, simplify data using a pooling layer, and then send it to neurons in the next layer. Finally, through the completely connected layer, which is the categorized category, the results are received in the output layer.

DL analyzes a vast quantity of data using a deep network with several hidden levels and learns features through training rather than using other ways to discover features through people, resulting in greater recognition efficiency and accuracy [26]. Because the number of neurons in the hidden layer of an automatic encoder is minimal, it is usually essential to compress the input data in the hidden layer. When the input data are a random sequence, the automatic encoder has a hard time extracting the data effectively. At this point, some constraints must be added to the automated encoder, resulting in a sparse automatic encoder. The above ideas underpin the deep network, while DL allows the deep network to learn data on its own. DL can learn features from unlabeled data as input. This process does not require human supervision and is carried out in an unsupervised manner. In the field of image processing, the convolutional neural network (CNN) is the most widely utilized DL network. The convolution layer for convolution, pooling layer for feature screening, and fully connected layer for feature fusion are the three network layers that make up CNN's basic structure. Each layer's convolution layer is made up of a large number of neurons. A weight value will be obtained between neurons during training or learning. The entire connection layer operates in the same way as a typical NN with a single hidden layer. By connecting weight and bias, the input layer is connected to the hidden layer, and the hidden layer is connected to the output layer. The initial step of DL training is unsupervised learning from the bottom-up, and the second step is supervised learning from the top-down [27]. A vast volume of unlabeled data can be input into a realistic application. The shallow layer can learn the data's structure and train it layer by layer. The characteristics are constantly abstracted, and the characteristics or traits can finally be appraised.

*3.2. Extraction of Movement Features of Music and Folk Dance Techniques.* There are many underlying features of music and dance movements, so it is the main work of this section to choose music and movement features with a high correlation for national dance technical movements and music correspondence. Firstly, the data of music and dance are preprocessed to prepare the data for feature analysis and extraction. There are two kinds of music files: MIDI files and audio files. MIDI music can be expressed as a music matrix, with each row representing the information of a note and each column representing an attribute of a note. There are many data formats in audio format, including lossy file format and lossless file format. Most of the dance accompaniment music uses audio files, and compared with MIDI files, the audio files can extract richer music features. This study takes audio files as an example to extract features. The preprocessing of audio files includes sampling and quantization, windowing, and framing. Frame rate should be considered when analyzing audio waveform. There will be problems when analyzing the whole time signal, and the analysis results describe the average of global features. To get the local features and dynamic changes in features, the analysis should be carried out in a short-time window, and the window moves along the time signal in a time sequence. The audio signal is a nonstationary random signal, and its characteristics will change with time, but the audio signal has short-term stationary characteristics. The short-term stability of audio signal makes it feasible to extract the underlying features. Each window is called a frame. The frame rate is the number of frames per second or hertz. To extract the underlying features of audio signals, it is necessary to window and frame music pieces.

Assuming that the time domain signal of the audio file is $x(l)$, the window function is $w(m)$, and the audio signal of the $n$ frame after windowing is $x_n(m)$; there are

$$x_n(m) = w(m)x(n+m) 0 \le m \le N-1, \quad n = 0, 1T, 2T, \ldots,$$
$$w(m) = \begin{cases} 1, & m = 0, 1, \ldots, N-1, \\ 0, & \text{other}. \end{cases} \tag{1}$$

Among them, $N$ is the frame length and $T$ is the displacement of the frame. The short-term energy of the $n$th frame is defined as follows:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m), \tag{2}$$

where $N$ represents the number of sampling points in the audio frame and $x_n(m)$ represents the value of the $m$th sampling point in the $n$th frame of the audio signal.

The dancer's bodily movement process must be captured in the action capture of national dance. The motion data produced by the motion capture system are high-dimensional data that cannot be used to calculate features directly. The coordinates of the human body model's root node are saved in a world coordinate system; however, the coordinates of other joint locations are defined in terms of displacement and rotation angle relative to the parent node, which is a local coordinate representation. Static elements, such as movement distance, arm form, and footstep imprint, are utilized to characterize the posture features of human figures. You must transfer motion data from the local coordinate system to the global coordinate system to recover human motion. As a result, various joint sites other than the root node must be transformed before motion information can be extracted. Finally, concrete models as carriers of action data must be used to demonstrate the spatial representation of activities. Because accurate and relevant models have a substantial impact on the binding of motion capture data, establishing a role model is a crucial procedure in and of itself. When picking action segments from the action database for input music to match, the length of the action segments must be more than or equal to the length of the music segments, according to the music and action feature matching model suggested in this work. The component with the same music length is then intercepted and matched with the music segment in the action segment. The principle of interception is that the matching between an action and a music segment with the same length of music is better than other actions with the same length of music. The motion and music processing flow based on DL optimization is shown in Figure 1.

Let the action features extracted from the action clip $N_i$ be recorded as follows:

$$\text{MotionFeature}(\text{f}) = \begin{bmatrix} F_R^{\text{Motion}}(f) \\ F_I^{\text{Motion}}(f) \end{bmatrix}, \quad f \in N_i. \tag{3}$$

Among them, $f$ is the frame number of the action segment $N_i$ and $F_R^{\text{Motion}}(f)$ and $F_I^{\text{Motion}}(f)$ are the rhythm and intensity features of the action segment, respectively. The instantaneous intensity of the current action clip $N_iE$ is as follows:

$$I(f) = v_{\text{foot}}(f) + k \cdot v_{\text{root}}(f). \tag{4}$$

Among them, $v_{\text{foot}}(f)$ represents the speed of the vertical displacement of the foot in the current action; $v_{\text{root}}(f)$ represents the speed of the horizontal displacement of the root joint; and $k$ is the weight parameter for the vertical displacement speed of the foot and the horizontal movement speed of the center of mass. The intensity feature extraction of action clip $N_i$ is as follows:

$$F_I^{\text{Motion}}(f) = \sum_{i=R_s^f}^{R_e^f} \frac{I(i)}{R_e^f - R_s^f + 1}, \tag{5}$$

where $R_s^f$ represents the start frame number of the same rhythm period where the action of the current $f$ frame is

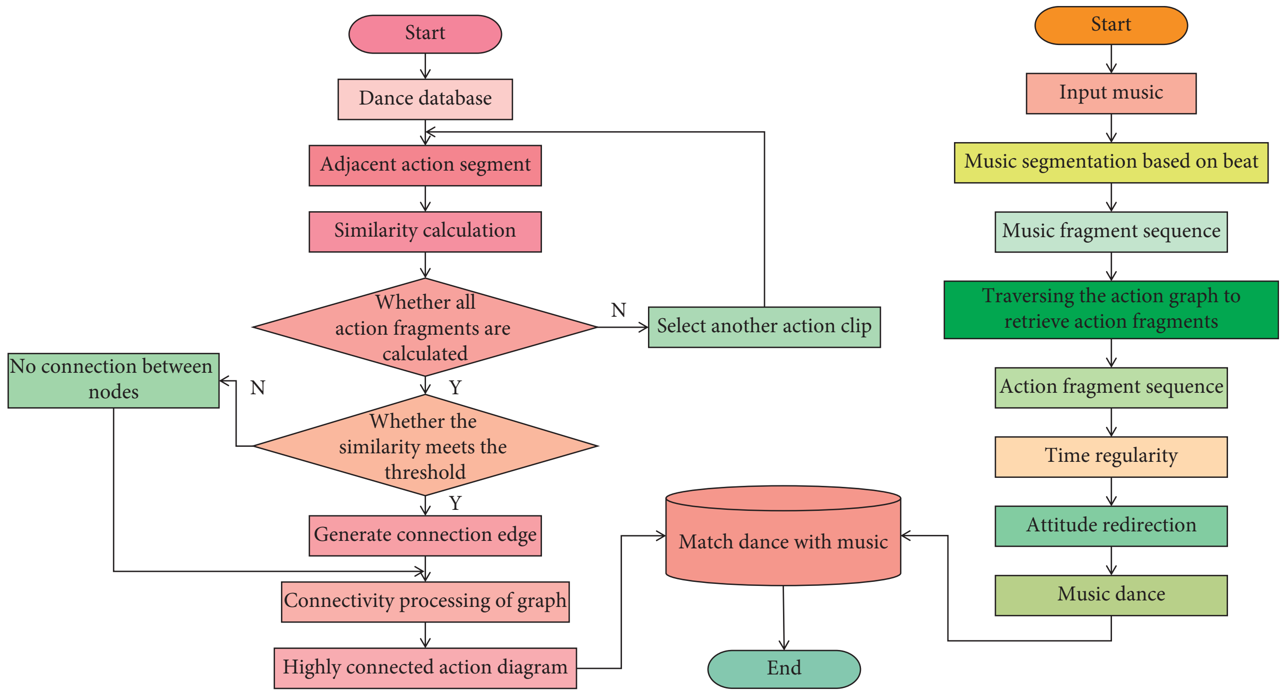


Figure 1: Schematic diagram of processing flow.

located and $R_e^f$ represents the end frame number of the same rhythm cycle where the action of the current $f$ frame is located.

In the dance-music segment matching, music features and action features are corresponding. To maintain the correspondence of features, the segmentation method of action segments refers to the windowing and framing method of music segments. Good feature expression is very important for the final accuracy of the algorithm. The calculation and testing of various pattern recognition algorithms are mainly embodied in the feature extraction stage, which takes the most time. Therefore, feature extraction is very important for the accuracy of dance-music matching. The segmentation algorithm in this study can segment the motion simply and effectively and at the same time ensure that the motion length after segmentation meets the music matching requirements. After the preprocessing of music data and action data is completed, the features are analyzed and extracted based on music segments and action segments.

*3.3. Optimization of Matching between Technical Movements and Music of Folk Dance Based on DL.* In establishing the principle model of dance technical action system, the collected music data are segmented, and the action segments in the dance action database are connected and organized to obtain the underlying features of historical music and dance actions, carry out correlation analysis to extract some feature pairs, and calculate the correlation coefficient between music and dance movement features. In the training stage, the system trains the corresponding relationship between music and movements. Firstly, the training parameters are set to determine the implicit buckling and the number of variables in each layer. To obtain a good network model, it is necessary to constantly train and test, adjust the relevant step length and times in time to improve the accuracy of the test, and determine the number of hidden layers and training parameters of the network through training. Then, the score dance is divided synchronously according to the beat, and music segments and action segments of equal length are obtained, and the combination of music segments and action segments is taken as a training example. After extracting music and motion features, each music segment and motion segment can be described by feature vectors. Table 1 gives the parameter settings of DL network in this study.

*Methods.* Traditionally, the bottom features of the style of music-dance bottom are obtained, and then, the correlation analysis of the bottom features of music and dance is carried out to remove redundant feature pairs to complete the matching. Its disadvantage is that it cannot extract the movement-music segment combination and cannot perform beat alignment and dance movement-music coefficient calculation, which leads to poor synchronization between matching dance movements and music changes. Therefore, this study puts forward a dance technical action-music matching optimization method based on DL. Considering the practical application, the common feature matching of music and movements can be divided into rhythm matching and intensity matching. The data of music and dance movements are both time series, so it is meaningless to only consider the overall statistical features when matching features. Dancing with music, the most intuitive match between music and action is that the characteristics of each time point are one-to-one correspondence. Through the motion segment retrieval, we can get a series of motion

Table 1: Parameter settings of DL network.

| Layer name | Patch size | Input size |
|---|---|---|
| Conv1 | $7 \times 7/2$ | $224 \times 224 \times 2$ |
| Conv2 | $5 \times 5/2$ | $117 \times 117 \times 96$ |
| Conv3 | $3 \times 3/1$ | $56 \times 56 \times 256$ |
| Conv4 | $3 \times 3/2$ | $14 \times 14 \times 512$ |
| Conv5 | $3 \times 3/1$ | $14 \times 14 \times 512$ |
| Conv6 | $3 \times 3/2$ | $14 \times 14 \times 512$ |
| Conv7 | $3 \times 3/1$ | $36 \times 36 \times 192$ |
| FullConnect1 | Logits | $7 \times 7 \times 512$ |
| FullConneet2 | Logits | $1 \times 1 \times 2048$ |
| Softmax | — | $1 \times 1 \times 500$ |

segments that match the music segments, and then, we need to synthesize the music and motion segments. Folk dance technical movement-music synthesis processing includes regular length of music and action segments and reorientation of human posture in action segments.

Suppose the abrupt point function value of the *n*th time window is $D_s(n)$. The extreme value detection on the abrupt point function of each time window is performed, the obtained extreme point sequence is aligned according to the beat, and the final abrupt point function value $D_{(n)}$ is expressed as follows:

$$D_{(n)} = D_H(n) \times D_s(n). \tag{6}$$

Assuming that $T_n$ represents the position of the music beat in the current stage, the position of the next music beat can be estimated through the music beat cycle:

$$T'_{n+1} = T_n + \tau_{\max}. \tag{7}$$

Among them, $T'_{n+1}$ represents the predicted value and $T_{n+1}$ represents the true beat position. The function $Di\ st_m(M_i, M_j)$ is introduced to measure the distance between $M_i$ and $M_j$, and the following formula is used to estimate the matching error of dance movement and music:

$$L\left(A_i, M_j\right) \underset{=}{\Delta} 1 - \exp\left(-Di\ st_m\left(M_i, M_j\right)\right). \tag{8}$$

The objective function of dance action matching optimization is expressed by the following formula:

$$F(n) \underset{=}{\Delta} \sum_{j=1}^{n} R\left(j, u_j\right) + \gamma \sum_{j=2}^{n} T\left(u_{j=1}, u_j\right), \tag{9}$$

where $R(j, u_j)$ represents the simplified representation of $S(A_j, M_{uj})$, which represents the matching error. $T(u_{j=1}, u_j)$ represents the smoothness score of the transition from dance action segment $M_{uj-1}$ to dance action segment $M_{uj}$. $Y$ stands for balancing the weight between the two.

Combined with the DNN algorithm, the corresponding relationship between different music and dance movements is trained, and the accuracy of the corresponding relationship is used as a fitness function to obtain an optimal corresponding relationship between music and dance movements, and the principle model of dance technical movement system is established. The correspondence between music and motion matching calculation is complex, and the matching feature pairs contained in the correspondence between different kinds of dances and music are quite different. According to the types of national dance, it is of great significance to determine the corresponding relationship between technical movements of national dance and music for the realization of music-driven automatic generation system of dance movements. Before feature matching, firstly, based on motion capture data, a motion database is established to store the rhythm and intensity characteristics of each motion segment. Then, when inputting a piece of music, first the characteristics of the music are analyzed and the music segments and the corresponding rhythm and intensity characteristics are obtained. The matching framework of folk dance technical movements and music is shown in Figure 2.

In the music-action correspondence training process, it has been discovered that as the number of matching feature pairs in the correspondence increases, the accuracy of the correspondence between music and action improves. The accuracy of correspondence between music and action does not rise noticeably when the number of matching feature pairs increases to a specific amount, indicating that sufficient correspondence accuracy can be achieved with a small number of matching feature pairs. The main feature of music and activity matching is rhythm. Simultaneously, feature matching considers the connectability of adjacent segments. Finally, the action that best fits the target audio is filtered out of the connectable action sequence of rhythm matching based on the intensity characteristic. This study uses DNN to train the correspondence between diverse folk dance movements and music in order to acquire the relationship between matching accuracy and operation speed. The key is the fitness function design, which starts with improving computation accuracy and employs correspondence accuracy as the fitness function. The degree of synchronization of rhythm points is taken into account when matching music and movements. The synchronization of dance and music is primarily represented in the one-to-one correlation between the rhythm points of music and activity in time, and the degree of synchronization can be assessed by the number of rhythm points that match. At the same time, the action is allowed to expand and compress on a tiny scale.

## 4. Result Analysis and Discussion

To verify the comprehensive effectiveness of the DL-based dance technique movement and music matching optimization method proposed in this study, simulation is required. The implementation language of the system is C++, and the compilation tool is MATLAB. The implementation environment is Windows, and the database server used is MySQL. The experimental data are taken from the dance motion capture database provided by a university for folk dance, and the music data set is selected from the George data set. The sampling frequency is 110 frames/second; the music data in the music data set use a 44 kHz sampling rate and 16 bit quantization bits during the sampling process. Each training data set contains 2000 training examples of dance movements

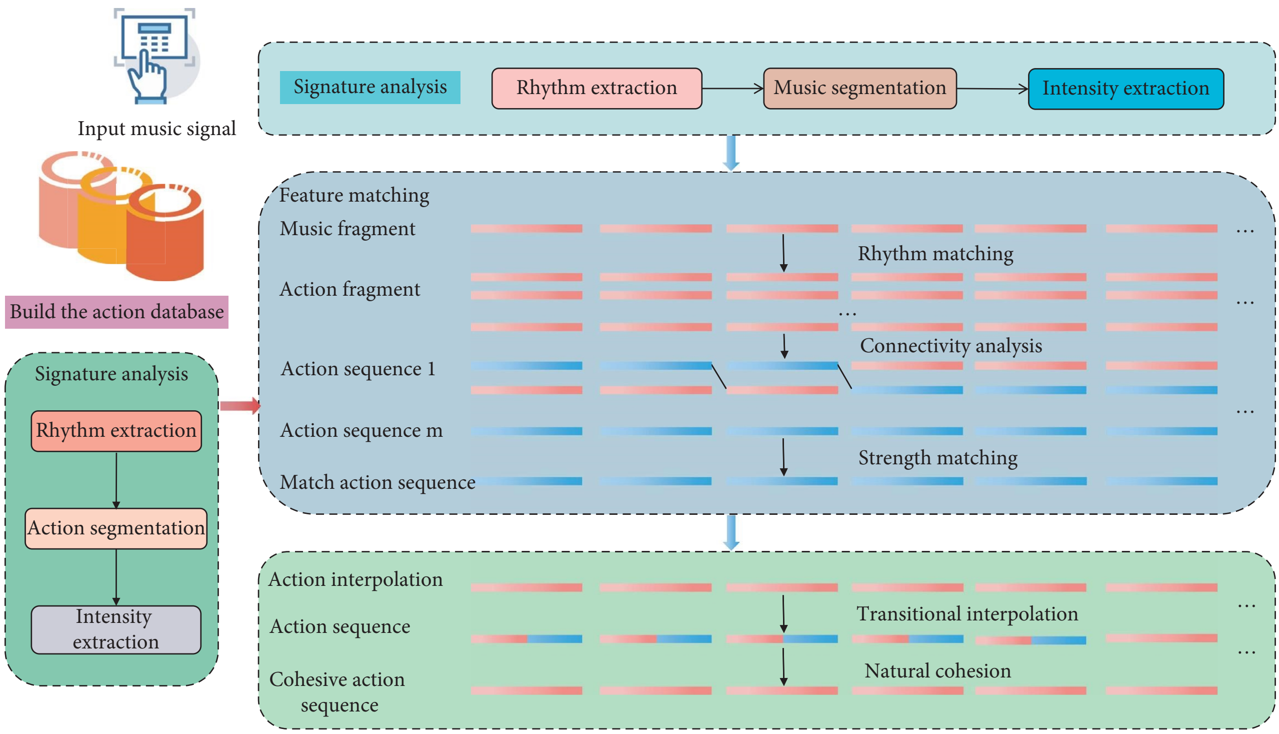


FIGURE 2: Matching framework of folk dance technical movements and music.

TABLE 2: Data set parameters.

| Parameter name | Parameter value |
|---|---|
| Total number of categories | 110 |
| Total video | 2000 |
| Average video duration | 5.13 s |
| Total video duration | 1450 mins |
| Sampling rate | 44 kHz |
| FPS | 25 fps |
| Resolution | 320 × 240 |
| Digitalizing bit | 16 bit |

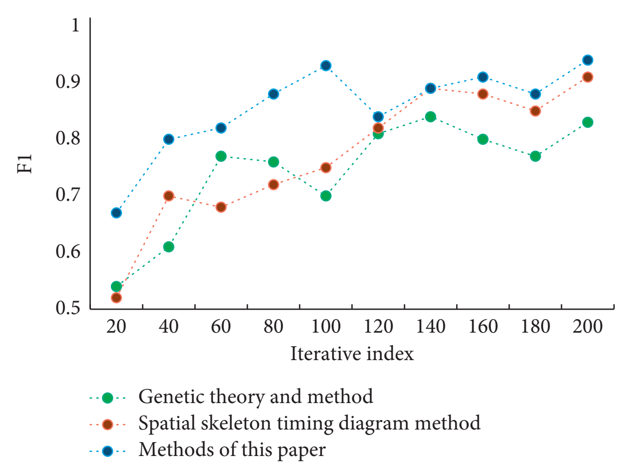


FIGURE 3: $F1$ value comparison chart.

performed by real people. The data set parameters are shown in Table 2.

The joints selected by the human model include 18 positions, including head, left and right shoulders, left and right elbows, left and right wrists, pelvis, left and right knees, and left and right ankles. In the training stage, when determining the corresponding relationship between technical movements and music of national dance, this study uses the training data set of national dance including 20 dominant music. Among them, the length of each piece of music is about 5 minutes. Each piece of music is divided according to the beat, and the length of each piece of music is about 1.5 seconds. The error experiments are carried out, and the $F1$ values of different methods are shown in Figure 3.

In this study, the error is kept as small as possible through continuous iteration, and the same parameter matrix is used to constrain the encoding and decoding of data, in order to reduce the number of parameters and control the complexity of the model. The complexity of the model refers to the time complexity and the space complexity. Time complexity determines the training or prediction time of the model. If the complexity is too high, it will take a lot of time to train and predict the model, and it is impossible to quickly verify ideas, improve the model, and predict it quickly. Using this method, genetic theory method, and spatial skeleton timing diagram method, respectively, the optimization experiment of dance action matching in music choreography is carried out, and the synchronization of dance action matching in music choreography is compared with different methods. The comparison results are described in Figure 4.

It can be seen that the synchronization of matching using this method is the best. In the stage of automatic generation of dance movements, six pieces of music different from the

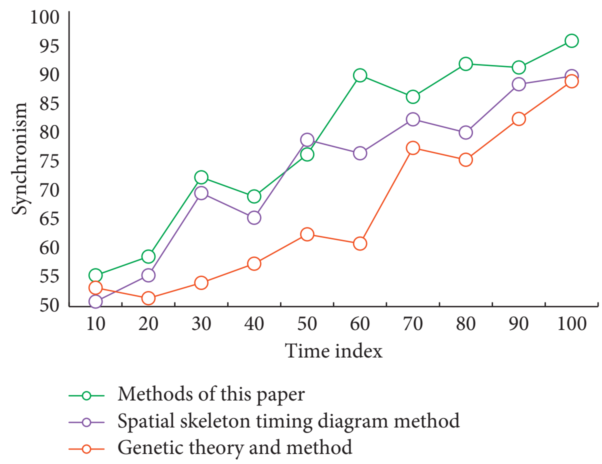


Figure 4: Synchronization comparison results.

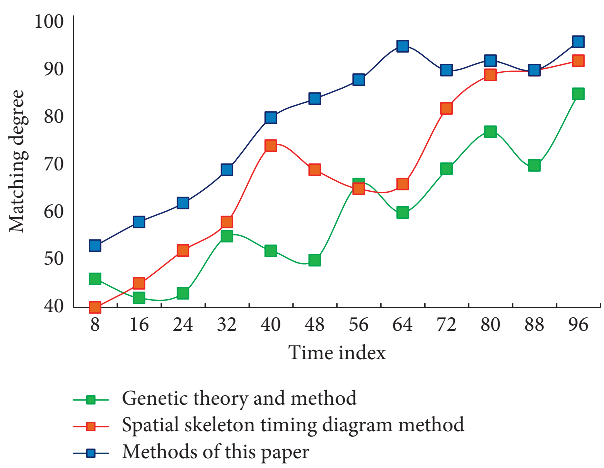


Figure 5: Comparison of the matching degree between dance movements and music of different methods.

training data set are used to compose the test data set for dance generation. The purpose of this is to use different music to objectively evaluate the effect of music-driven dance movements. Considering the dance effect, the dance can keep its original rhythm after the path editing algorithm is used to edit the path, while the original rhythm of the dance is basically lost after the path editing algorithm based on keyframe interpolation is used to edit the path. Therefore, the path editing algorithm provided in this study, under the condition that the movement path does not change much, the dance can really and naturally move under the new path without obvious slippage. We use this method, the genetic theory method, and the spatial skeleton timing chart method to carry out the matching experiment between dance movements and music and compare the matching degree of dance movements and music with the three methods. The comparison results are described in Figure 5.

By analyzing the data in Figure 5, it can be seen that the matching degree between dance movements and music using this method is better than that of the genetic theory and spatial skeleton timing diagram. This is mainly because the method in this study firstly integrates the theory of music

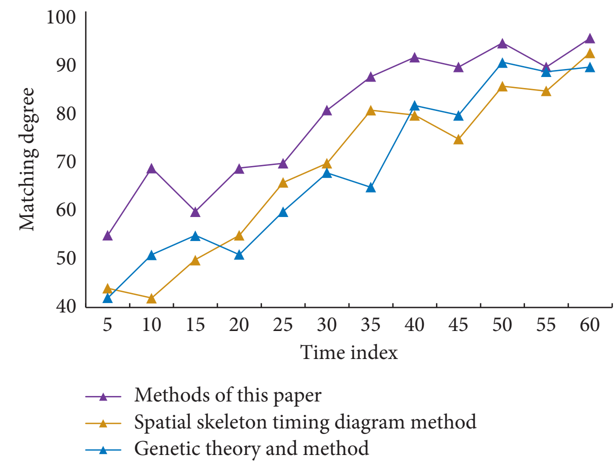


Figure 6: Comparison of the matching degree of dance movements in different methods.

beat extraction to divide the synchronized dance movements and music data and then obtains a plurality of movements with shorter length-music fragment combinations. Then, the correlation coefficient between dance movements and music pieces is calculated, and the objective function of dance movements matching optimization is obtained. Then, the objective function is optimized and solved, so that the matching degree between dance movements and music in this method is better. In the training stage, after completing the corresponding study of the technical action of folk dance and music, the optimal corresponding relation is selected according to the accuracy of the corresponding relation. In the stage of automatic generation of dance movements, dance movements are generated using the dance-music correspondence in the experimental data set for specific kinds of dances such as national dances. Different methods are used to optimize the matching of dance movements in music choreography. The results are shown in Figure 6.

By analyzing Figure 6, we can see that the matching degree of dance movements in music choreography using this method is better than that of the genetic theory method and spatial skeleton timing diagram method. DNN adopts the weight sharing strategy; that is, in the convolution process of the convolution layer, the same convolution kernel is used to convolve different receptive fields of the image, which further reduces the network structure parameters, reduces the complexity of the network model, makes the network easier to train and learn, and improves the learning performance of the network. To evaluate the overall performance of the match between the bottom features of folk dance technique and music, this study compares the accuracy of dance synthesis between the bottom features match and the top statistical features match. The calculation of dance synthesis accuracy is obtained by comparing synthetic dance with live dance through the formula of dance synthesis accuracy. The matching accuracy of different methods is shown in Figure 7.

It can be seen that the accuracy of dance-music matching in this study is higher than that of the genetic theory method and spatial skeleton timing diagram method. It has certain

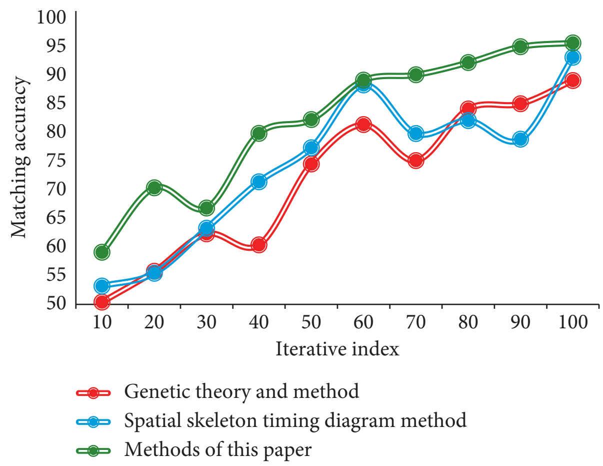


Figure 7: Matching accuracy of different methods.

advantages and accuracy and achieves the expected effect. The results of a large number of experiments in this section show that the matching method of dance movements and music optimized by DL can reach 95.78% accuracy, and the synchronization of dance-music matching can reach 96.17%. This algorithm can effectively change the path of the dance, make it truly and naturally move along the new path, and keep the rhythm of the original dance without serious foot slip and other distortion problems. The dance matched by the bottom feature map is more in line with people's aesthetics than the dance generated by the high feature map.

## 5. Conclusions

Minority cultures abound in China, and ethnic dance, as a vital component of many cultures, is an important section of intangible cultural heritage. The match between folk dance and music has begun to be digitized as a result of the rapid development of diverse technologies. The fit between stage and music can be improved by carefully matching technical moves and folk dance music. The most significant element in a match between technical motions and folk dance music is to extract features properly. The DL algorithm is currently one of the most efficient ways of extracting video information. This research proposes a DL-based optimization approach for matching folk dance technical movements with music. To investigate the performance of the method suggested in this research, a vast number of experiments and analyses were conducted. The simulation results demonstrate that the matching approach optimized in this study achieves 95.78 percent accuracy and 96.17 percent synchronization of dance-music matching. The proposed method is capable of accurately synchronizing and matching music and movement changes. The research presented in this study has aided in the promotion of the use of contemporary digital tools in the study of dance art. It is the result of reciprocal promotion, collaboration, and development of culture, science, and technology, and it aids in the promotion of the cultural digital construction project. However, because of my limited knowledge and research time, there are still some flaws in the music and dance feature matching approaches. The next stage in this research will be to do a more in-depth investigation into automatic feature matching across modes.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## Acknowledgments

This study was supported by Exploring the Interdisciplinary Promotion Model of Music Humanities Courses under the Background of Quality Education (number: JR2015Y09) and a Comparative Study on Music Education for Visually Impaired Students in Mainland China and Hong Kong, Macao, and Taiwan (number: 20210055).

## References

[1] J.-K. Tsai, C.-C. Hsu, W.-Y. Wang, and S.-K. Huang, "Deep learning-based real-time multiple-person action recognition system," *Sensors*, vol. 20, no. 17, p. 4758, 2020.

[2] M. Majd and R. Safabakhsh, "A motion-aware ConvLSTM network for action recognition," *Applied Intelligence*, vol. 49, no. 7, pp. 2515–2521, 2019.

[3] N. Dawar and N. Kehtarnavaz, "Action detection and recognition in continuous action streams by deep learning-based sensing fusion," *IEEE Sensors Journal*, vol. 18, no. 23, pp. 9660–9668, 2018.

[4] H. Spooner, J. B. Lee, D. G. Langston, J. Sonke, K. J. Myers, and C. E. Levy, "Using distance technology to deliver the creative arts therapies to veterans: case studies in art, dance/movement and music therapy," *The Arts in Psychotherapy*, vol. 62, pp. 12–18, 2019.

[5] Z. Wang, "Music choreography algorithm based on feature matching and fragment segmentation," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9274098, 9 pages, 2022.

[6] Z. Huang, Y. Liu, C. Zhan, C. Lin, W. Cai, and Y. Chen, "A novel group recommendation model with two-stage deep learning," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021.

[7] X.-B. Jin, W.-T. Gong, J.-L. Kong, Y.-T. Bai, and T.-L. Su, "A variational bayesian deep network with data self-screening layer for massive time-series data forecasting," *Entropy*, vol. 24, no. 3, p. 335, 2022.

[8] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[9] J. Zhou, Y. Wang, and W. Zhang, "Underwater image restoration via information distribution and light scattering prior," *Computers & Electrical Engineering*, vol. 100, Article ID 107908, 2022.

[10] J. Chen, Y. Zhang, L. Wu, T. You, and X. Ning, "An adaptive clustering-based algorithm for automatic path planning of

heterogeneous UAVs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 2, pp. 1–12, 2021.

[11] J. Zhou, X. Wei, J. Shi, W. Chu, and W. Zhang, "Underwater image enhancement method with light scattering characteristics," *Computers & Electrical Engineering*, vol. 100, Article ID 107898, 2022.

[12] J. Zhang, J. Sun, J. Wang, and X.-G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427–8440, 2021.

[13] D. Yao, Z. Zhi-li, Z. Xiao-feng et al., "Deep hybrid: multi-graph neural network collaboration for hyperspectral image classification," *Defence Technology*, 2022.

[14] J. Zhang and H. Hu, "Deep spatiotemporal relation learning with 3D multi-level dense fusion for video action recognition," *IEEE Access*, vol. 7, Article ID 15222, 2019.

[15] F.-P. An, "Human action recognition algorithm based on adaptive initialization of deep learning model parameters and support vector machine," *IEEE Access*, vol. 6, Article ID 59405, 2018.

[16] S. Zhou, J. Qiu, and A. Solanki, "Improved SSD using deep multi-scale attention spatial–temporal features for action recognition," *Multimedia Systems*, vol. 5, pp. 1–9, 2021.

[17] H. Wei, R. Jafari, and N. Kehtarnavaz, "Fusion of video and inertial sensing for deep learning-based human action recognition," *Sensors*, vol. 19, no. 17, p. 3680, 2019.

[18] E. P. Ijjina and K. M. Chalavadi, "Human action recognition in RGB-D videos using motion sequence information and deep learning," *Pattern Recognition*, vol. 72, pp. 504–516, 2017.

[19] A. Akula, A. K. Shah, and R. Ghosh, "Deep learning approach for human action recognition in infrared images," *Cognitive Systems Research*, vol. 50, no. 8, pp. 146–154, 2018.

[20] S. Saha and A. Konar, "Topomorphological approach to automatic posture recognition in ballet dance," *IET Image Processing*, vol. 9, no. 11, pp. 1002–1011, 2015.

[21] A. Parsons and L. Dubrow-Marshall, ""I'm able to put my thoughts into picturing them physically" - phenomenological experiences of Dance Movement Psychotherapy in a Secondary School: unexpected empowerment over external contingency," *The Arts in Psychotherapy*, vol. 64, pp. 1–8, 2019.

[22] Y. He, "Athlete human behavior recognition based on continuous image deep learning and sensors," *Wireless Networks*, vol. 6, pp. 1–12, 2021.

[23] L. Wang, L. Ge, R. Li, and Y. Fang, "Three-stream CNNs for action recognition," *Pattern Recognition Letters*, vol. 92, no. 6, pp. 33–40, 2017.

[24] C. Chen, T. Wang, D. Li, and J. Hong, "Repetitive assembly action recognition based on object detection and pose estimation - ScienceDirect," *Journal of Manufacturing Systems*, vol. 55, pp. 325–333, 2020.

[25] L. Chen, Z. Song, J. Lu, and J. Zhou, "Learning principal orientations and residual descriptor for action recognition," *Pattern Recognition*, vol. 86, pp. 14–26, 2019.

[26] S. Barnet-Lopez, S. Pérez-Testor, J. Cabedo-Sanromà, G. R. Oviedo, and M. Guerra-Balic, "Dance/movement therapy and emotional well-being for adults with intellectual disabilities," *The Arts in Psychotherapy*, vol. 51, pp. 10–16, 2016.

[27] P. Paivi and L. Raimo, "Change in body image among depressed adult outpatients after a dance movement therapy group treatment," *The Arts in Psychotherapy*, vol. 59, pp. 34–45, 2018.