Hindawi
Scientific Programming
Volume 2022, Article ID 1838468, 10 pages
https://doi.org/10.1155/2022/1838468

Research Article

Taekwondo Action Recognition Method Based on Partial Perception Structure Graph Convolution Framework

Jianqiao Liang[1] and Guocai Zuo[1,2]

[1]*Shenyang Institute of Urban Construction, Shenyang, Liaoning 110167, China*
[2]*Hunan Software Vocational and Technical University, Hunan, Xiangtan 411100, China*

Correspondence should be addressed to Guocai Zuo; zuoguocai@hnsoftedu.com

Received 22 December 2021; Revised 4 January 2022; Accepted 13 January 2022; Published 21 February 2022

Academic Editor: Baiyuan Ding

Action recognition in Taekwondo competitions and training is an important task, which can provide a very valuable reference factor for technicians, athletes, and coaches. We propose a graph convolution framework with part of the perception structure to recognize, decompose, and analyze Taekwondo actions. Taking advantage of the long short-term memory of a part of the perception structure, the recognized Taekwondo actions are marked in time series, and then features are extracted from the graph convolution level to obtain the spatial and temporal associations between joints. Predict the action category and perform score matching based on the manual tag database. Finally, it is verified on our self-made Taekwondo competition data set. Our method has an average accuracy of 90% in action recognition, and an average action score matching rate of 74.6%. The accuracy of action recognition is high, which provides great assistance to Taekwondo e training and competitions.

1. Introduction

With the rapid development of science and technology, the sports industry urgently needs the intervention of artificial intelligence technology. This field has gradually attracted a lot of research. Previously, sports training was done in a more traditional manner. When the coach explains the training essentials to the athletes, the manner they play the video is not obvious, lacking in sensibility and interaction for the athletes [1]. With the upgrading of computer vision technology, human motion recognition began to be applied to the sports training industry. The development of these technologies has directly promoted the interpretation, prediction, and interaction of training actions in the sports industry. They have become a factor in the successful application of techniques and tactics in the process of sports training [2]. Taekwondo competitions have been updated day by day, and new technical systems have been introduced for training and practical exercises, which have a direct connection to Taekwondo athletes to obtain better results. Different from the development of the scoring system, the model of athletes adapting to technical and tactical methods to achieve victory has gradually become standard [3]. It is important to understand the technical actions of players in taekwondo competitions. The number of actions that occur in taekwondo battles is very high [4], but only a few achieve the main goal, which is to score [5, 6]. The main focus of understanding the scoring actions and thus promoting victory is not unique to Taekwondo. It has also been the research object of other fighting sports such as karate [7], boxing [8], fencing [9], and judo [10].

In order to be able to recognize taekwondo actions, in-depth analysis of the scoring points of each action is required, to provide athletes with skills during the training process. This paper uses an action recognition algorithm to recognize and decompose the actions of athletes in the Taekwondo competition. However, action recognition has higher requirements for scenes and characters, so in most cases, the tasks faced are more complicated. Therefore, most scholars have begun to study motion recognition methods based on bone joint points. The bone joint point–based action recognition

method and the image-based action recognition method are very different in principle, and the recognition effect is also very different [11]. Bone-based methods mainly rely on human skeleton data, and image recognition relies on a variety of pattern information, which is more susceptible to nonobjective factors. With the development of convolutional neural networks, human bone keypoint detection algorithm technology has made great breakthroughs in accuracy and precision. Skeleton data, as the mainstream input for action recognition, provide a more solid foundation for action recognition algorithms. Most researchers prefer to use bone data instead of RGB data, because bone data are compact and low in cost, and are not affected by nonobjective factors.

In the previous neural network action recognition tasks, the spatial relationship between joints has not been used, and features can only be extracted from the time level. Some researchers started to focus on this problem and validated various neural network models through experiments and finally found that long short-term memory (LSTM) neural network can obtain the spatial relationship between joints. The LSTM neural network can divide the bones and joints into independent parts, then implement partial feature perception for each part, so as to extract the features of the part, and finally obtain the spatial relationship between joints through the association between the partial features. Each part corresponds to an independent LSTM network, and finally, all the outputs are combined. The proposal of this method improves the use of space. However, the network is subject to manual network architecture predetermined rules, resulting in poor overall network robustness. The graph convolutional network (GCN) is used to recognize actions [12], which enriches the feature capture at the time and space level. Although GCN is often used in the analysis and classification of social networks, it is extraordinary in processing arbitrary graph structures [13]. The application of CNN in action recognition only stays at the 2D and 3D levels. The GCN can learn feature information from adjacent joints in the skeleton data. Some researchers have focused on designing a GCN architecture suitable for bones. For example, Li proposed a spatial temporal graph convolutional network (ST-GCN), which first modeled the bone joint point data, performed spatiotemporal convolution on all joint points, and obtained the spatial correlation and time correlation, and then used filters for feature capture. Gao et al. [14] used GCN to obtain spatiotemporal features from the bone joint point, which gave the model strong expressive ability and good generalization ability.

Based on the previous investigations and experiments, the purpose of this research is to identify and decompose the scoring techniques in the Taekwondo competition. The analysis results will provide valuable references for technicians, coaches, and athletes; help improve Taekwondo training methods; and optimize skills and tactics systems [15]. Therefore, to extract the important features of bone joint point-based Taekwondo action, a new approach combining LSTM and ST-GCN is provided. According to the temporal and spatial relationship between human bones, each frame is coded in sequence. Through the coded bone joint point sequence, the dynamic changes can be mapped in real time. The LSTM algorithm recognizes the network topology features as an input.

2. Related Work

In the study of motion recognition, compared with RGB data, bone data are not affected by a nonobjective environment and have high robustness, so it is widely used. Skeleton diagrams belong to non-Euclidean space and are fundamentally different from grid data. The skeletal data are arranged in a grid to use the structure of the neural network to obtain powerful feature learning capabilities [16, 17]. However, Monti et al. [18] raised a problem that in the vertex domain, meaningful operators cannot be fully expressed. Therefore, most current research points to GCNs because the operators defined by GCN networks are not in the vertex domain, but in the non-Euclidean space. Yan et al. proposed a bone-based action recognition method for the first time [19]. Gao et al. proposed a GCN network based on sparse regression to take advantage of the dependence between joints [14]. Shi et al. proposed a dual-stream method that uses the GCN architecture to capture the second-order information between the joints while acquiring the joint information, and then uses the classification strategy to filter the features [20]. Our method still uses graph convolution, as graph topology matrix can help us better traverse the skeleton task.

NAS (neural architecture search) [21] is a crucial component of automated machine learning, which automatically builds neural networks. At present, there have been a large number of literature focusing on NAS research, such as the reinforcement learning and black-box optimization researched by Zoph; the evolutionary search method researched by Real et al. [22]; the gradient-based processing proposed by Liu et al. [23]. In addition, Liu has unique insights in search space network design; in the realm of semantic segmentation and picture automated classification architecture design, Saxena et al. have more in-depth research. Although the method based on RGB data is not robust, Peng et al. still have a very high right to speak in the field of RGB data processing [24]. At present, Pham et al. proposed to optimize the graph neural network of ENAS to achieve inductive learning and citation, and the effect is also quite good [25]. Compared with our tasks, there is a certain gap, because its goal is to find a network with only two to three layers of conversion, propagation, and aggregation functions so that the post-decomposition work of Taekwondo can be effectively completed.

Graph convolutional neural networks have great advantages in irregular data and biological data processing, and they have gradually been applied to social networks and have begun to show their prowess. At the beginning of the definition of GCN, Defferrard proposed a spectral domain method, which aims to decompose the Fourier domain model at the time level [26]. Monti proposed a node domain method to realize the free switching of operators between graph nodes and leader nodes. However, the above studies are difficult to simulate the global structure of GCN. In order to further optimize GCN, Veli introduces an attention mechanism, which realizes automatic selection of key information [27] in all outputs. Since then, Velickovic and others have continued to explore the road of optimizing

GCN and proposed a node classification method of attention mechanism graph with better results. Sankar also researched the attention mechanism. He said that the attention mechanism was introduced into the dual dimensions of time and space, and in the process of achieving self-attention, he predicted better characteristics. Nevertheless, our approach is unique. We need to construct a dynamic display of joint points through the correlation characteristics between Taekwondo actions. The other is the problem of calculating the weights of different representations or different frames.

3. Method

3.1. Taekwondo Action and Scoring. Taekwondo is a martial art mainly based on legwork. In actual combat, the use of footwork can ensure that the power of the legs is fully utilized, which is of great significance for achieving victory in actual combat. Taekwondo mainly uses the hind legs as the core strength, so the footwork of Taekwondo has distinct characteristics. The athlete's center of gravity needs to fall between the feet or the front foot, and most of the body posture is to stand sideways to protect the body and the vital parts below so that the back legs can be turned by twisting the waist and turning the hips to increase the strength and speed of hitting. In our research, we mainly analyze the scoring points of eleven taekwondo actions, which are a front kick, push kick, cross kick, downward kick, side kick, hook kick, back kick, backspin kick, single leg kick, double leg kick, and double flying kick. According to relevant research, taekwondo is a sport mainly based on leg actions. Therefore, this paper, as the first stage of Taekwondo action decomposition research, will take leg action as the main research object. In the later research, the action decomposition research of other body parts such as the hand will be gradually increased. The main Taekwondo leg actions are shown in Figure 1.

In the scoring system of Taekwondo, a precise and powerful hit to the effective scoring area using the allowed techniques is required to score in the competition. 3 points are awarded for headbutts, 2 points for spin kicks and back kicks, 1 point for other techniques, and no additional points for the referee's reading. The maximum score for a technique is 3 points, and the valid scoring areas include the abdomen and both ribs, and the areas of the face that are allowed to be attacked. If the permitted technique is used to hit a nonvalid scoring area protected by protective gear, the knockdown will be scored. The specific scoring rules are shown in Table 1.

3.2. Graph Convolutional Network. We will go through the search-based GCN in detail in this part. To keep the article self-contained, we go over how to utilize GCN to represent spatial maps briefly.

Consider the undirected graph $G = \{V, E, A\}$ composed of $n = |V|$. The nodes connected by $|E|$ are encoded in the adjacency matrix $A \in R^{n\times n}$. Let $X \in R^n$ X be the input representation of G, and $\{x_i, \forall i \in V\}$ be its n elements. Then, in order to model the representation of G, the graph is Fourier transformed, so that the transformed signal, such as in Euclidean space, can then process basic calculation formulas, such as filtering. For this reason, the graph Laplacian L whose normalization is defined as $L = I_n - D^{-1/2}AD^{-1/2}$, and $D_{\text{ii}} = \sum_j A_{\text{ij}}$ is used for Fourier transform. Then, the graph filtered by the operator g_θ, parameterized by θ, can be expressed as

$$Y = g_\theta(L)X = Ug_\theta(\Lambda)U^T X, \tag{1}$$

where Y is the extracted graphic feature. U is the Fourier basis, which is a set of orthogonal eigenvectors of L, so $L = U\Lambda U^T$, where Λ is the corresponding eigenvector. However, multiplying with the eigenvector matrix is expensive. The computational burden of this nonparametric filter is $O(n^2)$. As suggested by Hammond et al. [28], the filter g_θ can be a good approximation of the Chebyshev polynomial of order R.

$$Y = \sum_{\text{r}=0}^{R} \theta_r' T_r(\widehat{L})X \tag{2}$$

where θ_r' represents the Chebyshev coefficient. Chebyshev polynomial $T_r(\widehat{L})$ is recursively defined as

$$T_r(\widehat{L}) = 2\widehat{L}T_{r-1}(\widehat{L}) - T_{r-2}(\widehat{L}), \tag{3}$$

$T_0 = 1$ and $T_1 = \widehat{L}$. Here $\widehat{L} = 2L/\lambda_{\max} - I_n$ is normalized to [−1, 1]. For equation (2), the work in Kipf's research [29] sets $R = 1$, $\lambda_{\max} = 2$ and adapts the network to this change. In this way, the first-order approximation of the spectrogram convolution is formed. So,

$$Y = \theta_0' X + \theta_1'(L - I_n)X = \theta_0' X + \theta_1'\left(D^{-1/2}AD^{-1/2}\right)X. \tag{4}$$

Similarly, θ_r' can also be approximated by a uniform parameter θ, that is, $\theta = \theta_0 = -\theta_1$, so that the training process can adapt to the approximation error, then

$$Y = \theta\left(I_n + D^{-1/2}AD^{-1/2}\right)X. \tag{5}$$

The computational cost is $O(|E|)$. Multiple GCN layers can be stacked to obtain advanced graph features. For simplicity, in the following sections, we set $L = I_n + D^{-1/2}AD^{-1/2}$. In general, $X \in R^{n\times C}$ and multichannel. Therefore

$$Y = LX\theta. \tag{6}$$

Yan proposed the ST-GCN method in 2018. This method uses the bone joint matrix as input to obtain the spatiotemporal features between the joint points. The method we propose takes the function of the module as the demarcation point, separately obtains the spatiotemporal features before the skeleton node, and maps into a dynamic network, abandoning the predefined graph.

3.3. Inception Structure. A huge proportion of classic scholars have demonstrated that the convolutional neural network will cause the receptive field to have the characteristics of scaling invariance [30]. Therefore, in the application process, the network should be optimized, and multi-

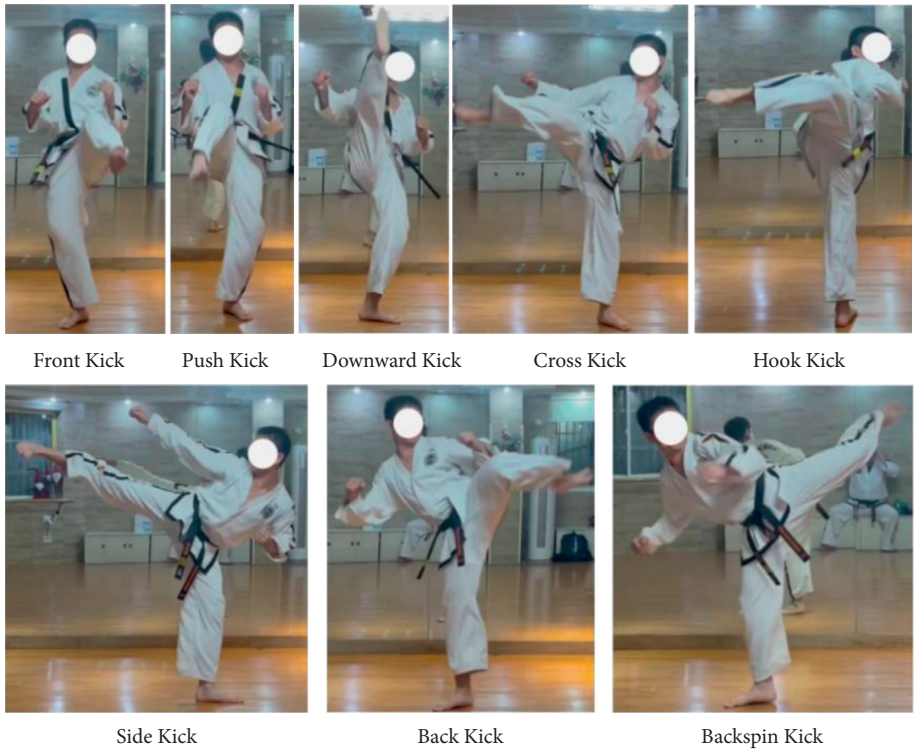

Figure 1: Regular taekwondo action.

Table 1: Taekwondo action scoring and requirements.

Score	Requirement
1	Attack with hands to the middle or high section; attack with feet to the middle section; perfect defense
2	Attack with the feet to the high section; jump into the air and hit the high section with hands (foot off the ground); jump with the feet to attack the middle section
3	Jump up and attack to the high section; jump up to 180 degrees and turn and kick to the middle section; jump up to 180 degrees or more and turn around and attack to the high section
4	Jump up 180-degree turnaround leg kick to high section; jump up 360 degrees or more turnaround leg kick to the middle section
5	Jump 360 degrees or more, turn and kick to the high section

scale filters should be appropriately combined selectively to achieve good generalization performance.

Szegedi et al. [31] found a structure, which they called inception, which can fuse convolution kernels of different scales, as shown in Figure 2. When the previous layer's output becomes the input, the extracted feature map will be shunted, and the convolution kernels of different sizes will obtain feature information of different scales. However, this method also has a disadvantage, that is, too many kernels can easily cause gradient explosion. To deal with the issue, Lin et al. [32] add one-dimensional convolution (conv1 × 1) before the original structure, followed by the activation function ReLU. In this way, the one-dimensional convolution adds the receptive fields corresponding to each other, greatly reducing the amount of calculation, and also realizing the feature fusion between different channels.

Considering that the trained data are all single-channel data, they are represented as 2D matrix in the neural network. And, each part of the human bone data that we take

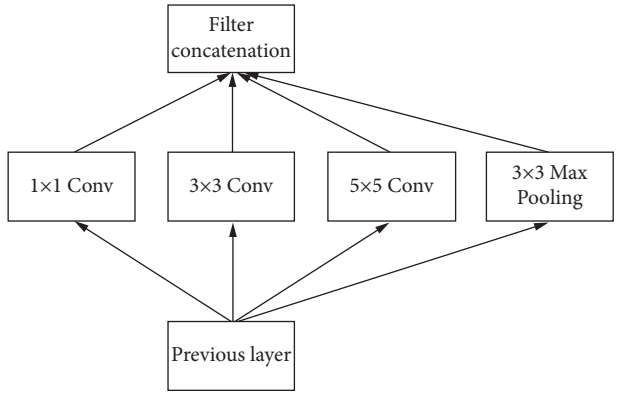

Figure 2: Inception structure.

contains three adjacent joints, forming a three-dimensional data with a column number of 3. A time series of joint coordinates is represented by every column of the matrix.

A convolution kernel with a scale of $m \times n$ has a span of scale m in the time dimension, and arranges n of each scale in a 3D sequence. Consider it a single-joint temporal convolution. From the perspective of the $m \times n$ convolution kernel, m is the time window spanning the entire time sequence, and has the following representation:

$$h^{m\times n} = w_0 p_t^k + w_1 p_{t+1}^k + \cdots + w_{m-1} p_{t+m-1}^k. \tag{7}$$

When the $m \times n$ convolution kernel is applied to extract spatial and temporal features from the action labels, the following output forms can be adopted:

$$h^{m\times n} = \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} w_{ij} p_{t+i}^{k+j}. \tag{8}$$

The pixels in the image structure have continuity, but the graph structure does not have this feature. Therefore, pooling operations cannot be performed between different joint point coordinates. Therefore, in the inception structure, the pooling operation only appears in the temporal convolution dimension, and the size of the pooling operation is $(s \times 1)$. For a single joint, its trajectory can be regarded as continuous. In this case, in the temporal dimension, the pooling process can be carried out.

3.4. Partially Aware Network. Long short-term memory neural network (LSTM) was proposed by Hochreiter and Schmidhuber [33] in 1997.

LSTM is a derivative of Recurrent Neural Network (RNN). Since 2010, it has been proven that RNN has been successfully applied to speech recognition [34], language modeling [35], and text generation [36]. However, the disappearance of gradients and explosions makes RNN difficult to apply to long-term dynamics research. As an improved network of RNN, LSTM can handle this problem well. LSTM gives the network a lot of freedom, so that the network memory unit has an adaptive solution to learn and update information, which greatly improves the performance of some perception networks.

Assume that $X = (x_1, x_2, \ldots, x_n)$ represents an input sentence composed of word representations of n words. In every position t, the RNN produces a hidden layer h in the middle denoted as y_t, and the hidden state h_t uses a nonlinear activation function to update the previously hidden layers h_{t-1} and the input x_t, as follows:

$$\begin{aligned} y_t &= \sigma\left(W_y h_t + b_y\right), \\ h_t &= f\left(h_{t-1}, x_t\right), \end{aligned} \tag{9}$$

where W_y and b_y are the parameter matrices and vectors learned during the training process, and σ represents the element-wise softmax function.

The LSTM unit includes an input gate i_t, a forget gate f_t, an output gate o_t, and a memory unit c_t to update the hidden state h_t, as follows:

$$\begin{aligned} i_t &= \sigma\left(W_i x_t + V_i h_{t-1} + b_i\right), \\ f_t &= \sigma\left(W_f x_t + V_f h_{t-1} + b_f\right), \\ o_t &= \sigma\left(W_o x_t + V_o h_{t-1} + b_o\right), \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tanh\left(W_c x_t + V_c h_{t-1} + b_c\right), \\ h_t &= o_t \odot \tanh\left(c_t\right), \end{aligned} \tag{10}$$

where $\odot$ is a kind of function which is similar to the multiplication operate, V represents a matrix related to weight, and b represents the learning vector. To increase the model's performance, morpheme training was carried out on two LSTMs. The first one is a morpheme that begins on the left and works its way to the right; the next one is a reverse duplicate of a character. Before passing to the next layer, the outputs of the forward and reverse passes are combined in series. Finally, the prediction value is observed using the activation function.

After understanding the partial perception algorithm LSTM, I was inspired by it, because in the human body recognition process, the human skeleton will be divided into multiple parts. Each part is an interconnected joint. These parts composed of joints are made by hand, for the graph convolution to be able to explore the relationship between these parts and extract the corresponding spatial features of the joint points. To obtain the information of a point in GCN, it is necessary to start from the field of that point. According to the adjacency matrix in the field, the skeleton data are automatically segmented, and then all the feedback information is input to the next joint point to complete the capture of the feature points of the entire human skeleton. Through this operation, the defects of manual design features are avoided, and the spatial features on the time series are obtained.

If an ordinary convolutional neural network is used, all parts will be merged into a whole for feature extraction of convolution operations. Partial perception networks can divide joints into different departments and capture individual features for each part. Separately extracting features in this way helps to explore the connection between parts, that is, the spatial temporal relationship between joints, as shown in Figure 3.

3.5. Network Structure. Our Taekwondo action recognition network structure is shown in Figure 4. When the athlete's bone features are input to the Part Dividing layer, all joints are grouped and merged, and each 3 joints share one part. A channel matrix is formed when all of the parts are put together, and then the three-channel spatiotemporal features are extracted from these channel matrices through the graph convolutional layer. Finally, all the extracted spatial temporal features are input into the fully connected layer, as shown in Figure 5.

To obtain the actions of the Taekwondo athletes, we stacked 7 GCN units. Each GCN unit is made up of a partial perceptual GCN layer, a BatchNorm layer, and a ReLU activation layer. The convolutional layer in each unit is filled

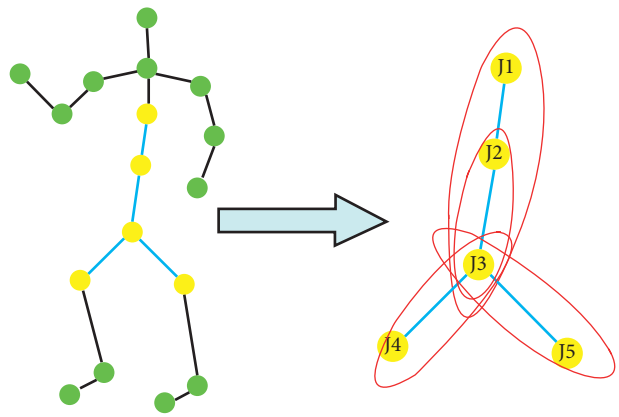

Figure 3: Spatial relationship of human skeleton joints.

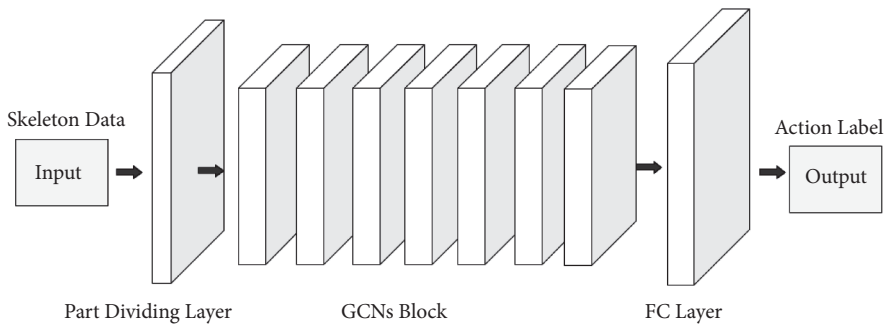

Figure 4: Global network structure.

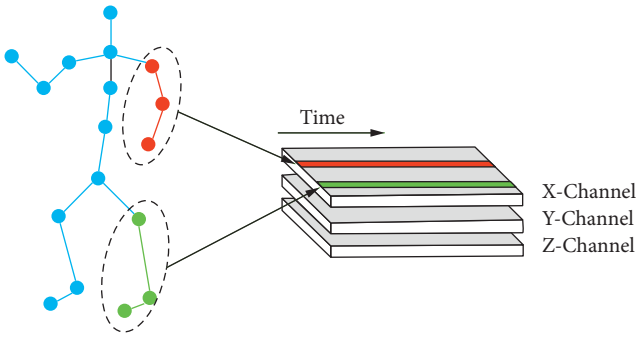

Figure 5: Part dividing layer structure.

according to requirements, with a step size of 1, as shown in Figure 6. Therefore, the size of the bone data will not change from input to output through these convolutional layers.

According to the preliminary test, the 7, 5, and 3 layers are supplemented with the initial structure. Then, set the first two layers to (3×1) filters. The third layer is the Inception structure, supplemented by (1×1), (3×1), and (3×3) filters. The fourth and sixth layers are supplemented by (3×1) filters. The fifth and seventh layers are consistent with the first layer structure. The global network structure is shown in Figure 4.

4. Experiments

4.1. Data set. Taekwondo is a kind of sports, and there is no dedicated data set for taekwondo action recognition in the world. To verify the performance of our method, we applied to the relevant departments for the data of the Chinese Taekwondo competition. In data processing, we use Openpose to preprocess the action video data. After the preprocessed data, we manually proofread the Taekwondo action, and then perform a series of tasks such as label production and action score matching. The following experiments are all based on this data set. A total of 45,000 training sets and 5,000 test sets are prepared.

4.2. Training. All experiments in this paper are executed on the Ubuntu framework, and configure python 3.7 as the language environment version. The experimental hardware environment uses RTX 2080 GPU, Intel i7-7700 CPU, 50 GB memory, and Pycharm Community 2021 as development tools. The Max-Margin criterion is used during training. It concentrates on the model's decision boundary's robustness and proposes a probabilistic-based estimating approach as an alternative. To avoid the difficulty of over-fitting during the training of the deep network, the dropout method was used in the training. The relevant training parameters are shown in Table 2:

4.3. Experimental Result. This paper proposes a method of convolutional time-space diagram to recognize Taekwondo athletes' actions. It mainly involves two parts of innovation, namely, the GCN network and part of the perception structure. In order to verify their respective effects, ablation

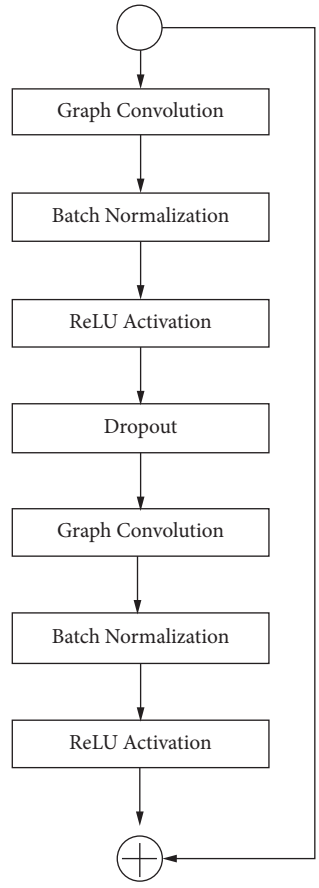

Figure 6: GCN block internal structure.

Table 2: Training parameter settings.

Parameter	Value
Epoch	70
Regularization	0.0001
Dropout rate	0.5
Initial learning rate	0.05
Hidden unit number	200
Weight attenuation coefficient	0.0005
Momentum	0.9

experiments were carried out. First, replace the network structure with a pure GCN network, named after the letter *G*, and build a GCN module performance verification experimental group. Secondly, replace the GCN module in the network with an LSTM structure and name it with the letter *L* to construct an experimental group to verify the performance of the LSTM module. Our method will be verified on a self-made Taekwondo competition data set, and the original GCN network will be compared with our method. From the level of joint recognition accuracy (Joint), bone recognition accuracy (Bone), accuracy (Acc), and parameter number (Param), the experimental results are shown in Table 3.

Table 2 shows that the GCN approach is effective, the overall accuracy is increased by 3.7%, and the number of parameters is also lowered accordingly. Through the partial perception structure method, the overall accuracy is increased by 7.5%, and the portion of parameters is lowered by half. The experimental results show that part of the perceptual structure contributes more to the overall performance improvement. Although the overall performance improvement effect of GCN is not as good as the part of the perceptual structure, it is indispensable at the level of capturing global information. The two methods contrast with each other. The experimental results demonstrate that our proposed GCN method based on partial perception structure is effective.

In order to further verify the effectiveness of this method, this paper compares four different types of bone-based action recognition methods, namely, Dynamic Skeleton [37], P-LSTM [16], TCN [38], and ST-GCN [12]. Among them, Dynamic represents hand-made action recognition methods; TCN represents CNN-based action recognition methods; P-LSTM represents RNN-based action recognition methods; and ST-GCN represents GCN-based action recognition methods. The above 4 methods and the IST-GCN method in this paper are verified on the self-made Taekwondo competition data set. The experimental results are shown in Table 4.

Table 3 shows that the verification experiment of the self-made Taekwondo competition data set yielded positive results, the GCN-based action recognition method is better than other types of action recognition methods, which proves that the graph convolutional network has a huge

Table 3: Results of ablation experiments.

Method	Joint (%)	Bone (%)	Acc (%)	Param (M)
ST-GCN	81.1	81.8	82.1	3.14
G	84.9	85.2	85.8	2.36
L	88.8	89.1	89.6	1.61
Ours	90.9	91.5	92.2	1.36

Table 4: Comparison of different types of action recognition methods.

Method	Action recognition accuracy (%)	Accurate scoring accuracy (%)
Dynamic skeleton	61.2	46.2
P-LSTM	63.9	51.3
TCN	75.3	64.1
ST-GCN	82.5	69.3
Ours	90.7	74.6

Table 5: Taekwondo competition data recognition and analysis results.

Taekwondo action	Usage frequency	Score frequency	Total score	Action recognition accuracy (%)	Score match correct rate (%)
Front kick	1845	25	53	86	69
Cross kick	866	94	186	90	74
Push kick	1566	25	49	86	71
Downward kick	180	9	12	91	77
Double flying kick	102	15	31	92	65
Single leg kick	131	299	33	93	69
Hook kick	21	2	8	88	73
Back kick	55	6	16	86	76
Backspin sick	16	0	0	95	69
Side kick	23	0	0	96	72

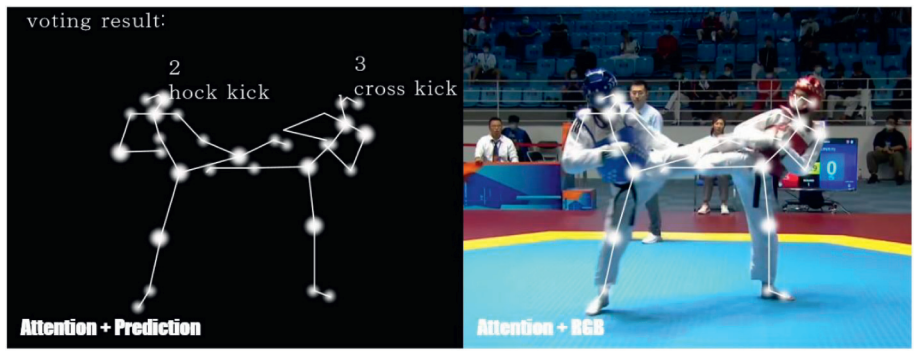

Figure 7: Action recognition and scoring effect.

advantage in the realm of action recognition. The effect of taekwondo action recognition by this method is shown in Figure 7, and the action category and corresponding score are, respectively, marked.

Through our method, all data of taekwondo competitions are identified and decomposed. The most frequently used technical actions in the game include: forward kick, side kick, push kick, down split, back kick, back spin kick, whirlwind kick, side kick, and other techniques. Perform statistical analysis on the identification data and accuracy of each technology, and match the scores. The results are shown in Table 5.

The results of taekwondo action recognition by our method are shown in Table 2. From the above data and analysis, it can be seen that in terms of the technical use of high-level Taekwondo athletes in simulated competitions in the team, the front kick and cross kick techniques are the most commonly used and best skills for athletes, whether it is the technique usage rate or the technical score rate. All of them are among the best. At the same time, rotating technology is more difficult to use and has higher requirements for specific physical fitness, so the utilization rate is generally low. The accuracy of Taekwondo technical action recognition is maintained at 80%, and the score matching

rate is relatively low. Because there are a lot of nonhuman factors in the Taekwondo action, there is a certain error in the score matching rate. In the next study, we will focus on the accuracy of the score matching rate.

5. Conclusion

In this paper, we examine the current state of work and research in the realm of action recognition before moving on to research on Taekwondo action recognition. Taekwondo is a sport, and its action recognition is an important task, which can provide a very valuable reference factor for technicians, athletes, and coaches. But action decomposition and score matching are a difficult point in this technique. This paper proposes a graph convolution framework of part of the perception structure to recognize, decompose, and analyze Taekwondo actions. Taking advantage of the long short-term memory of part of the perception structure, the recognized Taekwondo actions are marked in time series, and then features are extracted from the graph convolution level to obtain the spatial and temporal associations between joints. Predict the action category and perform score matching based on the manual tag database. In general, our method has an average accuracy of 90% in action recognition, and an average action score matching rate of 74.6%.

Because of the low matching rate of Taekwondo scores in this paper, in the future, we hope to use a dual-stream graph convolutional network to process actions, so that the secondary feature sequence of the action can be obtained, to get the second match between the action and the score, and further improve the score matching rate. Deeper picture feature extraction is also a key direction for our future research.

Data Availability

The data set can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Hunan Province (no. 2020JJ7007).

References

[1] A. Hernandez-Mendo, J. Castellano, O. Camerino, G. Jonsson, Á. Blanco-Villaseñor, and M. T. Antonio Lopes, "Anguera. observational software, data quality control and data analysis," *Revista de Psicología del deporte*, vol. 23, no. 1, pp. 111–121, 2014.

[2] D. G. Liebermann, L. Katz, M. D. Hughes, R. M. Bartlett, J. McClements, and I. M. Franks, "Advances in the application of information technology to sport performance," *Journal of Sports Sciences*, vol. 20, no. 10, pp. 755–769, 2002.

[3] U. Moenig, "La evolución de las técnicas de patada en taekwondo," *Revista de Artes Marciales Asiáticas*, vol. 6, no. 1, pp. 117–140, 2012.

[4] C. González, X. Iglesias, and M. Anguera, "Tactical moves in top level competition taekwondo combat. A descriptive study," in *Proceedings of the Scientific Congress on Martial Arts and Combat Sports*, pp. 48-49, Associaçao para o desenvolvimiento e investigaçao de Viseu Viseu, Portugal, Europe, 2011.

[5] M. Kazemi, C. Casella, and G. Perri, "2004 Olympic tae kwon do athlete profile," *Journal of the Canadian Chiropractic Association*, vol. 53, no. 2, p. 144, 2009.

[6] M. Kazemi, J. Waalen, C. Morgan, and A. R. White, "A profile of Olympic taekwondo competitors," *Journal of Sports Science & Medicine*, vol. 5, p. 114, 2006.

[7] A. Riveiro-Bozada, O. García-García, V. Serrano-Gómez, J. Antonio, L. Lopez, and A. Hernández-Mendo, "Influence the level of competition in technical actions performed point female Shiai Kumite-karate," *Cuadernos de Psicología del Deporte*, vol. 16, no. 1, pp. 51–67, 2016.

[8] M. Pic and G. K. Jonsson, "Professional boxing analysis with T-Patterns," *Physiology & Behavior*, vol. 232, Article ID 113329, 2021.

[9] R. Tarragó, X. Iglesias, D. Lapresa, M. Teresa Anguera, L. Ruiz-Sanchis, and Javier Arana, "Analysis of diachronic relationships in successful and unsuccessful behaviors by world fencing champions using three complementary techniques," *Anales de Psicología*, vol. 33, no. 3, p. 471, 2017.

[10] K. Ito, N. Hirose, M. Nakamura, N. Maekawa, and M. Tamura, "Judo kumi-te pattern and technique effectiveness shifts after the 2013 international judo federation rule revision," *Archives of Budo*, vol. 10, no. 1, 2014.

[11] C. Chen, R. Jafari, and N. Kehtarnavaz, "UTD-MHAD: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in *Proceedings of the 2015 IEEE International conference on image processing (ICIP)*, pp. 168–172, IEEE, Quebec City, Canada, 2015.

[12] L. Chaolong, C. Zhen, Z. Wenming, C. Xu, and J. Yang, "Spatio-temporal graph convolution for skeleton based action recognition," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[13] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," 2015, https://arxiv.org/abs/1511.05493.

[14] X. Gao, W. Hu, J. Tang, J. Liu, and Z. Guo, "Optimized skeleton-based action recognition via sparsified graph regression," in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 601–610, 2019.

[15] L. Quevedo Junyent, A. Padrós Blázquez, J. Solé i Fortó, and G. Cardona Torradeflot, "Entrenament perceptivocognitiu amb el Neurotracker 3D-MOT per potenciar el rendiment en tres modalitats esportives," *Apunts Educació Física i Esports*, vol. 119, no. 119, pp. 97–108, 2015.

[16] A. Shahroudy, J. Liu, T. T. Ng, and G. Wang, "Ntu rgb+ d: a large scale dataset for 3d human activity analysis[C]," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, 2016.

[17] S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, "An end-to-end spatio-temporal attention model for human action recognition from skeleton data[C]," in *Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press*, Palo Alto, CA, USA, 2017.

[18] F. Monti, D. Boscaini, J. Masci, and E. Rodolà, "Geometric deep learning on graphs and manifolds using mixture model cnns[C]," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, 2017.

[19] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in

Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press, Palo Alto, CA, USA, 2018.

[20] L. Shi, Y. Zhang, J. Cheng, and H. Lu, "Two-stream adaptive graph convolutional networks for skeleton-based action recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12026–12035, 2019.

[21] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," 2016, https://arxiv.org/abs/1611.01578.

[22] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," *Proceedings of the aaai conference on artificial intelligence*, vol. 33, no. 1, pp. 4780–4789, 2019.

[23] H. Liu, K. Simonyan, and Y. Yang, "Darts: differentiable architecture search," 2018, https://arxiv.org/abs/1806.09055.

[24] W. Peng, X. Hong, and G. Zhao, "Video action recognition via neural architecture searching," in *Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP)*, pp. 11–15, IEEE, 2019.

[25] H. Pham, M. Guan, B. Zoph, Q. Le, and J. Dean, "Efficient neural architecture search via parameters sharing," in *Proceedings of the International Conference on Machine Learning*, pp. 4095–4104, PMLR, 2018.

[26] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in Neural Information Processing Systems*, vol. 29, pp. 3844–3852, 2016.

[27] P. Veličković, G. Cucurull, A. Casanova, A. Romero, L. Pietro, and Y. Bengio, "Graph attention networks," arXiv preprint arXiv:1710.10903, 2017.

[28] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[29] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks[J]," arXiv preprint arXiv: 1609.02907, 2016.

[30] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, no. 6, pp. 1233–1258, 1987.

[31] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[32] M. Lin, Q. Chen, and S. Yan, "Network in network," arXiv preprint arXiv:1312.4400, 2013, 2013.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] O. Vinyals, S. V. Ravuri, and D. Povey, "Revisiting recurrent neural networks for robust ASR[C]," in *Proceedings of the 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 4085–4088, IEEE, 2012.

[35] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *Proceedings of the INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*, vol. 2, no. 3, pp. 1045–1048, Makuhari, Japan, 2010.

[36] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, Bellevue, Washington, USA, 2011.

[37] J. F. Hu, W. S. Zheng, J. Lai, and J. Zhang, "Jointly learning heterogeneous features for RGB-D activity recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5344–5352, 2015.

[38] T. S. Kim and A. Reiter, "Interpretable 3d human action analysis with temporal convolutional networks," in *Proceedings of the 2017 IEEE conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1623–1631, IEEE, Honolulu, HI, USA, 2017.