


Research Article

Improved Deep Learning Method for Intelligent Analysis of Sports Training Posture

Xianyou Yan 

Department of Physical Education, Henan University of Animal Husbandry and Economy, Zhengzhou 450046, China

Correspondence should be addressed to Xianyou Yan; 80448@hnuah.edu.cn

Received 19 May 2022; Revised 22 June 2022; Accepted 4 July 2022; Published 9 August 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Xianyou Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motion recognition based on human bones has attracted extensive attention in recent years because of its simplicity and robustness. Considering the causality of human movement, this paper proposes an improved deep learning method for posture analysis in sports training. In order to deal with the complex situation of calculating joint torques as weights, the edge weights and convolution weights of bone maps are used as auxiliary information networks according to the causality of joint distribution. Thus, the stronger driving force of joint weights in the neural network is improved, the low importance of joint attention is reduced, and the high importance of joint attention is enhanced. Experiments on three public motion recognition datasets show that the proposed method can distinguish similar motions effectively compared with the mainstream methods. Besides, experiments on a challenging UCF (University of Central Florida) sports dataset show that the proposed method can effectively enhance the motion features and improve the accuracy of recognition.

1. Introduction

To analyze the problem of motion recognition from multidata dimensions, the existing methods use visual appearance, depth information, optical flow, and even sound for fusion and auxiliary recognition. Literature [1] proposed that low redundancy and high separable joint information representation can significantly improve the performance of motion recognition. Literature [2] uses 3D joint coordinates to analyze motion pattern recognition motions, and the motion information extraction method adopted is simple and efficient. However, the spatial relationship between joints is ignored in this method; thus the accuracy is limited. To solve this problem, relative distance and angle coding joints were adopted in literature [3] to improve accuracy, but the recognition results relying only on manual features were not satisfactory. Deep learning model uses nonlinear neural network to extract deep-level motion features to improve accuracy [4]. Based on the excellent spatial feature extraction ability of CNN (convolutional neural network), the bone sequence is encoded as a pseudoimage in literature [5], and

its depth features are extracted based on CNN to improve the recognition effect. However, the lack of time domain information of the encoded image results in limited improvement of accuracy. In view of this problem, RNN (recurrent neural network) [6], which has a good temporal modelling ability, can recognize motions with a high accuracy. However, the inherent defect of gradient dispersion in RNN makes it difficult to learn long-term historical information.

The above recognition method based on deep network processes each image frame by frame and lacks the mining of key images and parts. However, motion sequences usually have large information redundancy, which makes the relevant methods have poor real-time performance and lack of highly separable information, resulting in limited improvement of accuracy. Based on this, a model based on spatiotemporal attention mechanism was proposed in literature [7]. It uses the space-time attention mechanism to extract bone features and assigns corresponding weights to joints based on their importance to enhance the influence of key images and parts, so as to improve the accuracy of

motions [8]. However, this method only considers joint coordinates and ignores spatial topological information, so the accuracy is limited.

Deep learning-based methods are used to study posture recognition in sports training from multiple perspectives, and good recognition performance is achieved [9]. However, these methods still have some limitations, such as ignoring the correlation between bone joints in the human body structure and not considering the weight changes of joints in different movements. Joints are considered to be the ends of a rigid body, and different joints play different roles in different training postures, so the key joints should have more weight in determining the type of movement. The joints of the human body are simplified to the multirigid body model, and the torques of each joint are calculated by solving partial differential equations, and then different weights are assigned to each joint according to the torques of each joint. This method is not suitable for complex graph convolutional networks because of too many equations and too much calculation. However, the main disadvantage of the local attention model is that it only uses the local variation of the motion sequence to get the attention weight, and it is difficult to get the accurate attention weight.

Therefore, this paper proposes a human skeletal motion recognition method combining causality and spatiotemporal graph convolution network. Firstly, the causal coefficients of the joint are calculated according to the joint coordinate sequence, and the causal coefficient matrix is constructed. Then, the causal coefficient matrix is applied to the graph convolution network, which assigns different weights according to the importance of joints in the process of motion. In order to effectively learn dynamic features and improve the accuracy of recognition, it is necessary to pay attention to the joints with greater influence in the process of motion and ignore the joints with less influence.

This paper has two main contributions:

- (1) Considering the causality in human movement, a spatiotemporal graph model is constructed for bone data, and a motion recognition method combining causality and spatiotemporal graph convolution network is proposed.
- (2) Aiming at the complex situation of calculating joint torques to obtain weights, a method of calculating joint weights based on causality was proposed, and edge weights were assigned to bone graph according to the causal relationship between joints. The weights are used as auxiliary information to enhance the convolutional network to improve the weights of some joints with strong driving forces in the neural network, so that the neural network can reduce the joint attention of low importance and enhance the joint attention of high importance.

This paper consists of five main parts: the first part is the introduction, the second part is state of the art, the third part is methodology, the fourth part is result analysis and discussion, and the fifth part is the conclusion.

2. State of the Art

2.1. Bone-Based Training Motion Recognition. Traditional bone motion recognition based on manual features captures the dynamics of joint motion by manually designing different feature extraction methods. For example, literature [10] established time-domain hierarchical covariance matrix descriptors to represent the motion trajectories of joints. Literature [11] used the relative position of joints as features. In literature [12], rotation and translation between various parts of the body are used to extract features, and then traditional machine learning algorithms are used to classify features, so as to classify motions. Because deep neural networks can better learn feature representation, some research studies on bone-based motion recognition has shifted from manual feature design to deep learning-based methods.

2.2. Training Motion Recognition Method Based on Deep Learning. Deep learning-based methods are divided into two stages. In the early stage, researchers use RNN or Temporal CNN to learn the motion recognition model in an end-to-end manner. Most of these methods directly take the bone coordinate sequence as the input feature or convert the bone coordinate sequence into grayscale image and then input it into the network for classification. However, RNN and CNN cannot completely represent bone structure. According to the natural structure of the human body, the graph model is more suitable for the representation of bone data. Therefore, literature [13] first applied a graph convolution network to bone-based motion recognition and proposed an ST-GCN network model. Later, on the basis of ST-GCN, 2S-AGCN was proposed in literature [14], which improved the GCN module so that it could adaptively learn the topological structure of the graph. Besides the bone data, the bone information that had never been noticed before was also used as the second information flow to improve the recognition effect. Literature [15] used SGR component to find connectivity between joints in spatial subgroup clustering and measure correlation of joint time trajectory. Literature [16] extended the skeleton diagram structure to capture potential dependencies specific to the motion. However, the above related algorithms based on bones only consider the information of bone depth and ignore the appearance features of effective expression of training movements.

3. Methodology

As shown in Figure 1, this paper proposes a human motion recognition model that integrates causality and spatiotemporal graph convolution network.

Firstly, the model calculates the causality of the joint according to the joint coordinate sequence and constructs the causal coefficient matrix. Then, the matrix is applied to the ST-GCN graph convolution network to extract deep features from bone data, so as to effectively learn dynamic features and increase the accuracy of motion recognition.

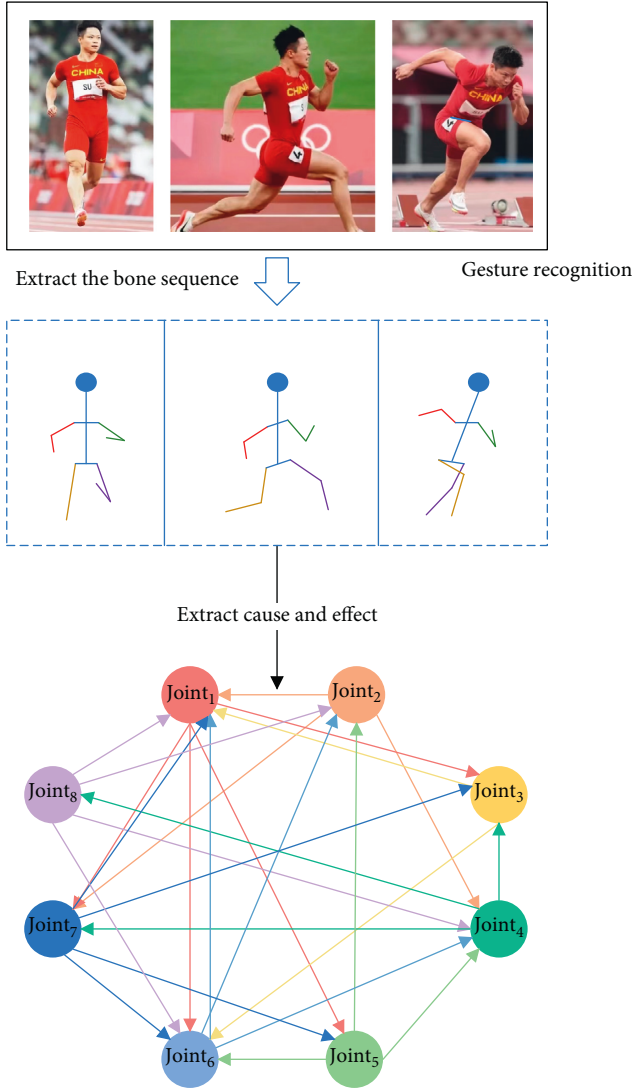


FIGURE 1: Motion recognition model combining causality and spatiotemporal graph convolution network.

3.1. Space-Time Map Model. Let the bone sequence space-time map be $A(Q, E)$ with length N frames, and point set Q contains joints at all times. Edge set E consists of two subsets, one of which is the interskeleton connection E_s of each frame (blue line in Figure 2). E_s reflects spatial attributes. The other edge subset is E_T (red line in Figure 2), which reflects temporal ownership. The top feature of $A(Q, E)$ is the joint coordinate vector $F(q_{nx})$, and the vertex q_{nx} represents the coordinates of joint x in the n th frame.

3.2. Graph Convolution. By giving $A(Q, E)$, the q_{nx} graph is convoluted.

$$f_{out}(q_{nx}) = \sum_{q_{ny} \in H(q_{nx})} \frac{1}{K_{nx}(q_{ny})} f_{in}(q_{ny}) \cdot m(l_{nx}(q_{ny})), \quad (1)$$

where $H(q_{nx}) = \{q_{ny} | d(q_{ny}, q_{nx}) \leq D\}$ is the convolution sampling region of q_{nx} . f_{in} is the input feature of q_{nx} . M is the weight function that provides the weight vector for the

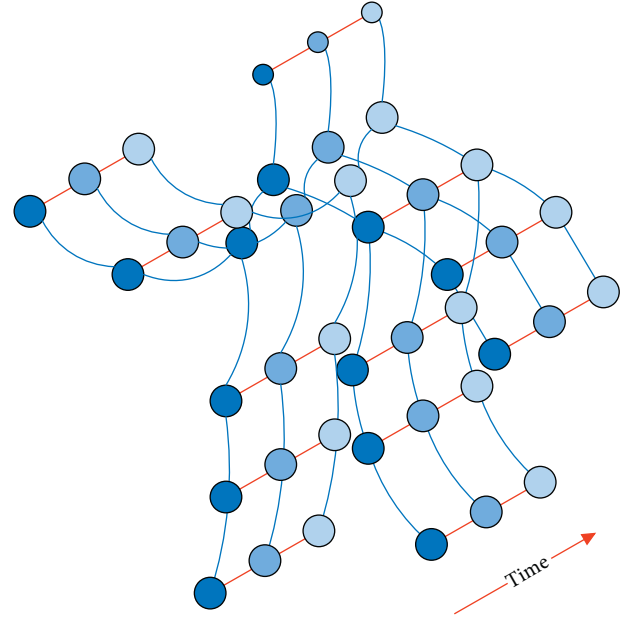


FIGURE 2: Schematic diagram of skeleton spatiotemporal graph.

input feature. Since the size of the traditional convolution sampling area is fixed, and the number of weight vectors is equal to the size of the sampling area, and the number of vertices in H is changing, the mapping vertices need to correspond to the weight vector in graph convolution. Mapping $l_{nx}: H(q_{nx}) \rightarrow \{0, \dots, Z-1\}$. The adjacent nodes are mapped to subset labels, and each neighbor node finds the corresponding weight vector according to subset labels. Usually, Z is set to 3, which will be divided into 3 subsets. The first subset (S_1) is the vertices themselves (orange nodes in Figure 3). The second subset (S_2) is a centripetal subset that contains nodes closer to the body's center of gravity (blue nodes in Figure 3). The third subset (S_3) is a centrifugal subset that contains nodes farther from the center of gravity (green nodes in Figure 3).

Graph convolution network is usually composed of multilayer spatiotemporal graph convolution layers; each layer carries out spatial graph convolution and temporal graph convolution in turn to extract high-level features of the graph, and finally carries out pooling and Softmax processing.

The feature of $A(Q, E)$ is represented by the (C, N, T) tensor, where C is the number of channels. N is the length of time, and T is the number of vertices. According to equation (1), the graph convolution formula on multidimensional tensors is shown in equations (2) and (3).

$$f_{out} = \sum_q^{Z_q} M_z(f_{in} G_z), \quad (2)$$

$$G_z = \Lambda_z^{-1/2} \overline{G}_z \Lambda_z^{-1/2} \in \mathbb{R}^{T \times T}, \quad (3)$$

where G_z is the adjacency matrix, Z_q represents the convolution kernel size; its element G_z^{xy} represents whether the vertex q_y is in the subset of q_x , and $M_z \in \mathbb{R}^{C_{opn} \times C_{xt} \times 1 \times 1}$ is the weight vector. The matrix G_z determines whether the

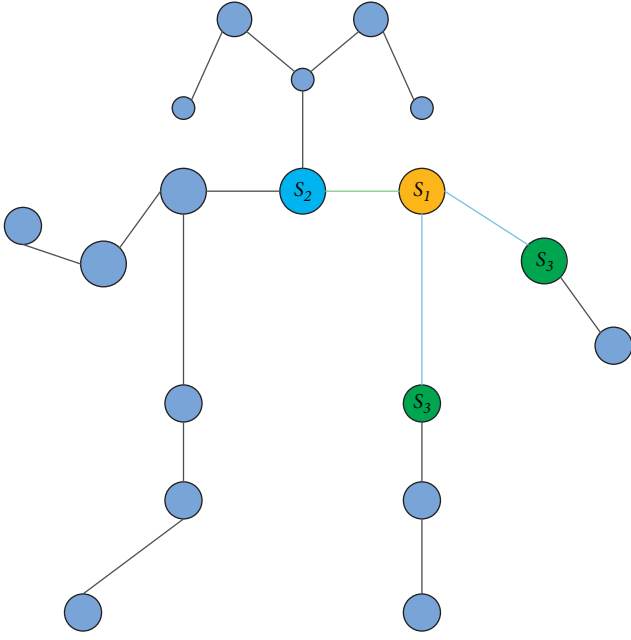


FIGURE 3: Subset mapping strategy.

vertex features are computed in conjunction with the weight vectors. Since G_z is defined in advance according to the bone structure, that is, two different motion videos, as long as the structure of human skeleton extracted is the same, G_z is the same in the convolution of the graph. Since the importance of joints in different human motions is different, it is necessary to reflect the difference of joint importance in the adjacency matrix G_z . In order to solve this problem, causality is integrated into the graph convolutional network to enhance the motion recognition effect.

3.3. Motion Recognition Integrating Causality. The multirigid body model of the human body is modelled by the Lagrange method, and joint torques are calculated by partial differential equations. The rigid body model has linkage; one rigid body will drive other rigid bodies to move through the rotating shaft. This linkage relationship obviously has causality. When a joint with a large torque exerts force, it will drive other joints more, and the exerting joint is the “cause” of the driven joint. Because the dynamic system of the human body is very complex and there are many equations after modelling, it is too complicated to calculate the interaction degree between joints through equations. Moreover, the multirigid body model can only calculate the joint torques at a single moment, and the calculation of joint torques in continuous time is more complicated. Therefore, it is difficult to analyze the strength and causality of joint interaction through the multirigid body model.

3.3.1. Calculation of Joint Causality. Convergent cross mapping (CCM) is a method to calculate the special correlation of time series in complex systems. This method is used to measure the causal relationship between the estimated value and the similarity test variable of I . It can

overcome the difficulties of complex multirigid body model modelling and high computational complexity. And it can quickly test the joint causal relationship, which has been widely used. This paper allocates edge weights based on CCM, and the steps are as follows:

(1) Build shadow flow

The e -dimensional delay vector is composed of I_n historical points of the point, which describes the change of the joint over a period of time and can be expressed as follows:

$$\tilde{I}_n = (I_n, I_{n-\tau}, I_{n-2\tau}, \dots, I_{n-(E-1)\tau}). \quad (4)$$

The shadow flow \tilde{I} of joint coordinate sequence I is the set of delay vectors I_n of each point I , τ is the step.

(2) Find the nearest point and create the weight

For \tilde{I}_n , find other $E+1$ time points that are most similar to the joint position changes at time n and calculate the Euclidean distance between \tilde{I}_n and other delay vectors. The calculation formula is shown in equation (4).

$$d_x = D(\tilde{I}_n, \tilde{I}_{\tilde{n}}). \quad (5)$$

Select $E+1$ delay vector with the smallest distance as the nearest neighbor point, and get the set of time points $\{\tilde{n}_1, \tilde{n}_2, \dots, \tilde{n}_{E+1}\}$ and distance sets $\{d_1, d_2, \dots, d_{E+1}\}$. Calculate the weight of \tilde{I}_n for the distance set as follows:

$$m_x = \frac{p_x}{T}, \quad (6)$$

$$p_x = e^{-d_x/d_1}, \quad (7)$$

$$T = \sum_{y=1}^{E+1} p_y. \quad (8)$$

(3) Calculate the estimated value of X to Y and the correlation coefficient

Using the weight of each point, the weighted sum of each point J_n in the coordinate sequence J is carried out, thus obtaining the estimate of I to J .

$$J_n | \tilde{I} = \sum_{x=1}^{E+1} m_x J_{\tilde{n}_x}. \quad (9)$$

In this paper, the ability of I historical information to estimate J is measured by calculating Pearson phase relation C_{JI} of original coordinate sequence J and estimated value $J|\tilde{I}$.

$$C_{JI} = [\rho(J, J|\tilde{I})]^2. \quad (10)$$

(4) Calculate the edge weight matrix

For a pair of joints I and J , two correlation coefficients C_{IJ} and C_{JI} can be obtained, which are I estimation J and J estimation I , respectively. Assuming that the space-time graph has T joints, the

causality coefficient matrix C can be obtained by calculating the causality coefficient of two joints.

$$C = \sum_{x=0}^{T-1} \sum_{y=0}^{T-1} \text{cor}(l_y, l_x). \quad (11)$$

x and y are the joint numbers. In this paper, each line of coefficient matrix C is normalized by Softmax, and the normalized coefficients are embedded as edge weights of $A(Q, E)$.

3.3.2. Edge Weight Embedding. Inspired by the attention mechanism of CNN and RNN, this paper changes the edge weight of $A(Q, E)$ edge set according to the causal coefficient. In order to assign different weights to different edges in $A(Q, E)$, equation (2) is modified as follows:

$$f_{out} = \sum_q^{Z_q} M_z(f_{in}(G_z \odot \hat{C})), \quad (12)$$

where \odot represents the product of each element of the matrix. In equation (12), G_z indicates whether two vertices are connected in a subset by a matrix element value of 0 or $1/|S_{xz}|$. However, in the process of movement, the strength of interarticular force is different. Since the matrix G_z was $T \times T$, the edge weight matrix \hat{C} was $T \times T$, the matrix size was also $T \times T$ when \hat{C} and G_z were applied element by element. Therefore, replacing G_z in equation (2) with $(G_z \odot \hat{C})$ will not cause the problem of matrix size mismatch.

As shown in Figure 4, the weight of the upper part of the bone marked \hat{C} corresponds to the movements of the lower part. Because the main movement part is the arm, and the movement range is large, while the movement range of the lower limb joints is small, the edge weight between the upper limb joints is correspondingly larger than the lower limb joints.

In this paper, the CCM method is used to test the characteristics of time variable causality and calculate the causality between each vertex. Joint causality is transformed into edge weights, and then edge weights are embedded into graph adjacency matrix by by-element product, thus reducing the joint attention of low importance and enhancing the joint attention of high importance.

3.3.3. Network Structure. ST-GCN is used as the basic network in this paper. In order to apply edge weight matrix, this paper adds edge weight matrix as input and multiplies edge weight matrix with adjacent matrix element by an element before convolution operation. The convolution layer structure of space-time graph is shown in Figure 5.

4. Result Analysis and Discussion

In this section, the experiment is firstly carried out on three public motion recognition datasets based on NTU RGB-D, Northwestern-UCLA, and SBU Interaction Dataset. The experiment was then performed on the UCF Sports dataset.

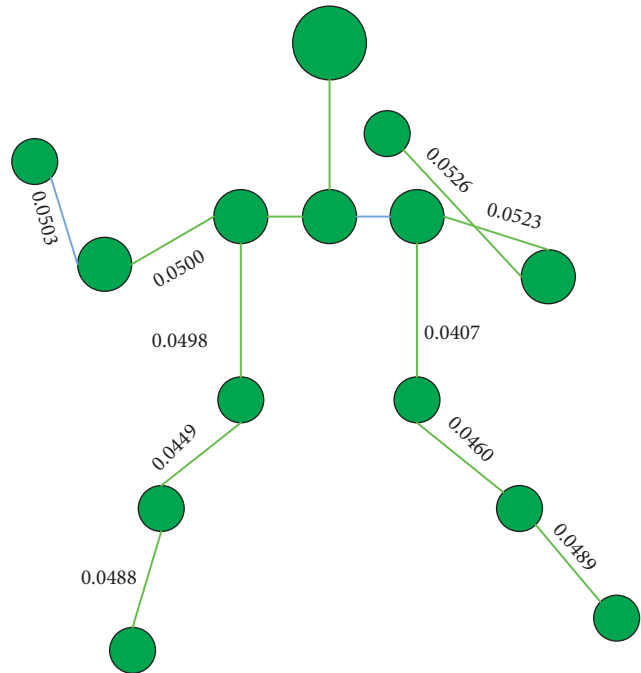


FIGURE 4: Schematic diagram of edge weight.

4.1. Experiment on General Motion Recognition Database.

This experiment is based on the framework of the space-time graph convolution network. The processor is Intel Core(TM) I7-7700, main frequency 3.60 GHz, 32 GB memory, and NVIDIA Ge Force GTX 1070. The number of neurons in each layer was 128, the apparent feature extraction radius was 5 pixels, and the initial learning rate was 0.002. The equilibrium factor $\lambda = 10^{-5}$ and the batch size is 64. Dropout = 0.45 to prevent overfitting.

4.1.1. NTU RGB-D Dataset. This dataset consisted of 40 subjects using 3 Kinect V2 cameras to collect 60 kinds of movements, 56,880 video clips, and 3D bone data sequences from -45° , 0° and 45° angles. It includes individual daily training movements (such as rope skipping, running, and squatting), interactive training of characters (such as barbell training, dumbbell training, and elastic belt training), and interactive training of pairs (such as drug ball relay, double flat back stretching, and double cross high-five)

In the cross-subject experiment, 40 types of subjects were divided into training and test sets, numbered as 1, 2, 4, 5, 8, 9, 13, 14, 15, 16, 17, 18, 19, 25, 27, 28, 31, 34, 35, 38, and the rest were test sets. In the cross-view experiment, the first camera was selected as the test set, and the rest were training sets.

In this section, the accuracy and loss curves corresponding to the training set and test set in the iterative training of cross subject and cross view are shown in Figures 6 and 7. It can be seen from Figures 6 and 7 that the accuracy of the model increases with the increase of training times and tends to be stable and the loss value converges when the iteration reaches 220 times. Based on the NTU RGB-D dataset, the accuracy of cross subjects and cross

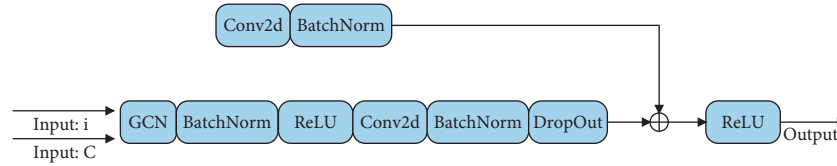


FIGURE 5: Spatiotemporal graph convolutional block.

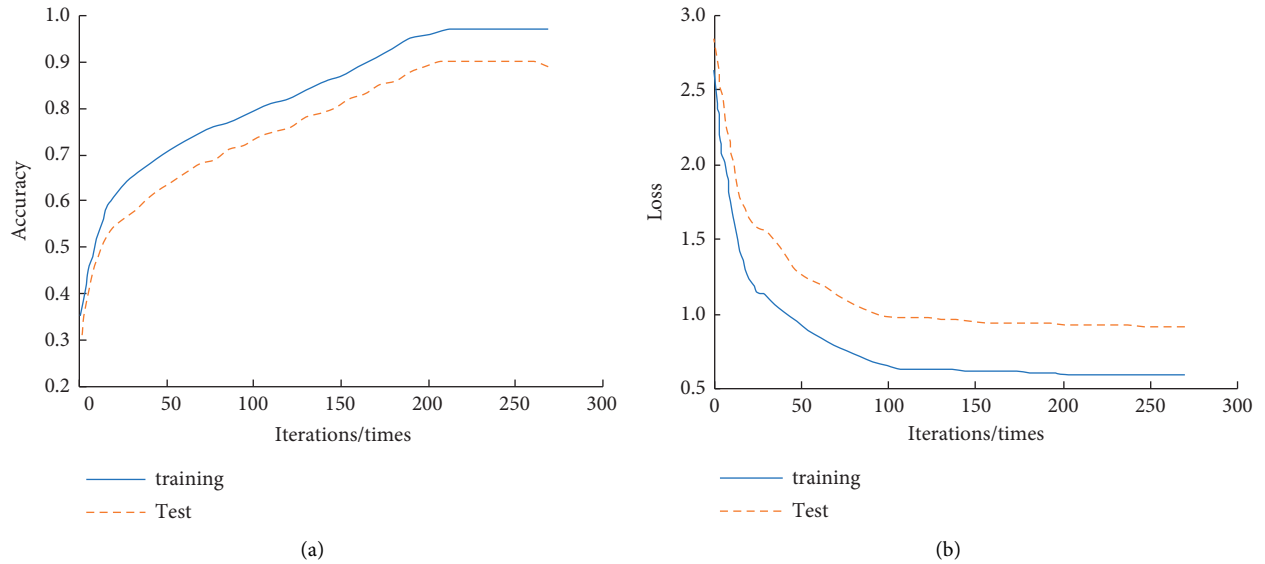


FIGURE 6: Accuracy and loss values in cross subject.

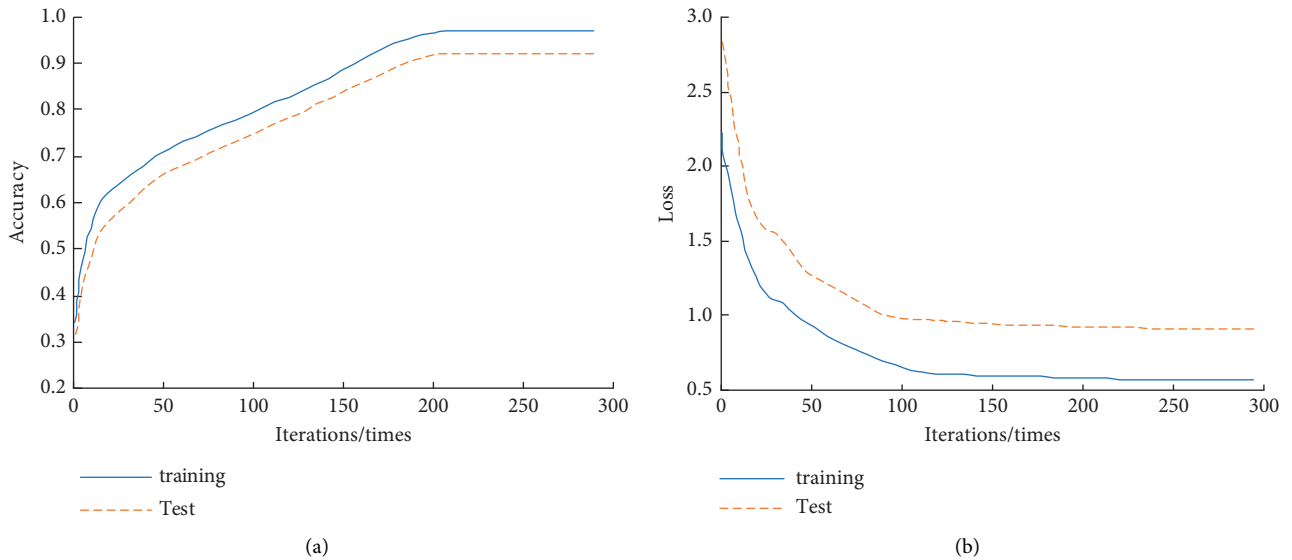


FIGURE 7: Accuracy and loss values in cross view.

angles obtained by the proposed method and the mainstream method is shown in Table 1.

From Table 1, it is clear that literature [17] based on variable parameter related skeletons and dynamic skeletons based on 3D geometric relationships) does not take into account deep space-time information, resulting in low accuracy. Literature [18] mapped joints to 3D space and

extracted depth features through 3D CNN, thus effectively improving accuracy to 67.96% and 73.69%. However, it does not consider the time domain information of bone recognition. Literature [19] considered the relative motion trend of interframe and intraframe joint links. However, this method only considers joint characteristics and ignores topological relations, so the accuracy is limited. Literature

TABLE 1: Cross subject and cross view accuracy obtained by each algorithm for the dataset.

Data	Methods	Cross subject %	Cross view %
Bone sequence	Literature [17]	51.19	53.87
	Literature [18]	67.96	73.69
	Literature [19]	78.91	88.44
	Literature [20]	85.33	90.82
	Proposed method	89.84	91.12

TABLE 2: Experimental results of northwestern-UCLA dataset.

Data	Features	Methods	Accuracy (%)
Bone sequence	Manual extraction	Literature [17]	55.70
		Literature [18]	75.40
	Spatiotemporal graph convolution network	Literature [19]	79.72
		Literature [20]	85.40
		Proposed method	87.95

[20] encodes spatial relations between joints to improve accuracy. However, they lack appearance features, thus limiting recognition ability. The proposed method uses causality as auxiliary information to enhance the graph convolutional network, thus improving the accuracy to 89.84% and 91.12%. It shows that the proposed method has high accuracy in complex scenarios.

4.1.2. Northwestern-UCLA Dataset. The Northwestern-UCLA dataset consisted of 1,494 sequences of 10 subjects performing 10 types of exercises: bends, presses, hard pulls, push-ups, planks, squats, pull-ups, pull-ups, dumbbell exercises, and barbell exercises. The dataset was collected from three different perspectives. The first two cameras were training data, and the rest were test data.

As shown in Table 2, literature [17] assumed that the skeleton was perpendicular to the ground for projection clustering discrimination, ignoring the spatial relationship of the skeleton, resulting in low accuracy. Reference [18] is better than reference [17] because it is based on variable parameter-associated skeleton to represent motions, but it ignores bone dynamic information. Literature [19] obtained 79.72% accuracy by considering the temporal characteristics of joints, but it was difficult to distinguish similar movements due to the lack of appearance information. In literature [20], multiperspective dynamic images are extracted to cope with spatial changes and take appearance features into consideration, but they lack timing features. This method integrates causality and represents important dynamic information of joint effectively, and extracts color texture information based on the heat map to obtain highly separable expression of the motion. The accuracy is improved to 87.95%, which is 8.23% and 2.85% higher than literature [19] and literature [20], respectively. It indicates that the proposed method has high recognition ability under the condition of different perspectives and diversified topics.

4.1.3. SBU Interaction Dataset. The SBU interaction dataset contains the following 5 types of interaction: double squats, double single-leg squats, high-five push-ups, reverse

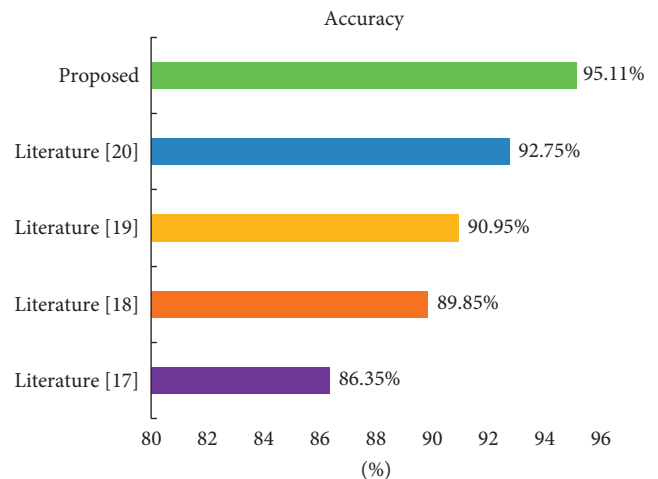


FIGURE 8: Experimental results of SBU interaction dataset.

crunches, and boxing, which are divided into 5 cross sets. Four of them were selected as training sets and the rest as test sets. The average value of the verification results of each cross set was taken as the final accuracy.

See Figure 8 for a comparison of the accuracy of the 5 algorithms. It shows that the accuracy of the proposed method can reach 95.46%, which is more accurate than the other four methods, indicating that the accuracy of the proposed method is relatively high under a small sample dataset.

4.1.4. Ablation Experiment. In order to further verify the effectiveness of the proposed method, based on the above dataset, the influence of the proposed method on accuracy was studied by integrating causality and spatiotemporal graph convolutional network (see Table 3). As can be seen from Table 3, compared with the spatiotemporal graph convolutional network model, the accuracy of the model in this paper has been improved by 12.90%, 7.29%, 8.15% and 3.13%, respectively. In this model, causal coefficients are added as edge weights, and causality is used as auxiliary information to enhance graph convolutional networks.

TABLE 3: Experimental results of different models.

Dataset	Spatiotemporal graph convolution network (%)	Proposed method (%)
NTU (cross subject)	77.06	89.96
NTU (cross view)	83.95	91.24
Northwestern-UCLA	78.81	86.96
SBU	93.56	96.69



FIGURE 9: Sample frames of each motion in the UCF motion dataset. (a) Barbell training. (b) Running. (c) Pull-ups. (d) Jumping rope. (e) Swimming. (f) Cycling. (g) Flying swallow. (h) Plank support.

Therefore, it can better highlight the main joints in the process of human movement, and its effect is far better than other methods.

4.2. Experiments on the Dataset of Motion Training Posture.

The UCF Sports dataset was used, which consists of eight training movements, including running, rope skipping, swimming, swallows, cycling, pull-ups, and planks. These movements are in a real moving environment, showing variations in background, lighting conditions and occlusion, making it a challenging dataset. A sample frame for each motion is shown in Figure 9.

5. Conclusion

Considering the weight of joints in human motion from the perspective of causality, this paper proposes a human motion recognition method combining causality and spatiotemporal graph convolution network. Inspired by the attention mechanism in RNN and CNN, causality is used as auxiliary information in this paper to enhance the graph convolutional network, so as to effectively improve the weight of some joints with strong driving force in the neural network. Experiments on three public motion recognition datasets show that the accuracy of the proposed method is significantly higher than that of the existing mainstream recognition methods. It is proved that the proposed method can effectively learn dynamic features and increase the

accuracy of motion recognition. In the future, this paper will try to integrate other modal information, such as RGB and skeleton data, and combine skeleton-based motion recognition and pose estimation methods in a unified framework.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Henan University of Animal Husbandry and Economy.

References

- [1] J. Chen, Y. Sun, and S. Sun, "Improving human Activity recognition performance by data fusion and feature Engineering," *Sensors*, vol. 21, no. 3, p. 692, 2021.
- [2] C. N. Phyo, T. T. Zin, and P. Tin, "Deep learning for recognizing human Activities using motions of skeletal joints," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 2, pp. 243–252, 2019.
- [3] E. K. Kumar, P. V. V. Kishore, A. S. C. S. Sastry, M. T. K. Kumar, and D. A. Kumar, "Training CNNs for 3-D

- Sign Language recognition with color texture coded joint Angular Displacement maps,” *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 645–649, 2018.
- [4] S. Z. Gurbuz and M. G. Amin, “Radar-based human-motion recognition with deep learning: Promising Applications for Indoor Monitoring,” *IEEE Signal Processing Magazine*, vol. 36, no. 4, pp. 16–28, 2019.
- [5] W. Xi, G. Devineau, F. Moutarde, and J. Yang, “Generative model for skeletal human movements based on conditional DC-GAN applied to pseudo-images,” *Algorithms*, vol. 13, no. 12, p. 319, 2020.
- [6] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal recurrent neural network for emotion recognition[J],” *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 839–847, 2018.
- [7] W. Du, Y. Wang, and Y. Qiao, “Recurrent spatial-temporal attention network for motion recognition in videos[J],” *IEEE Transmotions on Image Processing*, vol. 27, no. 3, pp. 1347–1360, 2017.
- [8] C. Ding, K. Liu, F. Cheng, and E. Belyaev, “Spatio-temporal attention on manifold space for 3D human action recognition,” *Applied Intelligence*, vol. 51, no. 1, pp. 560–570, 2021.
- [9] B. Hu and J. Wang, “Deep learning based Hand Gesture recognition and UAV Flight Controls,” *International Journal of Automation and Computing*, vol. 17, no. 1, pp. 17–29, 2020.
- [10] J. L. Liu and K. Li, “Design of an Intelligent Symptom Differentiation and Electrical Stimulation Rehabilitation system,” *Journal Européen des Systèmes Automatisés*, vol. 53, no. 5, pp. 681–693, 2020.
- [11] A. Göpfert, M. Van Hove, A. Emond, and J. Mytton, “Prevention of sports injuries in children at school: a systematic review of policies,” *BMJ open sport & exercise medicine*, vol. 4, no. 1, Article ID e000346, 2018.
- [12] A. Mathis, P. Mamidanna, K. M. Cury et al., “DeepLabCut: markerless pose estimation of user-defined body parts with deep learning,” *Nature Neuroscience*, vol. 21, no. 9, pp. 1281–1289, 2018.
- [13] K. Hu, Y. Ding, J. Jin, L. Weng, and M. Xia, “Skeleton motion recognition based on multi-scale deep spatio-temporal features,” *Applied Sciences*, vol. 12, no. 3, p. 1028, 2022.
- [14] J. Zhang, G. Ye, Z. Tu et al., “A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition,” *CAAI Transactions on Intelligence Technology*, vol. 7, no. 1, pp. 46–55, 2022.
- [15] B. Li, X. Li, Z. Zhang, and F. Wu, “Spatio-temporal graph Routing for skeleton-based action recognition,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 8561–8568, 2019.
- [16] D. Avola, M. Cascio, L. Cinque, G. Foresti, C. Massaroni, and E. Rodolà, “2-D skeleton-based motion recognition via two-branch stacked LSTM-RNNs[J],” *IEEE Transmotions on Multimedia*, vol. 22, no. 10, pp. 2481–2496, 2019.
- [17] L. Cai, C. Liu, R. Yuan, and H. Ding, “Human action recognition using Lie Group features and convolutional neural networks,” *Nonlinear Dynamics*, vol. 99, no. 4, pp. 3253–3263, 2020.
- [18] G. Yao, T. Lei, J. Zhong, and P. Jiang, “Learning multi-temporal-scale deep information for action recognition,” *Applied Intelligence*, vol. 49, no. 6, pp. 2017–2029, 2019.
- [19] X. Jiang, K. Xu, and T. Sun, “Motion recognition scheme based on skeleton representation with DS-LSTM network[J],” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 2129–2140, 2019.
- [20] A. Banerjee, P. K. Singh, and R. Sarkar, “Fuzzy Integral-based CNN classifier fusion for 3D skeleton motion recognition[J],” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2206–2216, 2020.