Available online at www.HighTechJournal.org

ISSN: 2723-9535

HighTech and Innovation Journal

Vol. 4, No. 2, June, 2023

A Study of Dance Movement Capture and Posture Recognition Method Based on Vision Sensors

Qun Wang [1], Gang Tong [1*], Sichao Zhou [1]

[1] *Department of Sports and Arts, Hebei Sport University, Shijiazhuang, Hebei 050041, China.*

Received 24 February 2023; Revised 09 May 2023; Accepted 17 May 2023; Published 01 June 2023

Abstract

With the development of technology, posture recognition methods have been applied in more and more fields. However, there is relatively little research on posture recognition in dance. Therefore, this paper studied the capture and posture recognition of dance movements to understand the usability of the proposed method in dance posture recognition. Firstly, the Kinect V2 visual sensor was used to capture dance movements and obtain human skeletal joint data. Then, a three-dimensional convolutional neural network (3D CNN) model was designed by fusing joint coordinate features with joint velocity features as general features for recognizing different dance postures. Through experiments on NTU60 and self-built dance datasets, it was found that the 3D CNN performed best with a dropout rate of 0.4, a ReLU activation function, and fusion features. Compared to other posture recognition methods, the recognition rates of the 3D CNN on CS and CV in NTU60 were 88.8% and 95.3%, respectively, while the average recognition rate on the dance dataset reached 98.72%, which was higher than others. The experimental results demonstrate the effectiveness of our proposed method for dance posture recognition, providing a new approach for posture recognition research and making contributions to the inheritance of folk dances.

Keywords: Vision Sensor; Dance; Movement Capture; Gesture Recognition; Kinect V2.

1. Introduction

With the continuous updating and progress of multimedia technology, music and video files have become increasingly important forms for carrying and disseminating information in addition to text and images. This has led to an increasing amount of data on the network, making the processing of these files more complex. Video files have a wide range of applications in security monitoring, intelligent navigation, etc. [1]. In order to effectively utilize these video files, computer vision technology is gradually developing [2]. The term “computer vision” refers to the use of computers for analyzing and recognizing human movements and postures in video files, supporting research in areas such as motion analysis and human-computer interaction. With continuous advancements in sensor technology, various sensors can be used to capture human movements [3], acquire a large amount of data on human behavior, and analyze this data to achieve recognition of different postures [4].

Currently, numerous methods have been applied for analyzing human movements [5]. Balmik et al. [6] designed a 7-layer 1D convolutional neural network to achieve the recognition of human movements. The experimental results

* Corresponding author: 2010009@hepec.edu.cn

http://dx.doi.org/10.28991/HIJ-2023-04-02-03

showed that it achieved an accuracy rate of 95%, outperforming the hidden Markov model. Pribadi et al. [7] used wearable sensors to collect hand movement data from welders, extracted features such as spectral peaks and spectral power, and employed a multilayer perceptron for recognition. The results showed that this algorithm accurately recognized welders' hand movements. Rotoni et al. [8] used MPU-6050 triaxial accelerometers to collect limb data from infants and identified irregular limb movements associated with cerebral palsy. Asmaul Husna et al. [9] converted students' gymnastics learning videos into digital images, employed Histogram of Oriented Gradients (HOG) to recognize students in frames, and used Principal Component Analysis (PCA) to distinguish various gymnastics movements. The experimental results showed that this method achieved a 96.09% accuracy rate.

Li et al. [10] investigated the effectiveness of M-mode ultrasound in identifying wrist and finger movements in the human body. Thirteen movements were tested on eight subjects, and a support vector machine (SVM) and a backpropagation neural network (BPNN) were used to classify the movements. The results showed that the average classification accuracy of the SVM classifier and the BPNN classifier using M-mode ultrasound reached 98.83 ± 1.03% and 98.70 ± 0.99%, respectively. In another study, Liu et al. [11] discussed the design and implementation of a human motion capture system based on microelectro mechanical systems and Zigbee network. Testing revealed that this method could accurately recognize human motion states, with an efficiency improvement of 10% compared to existing research, along with an increase in accuracy by nearly 15%.

Kurban et al. [12] utilized motion sequence information from masked depth video streams obtained from RGB-D data for action recognition and tested the proposed method on BodyLogin, NATOPS, and SBU Kinect datasets, finding that it provided higher performance and better motion representation. Ding et al. [13] introduced a temporal segment graph convolutional network that divides the entire skeleton sequence into different sub-sequences for recognition and demonstrated the effectiveness of this approach through experiments on the NTU-RGB+D and Kinetics datasets. Li et al. [14] developed a detection model using the You Only Look Once v5 algorithm and integrated it with the Openpose algorithm to recognize safe and unsafe human behaviors in videos. They achieved an accuracy of 0.9467 in experimental settings. Under the influence of rapid social development and change, dance, especially folk dance, faces increasing challenges in preservation and inheritance. The teaching of folk dance has important practical value for better recording and protecting folk dances and promoting their dissemination and inheritance.

However, in current folk dance teaching, students often learn by watching videos or receiving one-on-one guidance from teachers, resulting in low efficiency. If posture recognition technology can be applied to folk dance training, it would have certain significance for the teaching and training of folk dances. However, in the current field of posture recognition, although there is some involvement with sports, there is little research on folk dance. Moreover, motion capture methods based on wearable sensors can also affect the execution of dance movements. Therefore, posture recognition for folk dance poses a high level of difficulty. This article proposes a method for capturing folk dance movements based on Kinect visual sensors. Features were extracted from human skeletal joint data, and a three-dimensional convolutional neural network (3D CNN) model was used to classify different dance postures. Experimental analysis proved the effectiveness of this method for recognizing dance postures, providing a reference for the application of Kinect visual sensors and posture recognition technology in teaching folk dances as well as contributing to the protection and inheritance of folk dances.

2. Dance Movement Capture Based on the Kinect Vision Sensor

Visual sensors can capture human movements through cameras [15] and recognize posture. In this paper, the Kinect visual sensor was used for dance motion capture. Unlike other motion sensors, the Kinect visual sensor does not require attachment to the body for human-computer interaction. The Kinect V2 sensor [16] was used in this study. Compared with V1, V2 can capture information about 25 three-dimensional skeletal points, accommodate up to six people, and offer enhanced interactivity. The Kinect V2 SDK is a development package used in conjunction with Visual Studio 2012 or later compilers. It can be connected to a computer via USB to access data sources such as color, depth, and skeletal data from the Kinect device, which is convenient for developers conducting research. Therefore, in the Windows 10 environment, combined with Visual Studio 2017, the SDK development package, and the Kinect V2 visual sensor, this paper used the C# programming language to study dance movement capture and posture recognition methods by accessing the Kinect V2 data.

Kinect V2 is capable of providing 3D coordinate information for 25 skeletal joints at a rate of 30 frames per second, as shown in Figure 1 and Table 1.

Table 1. Names of the 25 skeletal joint points

0	Spine base	8	Shoulder right	16	Hip right
1	Spine mid	9	Elbow right	17	Knee right
2	Neck	10	Wrist right	18	Ankle right
3	Head	11	Hand right	19	Foot right
4	Shoulder left	12	Hip left	20	Spine shoulder
5	Elbow left	13	Knee left	21	Hand tip left
6	Wrist lest	14	Ankle left	22	Thumb left
7	hand left	15	Foot left	23	Hand tip right
				24	Thumb right

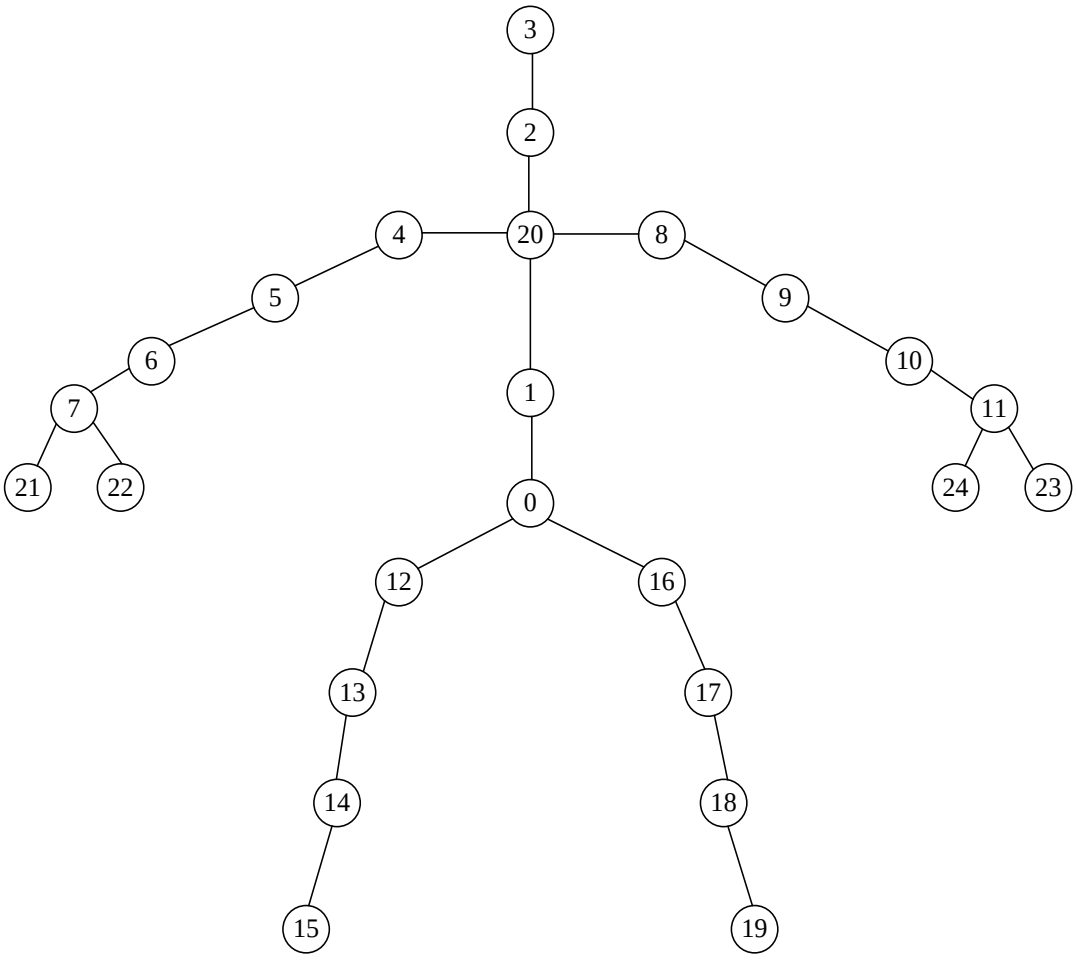

Figure 1. 25 skeletal joint points

Kinect V2 captures the body target through camera-based skeleton joint point tracking, converts it into a depth image, and then utilizes the skeleton tracking system. The SDK provides body tracking which is used to eliminate the background outside of the human body in order to obtain a grayscale image. A random forest algorithm is employed to identify human body parts and connect joint points for positioning skeletal points. Finally, skeletal points are connected to create a model of the human body's skeletal joint points.

This paper examines the recognition of various dance gestures in the folk dance "Drolma" using the Kinect V2 visual sensor to capture movements. The study collected movement data from 100 dancers who were skilled in "Drolma", with each dancer performing the dance three times and three postures captured for subsequent posture recognition. Figures 2 to 4 show skeletal point data from three postures collected from a specific dancer.

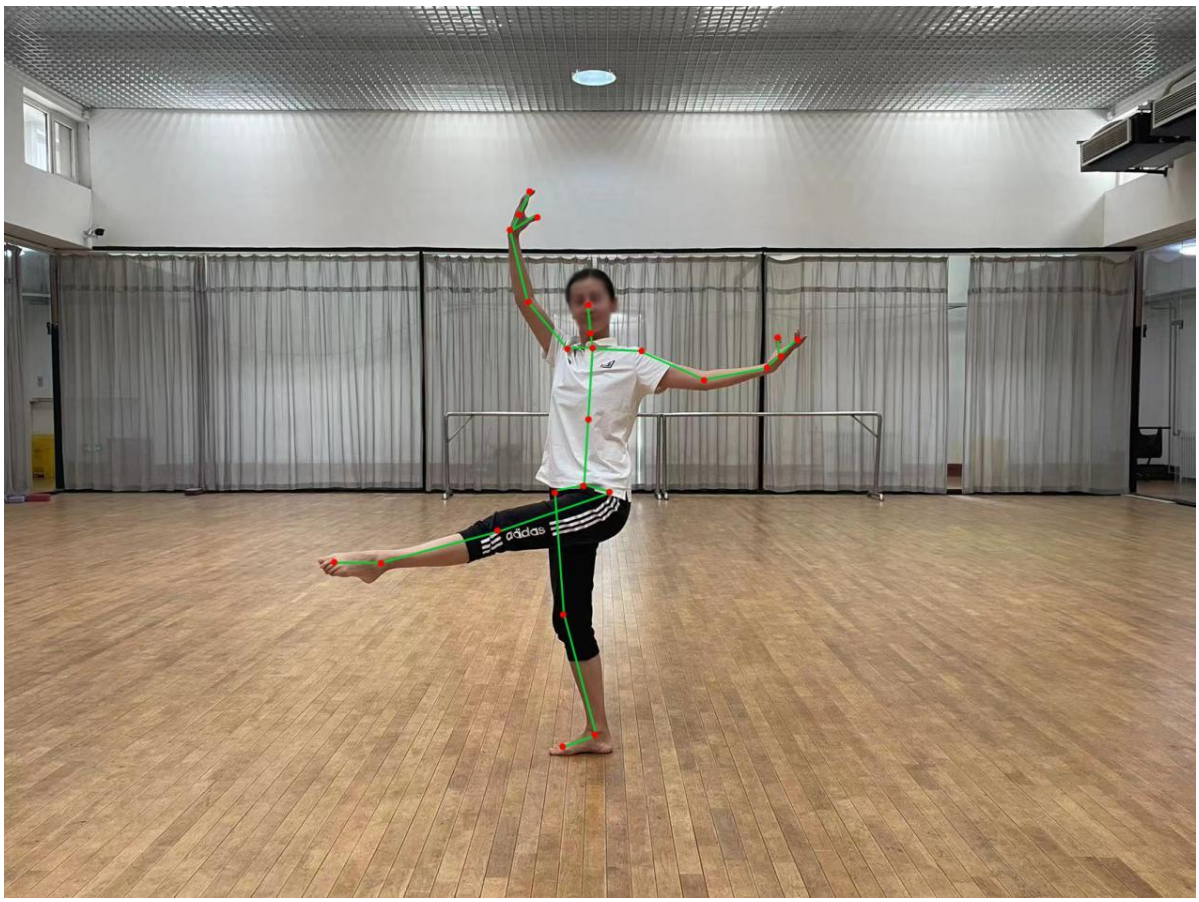

Figure 2. The first gesture of "Drolma"

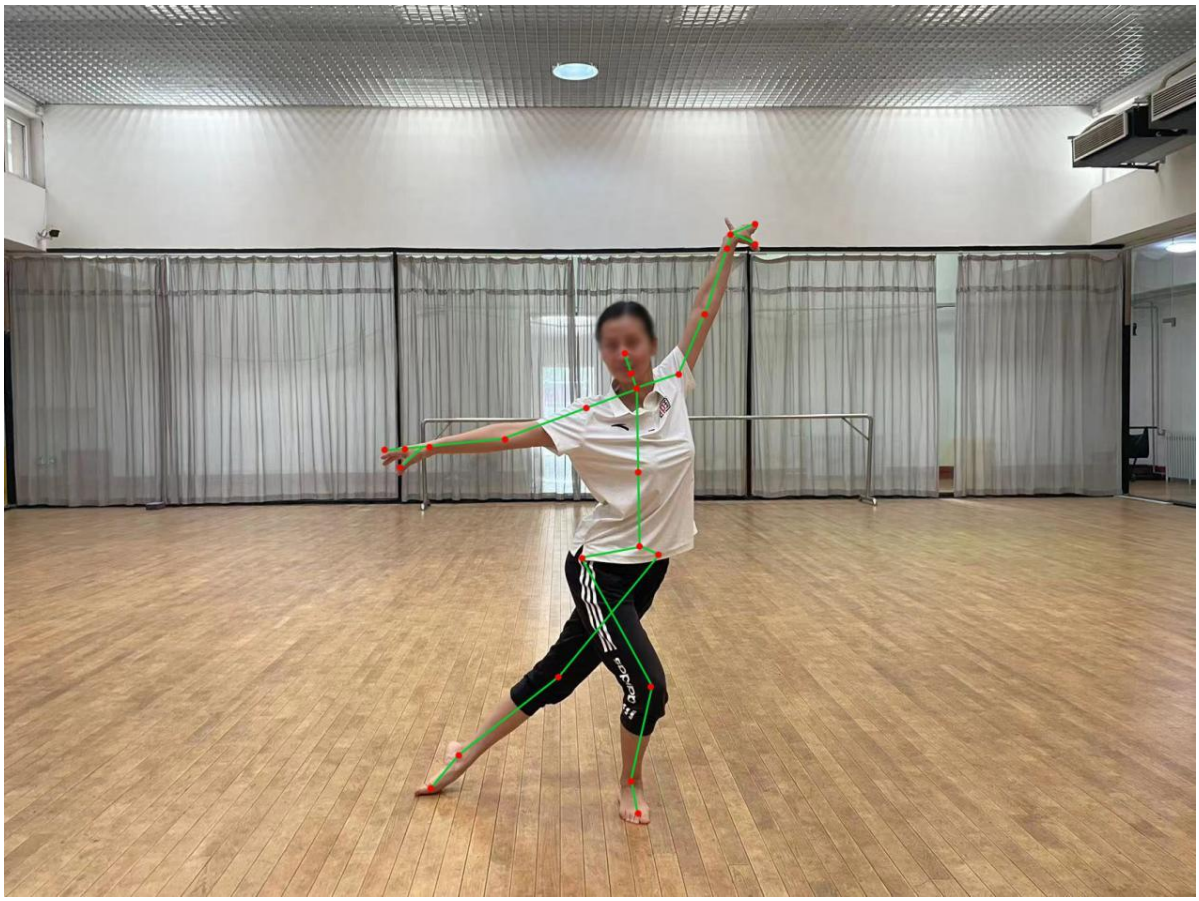

Figure 3. The second gesture of "Drolma"

Figure 4. The third gesture of "Drolma"

3. The Convolutional Neural Network-Based Posture Recognition Method

The human skeletal data is represented by joint coordinates. To improve the effect of dance posture recognition, this article incorporates joint velocity features in addition to joint coordinates as input for the subsequent pose recognition algorithm. It is assumed that in the skeleton sequence data acquired by Kinect V2, all joint points of the skeleton in each frame are represented as: $V = \{X_t^c | t = 1,2,\cdots,T; c = 1,2,\cdots,V\}$, where X_t^c refers to the c-th joint point of the t-th frame, the 3D position coordinates of X_t^i can be written as: $p_{t,i} = \left(x_{t,i}, y_{t,i}, z_{t,i}\right)^T$. Then, its joint velocity can be written as:

$$v_{t,i} = p_{t,i} - p_{t-1,i} = \left(x_{t,i} - x_{t-1,i}, y_{t,i} - y_{t-1,i}, z_{t,i} - z_{t-1,i}\right)^T \quad (1)$$

where $p_{t-1,i}$ is the coordinates of joint point X_{t-1}^i of the t-th frame.

The position and velocity features of the skeletal joints were mapped into a high-dimensional space by two fully connected (FC) layers, yielding $\widehat{p_{t,i}}$ and $\widehat{v_{t,i}}$. Taking the joint position as an example, the operation is as follows:

$$\widehat{p_{t,i}} = \sigma\left[W_2\left(\sigma\left(W_1 p_{t,i} + b_1\right)\right) + b_2\right] \quad (2)$$

where σ is the RELU activation function, W_1 and W_2 are the weights of the two FC layers, and b_1 and b_2 are biases.

After fusing these two features, the general feature of the input to the posture recognition algorithm is obtained:

$$z_{t,i} = \widehat{p_{t,i}} + \widehat{v_{t,i}} \quad (3)$$

A CNN was used for posture recognition. CNN is a kind of network featured by local connectivity and weight sharing [17], which has various applications in image and video processing [18]. In video processing, temporal features are also crucial in addition to spatial features. To fully utilize the spatio-temporal feature information in the Kinect V2 skeletal data, this paper employed a 3D CNN. In a 3D-CNN, the calculation formula of V_{ij}^{xyz} of coordinates (x, y, z) in the j-th feature map of the i-th layer is:

$$V_{ij}^{xyz} = f\left(b_{ij} + \sum_r \sum_{l=0}^{l_i-1} \sum_{m=0}^{m_i-1} \sum_{n=0}^{n_i-1} \omega_{ijr}^{lmn} v_{(i-1)r}^{(x+l)(y+m)(z+n)}\right) \quad (4)$$

where ω_{iir}^{lmn} is the value of the convolution kernel connecting the m-th feature map in the previous layer, n_i is the time dimension of the convolution kernel, and f is the activation function. The following activation functions are often employed.

(1) Sigmoid function: $f(x) = \frac{1}{1+e^{-x}}$

(2) $tanh$ function: $\tanh(x) = 2\sigma(2x) - 1$

(3) RELU function: $f(x) = \max(0, x)$

The process of 3D pooling is similar to 2D pooling, except that a time dimension, i.e., the number of image frames. The formula for maximum pooling is written as:

$$V_{x,y,z} = \max_{0\le i\le s_1, 0\le j\le s_2, 0\le k\le s_3} \left(O_{x\times s+i, y\times t+j, z\times r+k}\right) \quad (5)$$

where $V_{x,y,z}$ is the pooling output, s, t, and r are the sampling step length in three directions, and O is the 3D input vector.

Finally, the classification of the model was implemented in the Softmax layer, ensuring that the probability of the correct category converges to 1 and that the sum of all category probabilities is 1. The overall flow of the designed dance posture recognition method is illustrated in Figure 5.

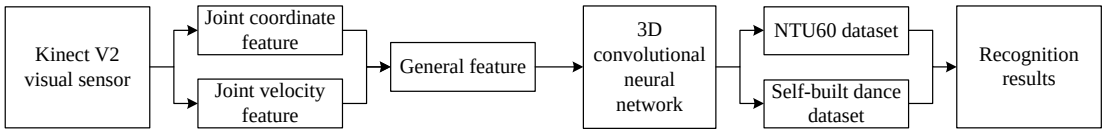

Figure 5. The flow of dance posture recognition

The model parameters of the 3D CNN for dance posture recognition are listed in Table 2.

Table 2. 3D CNN model parameters

Network layer	Size
Convolutional layer	3×3×3
Convolutional layer	3×3×3
Maximum pooling	2×2×2
Convolutional layer	3×3×3
Convolutional layer	3×3×3
Maximum pooling	2×2×2
FC layer	1×2048
FC layer	1×512

4. Results and Analysis

The model was implemented on the Python 3.6 platform using the Keras deep learning framework. The 3D CNN was trained using the Adam optimizer and the cross-entropy loss function, with an initial learning rate of 0.001 and a total of 120 iterations. Experiments were conducted on two datasets to evaluate the effectiveness of the 3D CNN posture recognition method.

(1) NTU60 RGB+D dataset [19]: the data are collected from Kinect V2 and used to evaluate the human skeletal behavior recognition model. It consists of 60 categories of movements performed by 40 actors. The dataset evaluation includes two types: ① cross-subject (CS), where the training and test sets are divided according to actor ID; ② cross-view (CV), where the training and test sets are divided based on camera viewpoint.

(2) Self-built folk dance dataset: the data are collected from Kinect V2, as described in the section on dance movement capture based on the Kinect vision sensor. It consists of three postures performed by 100 dancers. As the dance was repeated three times, there were a total of 900 samples. The training and test sets were randomly divided according to 2:1.

The algorithm was evaluated in terms of the recognition rate:

$$acc = \frac{N_c}{N} \times 100\% \quad (6)$$

where N is the total number of postures in the dataset and N_c is the number of postures correctly identified by the algorithm.

To prevent overfitting, a dropout layer was added after the first CNN layer with varying dropout rates of 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 to compare recognition rates on the NTU60 RGB+D. The results are demonstrated in Table 3.

Table 3. Effect of the dropout rate on recognition rate

	0.2	0.3	0.4	0.5	0.6	0.7
CS/%	86.77	87.91	88.73	88.55	86.71	81.26
CV/%	88.12	91.24	95.15	94.71	94.32	93.48

From Table 3, it can be observed that the recognition rates of the algorithm varied with changes in dropout rate. Comparing the results, when the dropout rate was set to 0.4, the algorithm achieved the highest recognition rates on both NTU60 RGB+D datasets, reaching 88.73% and 95.15% respectively. This indicated that the algorithm performed optimally at this dropout rate. Therefore, a dropout rate of 0.4 will be used in subsequent experiments.

To determine the most suitable activation function, the recognition rates under different activation functions were compared, and the results are presented in Figure 6.

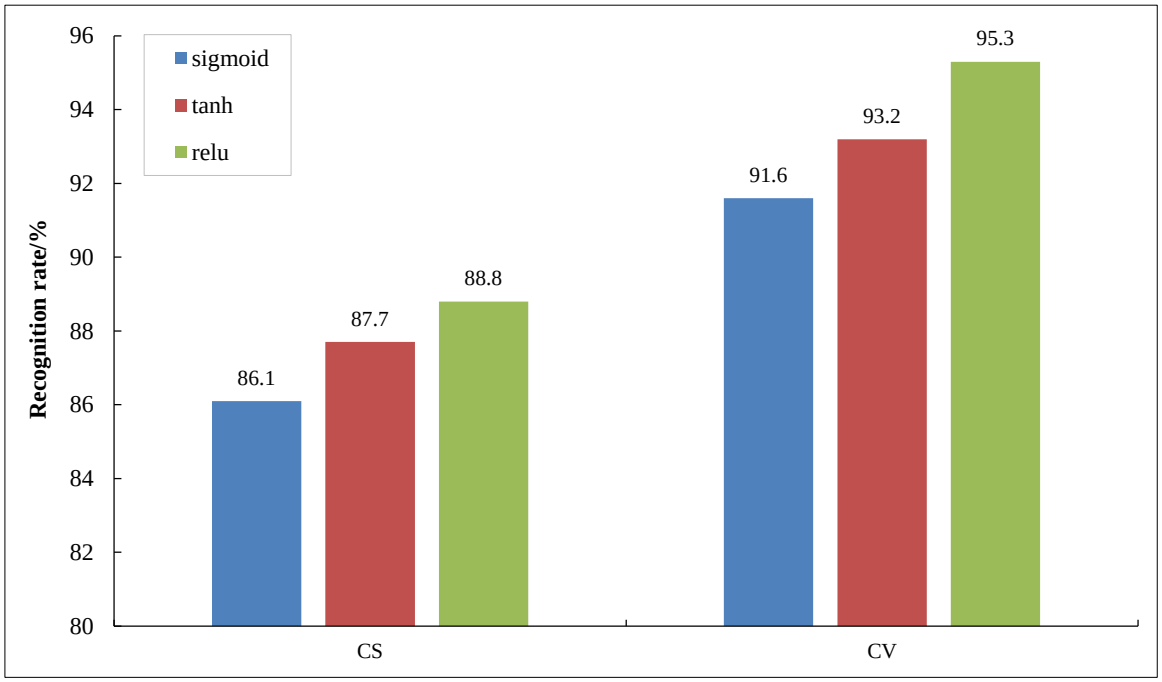

Figure 6. The effect of the choice of activation function on the recognition rate

From Figure 6, the recognition rate of the 3D CNN for CS was 86.1% with sigmoid activation function and 87.7% with $tanh$, indicating an improvement of 1.6%. However, when using ReLU as the activation function, the recognition rate increased to 88.8%, which was 2.7% higher than both sigmoid (2.7%) and $tanh$ (1.1%). Then, for the CV, the recognition rates of the three activation functions were as follows: sigmoid < $tanh$ < ReLU. The recognition rate of ReLU was 95.3%, which was 3.7% higher than that of sigmoid and 2.1% higher than that of $tanh$. Among these activation functions, sigmoid has a non-zero mean output, making it prone to the problem of gradient dispersion, while $tanh$ has a zero mean output, resulting in a higher recognition rate. As a piecewise function, Relu does not suffer from gradient dispersion and converges quickly. Therefore, it was used as the activation function in the following experiments.

Table 4 shows the results on the NTU60 dataset, taking into account the impact of input features on the 3D CNN.

Table 4. The impact of input features on the recognition rate

	CS/%	CV/%
Using the joint coordinate feature only	77.9	84.2
Using the joint velocity feature only	78.2	85.3
General feature	88.8	95.3

From Table 4, it can be observed that when using only a single feature as the input for the 3D CNN, the recognition rate of the algorithm was relatively low. However, when using the general feature as the input feature, there was a significant improvement in the recognition rate of the algorithm, which was about 10% higher than that of a single feature. This proved the reliability of the proposed feature fusion method and its ability to extract more features from human skeletal joint data to improve recognition accuracy.

The trained 3D CNN successfully recognized different folk dance postures, and the recognition rates for each posture are shown in Figure 7.

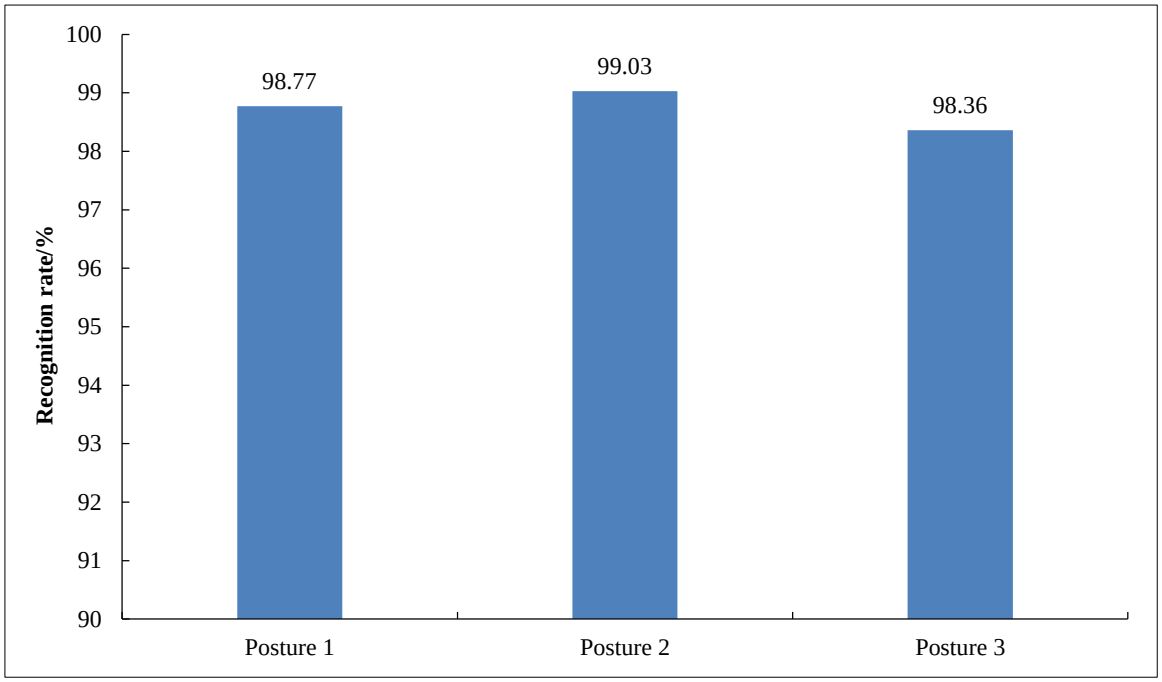

Figure 7. Recognition rate of 3D CNN for different postures in the dance set

From Figure 7, it was found that the recognition rate of the self-built folk dance dataset using the 3D CNN was high, exceeding 95%. Among these postures, posture 2 achieved the highest recognition rate of 99.03%, while posture 3 had the lowest recognition rate of 98.36%. The average recognition rate for these three postures was calculated at 98.72%. In comparison to the NTU60 dataset, the dance set exhibited a higher recognition rate with the use of the 3D CNN model. This could be attributed to two factors: firstly, there were fewer categories of postures in the dance set; secondly, their complexity level was relatively lower.

To further demonstrate the performance of the 3D CNN for posture recognition, it was compared with other methods:

(1) Spatial-temporal graph convolutional network (ST-GCN) [20],

(2) Attention enhanced graph convolutional LSTM network (AGC-LSTM) [21],

(3) Two-stream adaptive graph convolutional network (2s-AGCN) [22],

(4) Semantics-guided neural network (SGN) [23].

The comparison results of these methods for the NTU60 RGB+D are demonstrated in Table 5.

Table 5. The recognition rate of different methods for the NTU60 RGB+D

	CS/%	CV/%
ST-GCN	81.5	88.3
AGC-LSTM	87.5	93.5
SGN	88.7	94.3
2s-AGCN	88.5	95.1
3D CNN	88.8	95.3

From Table 5, it can be observed that, compared to the current posture recognition models, the 3D CNN achieved higher recognition rates on both the CS and CV datasets. Firstly, on the CS dataset, the 3D CNN achieved a recognition rate of 88.8%, which was an improvement of 7.3% over the ST-GCN, 1.3% over the ACG-LSTM, 0.1% over the SGN, and 0.3% over the 2s-AGCN. In the CV dataset, the 3D CNN achieved a recognition rate of 95.3%, which was 7% higher than the ST-GCN, 1.8% higher than the ACG-LSTM, 1% higher than the SGN, and 0.2% higher than the 2s-AGCN. Overall, the 3D CNN obtained the best results on the NTU60 RGB+D dataset, demonstrating its superiority in posture recognition.

The comparison results of these methods for the dance set are demonstrated in Table 6.

Table 6. The recognition rate of different methods for the danceset

	Posture 1/%	Posture 2/%	Posture 3/%	Average value/%
ST-GCN	96.89	96.37	96.32	96.53
ACG-LSTM	97.39	96.08	96.56	96.68
SGN	97.16	97.17	96.88	97.07
2s-AGCN	97.21	97.33	97.58	97.37
3D CNN	98.77	99.03	98.36	98.72

From Table 6, it can be observed that the 3D CNN achieved a high recognition rate for different postures and outperformed the other methods. In terms of average recognition rate across three postures, the 3D CNN achieved 98.72%, which was a 2.19% improvement over the ST-GCN, a 2.04% improvement over the ACG-LSTM, a 1.65% improvement over the SGN, and a 1.35% improvement over the 2s-AGCN. These results indicated that the use of a 3D CNN was more suitable for recognizing folk dance postures compared to these posture recognition models.

5. Conclusions

This article mainly focuses on the posture recognition of folk dances. The Kinect V2 visual sensor was used to capture the movements of folk dance and obtain human skeletal joint data. The fused joint coordinates and velocities were used as a general feature, and a 3D CNN was designed to achieve recognition of different postures. It was found through the comparative experiment on the NTU60 RGB+D dataset that:

- When the dropout rate of the 3D-CNN was 0.4, ReLU was used as the activation function, and the fused feature was used as the input, the recognition performance of the algorithm was the best;
- The 3D CNN achieved a recognition rate of over 95% in recognizing the three postures in the dance set;
- Compared with posture recognition models such as the ST-GCN and ACG-LSTM, the 3D CNN achieved higher recognition rates for the NTU60-CS and NTU60-CV, which were 88.8% and 95.3%, respectively;
- Compared with posture recognition models such as the ST-GCN and ACG-LSTM, the 3D CNN achieved higher recognition rates for the three postures in the dance dataset, with an average recognition rate of 98.72%.

The experimental results have demonstrated the reliability of the method proposed in this paper for posture recognition, which can effectively recognize different folk dance postures with high accuracy and can be promoted and applied in practice. However, there are still some shortcomings in this method, such as whether the proposed algorithm can be further optimized, the small scale of the dance set used in experiments, whether the algorithm can maintain its accuracy in more complex dance posture recognition tasks, and whether its real-time performance meets practical requirements. These issues need to be considered in future work.

6. Declarations

6.1. Author Contributions

Conceptualization, Q.W. and G.T.; methodology, Q.W.; validation, Q.W. and G.T.; resources, Q.W. and S.Z.; data curation, G.T. and S.Z.; writing—original draft preparation, Q.W. and G.T.; writing—review and editing, Q.W. and G.T.; project administration, S.Z.; funding acquisition, Q.W., G.T. and S.Z. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

Data sharing is not applicable to this article.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. References

[1] Senthil Murugan, A., Suganya Devi, K., Sivaranjani, A., & Srinivasan, P. (2018). A study on various methods used for video summarization and moving object detection for video surveillance applications. Multimedia Tools and Applications, 77(18), 23273–23290. doi:10.1007/s11042-018-5671-8.

[2] Fuggle, N. R., Lu, S., Breasail, M. Ó., Westbury, L. D., Ward, K. A., Dennison, E., Mahmoodi, S., Niranjan, M., & Cooper, C. (2022). OA22 Machine learning and computer vision of bone microarchitecture can improve the fracture risk prediction provided by DXA and clinical risk factors. Rheumatology, 61(Supplement_1). doi:10.1093/rheumatology/keac132.022.

[3] Mahbubur Rahman, M., & Gurbuz, S. Z. (2021). Multi-Frequency RF Sensor Data Adaptation for Motion Recognition with Multi-Modal Deep Learning. 2021 IEEE Radar Conference (RadarConf21). doi:10.1109/radarconf2147009.2021.9455204.

[4] Moreau, P., Durand, D., Bosche, J., & Lefranc, M. (2020). A motion recognition algorithm using polytopic modeling. 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT). doi:10.1109/codit49905.2020.9263883.

[5] Lee, K.-T., Yoon, H., & Lee, Y.-S. (2018). Implementation of smartwatch user interface using machine learning based motion recognition. 2018 International Conference on Information Networking (ICOIN). doi:10.1109/icoin.2018.8343229.

[6] Balmik, A., Paikaray, A., Jha, M., & Nandy, A. (2022). Motion recognition using deep convolutional neural network for Kinect-based NAO teleoperation. Robotica, 40(9), 3233–3253. doi:10.1017/S0263574722000169.

[7] Pribadi, T. W., & Shinoda, T. (2020). Hand Motion Recognition of Shipyard Welder Using 9-DOF Inertial Measurement Unit and Multi-Layer Perceptron Approach. IOP Conference Series: Earth and Environmental Science, 557(1). doi:10.1088/1755-1315/557/1/012009.

[8] Rotoni, G. M., Unabia, S. A., & Villaverde, J. F. (2020). Wireless Accelerometer-based Motion Recognition Sensors for Limb Movement Analysis in Babies. Proceedings of the 2020 10th International Conference on Biomedical Engineering and Technology. doi:10.1145/3397391.3397399.

[9] Asmaul Husna, R., Achmad, A., Ilham, A. A., Zainuddin, Z., & Jaya, A. K. (2020). Early Childhood Gymnastic Motion Recognition System Using Image Processing Technology. 2020 27th International Conference on Telecommunications (ICT). doi:10.1109/ict49546.2020.9239482.

[10] Li, J., Zhu, K., & Pan, L. (2022). Wrist and finger motion recognition via M-mode ultrasound signal: A feasibility study. Biomedical Signal Processing and Control, 71, 103112. doi:10.1016/j.bspc.2021.103112.

[11] Liu, Q. (2022). Human motion state recognition based on MEMS sensors and Zigbee network. Computer Communications, 181, 164–172. doi:10.1016/j.comcom.2021.10.018.

[12] Kurban, O. C., Calik, N., & Yildirim, T. (2022). Human and action recognition using adaptive energy images. Pattern Recognition, 127, 108621. doi:10.1016/j.patcog.2022.108621.

[13] Ding, C., Wen, S., Ding, W., Liu, K., & Belyaev, E. (2022). Temporal segment graph convolutional networks for skeleton-based action recognition. Engineering Applications of Artificial Intelligence, 110, 104675. doi:10.1016/j.engappai.2022.104675.

[14] Li, J., Zhao, X., Zhou, G., & Zhang, M. (2022). Standardized use inspection of workers' personal protective equipment based on deep learning. Safety Science, 150. doi:10.1016/j.ssci.2022.105689.

[15] Lee, T. J., Kim, C. H., & Cho, D. I. D. (2019). A Monocular Vision Sensor-Based Efficient SLAM Method for Indoor Service Robots. IEEE Transactions on Industrial Electronics, 66(1), 318–328. doi:10.1109/TIE.2018.2826471.

[16] Ayed, I., Jaume-I-capó, A., Martínez-Bueso, P., Mir, A., & Moyà-Alcover, G. (2021). Balance measurement using microsoft kinect v2: Towards remote evaluation of patient with the functional reach test. Applied Sciences (Switzerland), 11(13), 6073. doi:10.3390/app11136073.

[17] Akhtar, M. B. (2022). The use of a convolutional neural network in detecting soldering faults from a printed circuit board assembly. HighTech and Innovation Journal, 3(1), 1-14. doi:10.28991/HIJ-2022-03-01-01.

[18] Kurdthongmee, W., Kurdthongmee, P., Suwannarat, K., & Kiplagat, J. K. (2022). A YOLO Detector Providing Fast and Accurate Pupil Center Estimation using Regions Surrounding a Pupil. Emerging Science Journal, 6(5), 985-997. doi:10.28991/ESJ-2022-06-05-05.

[19] Basak, H., Kundu, R., Singh, P. K., Ijaz, M. F., Woźniak, M., & Sarkar, R. (2022). A union of deep learning and swarm-based optimization for 3D human action recognition. Scientific Reports, 12(1). doi:10.1038/s41598-022-09293-8.

[20] Yan, S., Xiong, Y., & Lin, D. (2018). Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). doi:10.1609/aaai.v32i1.12328.

[21] Si, C., Chen, W., Wang, W., Wang, L., & Tan, T. (2019). An Attention Enhanced Graph Convolutional LSTM Network for Skeleton-Based Action Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2019.00132.

[22] Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019). Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2019.01230.

[23] Zhang, P., Lan, C., Zeng, W., Xing, J., Xue, J., & Zheng, N. (2020). Semantics-Guided Neural Networks for Efficient Skeleton-Based Human Action Recognition. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr42600.2020.00119.