DLINE JOURNALS

Digital Signal Processing and Artificial Intelligence for Automatic Learning

Online ISSN: 2583-5009

DSP 2024; 3 (3)

https://doi.org/10.6025/dspaial/2024/3/3/97-103

Application of Dynamic Intelligent Simulation Technology in Dance Teaching

Ning Fu*, Xia Peng
Normal College, Xinxiang Vocational and Technical College, 453006
Xinxiang, Henan, China
funing20031317@163.com

ABSTRACT

This paper is based on the dance movement recognition model of 3D Convolutional Neural Networks (CNNs) and aims to explore the feasibility of applying intelligent technology to dance teaching. By investigating the current status of intelligent technology in dance teaching both domestically and internationally, the advantages of 3D CNNs in dance movement recognition are analyzed. In response to the needfor dance movement recognition, a model based on 3D CNNs is designed and validated using the MSRAction3D dataset. The experimental results show that the model achieves high recognition accuracy in 20 dance movement categories, proving its potential application in dance teaching. This model can provide dance teachers with accurate movement recognition and personalized guidance, improving teaching effectiveness.

Received: 10 June 2024
Revised: 18 August 2024
Accepted: 31 August 2024

Keywords: 3D Convolutional Neural Networks, Dance Movement Recognition, Simulation Technology, Dance Teaching

1. Introduction

In dance teaching, the application of dynamic intelligent simulation technology has attracted widespread attention. Traditional dance teaching typically relies on teachers' demonstrations and students' imitation, but this approach has limitations. Firstly, teachers cannot provide guidance simultaneously at different locations and times, restricting students' learning resources. Secondly, imitating the teacher's movements may be difficult for students, especially beginners. Therefore, dance teaching systems based on intelligent technology have become a research hotspot, including applying dynamic, intelligent simulation technology in dance movement recognition. This study aims to implement dance movement recognition based on 3D Convolutional Neural Networks (3D CNNs) to promote the development of intelligent dance teaching. Through this research, we aim to enhance the effectiveness of dance teaching and provide students with intuitive and vivid learning methods, allowing them to learn and imitate movements by observing virtual dancers generated by the system. Furthermore, we aim to achieve personalized and autonomous learning

by adjusting the difficulty of dance movements based on students' abilities and levels to help them improve their dance skills [1]. In addition, this study aims to provide dance teachers with auxiliary tools to demonstrate different dance movements and interact with students, thereby enhancing teaching effectiveness. Importantly, by deeply studying and applying the effectiveness of 3D CNNs in dance movement recognition, we will provide beneficial experience and references for the design and development of intelligent dance teaching systems, promoting the advancement of this field. Through these efforts, we hope to bring innovation and improvement to dance teaching and enhance learners' dance skills and experiences.

The content of this study is based on dance movement recognition using 3D CNNs. We will use this technology to extract key features from input dance video data and classify and recognize dance movements. By building and training the 3D CNN model, we will explore how to effectively capture the spatiotemporal information of dance movements and achieve accurate motion recognition in real-time or offline environments. Key tasks in the research include data collection and preprocessing, network model design and optimization, feature extraction, and selection and implementation of classification algorithms. Through these steps, we aim to achieve a high-accuracy dance movement recognition system and provide powerful auxiliary tools for dance teaching.

2. Current Application Status of Intelligent Technology in Dance Teaching

The application of intelligent technology in dance teaching has received extensive attention and exploration both domestically and internationally. The following is a summary of the current application status of intelligent technology in dance teaching.

Researchers have proposed *DeepPose*, a pose estimation method based on deep neural networks. It outputs highly accurate pose position coordinates through cascaded deep neural network regression. This method utilizes the latest advancements in deep learning to predict poses in a complete method [2]. Some researchers explored the introduction of convolutional networks into the pose estimation framework for learning image features and used image-based spatial models for pose estimation tasks. They designed sequential models consisting of convolutional networks to implicitly model long-term dependencies between variables [3]. This architecture can directly operate on confidence maps from previous stages to generate more accurate keypoint estimates without explicit graphical model predictions. Intermediate supervision and natural learning objectives were introduced to address the vanishing gradient problem and strengthen backpropagation gradients during training [4]. These models demonstrated advanced performance on MPII, LSP, and FLIC datasets. Researchers used Recurrent Neural Networks (RNNs) with Long-Short Term Memory (LSTM) to model skeleton data for 3D action recognition. LSTM-based methods showed excellent capabilities in time series modeling and achieved good results in 3D action recognition tasks [5]. To better utilize the kinematic relationships between body joints, researchers proposed a spatiotemporal LSTM-based RNN to model spatial correlations of skeletons. This network effectively learned spatial relationships between skeleton joints, improving action recognition performance [6]. In applying intelligent technology in dance teaching abroad, some progress has been made using technologies such as RNNs, attention models, and CNNs. These methods are significant for modeling and feature learning of skeleton data, contributing to improved accuracy and effectiveness of dance movement recognition.

Domestic researchers have researched dance movement recognition using deep learning techniques such as CNNs and RNNs. They trained deep learning models to extract features from dance video data to recognize various dance movements automatically. For example, a research team designed a dance movement classification model based on deep learning, which accurately identifies various dance movements and provides a powerful auxiliary tool for dance teaching. Domestic researchers have also begun to explore the application of Virtual Reality (VR) technology in dance teaching [7]. They use VR devices and sensors to create a virtual dance environment, immersing students in the experience and learning of dance movements. For example, a research team developed a VR-based dance teaching system where students can wear VR headsets to interact with virtual dance instructors, imitate various dance movements, and learn in real-time [8]. Accurately capturing and tracking students' body postures in dance teaching is critical. Domestic researchers are devoted to developing efficient and accurate human pose estimation and tracking technologies to monitor students' dance movements in real time and provide feedback. For example, a research team proposed a deep learning-based human pose estimation method to track students' skeleton joints, compare them with standard dance movements, and help students correct movements and improve their skills [9].

In summary, researchers have conducted extensive research on the application of intelligent technology in dance teaching both domestically and internationally. They have developed various dance teaching systems and tools using deep learning, virtual reality, and pose estimation technologies to enhance the effectiveness and experience of dance learning. These studies provide important support and promotion for developing domestic dance education and applying intelligent technology.

3. Dance Movement Recognition Model Based on 3D Convolutional Neural Networks

3.1. Theoretical Basis of 3D Convolutional Neural Networks

The 3D Convolutional Neural Network (CNN) is a deep learning model specifically designed for processing three-dimensional data with a temporal dimension, such as videos and volumetric data. Compared to traditional 2D CNNs, 3D CNNs can effectively capture the correlation between time and space, enabling the extraction of richer feature representations. In traditional 2D CNNs, the convolutional kernel slides over two-dimensional input data and performs convolution operations to extract local features. In contrast, in 3D CNNs, the convolutional kernel slides over three-dimensional input data, considering the dimensions of time, height, and width. This enables 3D CNNs to model dynamic changes in time sequences and better understand features along the time dimension.

The basic structure of 3D CNNs is similar to that of 2D CNNs, consisting of convolutional layers, pooling layers, and fully connected layers. The convolutional layers use multiple kernels to slide over spatial and temporal dimensions, extracting local features from the input data. The pooling layers reduce the dimensions of feature maps while preserving the most significant features. The fully connected layers map the features to the final output categories or predictions. A key feature of 3D CNNs is parameter sharing. Similar to 2D CNNs, the convolutional kernels in 3D CNNs share parameters across the entire input data, reducing the model's number of parameters. This parameter sharing enables extracting local features from the input data and maintains parameter sharing along the time dimension, effectively capturing temporal information of movements. To train 3D CNNs, a large annotated dataset is typically required. By matching input videos with corresponding labels, the backpropagation algorithm is used to update the parameters in the network to minimize the difference between predicted results and true labels [10]. As training progresses, 3D CNNs learn feature representations with good discriminative abilities and use them to classify or predict new samples during the testing phase. The differences between 2D and 3D CNNs are illustrated in Figure 1.

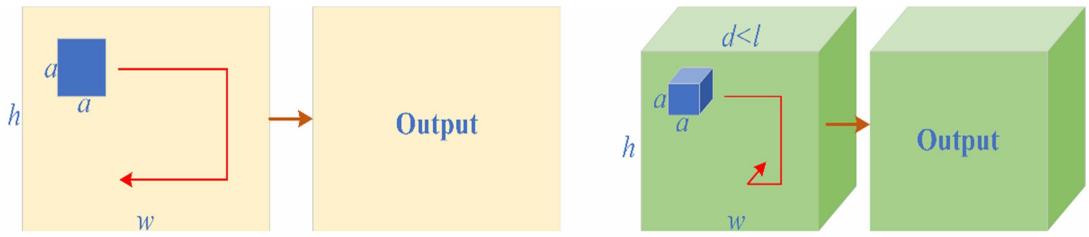

Figure 1. Differences between 2D Convolution and 3D Convolution]

In 3D CNNs, the convolution operation slides over not only the spatial dimension but also the temporal dimension. By sliding the convolutional kernel in the time, height, and width dimensions, 3D CNNs can capture features in time sequences and three-dimensional space. The convolution operation performs a weighted sum of the input data at each position and introduces a non-linear transformation through an activation function to extract local features. After each convolutional layer, pooling operations are usually applied to reduce the dimensions of feature maps. Pooling operations slide over the time, height, and width dimensions and aggregate features within each region, thereby reducing the size of the feature maps. Common pooling operations include max pooling and average pooling. Finally, fully connected layers map the features to the final output categories or predictions. The fully connected layers flatten the feature maps from the previous layer and input them into one or more fully connected neurons, achieving the mapping from features to categories. Training 3D CNNs typically requires a large annotated dataset. By matching input data with corresponding labels, gradient descent and backpropagation algorithms are used to update the parameters in the network to minimize the difference between predicted results and true labels.

3.2. Dance Movement Recognition Based on 3D Convolutional Neural Networks

3D CNNs have wide applications in video analysis, action recognition, behavior recognition, medical image processing, and other fields. 3D CNNs can capture dynamic spatial relationships and temporal information by utilising their ability to handle time-series data, leading to more accurate action or behavior classification. Furthermore, 3D CNNs are widely used in video content analysis, video retrieval, behavior monitoring, and other tasks, providing powerful feature learning and representation capabilities.

This study uses a 3D deep convolutional neural network consisting of four convolutional layers, two down-sampling layers, two fully connected layers, and one *Softmax* classification layer. The downsampling layers use Max-pooling with a kernel size of 3×3×3 and a stride of 1 (as shown in Figure 2).

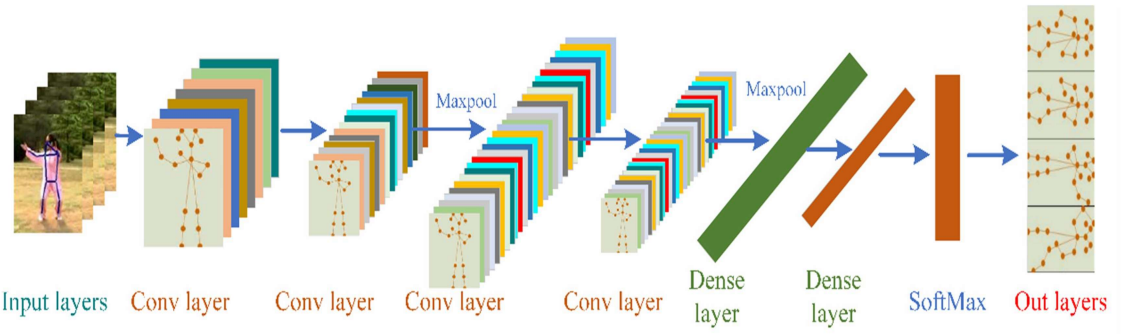

Figure 2. Framework of the 3D CNN Model

To capture motion information from multiple consecutive frames in both spatial and temporal dimensions, the unit's value at position coordinates (*i, j*) in the j^{th} feature map of the i^{th} layer is calculated as shown in Formula (1). Here, the time dimension of the 3D convolutional kernel is denoted as *t*, and the weight value of the convolutional kernel connected to position (*r, s*) with the r^{th} feature map is denoted as w.

$$V_{ij}^{xyz} = f\left(b_{ij} + \sum_r \sum_{l=0}^{l_i-1} \sum_{m=0}^{m_i-1} \sum_{n=0}^{n_i-1} \omega_{ijr}^{lmn} v_{(i-1)r}^{(x+l)(y+m)(z+n)}\right) \quad (1)$$

The *ReLU* function is the most commonly used activation function in deep learning models. This function can make the model's parameters sparse, thereby reducing overfitting and computational complexity. The *ReLU* activation function is defined as shown in Formula (2).

$$f(x) = \max(0, x) = \begin{cases} 0, (x \le 0) \\ x, (x > 0) \end{cases} \quad (2)$$

The calculation of max-pooling in the model is shown in Formula (3), where *i* represents the three-dimensional input vector, *V* represents the output after the pooling operation, and *s, t*, and *r*represent the sampling step size in the respective directions.

$$V_{x,y,z} = \max_{0 \le i \le s_1, 0 \le j \le s_2 0 \le k \le s_3} \left(\mu_{x \times s+i, y \times t+j, z \times r+k}\right) \quad (3)$$

The dance movement recognition model based on 3D CNN is a deep learning model that uses 3D CNN to recognize and classify dance movements automatically. By extracting features and modeling time sequences from the input dance video data, this model can effectively capture the spatiotemporal features of dance movements. The basic structure of this model includes convolutional layers, pooling layers, fully connected layers, and a classifier. The main components and functions of this model are as follows:

Convolutional Layers: The core part of 3D CNN. It uses multiple convolutional kernels to slide over the time, height, and width dimensions, extracting local features from the input data. The convolution operation performs a weighted sum of the input data at each position and introduces a non-linear transformation through the activation function to extract spatiotemporal features.

Pooling Layers are used to reduce the dimensions of feature maps and retain the most significant features. Pooling operations slide over the time, height, and width dimensions and aggregate features within each region. This helps reduce the size of the feature maps, lower the computational complexity, and preserve important spatiotemporal features.

Fully Connected Layers: Maps the extracted features to the final output categories or predictions. It connects each neuron and introduces weight parameters to map high-dimensional features to target categories.

Classifier: Performs classification on the features. The classifier can be a *softmax* function, support vector machine (SVM), etc. It matches the learned features with known dance movement categories and outputs the most likely class label.

4. Experimental Results Analysis of the Dance Movement Recognition Model Based on 3D CNN

To verify the effectiveness of the proposed method in dance teaching, the method is experimentally validated on the MSRAction3D dataset. The MSRAction3D dataset is a publicly available action recognition dataset provided by Microsoft Research. It aims to promote research in action recognition using depth sensor data. The dataset contains human action data captured by the Microsoft Kinect depth sensor. It includes 20 action categories covering various daily activities, such as waving, dancing, handshaking, and hugging. Each action category has multiple samples performed by different participants, resulting in 567 action samples. Each action sample consists of the 3D coordinates of human joints recorded by the Kinect sensor. The recorded data describes the posture and motion of the human body during the actions.

Firstly, action data, including the coordinate data of skeletal joints and corresponding action category labels, are extracted from the MSRAction3D dataset. The dataset is then divided into training, validation, and testing sets. Secondly, the dance movement recognition model based on 3D CNN is designed. The parameters, such as the number of layers, the size of convolutional kernels, pooling methods, and the structure of fully connected layers, are determined based on the characteristics and requirements of the MSRAction3D dataset. Next, the model is trained using the training set. By passing the input data into the model, the predicted results are calculated and compared with the true labels using a loss function to measure the difference between the predicted and true labels. The backpropagation algorithm is used to optimize the model's weights and parameters, gradually improving the prediction accuracy of the training set. Finally, the model's performance is evaluated using the validation set. The validation set's data is input into the trained model, and the predicted results are compared with the true labels to assess the model's generalization ability on unseen data. Based on the evaluation results from the validation set, adjustments and optimizations can be made to the model, such as adjusting hyperparameters and network structure. The experimental results are shown in Figure 3.

This study conducted experimental verification by applying the proposed method to the MSRAction3D dataset. The experimental results demonstrated that the dance movement recognition model based on 3D convolutional neural networks achieved satisfactory performance on this dataset. Analyzing recognition accuracy for 20 different action categories, we found that the method exhibited high performance in action recognition tasks. The experimental results showed a minimum recognition accuracy of 0.81, a maximum recognition accuracy of 0.98, and an average recognition accuracy of 0.91. This indicates that the method effectively recognizes different types of dance movements. Based on the analysis of experimental results, we can conclude that the dance movement recognition model based on 3D convolutional neural networks demonstrated good performance on the MSRAction3D dataset, providing strong support for its application in dance teaching. The high recognition accuracy means that the model can accurately distinguish and recognize different dance movements, providing reliable

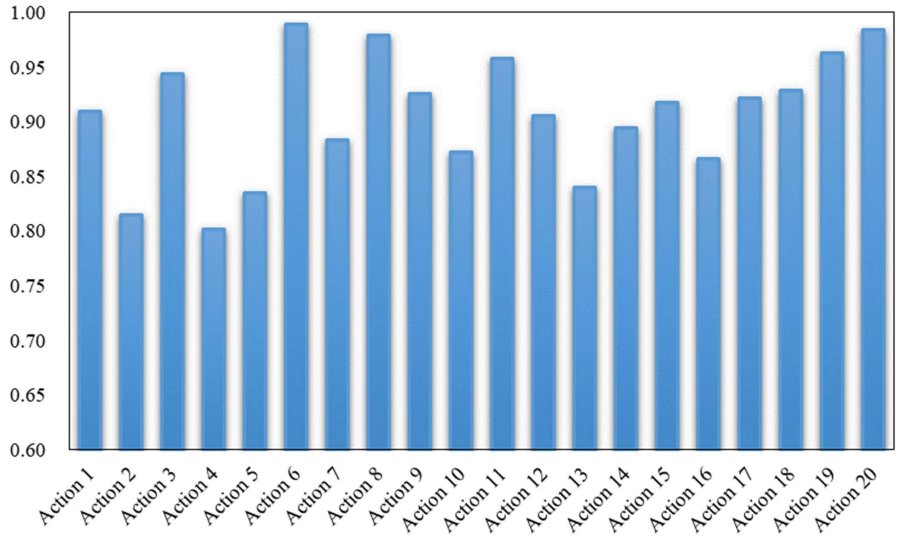

Figure 3. Recognition rate of 20 typical actions

5. Conclusions

This paper explored the application of the dance movement recognition model based on 3D convolutional neural networks in dance teaching. A survey of the current status of intelligent technology in dance teaching both domestically and internationally found that deep learning and 3D convolutional neural networks have great potential for application in dance movement recognition. Therefore, this paper aims to explore effective methods for applying 3D convolutional neural networks to dance movement recognition. The MSRAction3D dataset was selected for experimentation to verify the proposed method's effectiveness in dance teaching. The analysis of experimental results found that the dance movement recognition model based on 3D convolutional neural networks achieved satisfactory results on the MSRAction3D dataset. The dance movement recognition model based on 3D convolutional neural networks has significant potential for application in dance teaching. With this model, dance instructors can accurately analyze, assess, and guide students' dance movements, providing personalized guidance and feedback and enhancing the effectiveness and quality of dance teaching.

References

[1] Basak, H., Kundu, R., Singh, P. K., et al. (2022). A union of deep learning and swarm-based optimization for 3D human action recognition. *Scientific Reports, 12*(1), Article 5494.

[2] Pareek, P and Thakkar, A. (2021). A survey on video-based human action recognition: Recent updates, datasets, challenges, and applications. *Artificial Intelligence Review, 54*, 2259–2322.

[3] Segalin, C., Williams, J., Karigo, T., et al. (2021). The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. *Elife, 10*, Article e63720.

[4] Muhammad, K., Ullah, A., Imran, A. S., et al. (2021). Human action recognition using attention-based LSTM network with dilated CNN features. *Future Generation Computer Systems, 125*, 820–830.

[5] Herath, S., Harandi, M and Porikli, F. (2017). Going deeper into action recognition: A survey. *Image and Vision Computing, 60*, 4–21.

[6] Plizzari, C., Cannici, M and Matteucci, M. (2021). Skeleton-based action recognition via spatial and temporal transformer networks. *Computer Vision and Image Understanding, 208*, Article 103219.

[7] Li, X., Hou, Y., Wang, P., et al. (2021). Trear: Transformer-based RGB-D egocentric action recognition. *IEEE Transactions on Cognitive and Developmental Systems, 14*(1), 246–252.

[8] Abdulazeem, Y., Balaha, H. M., Bahgat, W. M., et al. (2021). Human action recognition based on transfer learning approach. *IEEE Access, 9*, 82058–82069.

[9] Huifang, Q., Jianping, Y and Yunhu, F. U. (2021). Review of human action recognition based on deep learning. *Journal of Frontiers of Computer Science & Technology, 15*(3), 438.

[10] Alakwaa, W., Nassef, M and Badr, A. (2017). Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *International Journal of Advanced Computer Science and Applications, 8*(8).