High-Order Graph Convolutional Network for Skeleton-Based Human Action Recognition

Zhimin Bai[1], Hongping Yan[1], and Lingfeng Wang[2(✉)]

[1] Institute of Information Engineering, China University of Geosciences, Beijing, China
{zmbai,yanhp}@cugb.edu.cn
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, China
lfwang@nlpr.ia.ac.cn

Abstract. Skeleton-based action recognition plays an important role in the field of human action recognition. Recently, with the introduction of Graph Convolution Network (GCN), GCN has achieved superior performance in the field of skeleton-based human action recognition. In this work, we propose a high-order GCN model. In this model, we introduce the expression of high-order skeletons and establish a new high-order adjacency matrix. Through this matrix, the relationship between skeleton nodes and non-neighbor nodes has being established. In addition, based on the degree of node association of different hierarchical neighborhoods, the value of the matrix expresses the importance of different hierarchies. As a result, the proposed model extracts the co-occurrence feature of the skeleton which is superior to the local features and improves the recognition rate. We evaluate our model on two human skeleton action datasets, Kinetics-skeleton and NTU RGB+D, and then further explore the influence of skeleton nodes based on different hierarchies on the recognition results.

Keywords: Human action recognition · High-order skeleton information · Graph convolution network

1 Introduction

Human action recognition is an important and challenging research field of computer vision, and has received extensive attention in recent years. At present, the RGB image sequence is the main research field of human action recognition [1–8]. However, it is greatly affected by the environment, such as light and background. In addition, human action recognition based on RGB image sequences is difficult to distinguish subtle motion differences between similar actions. Currently, with

This work is supported by the National Natural Science Foundation of China (Grant Number 61773377 and 61573352).

Z. Lin et al. (Eds.): PRCV 2019, LNCS 11857, pp. 14–25, 2019.
https://doi.org/10.1007/978-3-030-31654-9_2

the continuous development of software and hardware equipment, advanced algorithms for extracting skeleton sequences [9] and human action datasets based on skeleton [10] have been proposed. Based on this, skeleton-based human action recognition algorithm is proposed [12–14]. These Convolutional Neural Network (CNN) based models tend to be complex and difficult to obtain skeleton spatial features. For example, the TCN model [14] proposed by Kim et al. only considers the temporal information of the skeleton and ignores the spatial relationship between the skeleton nodes. In order to solve these problems, Yan et al. [18] proposed a new model, which breaks through the traditional CNN method and uses GCN to extract temporal and spatial information of the skeleton.

CNN has been able to efficiently process Euclidean data. It refers to grids, sequences, etc. For instance, images can be viewed as 2D grids data. There are many non-Euclidean data in reality, however, such as the human skeleton. Kipf et al. [23] formally proposed GCN to deal with non-Euclidean data, and also achieved good results in the field of human action recognition [18]. Compared with the traditional CNN method, the GCN is simpler and more precise.

At present, the GCN-based model extracts feature information through connections between nodes. This makes the feature representation of human skeleton simpler and more comprehensive than CNN. However, there is a correlation between multiple joints of the human skeleton when the person is moving. For example, in Fig. 1, when the person is drinking water, the wrist, elbow, shoulder, neck and head will have relative movement, even the interaction of the left and right arms is required to fully realize the behavior of drinking water. Therefore, when using GCN to implement skeleton-based human action recognition, only the skeleton nodes and their adjacent nodes are considered, that is limited.

(a) (b)

Fig. 1. People need to interact with multiple joints when drinking water, such as wrists, elbows, shoulders, neck and head.

In this paper, we propose a graph convolutional network model based on higher-order skeletons. Based on the current best network model ST-GCN [18], we consider the spatial-temporal information of the skeleton during motion and the kinematics theorem of the body. Furthermore, the expressions of higher-order skeleton nodes and higher-order adjacency matrices are introduced. In addition, we establish connections between non-neighboring nodes and express

the importance of joint points of different hierarchies by parameters, which is determined by the degree of correlation between the nodes. The main details and superiors of this work are listed as follows:

(1) Based on the ST-GCN network, we propose a high-order graph convolution network model based on skeleton. Through this model, the relationship between the high-order skeletons is expressed, and the co-occurrence characteristics of the skeleton are extracted. We verified the effectiveness of the method through experiments.
(2) We propose a new high-order neighborhood representation that achieves the importance of different neighborhood nodes by defining parameters and learnable weights. It reduces the noise and experimentally verifies that the method can improve the result.

2 Related Work

The current mainstream models and methods for human action recognition are based on RGB video, such as C3D [4,5], Two-stream [1–3] and Long Short Term Memory (LSTM) [6,7], etc. However, with the introduction of the human skeleton extraction method [9] and the establishment of related datasets [10], the skeleton-based human action recognition has gradually developed. Early traditional methods mainly used the sliding window [15] or relative position between joints [16] to obtain characteristic information of skeleton. With the popularization of deep learning in the field of computer vision, deep networks based on skeleton-based human action recognition is proposed. It is mainly divided into two methods: One is to convert the node coordinate information of the skeleton [11] or the distance between the joint points and the angle between the skeletons into a picture [12], and then extract features through the CNN. Li et al. [17] proposed a new end-to-end hierarchical feature learning network, which realizes the aggregation from the point level to the global co-occurrence feature. On the other hand, Song et al. [19] introduced the spatial-temporal attention mechanism based on the RNN neural network of LSTM and achieved good results. Liu et al. [20] optimized the spatial-temporal attention model and improved the performance of the network. Wang et al. [21] used two-stream RNN to realize the extraction of spatial-temporal features. Zhang et al. [22] proposed a new idea, which is a new viewpoint adaptive scheme. The coordinates of the skeleton are rotated to the appropriate angle of view, and the action recognition is performed through the RNN. Its has been greatly improved compared to the previous method.

With the emergence of many datasets in the form of graphs or networks, neural networks based on graph structure are an emerging topic in current deep learning research. In the past few years, many researchers have paid attention to the problem of generalizing neural networks to handle arbitrary graph structures. For example: Bruna et al. [24] first proposed the application of irregular grids on CNN, and proposed two methods: spatial domain and spectrum domain. For the spectral domain, Henaff et al. [25] proposed using a smooth kernel to implement

a local filter. For the spatial domain, Niepert et al. [26] proposed to use the CNN to efficiently process the graph structure data by labeling the graph nodes and then convolving the nodes according to the sequence. Thomas et al. [23] proposed an extensible semi-supervised learning convolutional neural network method to process graph-based data and formally propose GCN. Meantime, with the introduction of GCN, it provides new research directions for data application based on graph structure [27,28], such as skeleton-based human action recognition. Although CNN and LSTM perform well in skeleton-based human action recognition, they have problems such as complex models and difficulty in training. Therefore, Yan et al. [18] proposed using GCN to realize skeleton-based human action recognition (ST-GCN), which shows better performance than the most advanced model.

3 Method

Actions of the human body is a range of local motion centered on certain joint points. Therefore, an important step of human action recognition based on high-order skeleton is to divide the skeleton based on kinematics. In addition, since the establishment of the network framework is based on the graph convolution, it is necessary to transform the skeleton into the expression form of the graph structure, and then realize the representation, division and graph convolution method of the high-order skeleton.

3.1 Skeleton Graph Construction

(a) Spatial diagram (b) Spatial-Temporal diagram

Fig. 2. Skeleton spatial-temporal structure diagram. (a): The spatial relationship of the skeleton is represented by the connection between joint points. (b): The temporal relationship of the skeleton is represented by connecting the same joint points between consecutive frames.

The human skeleton is a typical graph structure. When describing the behavior of the human body, we need to obtain the spatial and temporal information

of the skeleton. Therefore, the joint point set of the skeleton can be expressed as $V = \{v_{tn} | t = 1, 2, \ldots, T, n = 1, 2, \ldots, N\}$, where T represents the number of video frames and N represents the number of joint nodes. First, one connects the joint points in the same frame, then each edge represents the spatial relationship of the joint points, as shown in Fig. 2(a). We use a subset to represent the spatial relationship of the edges, denoted as $E_s = \{v_{ti}v_{tj} | (i, j) \in S\}$, where S represents the natural connection of the human joint. Temporal relationship is established by connecting the same joint point between consecutive frames. The set of temporal relational edges can be expressed as $E_t = \{v_{ti}v_{(t+1)i} | t = 1, 2, \ldots, T-1, i = 1, 2, \ldots, N\}$. The set E of skeleton edges can be expressed as: $E = E_s \bigcup E_t$. Skeleton spatial-temporal relationship diagram is shown in Fig. 2(b).

The spatial relationship of the skeleton can be converted into an adjacency matrix. In t frame, if there is a connection between two nodes: $v_{ti}v_{tj} \in E_s$. It can be expressed as $A_{ij} = 1$ in the adjacency matrix A. In addition, taking into account the impact of the joint itself, we set $A_{ii} = 1$. If there is no association between nodes, i.e. $v_{ti}v_{tj} \notin E_s$ and $i \neq j$, then $A_{ij} = 0$. The temporal relationship of the skeleton, we recall, is constructed by connecting the same nodes of consecutive frames. Therefore, based on the spatial relationship, it can be easily extended to the spatial-temporal relationship. Suppose we operate on the skeleton in the time range θ, the spatial-temporal relationship $ST(v_{ti})$ can be expressed as:

$$ST(v_{ti}) = \{v_{qi} | d(v_{ti}, v_{tj}) < n, |t - q| \leq \lfloor \theta/2 \rfloor\} \tag{1}$$

3.2 Division Strategy

Skeleton High-Order Adjacency Matrix. According to the graph structure of the skeleton, we can establish the 1-order adjacency matrix of the skeleton. However, based on the kinematics of the human body, the human body needs multiple coordination of the body to complete the exercise. In this regard, we can obtain the co-occurrence characteristics of the skeleton joint by establishing a high-order adjacency matrix. To distinguish the joint points of different hierarchies, we use the shortest path length of the two joints, $d(v_i, v_j)$, to express the relationship between the joint points. Then the nodes in the n-order that affect each other need to satisfy: $d(v_i, v_j) \leq n$, where if $i = j$, set $d(v_i, v_j) = 1$. It is known that the adjacency matrix A represents the spatial relationship information between the node and the neighbor nodes. The n-order adjacency matrix A_n can be expressed by a 1-order adjacency that extends the expression of the spatial relationship to non-neighbor nodes:

$$A_n = A^n \tag{2}$$

Where A is the 1-order adjacency matrix, n means n^{th} order. The n-order adjacency matrix established by Eq. (2) implements the adjacency matrix parameter to represent the shortest distance between the joint points of the skeleton. The

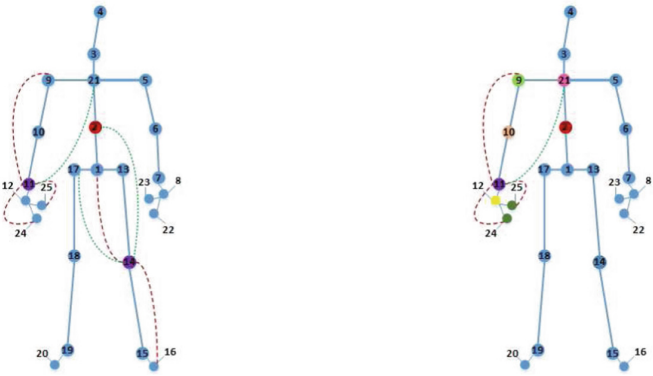
equation is as follows:

$$A_n^{ij} = \begin{cases} 0 & d(v_i, v_j) > n \\ d(v_i, v_j) & d(v_i, v_j) \leq n \end{cases} \tag{3}$$

High-Order Skeleton Division. Human action is based on the local motion of some joint nodes within a range, so it is necessary to divide the skeleton. Referring to the division method in [18], the division of the skeleton needs to consider the kinematics of the human skeleton. For simplicity, we only consider skeleton partitioning within a single frame. The division of the skeleton is mainly divided into two parts, as shown in Fig. 3. Firstly, the multi-order neighborhood of skeleton node is divided. We set $N_n(v)$ is the nth-order neighborhood of node v. Therefore, assuming that the 3-order node neighborhood of the skeleton is divided, the n-order neighbor nodes of node v can be expressed as:

$$N(v) = N_1(v) + N_2(v) + N_3(v) \tag{4}$$

where $N_1(v)$ includes node v itself.



(a) Dividing nodes by layer (b) Dividing nodes by kinematics

Fig. 3. Division strategy. Taking the division of the 3-order skeleton as an example, the skeleton is divided according to the division strategy. (a): The red dashed line indicates the connection of the 2nd-order neighborhood joint point, and the green dashed line indicates the connection of the 3rd-order neighborhood joint point. (b): The red node represents the center point. We further divide the nodes of different hierarchies, and different colors represent different divided regions.

We divided the skeleton according to the kinematics theory of the human body. Then, considering that all movements of the human body belong to centripetal or eccentric motion, we select the central node c of the skeleton as the center of the motion range. The nth-order neighborhoods of the skeleton are respectively divided, and a label map r is set for each partition. We divide the

skeleton according to the following method: (i). According to the law of motion, the first division should be the node itself. The corresponding mapping is $r = 0$. (ii). For a neighboring node in a hierarchy, if the distance from the node to the center point is closer or equal to the distance from the feature node to the center point, then the neighboring node belongs to the second partition. The corresponding mapping is $r = 1$. (iii). The remaining nodes, that is, the distance from the node to the central point is farther than the distance from the feature node to the central point, belonging to the third partition. The corresponding mapping is $r = 2$.

According to the partitioning strategy and label mapping, correspondingly, the adjacency matrix A_n can be divided. Assume that A_{N_i} is used to represent the adjacency matrix of the i-th neighbor node. Then, A_{N_i} is further divided according to the partitioning strategy, and finally the high-order adjacency matrix can be expressed as:

$$A_n = \sum_{i=1}^{n} \sum_{r=0}^{2} A_{N_i}^r \tag{5}$$

where $A_{N_i}^r$ represents the matrix after A_{N_i} is partitioned.

In addition, considering the different degrees of association between different hierarchies, we define a parameter Φ represent the importance of different hierarchies, the setting of which is related to the hierarchy of the node. In summary, the expression of the higher-order adjacency matrix is as follows:

$$A_n = \Phi(N_i) \sum_{i=1}^{n} \sum_{r=1}^{2} A_{N_i}^r \tag{6}$$

3.3 Spatial-Temporal Graph Convolution

We learn the spatial feature information of the skeleton on a single frame. The convolution operation acts on the node. If the n-order neighbor of v_{ti} is sampled, the sampling function is: $p(v_{ti}, v_{tj}) = \{v_{tj} | d(v_{ti}, v_{tj} \leq n)\}$.

Compared with the sampling function, the definition of the weight function has a relatively large change. The node spatial structure of the graph structure is not fixed, and the number of neighbor is different. The weight of the graph structure can be expressed in another form, that is, the adjacency matrix of the graph. The adjacency matrix is another expression of the spatial relationship of the graph structure. The parameter of the adjacency matrix is the weight of the graph convolution. Therefore, the weight function W can be written as: $W(v_{ti}, v_{tj}) = A_n(v_{ti}, v_{tj})$. Spatial graph convolution can be expressed as:

$$Fs_{out} = \sum_{v_{tj} \in p(v_{ti}, v_{tj})} f_{in}(p(v_{ti}, v_{tj})) \cdot A_n(v_{ti}, v_{tj}) \tag{7}$$

The convolution operation in the temporal direction is relatively simple. After determining the node information in the spatial direction, the convolution in the

time direction needs to implement the convolution of the same node in a certain period of time. We set the input of the convolution network to $X(v_{qi}, t)$. It represents v_{qi} node within the t frame range. Therefore, the convolution operation in the time direction is as follows:

$$Ft_{out}(v_{qi}) = \sum_{|t-q|=0}^{\theta/2} X(v_{qi}, t) \cdot w(t, 1) \tag{8}$$

In addition, we consider that even the nodes in the same neighborhood have different effects on motion. Therefore, when extracting the features in the spatial direction, we set a parameter matrix M that can be learned. Set all the parameters in M to 1, multiply with A_n by element, and learn the parameters of each node through deep learning network. Therefore, higher-order graph convolution operation can be expressed by:

$$f_{out} = \sum_{n} \Lambda_n^{-1/2} M \otimes A_n \Lambda_n^{-1/2} f_{in} \cdot W_t \tag{9}$$

where $\Lambda_n^{ij} = \sum_k A_n^{ik} + \alpha$, we set $\alpha = 0.001$ to avoid empty rows in A_n. W_t is a weight function in the time direction. f_{in} is the characteristic function of the input skeleton. $\otimes$ represents multiplication of the M matrix and the A_n matrix by elemental correspondence.

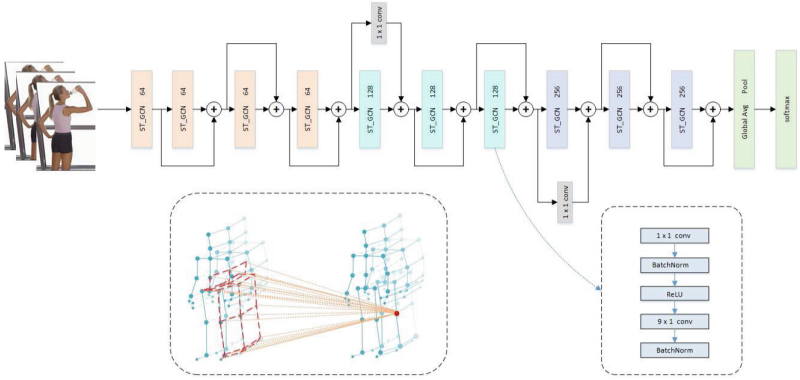


Fig. 4. High-order graph convolutional network model.

Before performing feature learning, the data needs to be pre-processed to fit the graph convolution network. The human skeleton data of a video is converted into a 3-dimensional tensor (C, T, V). C represents the number of channels, which corresponds to the 3-dimensional coordinates of the skeleton. T represents the number of video frames, and V represents the number of nodes. The extraction of spatial-temporal information is realized by convolution operations on spatial dimensions and temporal dimensions. The network frame of the spatial-temporal graph convolution is shown in Fig. 4.

4 Experiment

We verified that our approach can achieve better accuracy. We separately evaluated the human motion datasets based on 2D and 3D skeletons, and discussed the experimental results based on different high-order skeleton data. In order to make the experimental result data more objective, we set the same experimental environment for the experiments under each dataset.

4.1 2D Skeleton Data

The 2D skeleton data we used is Kinetics-skeleton, which is extracted from the Kinetics human action video dataset [29] via openpose [9]. Then use openpose to identify the 18 joint points of the skeleton, and extract the skeleton of the human behavior in each frame. The storage form of the skeleton node is (X, Y, C), and (X, Y) represents the 2D coordinates of the skeleton node, C represents the confidence score of the skeleton node. According to the recommendations of the dataset authors, we use $Top-1$ and $Top-5$ to evaluate their performance. The probability of correct classification. The performance of this dataset on the graph convolution network of the nth-order skeleton is shown in the following table. In this experiment, we set $batchsize = 100$, $epoch = 60$ and experiment with $n = 1, 2, 3, 4$ respectively.

Table 1. Action recognition performance for high-order skeleton based models on the Kinetics database. 2-order ST-GCN* means that parameters Φ that represent the importance of different orders are not considered

Method	Top-1	Top-5
Deep LSTM [10]	16.6	35.3
TCN [14]	20.5	40.4
ST-GCN [18]	31.6	53.7
1-order GCN	31.5	54.2
2-order GCN*	32.7	55.4
2-order GCN	33.3	56.2
3-order GCN	**33.8**	**56.4**
4-order GCN	33.7	56.0

Under this dataset, when only 1-order neighbor nodes are considered, our method is similar to st-gcn and the result is basically the same. If we only increase the order of the skeleton and regardless of the importance of different classes (the 2-order GCN* in Table 1), the results show that the accuracy rate will increase 1%. When we add a parameter that expresses the importance of the hierarchy, the accuracy increased by a further 0.5%.

4.2 3D Skeleton Data

The 3D skeleton dataset used in this study is NTU RGB+D [10], which is the largest dataset of behavior recognition research based on 3D skeleton data. The preservation form is the 3D coordinates (X, Y, Z) of the skeleton node, and the skeleton sequence includes 25 joint points whose center point is the joint point located at the center point of the human skeleton. This dataset divides all skeletons into two themes: X-sub and X-view. X-sub implements training and testing of different skeletons, that is, training with some actors and testing with other actors; X-view realizes skeleton training and testing from different perspectives, that is, training with two perspectives and testing with another perspective. We use $Top-1$ to evaluate its performance. The performance of the dataset in the high-order spatial-temporal graph convolution network is shown in the following table. In this experiment, we set $batchsize = 30$, $epoch = 100$ and experiment with $n = 1, 2, 3, 4$ respectively. In addition, we changed the center point of the 3D skeleton to the center of the human body (It is 1-order GCN in Table 2). The experimental results were improved compared with ST-GCN.

Table 2. Skeleton based action recognition performance on NTU-RGB+D datasets. 2-order ST-GCN* means that parameters Φ that represent the importance of different orders are not considered

Method	X-sub	X-view
Deep LSTM [10]	60.5	67.0
TCN [14]	74.2	82.9
C-CNN+MTLN [11]	79.0	84.2
ST-GCN [18]	79.5	86.4
1-order GCN	80.5	88.0
2-order GCN*	80.8	88.7
2-order GCN	**81.1**	**89.3**
3-order GCN	80.6	89.2

4.3 Discussion

The two datasets in experiments have very different natures. The 2D skeleton dataset is extracted by openpose and the 3D skeleton dataset is obtained by depth sensor. The number of skeleton nodes and the saved form are different, which makes the performance on the high-order spatial-temporal convolution network also very different. The experimental results based on 2D data show that before the 3-order, the accuracy rate is on the rise, and the accuracy is stable after the 3rd order. However, based on 3D data, it is bounded by 2-order. We suspect that this is due to the large noise generated by the 3D skeleton data annotation. Therefore, when the order is gradually increased, the error will

accumulate and accumulate more and more, which may affect the accuracy of the high-order skeleton data. However, based on the above results, we can still conclude that for any human skeleton data, when performing human behavioral motion, the feature information of a certain node should consider all neighbor information in the 2nd or 3rd order neighborhood.

5 Conclusion

In this work, we propose a high-order spatial-temporal graph convolutional network model, which is mainly to redefine the structure of the high-order skeleton and to divide the skeleton based on it. The divided skeleton is inputted into the Spatial-Temporal graph convolution network to realize the action recognition of the human body. The model shows good performance on different datasets, but it is difficult to express the skeleton nodes that are far away, which makes the expression of higher order features limited. In addition, certain behaviors of the human body are accomplished by the cooperation of various parts of the body. At this time, there are different degrees of correlation between different parts. We can learn this correlation through the network to get higher accuracy, which is reserved for future work.

References

1. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos (2014)
2. Feichtenhofer, C., Pinz, A., Zisserman, A.: Convolutional two-stream network fusion for video action recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
3. Wang, L., Qiao, Y., Tang, X.: Action recognition with trajectory-pooled deep-convolutional descriptors (2015)
4. Tran, D., Bourdev, L., Fergus, R., et al.: Learning spatiotemporal features with 3D convolutional networks (2015)
5. Carreira, J., Zisserman, A.: Quo vadis, action recognition? A new model and the kinetics dataset (2017)
6. Donahue, J., Hendricks, L.A., Rohrbach, M., et al.: Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 677–691 (2014)
7. Sharma, S., Kiros, R., Salakhutdinov, R.: Action recognition using visual attention. Comput. Sci. (2015)
8. Wang, Y., Zhou, W., Zhang, Q., et al.: Weighted multi-region convolutional neural network for action recognition with low-latency online prediction (2018)
9. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2D pose estimation using part affinity fields (2016)
10. Shahroudy, A., Liu, J., Ng, T.T., Wang, G.: NTU RGB+D: a large scale dataset for 3D human activity analysis. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1010–1019 (2016)
11. Ke, Q., Bennamoun, M., An, S., et al.: A new representation of skeleton sequences for 3D action recognition (2017)

12. Ding, N.Z., Wang, N.P., Ogunbona, P.O., Li, N.W.: Investigation of different skeleton features for CNN-based 3D action recognition. In: 2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW) (2017)
13. Liu, J., Shahroudy, A., Xu, D., Wang, G.: Spatio-temporal LSTM with trust gates for 3D human action recognition (2016)
14. Kim, T.S., Reiter, A.: Interpretable 3D human action analysis with temporal convolutional networks (2017)
15. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Learning actionlet ensemble for 3D human action recognition. In: IEEE Conference on Computer Vision Pattern Recognition (2012)
16. Du, Y., Wang, W., Wang, L.: Hierarchical recurrent neural network for skeleton based action recognition. In: CVPR, pp. 1110–1118 (2015)
17. Li, C., Zhong, Q., Xie, D., Pu, S.: Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation (2018)
18. Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition (2018)
19. Song, S., Lan, C., Xing, J., et al.: An end-to-end spatio-temporal attention model for human action recognition from skeleton data (2016)
20. Liu, J., Wang, G., Duan, L.Y., Abdiyeva, K., Kot, A.C.: Skeleton-based human action recognition with global context-aware attention LSTM networks. IEEE Trans. Image Process. **PP**(99), 1 (2017)
21. Wang, H., Wang, L.: Modeling temporal dynamics and spatial configurations of actions using two-stream recurrent neural networks (2017)
22. Zhang, P., Lan, C., Xing, J., et al.: View adaptive neural networks for high performance skeleton-based human action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **PP**(99) (2018)
23. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks (2016)
24. Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs (2014) arXiv:1312.6203
25. Henaff, M., Bruna, J., Lecun, Y.: Deep convolutional networks on graph-structured data. Comput. Sci. (2015)
26. Niepert, M., Ahmed, M., Kutzkov, K.: Learning convolutional neural networks for graphs. In: International Conference on Machine Learning, 2014C2023 (2016)
27. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting (2017)
28. Mahdi, K., Jianhui, W.: Spatio-temporal graph deep neural network for short-term wind speed forecasting. IEEE Trans. Sustain. Energy, 1 (2018)
29. Kay, W., Carreira, J., Simonyan, K., et al.: The kinetics human action video dataset (2017)