Recognition of typical skills and movements of classical dance based on attention machine graph convolutional network

Zhen Liu[a], Yimei Zhu[b*]

Zhuzhou Normal College, No. 89 Zhizhi Road, Yunlong Demonstration Zone, Zhuzhou City, Hunan Province, China

[a]370753411@qq.com, *[b]751533276@qq.com

Abstract. Classical dance is a form of dance with a profound historical and cultural heritage, and its techniques and movements are crucial for the performance and skill level of dancers. With the development of artificial intelligence technology, graph convolutional networks based on attention mechanisms have become an effective tool for identifying and analyzing typical techniques and movements in classical dance. This technology helps students improve their skills by capturing key movements of dancers, providing real-time feedback and personalized guidance. Teachers can use this to evaluate student performance, adjust teaching content and methods, and promote the progress of dance education. This article aims to explore the recognition of typical movements in classical dance using graph convolutional networks based on attention mechanisms.

Keywords: attention mechanism, Action recognition, Typical Techniques and Actions of Chinese Classical Dance.

1 Introduction

With the advancement of digital technology, classical dance typical techniques and action recognition based on attention machine graph convolutional networks have attracted widespread attention. Accurately identifying the skills and movements of dancers is crucial for improving learning outcomes in dance teaching[1-3]. The traditional human dance pose recognition methods have some shortcomings in practical applications, which fail to fully cover various application scenarios of dance pose segmentation[4]. The limitations are quite significant, and the description effect of action information is poor, which is easily affected by complex backgrounds[5-6]. Therefore, their recognition accuracy and detection accuracy are relatively low[7]. This paper explores a technique for integrating attention mechanisms into graph convolutional networks[8]. This approach focuses on identifying and understanding the spatial and temporal connections between crucial joints during dance movements. By doing so, it allows the network to concentrate more effectively on essential data, thereby enhancing its ability to recognize standard maneuvers in classical dance with greater accuracy[9-10]. This

M. Yu et al. (eds.), *Proceedings of the 2024 5th International Conference on Big Data and Informatization Education (ICBDIE 2024)*, Advances in Intelligent Systems Research 182,
https://doi.org/10.2991/978-94-6463-417-4_53

method not only has a positive promoting effect on the research of classical dance, but also has important significance for promoting the integration and development of artificial intelligence technology and humanities and arts.

2 Research Contents

Our approach is based on two core technologies: graph convolutional networks (GCN) and attention mechanisms. GCN can effectively process non-Euclidean data (such as human joints and limb movements in dance movements) and capture the spatial relationships in movements; while the attention mechanism enhances the model's sensitivity and identification ability of key movement features. Specifically, we employ three attention mechanisms: spatial attention, temporal attention, and channel attention to weight and highlight important spatial locations, key time frames, and feature channels of dance movements, respectively.

In this research, the attention mechanism employed is a soft attention model. The dimensions for both input and output of the attention module align with those of convolutional layers, allowing for seamless integration into the convolutional neural network architecture.

2.1 Channel Attention

The primary objective of the channel attention module is to empower the model with the capability to dynamically modify the importance assigned to each channel. This adjustment enhances its ability to detect features that are crucial for the task at hand. This helps the model learn during training in which channels are more important for the current task. Channel attention modules can often be embedded into different layers of deep learning networks to enhance the model performance for specific tasks. Based on this, we present a channel attention module, the architecture of which is depicted in Figure 1.

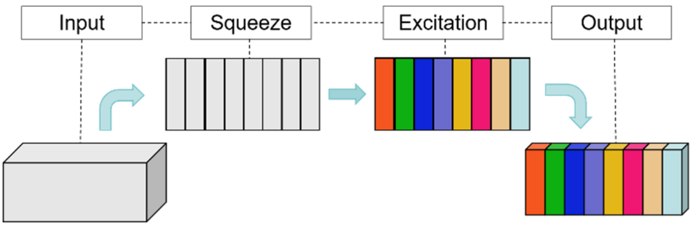

Fig. 1. Channel attention module.

The structure of the network, which incorporates the channel attention module, is depicted in Figure 2.

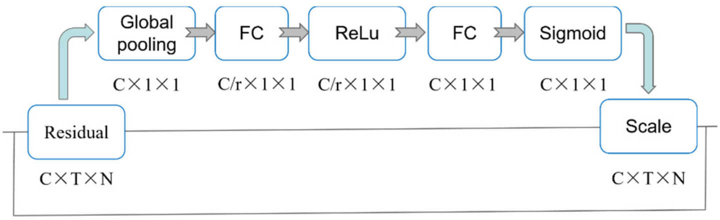

Fig. 2. Channel attention residual structure.

2.2 Spatial attention

The spatial attention mechanism helps the model focus on important spatial areas of the image or feature map.Following the integration of the spatial attention module, its resultant output underwent additional processing through convolution and pooling layers to derive more advanced features. Finally, the recognition results of classical dance movements were obtained through the fully connected layer and Softmax activation function. By introducing the spatial attention module, the quality of spatial feature extraction by the network was further improved.

The spatial attention module is expressed as Formula 1:

$$s = \sigma\left(w_s\left(AvgPool\left(f_{in}\right)\right)\right) \tag{1}$$

In this configuration, the input is processed by module $f_{in} \in R^{C \times T \times N}$, followed by $AvgPool$ which performs global average pooling across the time dimension. Subsequently, w_s applies a one-dimensional convolution operation, and σ signifies the use of a sigmoid activation function.

After the residual connection, the structure of the entire spatial attention module is depicted in Figure 3.

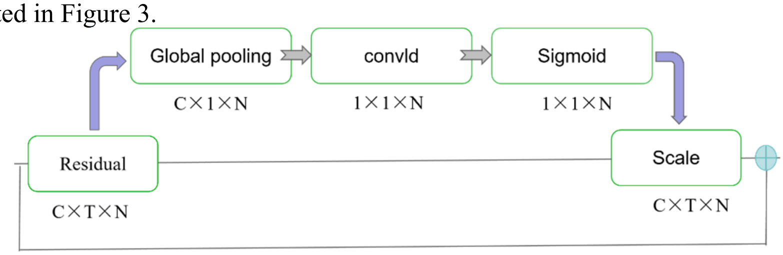

Fig. 3. Spatial attention residual structure.

2.3 Time attention

The temporal attention mechanism focuses on key frames in action sequences and is particularly important for action recognition. By assigning corresponding weights to different frames, the problem of which frames should be focused on in the time series is solved so that the network can better extract time series features. The configuration of the temporal attention module mirrors that of the spatial attention module. The temporal attention module is expressed as Formula 2.

$$s = \sigma\left(w_t\left(AvgPool\left(f_{in}\right)\right)\right) \tag{2}$$

The difference is that $AvgPool$ is a global average pooling operation that compresses the spatial dimension information. Finally, a temporal attention weight with dimensions of 1×T×1 was obtained.

3 Experiment and Analysis

In order to fully verify the effectiveness and feasibility of the classical dance typical skill action recognition model based on the attention mechanism graph convolution network. We established a standard library data set of typical techniques and movements of classical dance, and collected Chinese classical dance videos from multiple public dance video libraries and cooperative art groups to ensure that a wide range of classical dance is covered, as shown in Figure 4.

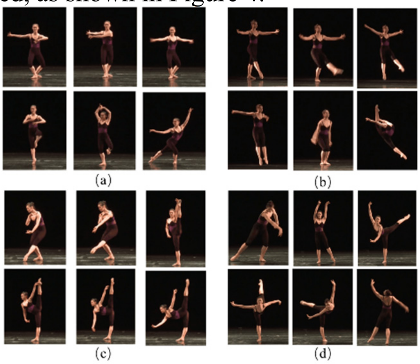

Fig. 4. Standard library data set of typical dance postures and movements of classical dance. (a) Spin; (b) Jump; (c) Control; (d) Turn over

Dance professionals annotate the video frame by frame, accurately mark the typical technical movements appearing in the video, and classify each movement. Unify the format and adjust the resolution of the collected videos, and extract the dancer's key point information through the existing pose estimation algorithm as the input of the graph convolution network. The data set contains different classical dance types, dancers of different genders and age groups, and performances in different environments to ensure the generalization ability of the model. Ensure that the data set contains a sufficient number of samples, at least several hundred samples for each typical skill move, to support the training needs of the deep learning model. 80% of the data is used for training, 10% is used for verification, and 10% is used for testing to ensure the independence of the test data to fairly evaluate the model performance.

Through a series of ablation studies, the model's attention mechanisms were systematically eliminated. This process aimed to assess the impact of channel, spatial, and temporal attention on recognizing classical dance movements. The experiment selected a dataset of typical movements in classical dance, which had two evaluation criteria: (1) cross subject (CS). In this evaluation, the division of training and validation datasets was based on the volunteers who performed the actions. (2) Cross-View (CV) Benchmark. For this benchmark, the division of training and validation datasets was determined by the camera angle. The model underwent training with the designated dataset, during which its parameters were fine-tuned according to the loss function's feedback. Subsequently, its efficacy was assessed on a testing dataset. A statistical analysis was conducted to juxtapose the performance metrics of each model under experimentation against those in the test set, with findings detailed in Table 1.

Table 1. Effectiveness of each attention module.

Attention module	Cross subject (%)	Cross view (%)
Full Model Group	87.5	94.7
No Channel Attention Group	87.1	93.8
No Spatial Attention Group	86.9	93.6
No Temporal Attention Group	86.7	93.5
No attention model group	84.5	92.6

It is evident that the inclusion of each attention module contributes to enhancing the model's performance, especially the difference between the two models with and without the attention mechansism, and the recognition accuracy of models with and without the attention mechanism for various actions in the classical dance typical skill video dataset (cross-object benchmark). The results are presented in Figures 5.

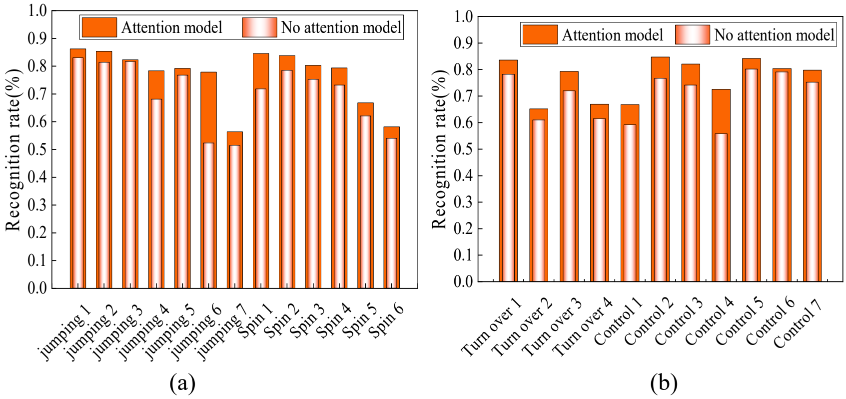

Fig. 5. Comparison of action recognition accuracy in typical classical dance technique data sets. (a) Jumping and spinning; (b) Turning over and controlling.

4 Conclusion

In summary, the dance action recognition method based on the attention GCN proposed in this study has stronger robustness and higher recognition accuracy compared to the performance of a single feature in the same group of classical dance typical skill action datasets. It can be applied more easily to practical recognition requirements.

Acknowledgment

This study was supported by the Science Research Project (22C1496) of the Education Department of Hunan Province.

References

1. Jia R, Zhao L, Yang R, et al. HFA-GTNet: Hierarchical Fusion Adaptive Graph Transformer network for dance action recognition[J]. Journal of Visual Communication and Image Representation, 2023: 104038.
2. Mallick T, Das P P, Majumdar A K. Posture and sequence recognition for Bharatanatyam dance performances using machine learning approaches[J]. Journal of Visual Communication and Image Representation, 2022, 87: 103548.
3. Rallis I, Voulodimos A, Bakalos N, et al. Machine learning for intangible cultural heritage: a review of techniques on dance analysis[J]. Visual Computing for Cultural Heritage, 2020: 103-119.
4. Liu L. Moving object detection technology of line dancing based on machine vision[J]. Mobile Information Systems, 2021, 2021: 1-9.
5. Chang X, Peng L. Visual sensing human motion detection system for interactive music teaching[J]. Journal of Sensors, 2021, 2021: 1-10.
6. Bek J, Arakaki A I, Lawrence A, et al. Dance and Parkinson's: A review and exploration of the role of cognitive representations of action[J]. Neuroscience & Biobehavioral Reviews, 2020, 109: 16-28.
7. Kong Y, Fu Y. Human action recognition and prediction: A survey[J]. International Journal of Computer Vision, 2022, 130(5): 1366-1401.
8. Li C, Zhang J, Yao J. Streamer action recognition in live video with spatial-temporal attention and deep dictionary learning[J]. Neurocomputing, 2021, 453: 383-392.
9. da Costa V G T, Zara G, Rota P, et al. Dual-head contrastive domain adaptation for video action recognition[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022: 1181-1190.
10. Shah A, Mishra S, Bansal A, et al. Pose and joint-aware action recognition[C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022: 3850-3860.