www.cad-journal.net
Computer-Aided Design AND APPLICATIONS

Interactive Experience Design of Traditional Dance in New Media Era Based on Action Detection

Yan Liang[1] and Feng Pan[2]

[1]School of Educational Science and Music, Luoyang Institute of Science and Technology, Luoyang, Henan 471023, China, liangyan2003@126.com
[2]Department of Sports training, Guangzhou Sport University, Guangzhou, Guangdong 510500, China, 11432@gzsport.edu.cn

Corresponding author: Feng Pan, 11432@gzsport.edu.cn

Abstract. In the stage of dance image creation, advanced sci & tech can lay a technical foundation for it, and at the same time give dance art digital characteristics. On the basis of successful motion capture and feature extraction, action detection automatically recognizes dance movements by analyzing the obtained dance movement feature parameters. The timely introduction of dance creation computer-aided design (CAD) system can continuously provide diversified visual modes for dance creators. This text takes the cross-border integration of dance art and sci & tech as the context, takes "interactive dance" under the new media technology as the breakthrough point, combines deep learning (DL) and CAD technology to realize dance action recognition and virtual reality (VR) digital modeling, and explores the design method of traditional dance interactive experience from the perspective of new media. Simulation results show that the accuracy and efficiency of this model are improved compared with traditional algorithms, and it can achieve good recognition accuracy with low parameters. The results verify the superiority of this model, which can improve the interactive experience of traditional dance performance and creation with the support of VR and CAD technology.

Keywords: New Media; Deep Learning; CAD; Dance Design; Action Detection
DOI: https://doi.org/10.14733/cadaps.2024.S7.241-255

1 INTRODUCTION

Due to the growth of society and art, the artistic environment of traditional dance is changing quietly. Many cultural memories have gradually faded, and the quantity of performing groups and audience groups has gradually decreased, which makes the development and inheritance of traditional dance in trouble. Using machine learning to identify musical genres and individual dancers from Motion capture is a challenging task, which requires a lot of data and complex algorithm support. Carlson et al. [1] collected a large amount of Motion capture data, including dancers' movements, postures, expressions, etc., and recorded corresponding music information, such as tracks, genres, etc. Then, it is necessary to preprocess and extract features from the data. This includes removing noise, calibrating Motion capture data, extracting effective features, etc. Next, machine learning algorithms can be used for pattern recognition. Common algorithms include support vector machines, decision trees, neural networks, etc. These algorithms can train models, predict music genres or identify specific dancers according to the Motion capture data of dancers. Finally, the model needs to be evaluated and optimized. This includes calculating the accuracy rate, recall rate, F1 score and other indicators, as well as performing operations such as Model selection and optimization. In general, using machine learning to identify musical genres and individual dancers from Motion capture is a complex and challenging task, which requires a large amount of data, complex algorithms and fine feature extraction techniques. In the stage of dance image creation, advanced sci & tech can lay a technical foundation for it, and at the same time give dance art digital characteristics. Choreographers may use virtual reality technology to create new dance movements and performance forms. Virtual reality technology can provide a new creative environment, allowing choreographers to imagine more unique and imaginative dance movements, and conduct experiments and performances on virtual stages. In addition, virtual reality technology can also provide a safer and more convenient training tool, allowing dancers to practice and perform in a virtual environment without worrying about mistakes or injuries. These new technologies can not only improve the creative efficiency and quality of choreographers, but also bring more shocking and unique learning and experience to the audience. Cisneros et al. [2] provided a new creative environment through virtual reality technology, allowing choreographers to imagine more unique and imaginative dance movements. And conduct experiments and performances on a virtual stage. Virtual reality technology can provide a more efficient and accurate training and creative tool, allowing dancers to practice and perform more quickly and accurately, while reducing the risk of errors and injuries. In order to obtain more interactive artistic expression, the timely introduction of dance creation CAD system can continuously provide diversified visual modes for dance creators and optimize the interactive communication platform between actors and choreographers. Faridee et al. [3] deployed IMUs and IoT devices on the body of dancers to capture the movements and postures of various parts of the body. At the same time, the collected data is sent to a computer or server for processing through wireless or wired transmission. Process and analyze the collected data to extract useful features such as angle, velocity, acceleration, etc. Then, machine learning or deep learning algorithms are used to train and model the data to achieve automated evaluation. Train and evaluate models using historical and real-time data to continuously improve evaluation accuracy and stability. At the same time, models can be used for real-time evaluation to provide timely feedback and guidance. Integrate automated evaluation systems into dance performances, dance teaching, and other scenarios, and deploy and configure them. At the same time, it is necessary to ensure the stability and security of the system to avoid Data breach and system attacks. As a subdivision of the use of action detection, the related music and dance fragments can be quickly obtained through the rapid recognition of dance movements. In dance action detection, bone data has attracted more and more attention because of its robustness and compactness. In traditional methods, the human structure is usually modeled by manually constructing or learning the characteristics of human connecting points. Iqbal and Sidhu [4] conducted a system validation analysis of AR dance training. Through virtual reality technology, create an intelligent coach that can simulate the behavior and language of a real coach, providing personalized guidance and services for dancers. Dancers can practice and perform on virtual stages without worrying about mistakes or injuries. At the same time, the audience can watch the performance through augmented reality technology and interact with the dancers. The acceptance of this system can be evaluated through a technical acceptance model. The technology acceptance model believes that users' acceptance of technology depends on their perception of its usefulness, ease of use, and security. The dance training system based on augmented reality and technology acceptance models has high acceptance potential, which can increase the training experience and interest of dancers through virtual reality technology and gamification. Simultaneously ensuring the safety and usability of the system, making more dancers willing to use this system for dance training.

CAD software can be used to visualize drawings and models of environmental art design for better design and communication. Jin and Yang [5] use CAD software to enable students to visualize designs more accurately and easily make modifications and updates. In the teaching of environmental art design, the application of CAD software can enable students to more accurately grasp design drawings and models, improving the efficiency and accuracy of design. At the same time, CAD software can also be easily modified and updated, making students more flexible in responding to changes and needs in design. Therefore, CAD software plays a very important role in environmental art and design teaching. The interdisciplinary Chinese dance pose aesthetic preference recognition based on neural network intelligent CAD is a transdisciplinarity method that uses neural network and intelligent Computer-aided design (CAD) technology, and combines art, computer science, neuroscience and other fields. Li et al. [6] collected a large number of Chinese dance pose image data, as well as the corresponding manually marked aesthetic scoring data. Design and train a neural network model for automatic aesthetic evaluation of dance pose images. Use intelligent Computer-aided design technology to optimize and improve dance postures to improve their aesthetic value. Conduct comprehensive analysis and research by combining knowledge and methods from multiple fields such as art, computer science, and neuroscience. Through this method, the aesthetic value of dance posture can be automatically evaluated and optimized, providing new tools and ideas for dance creation and performance. At the same time, this interdisciplinary research method can also provide new ways and means for the creation and evaluation of other art fields. Action detection is defined as being able to automatically identify the content of a specific video and further identify the problem of the motion occurring therein. By modeling the spatio-temporal information sequence in the video, analyzing the representation and motion information in it, and identifying the action of the target from a series of observations on behavior and environmental conditions, the mapping relationship between video content and action categories is established. In order to achieve better artistic expression and improve the interactive experience of dance creation and performance, the CAD system and VR technology of dance creation can be introduced in time to help dance creation and development. VR is an important frontier development direction of the new generation of IT. It refers to an interactive 3D dynamic scene and entity behavior system with multi-source information, which can simulate the real-world scene in the computer and immerse users in the environment. This text takes the cross-border integration of dance art and sci & tech as the context, takes "interactive dance" under the new media technology as the breakthrough point, combines DL and CAD technologies to realize dance action recognition and VR digital modeling, and explores the design method of traditional dance interactive experience from the perspective of new media.

Before the application of DL method, the traditional computer vision method was widely used in dance movement recognition. As a new productive force, AI has brought diversified innovation paths to the dance industry. As an important branch of AI, video-based action detection is to study how to identify specific dance movements from specified video sequences. On the basis of successful motion capture and feature extraction, action detection automatically recognizes dance movements by analyzing the obtained dance movement feature parameters. It is of great significance to construct a traditional dance action recognition model with excellent performance. This text mainly studies the optimization method of traditional dance interactive experience design from the perspective of new media:

⊖ In this text, random projection algorithm is used to reduce the dimension of feature vectors, and convolutional neural network (CNN) model is used to learn the reduced dimension training samples, and a dance action recognition model is constructed.

⊜ The original multi-scale spatial graph convolution module is improved by using the layered strategy, and the expansion convolution to time convolution module is introduced to obtain a wider effective receptive field without changing the convolution kernel size. Then the depth separable convolution is used to replace the standard convolution to reduce the parameter quantity and improve the training speed of the model.

Based on DL and CAD technology, this text explores the interactive experience design of traditional dance from the perspective of new media. The specific framework is as follows:

The first section mainly analyzes the significance of traditional dance interactive experience design in the current environment. In the second section, the related technologies of dance movement recognition are introduced, and the traditional dance interactive experience design method is put forward by combining the concept of human-computer interaction and CAD technology. The third section is the description and experiment of dance recognition model. The fifth section is the summary and prospect, which summarizes the contents and contributions of this text, and at the same time, points out the shortcomings of this text and the prospect of future work.

2 RELATED WORK

Different dance movements have different forms and rhythms, and require different animation effects to be used to express them. Coordination of dance movements: dance movements need to be coordinated with music, Stage lighting and other elements to achieve the best visual effect. Liu et al. [7] used CAD software to simulate and test animation effects to ensure their accuracy and feasibility. Use data analysis tools such as MATLAB, Python, etc. to analyze and optimize the data of 3D models and animation effects, in order to improve the performance and efficiency. To sum up, the integration of Scientific visualization design technology and dance art is a research hotspot in the field of Computer-aided design. Through CAD computer-aided analysis, this fusion can be better achieved and the effectiveness and efficiency of the performance can be improved. Liu and Yang [8] analyzed Computer-aided design of contemporary art dance. In the MVC three-layer Architectural pattern based on the B/S mode, the browser interacts with the view layer as the client, accesses the server through the HTTP protocol, and the server processes the requests of the control layer and the model layer, and returns the results to the browser. In the innovative dance teaching mode of Computer-aided design, MVC three-layer Architectural pattern is used to develop Web applications to achieve Visualization display of dance actions, editing and storage of dance actions and other functions. Through the MVC three-tier Architectural pattern, you can better organize the code, improve the Reusability and maintainability of the code, and better adapt to changes in requirements. SKlkanc network has important application value in the field of modern dance action recognition, and there is still much room for improvement. Through in-depth research and improvement of this network, we can better understand and analyze dance movements, providing more intelligent support and assistance for dance performance, dance teaching, and dance rehabilitation. Some researchers have introduced attention mechanisms to weight different features in the network, in order to better capture key features. Some researchers have also increased the network's ability to extract features by increasing network depth and broadening network width. In addition, some researchers have improved the robustness and generalization ability of networks by introducing techniques such as data augmentation and adversarial training. In IEEE Visualization and Computer Graphics, Liu et al. [9] discussed the latest application and research results of SKlkanc network in modern dance action recognition. Researchers in this field have continuously explored and attempted various improvement and optimization strategies based on the SKlkanc network, in order to further improve the performance and recognition accuracy of the network.

Pang and Niu [10] use computer vision technology to detect and track the position and posture of dance movements in dance videos. By detecting the keyframes of dance movements, the features of dance movements can be extracted and subsequently analyzed and recognized. After extracting the features of dance movements, artificial intelligence technology can be used for feature extraction and recognition. Through computer vision and image processing technology, dance movements can be Motion planning and simulation. Simultaneously planning and simulating the posture, speed, and trajectory of dance movements can generate more accurate and natural dance movements, thereby improving the effectiveness and quality of dance performances. Computer vision and image processing technology can evaluate and provide feedback on dance movements. By evaluating and analyzing the posture, speed, amplitude, and other aspects of dance movements, more objective and accurate feedback can be provided, helping dancers better understand and improve their dance movements. In summary, the application of computer vision and image processing in dance video motion recognition can be achieved using artificial intelligence technology. These applications can help dancers, coaches, and audiences better understand and analyze dance movements, providing more intelligent services and support. Dance action recognition based on Rgb-D data is a technology that uses Conjoint analysis of color images and depth images to recognize and classify dance actions. Shaikh and Chai [11] use depth sensors or depth cameras to collect depth information of dancers, while using color cameras to capture color images. Preprocess the collected depth data and color images, including denoising, correction, alignment, and other operations to obtain more accurate feature extraction. Use appropriate feature extraction algorithms to extract key features of dance movements from preprocessed depth data and color images, such as bone key points, posture, action sequences, etc. Integrate the features of depth data and color images to obtain a more comprehensive and accurate feature representation. Use machine learning or deep learning algorithms to classify and recognize the fused features, achieving automatic recognition and classification of dance movements. Tan and Yang [12] conducted a 3D computer dance motion assistance for binary boundary stereo surface image segmentation algorithm. By analyzing the effectiveness of the dance dataset using midpoints, the target algorithm for the iso-surface binary dataset was determined. By optimizing the dance motion design of the computer three-dimensional assistance system, the organic combination of image segmentation and MC algorithm has been achieved. That is to say, using image segmentation and MC algorithm for the design and optimization of dance movements can make dance movements more realistic and natural. Tasnim et al. [13] used 3D motion capture systems such as OptiTrack, Vicon, etc. to collect 3D bone and joint information of dancers, including angles, positions, velocities, etc. Pre-process the collected 3D bone joint information, including Data cleansing, noise removal, Outlier processing, etc. Extract effective features from preprocessed data, including bone joint angle, distance, velocity, etc. The model uses deep learning algorithms, such as Convolutional neural network (CNN), Recurrent neural network (RNN), etc., to learn and train the extracted features and establish a dance action recognition model. Use a test dataset to test the trained model and calculate indicators such as recognition accuracy and error. The use of 3D bone joint information based on deep learning dance motion recognition technology has high recognition accuracy and flexibility. It can adapt to the recognition and classification of various dance movements, providing more intelligent services and experiences for dancers, coaches, and audiences.

Extracting effective features from dance action data is another challenge. The accuracy of Feature selection and extraction directly affects the learning and recognition performance of the model. Therefore, it is necessary to select features that can fully describe dance movements and avoid redundancy and noise. Choosing a suitable deep learning model for dance action recognition is another challenge for Wang and Tong [14]. The selection and training of models need to consider factors such as data volume, data distribution, computational resources, and model complexity. At the same time, it is necessary to optimize the model parameters and structure to improve the accuracy and generalization ability of the model. Building real-time automated dance recognition and evaluation tools requires efficient processing and transmission of data, as well as ensuring the reliability and stability of recognition results. Therefore, algorithms and codes need to be optimized to improve the running speed and efficiency, and avoid Outlier and error identification. Extracting dance motion features from videos or images typically involves some image processing and computer vision techniques. Such as edge detection, shape description, motion tracking, etc. The extracted features need to be separable and descriptive in order to be recognized and understood by machines. Zhai [15] uses machine learning and data mining techniques to classify and cluster the extracted features. This involves the selection of appropriate data mining algorithms and parameters, as well as the preprocessing of data sets and Feature selection. In addition, the recognition of dance movements also needs to consider the diversity and variability of dance movements. Different dance types have different movements and styles, and different dancers in the same dance type also have different ways of expression and styles. Therefore, it is necessary to train a universal model to recognize various dance movements, while also considering personalized needs. The recognition of dance movements based on feature representation and attribute mining requires the support of a large number of datasets. The quality and size of the dataset both affect the performance and accuracy of the model. Therefore, a significant amount of time and effort is required to collect and process datasets. Zhang [16] uses Convolutional neural network and biological image visualization technology to analyze the body changes of high-level dance movements. First, supercomputing can provide powerful computing power and accelerate the training and reasoning process of Convolutional neural network. During the training phase, supercomputing can train multiple network models simultaneously to accelerate training speed and improve recognition accuracy. In the inference stage, supercomputing can provide high-speed computing power to process and analyze a large amount of dance motion data in real-time. Secondly, biological image visualization technology can present the body changes of dance movements in the form of images, in order to better understand and analyze the characteristics and patterns of dance movements. Dancesport action Automatic identification system needs a lot of calculation, including feature extraction, classification and recognition. Microprocessors and microsystems need to support parallel computing in order to quickly process large amounts of data. This can be achieved by using multi-core processors, GPUs, or other parallel computing technologies. Zheng et al. [17] analyzed that Dancesport action Automatic identification system based on computer vision and parallel computing faced many challenges. This includes the complexity of image processing and computer vision, the size and diversity of datasets, and real-time performance requirements. The future development direction may include more efficient algorithms and algorithm combinations, stronger computing power, more efficient data storage and access mechanisms, faster data transmission and communication protocols, etc.

3 INTERACTIVE EXPERIENCE DESIGN OF TRADITIONAL DANCE BASED ON DL AND CAD

3.1 Extraction of Dance Action Feature Data

Video-based dance movement recognition is a basic subject in computer vision research. Although the traditional dance movement recognition has a good recognition rate in some specific occasions, it needs manual design and feature extraction, and is greatly influenced by individuals and surrounding environment. As a new productive force, AI has brought diversified innovation paths to the dance industry. In dance action detection, bone data has attracted more and more attention because of its robustness and compactness. In traditional methods, the human structure is usually modeled by manually constructing or learning the characteristics of human connecting points. The early fusion modal information usually has different characteristics, which leads to the inconsistency of the spatial and temporal dimensions of the extracted feature vectors. This is an obstacle to the network fusion of potential multimodal information in low-level feature space. However, the late fusion lacks the correlation exploration of feature levels between modes. Therefore, when the information of different modes is different, it may be too simple to fuse only in the decision-making stage, and the complementary information of each mode can not be fully and effectively utilized. On the pixel block of the image, the pixel value in the pixel block is operated with the corresponding convolution kernel to obtain the output value of the corresponding block; Then select new pixel blocks step by step, move the convolution kernel, and operate in turn to get the output after convolution of the whole image.

In the task of dance movement recognition, data modes are generally divided into three categories: video data, depth image and bone movement sequence. According to the different data modes of the recognition task, different algorithms or models are designed to complete the recognition task. The core idea of the method based on whole feature representation for dance movement recognition is to extract the whole human contour; Using these three features to model; Action recognition is completed through the constructed model. The dance action detection stage based on skeleton is shown in Figure 1.

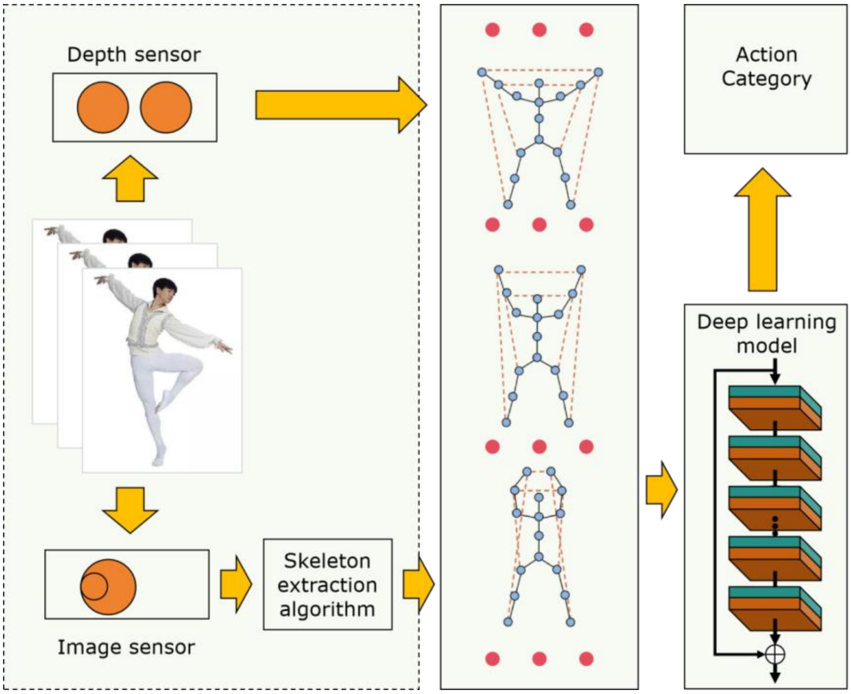

Figure 1: Skeleton based dance action detection stage.

The system mainly adopts skinning algorithm in establishing the action model, and its main principle is to obtain the skin vertex position in a new frame according to the human skin vertex by using linear mixed calculation, thus assisting to complete the decomposition and calculation of human actions:

$$v' = \sum_{i=1}^{n} w_i M_i D_i^{-1} v \tag{1}$$

$$\sum_{i=1}^{n} w_i = 1 \tag{2}$$

The vertex coordinates of human skin before and after degeneration are expressed by v and v' respectively.

Input into the network is the 2D or 3D coordinate information of connecting points, which is modeled and a Shi Kongtu is constructed. The connecting points are regarded as nodes in the graph, and the bones connected by adjacent connecting points are regarded as edges in the graph. In order to enhance the generalization ability of the CNN, prevent the over-fitting of the network and reduce the sensitivity of the network to noise in the stage of CNN training, 10% cut data sets are flipped, rotated and scaled in this text. The multi-order feature maps obtained by convolution operation of spatial maps of each order are connected and input into the time module.

The purpose of this module is to process the time sequence information and learn the relevant features of connecting point changes in time series to superimpose the time features for the maps.

3.2 Dance Movement Recognition and Interaction Model

In the past action detection methods, a single feature can only describe a certain aspect of human action in the video, and it can't effectively describe human action. Adjacency polynomials make the distant connecting points accessible. Through the training of large-scale raw data, the network can extract the features that can best express the raw data, and then make predictions or classification of samples.

With the appearance of CNN, more and more researchers apply it to the task of action detection. CNN has shown excellent performance in image recognition and other tasks, and can automatically extract features from low-level to high-level. Compared with the traditional method of manually extracting data features, CNN automatically extracts richer and more abstract features from the understanding of the object itself. CNN adopts the form of partial connection, and only some neurons in the network are connected. CNN generally consists of three parts, which are convolution layer for extracting features, pooling layer for reducing the size of feature graph and full connection layer. The model of dance video image enhancement processing based on CNN is shown in Figure 2.

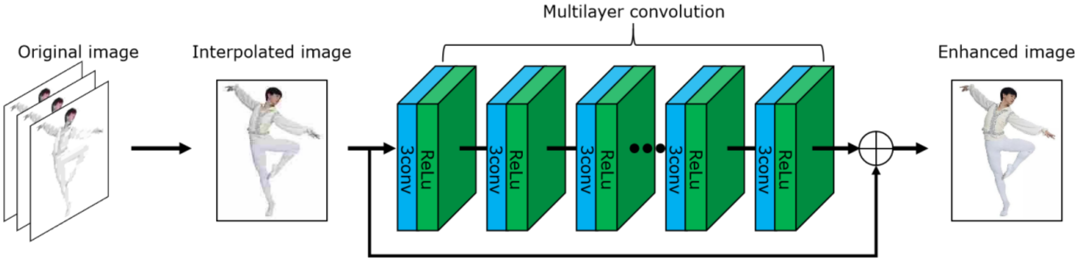

Figure 2: Dance video image enhancement processing model based on CNN.

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

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

The intensity characteristics of action fragments can be extracted from multiple aspects, including the vertical displacement velocity of the feet and the horizontal displacement velocity of the root joint. Specifically, the speed signal of foot vertical displacement can be collected through sensors and used as a part of the intensity feature of the action segment. At the same time, the horizontal displacement velocity signal of the root joint can also be collected through sensors and used as another part of the intensity feature of the action segment.

By synthesizing the signals from these two parts, the instantaneous intensity of the current action clip E can be obtained, which can be described by a formula:

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

The specific synthesis method can be designed and adjusted according to actual application scenarios and requirements:

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

In the middle stage of multi-modal information fusion, the feature information of video and skeleton is globally averaged and pooled channel by channel without reducing the dimension. Then, the cross-channel interaction of feature information is obtained by capturing the channel features of each channel and its K neighbors. In the skeleton topology diagram, the strength of the connection is represented by the thickness of the line segment between nodes, and the dotted line represents the connection other than the physical connection. In the stage of drawing, for the newly-built connections between nodes, several connections with the highest connection strength are selected to add, while for the existing connections, the thickness of the line segment is modified according to the change of connection strength.

Attention in both spatial and temporal dimensions acts on input features at the same time to enhance the representation ability of features. The elements in the two attentional diagrams reflect whether the concerned information exists in the joints of the corresponding frames, and the joints with the most abundant information in a specific frame can be distinguished from the whole skeleton sequence, thus helping to improve the recognition performance of the whole model.

Because the dynamic range distribution of different dimensional data is very different, this text normalizes each dimensional data to make it a unit vector with a modulus of 1, which is convenient for subsequent calculation. Let p_λ be the probability density function, $\lambda=\left[\lambda_1,\lambda_2,\lambda_3,\ldots,\lambda_M\right]$ be the M parameter vectors of p_λ ; $X_1=\left[x_t,t=1,2,3,\ldots,T_I\right]$ be the effective features of tennis training action videos. FF obeys independent distribution; d is the feature dimension after dimension reduction, including K Gaussian unit parameter sets:

$$\lambda=\left\{w_i,u_i,\Sigma_i\right\} \quad i=1,2,3,\ldots,k \tag{6}$$

Its Gaussian mixture model is:

$$p_\lambda\left(x_t\right)=\sum_{i=1}^{k} w_i p_i\left(x_t\right) \tag{7}$$

In the formula, w_i , u_i , Σ_i are the mixed weight, mean vector and covariance matrix, respectively; $p_i\left(x_t\right)$ is the i th Gaussian unit of x_t . According to the Bayesian formula, the calculation formula of the probability that x_t is assigned to the i th Gaussian unit is:

$$r_t(i)=w_i p_i\left(x_t\right)/\sum_{k=1}^{k} w_k p_k\left(x_t\right) \tag{8}$$

Then the gradients of x_t with respect to $\lambda=\left\{w_i,u_i,\Sigma_i\right\} \quad i=1,2,3,\ldots,k$ are respectively expressed as:

$$\frac{\partial \ell_\lambda(X)}{\partial w_i}=\sum_{t=1}^{T}\left[\frac{r_t(i)}{w_i}-\frac{r_t(1)}{w_1}\right] \tag{9}$$

$$\frac{\partial \ell_\lambda(X)}{\partial u_i^k}=\sum_{t=1}^{T} r_t(i)\left[\frac{x_t^k-u_i^k}{\left(\sigma_i^k\right)^2}\right] \tag{10}$$

$$\frac{\partial \ell_\lambda(X)}{\partial \sigma_i^k}=\sum_{t=1}^{T} r_t(i)\left[\frac{\left(x_t^k-u_i^k\right)^2}{\left(\sigma_i^k\right)^3}-\frac{1}{\sigma_i^k}\right] \tag{11}$$

In the formula, σ_i^k represents the standard deviation in the covariance matrix Σ_i.

For the modeling of bone sequence in time series, time chart is composed of continuous frames connected by corresponding connecting points in time dimension. Because the time window is predefined, there is no flexibility in identifying different actions. Therefore, a time convolution module including long, medium and short scales is proposed to obtain the time characteristics in different time domain receptive fields. In addition, the time domain receptive field used by adaptive adjustment and dynamic combination of weights is introduced, which can capture the temporal characteristics of bone data more effectively.

4 MODEL TESTING AND RESULT ANALYSIS

The Folk Dance data set used in the study is the dance data set made by the laboratory itself, and the motion capture device During the whole data set making process, the final scheme is discussed with dance experts according to the data set making plan to design four groups of traditional dance movements. This data set contains many kinds of dance movements, and the dance movements are complex, so it is challenging to identify dance movements. The training set is used to train the classifier, and after the training is completed, the model obtained through training is tested by using the test set, so as to assess the performance of the classifier, which is the feasibility of the algorithm. This text uses 7727 dance performance video samples with complex backgrounds and great differences in visual angles, and will select 4500 samples as the training set and other samples as the test set. PyTorchDL framework was adopted in the experiment. Select SGD as the optimizer. The loss function is cross entropy loss function.

In time series, it is equally important to deal with the time dimension. An action may contain multiple stages, and different sub-stages have different effects on the final recognition results. Before the formal training of the model, the parameters of the model are initialized. In the convolution layer and batch standardization, the convolution kernel parameters and batch standardization parameters are all standardized into random numbers with a mean value of 0.02 and a standard deviation of 1, and the bias parameters are set to 0. The model training situation is shown in Figure 3.

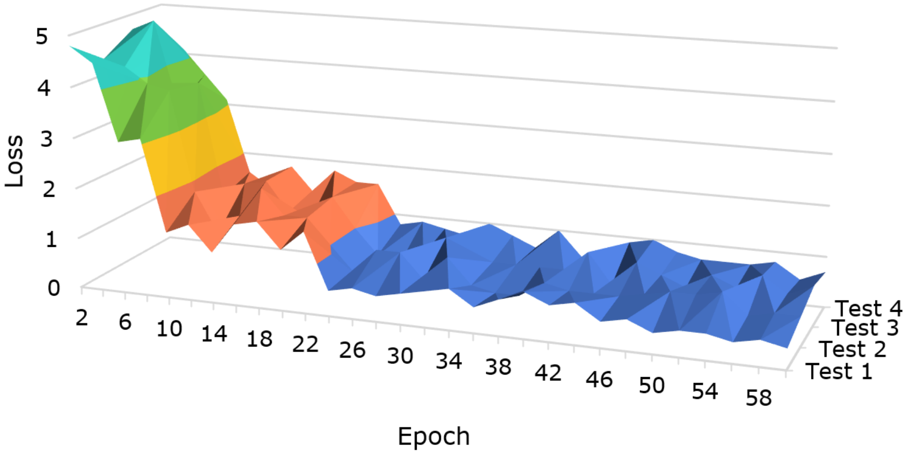

Figure 3: Model training situation.

Dance movements need to be described by corresponding feature parameters, and different movements may focus on different feature parameters, so it is necessary to express complex motion characteristics with the most effective and appropriate features. After encoding the spatio-temporal information, in order to make use of these representations, position attention generation is introduced to make full use of the captured spatial position information and time dependence, and effectively obtain the relationship between channels.

In order to express the function of adaptive module more intuitively, the learnable adjacency matrix obtained by model training is visualized. After the dance action feature passes through the classifier, the feature vector is mapped into a classification score vector, that is, the probability distribution vector of the sample corresponding to the dance action feature in the whole movement category. Taking the index position corresponding to the maximum probability value is the prediction classification result of the model. The classification accuracy of different algorithms is shown in Figure 4.

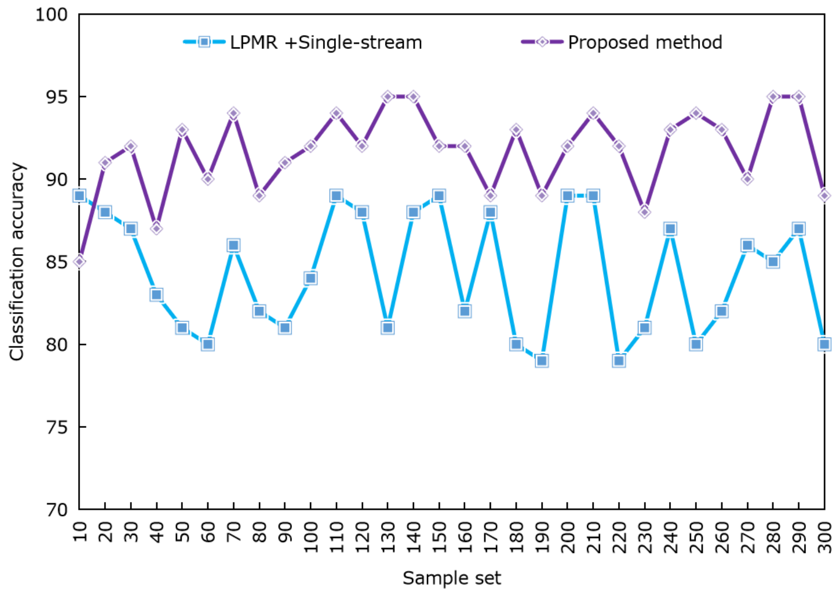

Figure 4: Classification accuracy of different algorithms.

In this text, both skeleton information and depth information are used, and the recognition rate is improved. The reason is that when the skeleton is not accurate, the depth information is used, thus eliminating the interference caused by the unstable skeleton part, thus improving the accuracy of recognition.

Given the feature map as input, two branches are connected to generate temporal features and spatial features respectively. Global pooling is usually used to capture global spatial information. Because the spatial information is compressed into the channel dimension by pooling operation, it is difficult to retain the spatial position structure information. Therefore, the global pooling is decomposed into two one-dimensional feature coding operations, so that the attention module can capture the long-distance interaction of position structure information in time and space. The results of dance performance action recognition by this method and the comparison method are shown in Table 1.

Dance type	*Training sample*		*Test sample*	
	LPMR +Single-stream	*Proposed method*	*LPMR +Single-stream*	*Proposed method*
Classical dance	90.69%	94.55%	83.44%	94.72%
Folk dance	85.38%	95.16%	84.65%	93.36%

Contemporary dance	87.37%	95.22%	82.28%	94.29%

Table 1: Correct rate of dance performance action recognition.

Figure 5 shows the error analysis results of the training set using the method used in this article and the LPMR+single stream method. Figure 6 shows the errors of different algorithms on the test set. Figure 7 shows the response times of different algorithms.

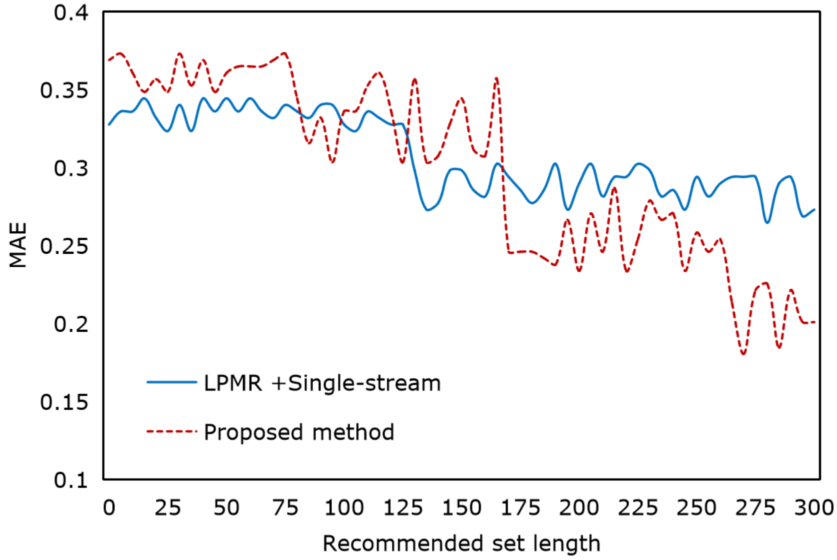

Figure 5: Error on training set.

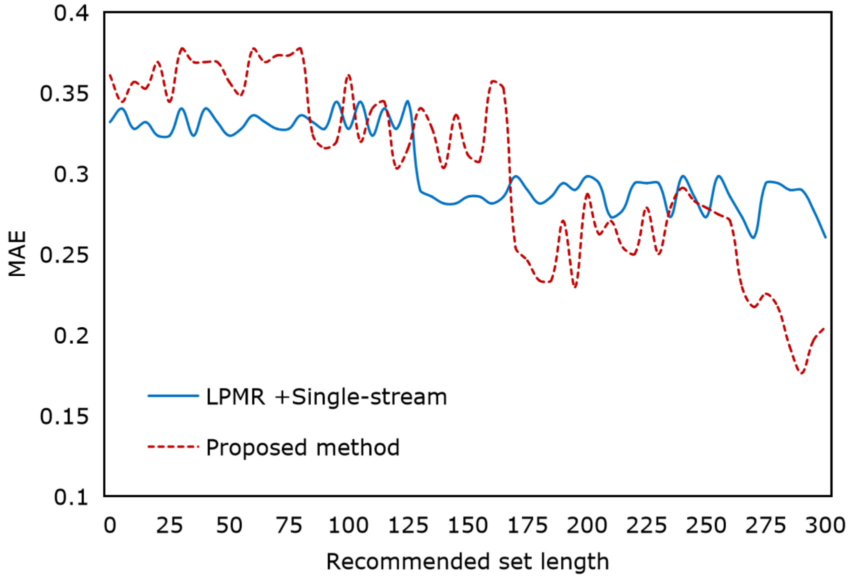

Figure 6: Error on test set.

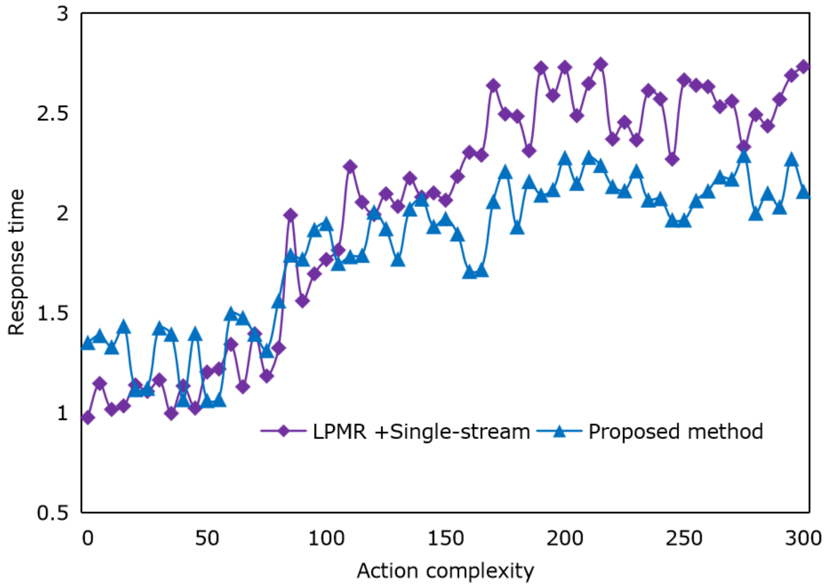

Figure 7: Response time of different algorithms.

With the introduction of nonlocal and self-attention blocks, due to their ability to build channels and spatial attention modules, nonlocal mechanisms can be used to capture space-time and channel information. The calculation time of different algorithms is shown in Table 2.

Dance type	***Training sample***		***Test sample***	
	LPMR +Single-stream	***Proposed method***	***LPMR +Single-stream***	***Proposed method***
Classical dance	6.77	5.36	8.31	5.59
Folk dance	8.52	4.45	6.78	4.08
Contemporary dance	7.78	5.44	7.62	3.48

Table 2: Dimension reduction time of dance performance action methods.

According to the comprehensive test results, it is not difficult to find that the accuracy and operation efficiency of this model are improved compared with the comparative model, which verifies the superiority of this model and can improve the interactive experience of traditional dance performance and creation with the support of VR and CAD technology. No matter what the function diagram of the corresponding virtual dance action design is, the working technicians should take the morphological and molecular changes of each limb joint of the designated dancing performance object when making a certain dance action as the key reference object for setting the parameters of the design scheme. Only the setting of various parameters and indicators in the virtual design environment can better follow the characteristics of the actual dancer's skeleton, and the dance action design scheme has its inherent feasibility in the practical performance.

5 CONCLUSIONS

The application of CAD function to the practical workflow of dance design work is an inevitable trend of dance design work gradually moving towards modern design mode. Due to the growth of depth sensor and its robustness to complex background, dance action detection based on bones has attracted wide attention. With the appearance of CNN, more and more researchers apply it to the task of action detection. Based on the cross-border integration of dance art and sci & tech, this text combines DL and CAD technologies to realize dance action recognition and VR digital modeling, and explores the design method of traditional dance interactive experience from the perspective of new media. Compared with the contrast model, the accuracy and operation efficiency of this model are improved, which verifies the superiority of this model and can improve the interactive experience of traditional dance performance and creation with the support of VR and CAD technology. In the action detection model based on skeleton points, this text uses prior knowledge in feature extraction, however, different feature extraction methods will have a great influence on the accuracy of the model. In the future, how to use end-to-end model design method to design neural network model will become an important topic.

Yan Liang, https://orcid.org/0009-0002-3640-6777
Feng Pan, https://orcid.org/0009-0000-7673-2791

REFERENCES

[1] Carlson, E.; Saari, P.; Burger, B.; Toiviainen, P.: Dance to your own drum: Identification of musical genre and individual dancer from motion capture using machine learning, Journal of New Music Research, 49(2), 2020, 162-177. https://doi.org/10.1080/09298215.2020.1711778

[2] Cisneros, R.-E.-K.; Wood, K.; Whatley, S.; Buccoli, M.; Zanoni, M.; Sarti, A.: Virtual reality and choreographic practice: The potential for new creative methods, Body, Space and Technology, 18(1), 2019, 1-32. https://doi.org/10.16995/bst.305

[3] Faridee, A.-Z.; Ramamurthy, S.-R.; Roy, N.: Happyfeet: Challenges in building an automated dance recognition and assessment tool, GetMobile: Mobile Computing and Communications, 22(3), 2018, 10-16. https://doi.org/10.1145/3308755.3308759

[4] Iqbal, J.; Sidhu, M.-S.: Acceptance of dance training system based on augmented reality and technology acceptance model (TAM), Virtual Reality, 26(1), 2022, 33-54. https://doi.org/10.1007/s10055-021-00529-y

[5] Jin, H.; Yang, J.: Using CAD software in teaching environmental art design, Computer-Aided Design and Applications, 19(S1), 2021, 173-183. https://doi.org/10.14733/cadaps.2022.S1.173-183

[6] Li, J.; Wu, S.-R.; Zhang, X.; Luo, T.-J.; Li, R.; Zhao, Y.; Peng, H.: Cross-subject aesthetic preference recognition of Chinese dance posture using EEG, Cognitive Neurodynamics, 17(2), 2023, 311-329. https://doi.org/10.1007/s11571-022-09821-2

[7] Liu, F.; Gao, Y.; Yu, Y.: Computer aided design in the diversified forms of artistic design, Computer-Aided Design and Applications, 19(S3), 2021, 33-44. https://doi.org/10.14733/cadaps.2022.S3.33-4-44

[8] Liu, F.; Yang, K.: Exploration on the teaching mode of contemporary art computer aided design centered on creativity, CAD and Applications, 19(S1), 2021, 105-116. https://doi.org/10.14733/cadaps.2022.S1.105-116

[9] Liu, Y.; Zhang, H.; Li, Y.; He, K.; Xu, D.: Skeleton-based human action recognition via large-kernel attention graph convolutional network, IEEE Transactions on Visualization and Computer Graphics, 29(5), 2023, 2575-2585. https://doi.org/10.1109/TVCG.2023.3247075

[10] Pang, Y.; Niu, Y.: Dance video motion recognition based on computer vision and image processing, Applied Artificial Intelligence, 37(1), 2023, 2226962. https://doi.org/10.1080/08839514.2023.2226962

[11] Shaikh, M.-B.; Chai, D.: Rgb-d data-based action recognition: A review, Sensors, 21(12), 2021, 4246. https://doi.org/10.3390/s21124246

[12] Tan, B.; Yang, F.: Dance movement design based on computer three-dimensional auxiliary system, Computer-Aided Design and Applications, 19(S3), 2021, 108-118. https://doi.org/10.14733/cadaps.2022.S3.108-118

[13] Tasnim, N.; Islam, M.-M.; Baek, J.-H.: Deep learning-based action recognition using 3D skeleton joints information, Inventions, 5(3), 2020, 49. https://doi.org/10.3390/inventions5030049

[14] Wang, S.; Tong, S.: Analysis of high-level dance movements under deep learning and internet of things, The Journal of Supercomputing, 78(12), 2022, 14294-14316. https://doi.org/10.1007/s11227-022-04454-y

[15] Zhai, X.: Dance movement recognition based on feature expression and attribute mining, Complexity, 2021(21), 2021, 1-12. https://doi.org/10.1155/2021/9935900

[16] Zhang, R.: Analyzing body changes of high-level dance movements through biological image visualization technology by convolutional neural network, The Journal of Supercomputing, 78(8), 2022, 10521-10541. https://doi.org/10.1007/s11227-021-04298-y

[17] Zheng, H.; Liu, D.; Liu, Y.: Design and research on automatic recognition system of sports dance movement based on computer vision and parallel computing, Microprocessors and Microsystems, 80(11), 2021, 103648. https://doi.org/10.1016/j.micpro.2020.103648