



Motion Analysis for Dragon Boat Athlete Using Deep Neural Networks

Chuanzhi Xu¹, Zeyong Wu², Bifeng Wu², and Yihua Tan²(✉)

¹ School of Physical Education, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People's Republic of China

² School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People's Republic of China
yhtan@hust.edu.cn

Abstract. In the training of sports, video-based motion analysis is important to automatically capture the action of trainees and provide training suggestions. Focusing on the dragon boating which mainly involves periodic rowing actions, this paper proposes a motion analysis method by comparing the action patterns of the athletes in the video and the expert athletes. First, taking a dragon boating video as input, the key points of human body of the dragon boat athlete in every frame are extracted with the deep convolutional neural network HRNET. Second, the extracted key points of athlete's body are constructed as a sequence to represent the motion of human body, and a mathematical analysis method is designed to obtain the related action parameters. Third, the parameters are compared with the standard actions of expert athletes to give the advice for action correction. Experimental results demonstrate that the proposed method can provide reliable suggestions for dragon boat athletes, which is robust to individual differences and non-uniform resolutions caused by different videos.

Keywords: Motion analysis · Deep convolutional neural network · Dragon boat · Key points of athlete's body

1 Introduction

Although current researches of athletes' training have covered different kinds of sports, the related technologies still heavily rely on manual analyses provided by professional workers. Considering the low efficiency and high cost, current manual analyses can hardly be applied in large-scale athletes' training which normally involves different levels. One direction to solve this problem is applying artificial intelligence technologies in the analyses of athlete's actions, which are designed to automatically capture the action of trainees and provide suggestions. Because videos can effectively record the actions of athletes, video-based motion analysis methods are the trend in the research of automatic action analyses.

Generally speaking, video-based motion analysis normally contains two steps. First, taking the video containing the training athletes as input, the key points of human body in

every frame are extracted, which are constructed as a sequence to represent the motion of athletes. Second, the analyses are performed on the sequence of the key points to give advice for the action corrections. However, both the two steps still have unsolved problems.

In the first step, existing methods can hardly achieve both high accuracy and efficiency when the extracting key points of human body in the scene with complex background. Two categories of the key points extraction methods are widely used in video-based motion analyses. The first category is to obtain the global features and consider the pose estimation problems as classification or regression problems [1, 2]. However, the accuracy of such methods is ordinary and this method can only solve the scenes with clean background. The second category is based on the graph structure [3], which expresses the characteristics of individual partial areas. The position of a single area is usually obtained by deformable part-based model, and the relationship between the key points of human body is optimized by considering the pairwise relationship. However, this type of method has an obvious disadvantage, that is, it is difficult to determine the topological model it depends on.

In the second step, existing methods normally analyze every action separately, then integrate them to obtain the final results. This kind of approach can hardly be adapted to the case that the variance of the key points of human body is high. The viewpoint changes lead to the key points which are extracted in a short time are normally unstable, so separately analyzing every action can hardly induce a reliable result.

Focusing on the dragon boating which mainly involves periodic rowing actions, this paper proposes a motion analysis method which can mitigate above two problems. In the first step, we use deep convolutional neural network HRNET [4] to improve the performance of the key points extraction. In the second step, we calculate the statistics for multiple periodic actions to improve the reliability of action analysis. This analysis process can extract various information about an athlete's actions, which is the base to give suggestions for the athlete. Overall, the proposed method has the advantages of high accuracy and efficiency in analyzing athletes' movements, which can reduce the time and cost in the training of dragon boating.

2 Related Work

Nowadays, motion analysis methods based on video analysis technology are mainly divided into two steps. The first step is to extract the human motion information in the video, which is mainly based on the method of extracting the key points of the human body; the second step is to analyze the extracted human action information, which are mainly based on the set rules or the action similarity.

2.1 Extraction of the Key Points of Human Body

In the field of motion analysis, the methods for the extracting key points of the human body are mainly divided into two categories. The first type of methods are to extract the key points of the human body frame by frame [5], and the second type of methods are to use the pose estimation algorithm to extract the key points of the human body.

The specific effect is shown in Fig. 1. Since the latter has greater advantages over the first method in terms of timeliness, accuracy and anti-interference, this paper only discusses the methods of the extracting key points of human body based on the pose estimation algorithm. Current pose estimation methods based on deep convolutional neural networks are mainly divided into top-down and bottom-up methods:

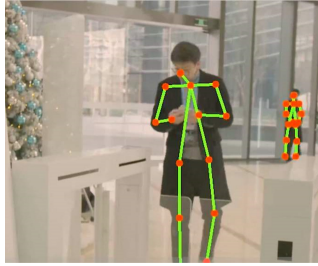


Fig. 1. Key points of human body extraction result

(1) The top-down methods use a detector to distinguish human bodies, within which extraction algorithm of the key points is performed. According to the style of the feature acquisition, there are regression [6–8] and detection [9–12] algorithms. As a typical regression algorithm, Diogo C. Luvizon [8] used specific functions to directly obtain the key points of human body from feature maps, establishing a completely differentiable framework. As for detection algorithm, the estimated heat map output by neural network is exploited to identify the points with the highest heat values as the key points, which is more robust. There are several typical algorithms, such as RMPE which is local multi-person pose estimation [10], Mask-RCNN [11], CPN that is cascaded pyramid network [12].

(2) On the other hand, the bottom-up method performs the key points detection within the whole image, then the key points are connected in terms of the relationships of connection and spatial position among them. The advantage of this method is that the detection time has no relation to the number of persons in an image. The disadvantage is that the dense key points are difficult to be distinguished whether they belong to one person or the another. Some representative algorithms include OpenPose proposed by Zhe Cao [13] and DeepCut proposed by Leonid Pishchulin [14].

2.2 Motion Analysis

When the action information of the persons in the video is obtained, the motion information can be analyzed. Motion analysis mainly solves two problems. One is the recognition of the type of action. Jiang et al. proposed a deep neural network model based on ResNeXt to recognize human actions in video [15]. The other one is to determine how correct the action is when comparing with the standard motion. There are rule-based and action analysis algorithms.

According to the rule-based action analysis methods, Li et al. used a cascaded convolutional neural network to extract facial feature points, and calculates various angle

parameters of head movement through the extracted feature points, then the corresponding angle value is compared with the set threshold to determine if it is an abnormal behavior [16]. Zhu et al. applied OpenPose to extract the key points of the human body, so as to obtain the position of the foot joint points, and obtain the distance from the specified safe position to determine whether the person is in safe position [17].

On the other hand, based on action similarity, Li et al. collected the badminton player’s swing action in the video by the badminton robot’s vision system, and compared it with the standard swing motion to analyze and evaluate the motion in the video [18]. Ji used the OpenPose algorithm to extract golf swing action parameters in the video, and then compared them with the professional golf swing action parameters to analyze and evaluate the standard degree of golf swing action in the video [19].

3 Method

The proposed method includes two steps: (1) convolutional neural network is applied to extract the key points of athlete’s body corresponding to the joints, such as elbow, wrist, shoulder, etc., (2) the key points are connected to analyze the motion information of athlete which could be further fed back to the coach for the dragon boat training.

3.1 Extraction of Key Points Using HRNET

Since the accurate extraction of the key points of athlete’s body is very crucial for the following motion analysis, we apply high-resolution network (HRNET [4]) to conduct the task. The main reason is that the model structure of HRNET output high resolution feature representation such that the locations of points are more accurate than those recovered from low resolution feature. The overall model structure is shown in Fig. 2. From the structure, we can imply that multi-resolution paths extract rich features as well as accurate locations of points.

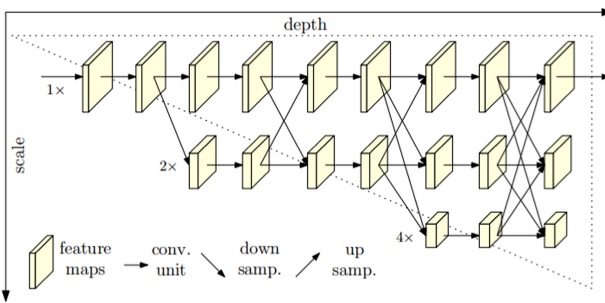


Fig. 2. Model structure of HRNET [4]

In the implementation of HRNET, we follow the design rules of ResNet so that the number of channels of each resolution is refined, and the depth information is embedded to all steps. Finally, a small net and a large net are set up with different parameters:

HRNET-W32 and HRNET-W48, where 32 and 48 represent the width of the high-resolution subnets in the last three steps. The widths of the three parallel subnets in the small network are: 64, 128, and 256, and the widths of the three parallel subnets in the large network are: 96, 192, and 384.

With the trained model, the key points of dragon boat athlete can be extracted from video. An extraction example is shown in Fig. 3. It can be seen that the key points of the human body in Fig. 3 have been clearly marked as blue dots corresponding to the joint location.

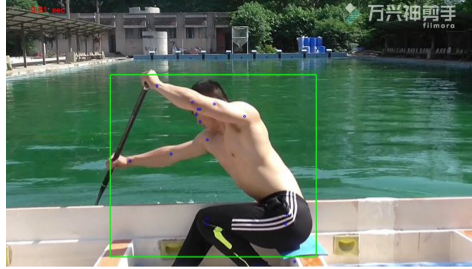


Fig. 3. Key points extracted from dragon boat athlete (blue dots). (Color figure online)

3.2 Extraction of Athlete Motion Information

For the rowing action in dragon boat, we mainly pay attention to the movement of the wrist, elbow, shoulder and hip joints. Therefore, we define the following angles to describe the rowing action:

- (1) Paddle angle: it refers to the acute angle between the paddle and the horizontal plane when the paddle is inserted into or taken out of water. Paddle angle can be calculated by connecting the two wrist joints;
- (2) Shoulder joint angle: it refers to the angle between the upper arm and the trunk;
- (3) Elbow joint angle: it refers to the angle between the forearm and the upper arm;
- (4) Hip joint angle: it refers to the angle between the trunk and the thigh.

These angles extracted from video are combined to characterize the details of row action, which can be used to judge the motion level of dragon boating athletes.

Basically, rowing action is a periodical motion in which the coordinates of the key point corresponding to a joint form a periodical wave. According to the observation that the location extent of a joint is limited, we can set the time interval between two neighboring maximum locations of the same key point as a cycle. In the following section, we give the specific procedure of finding the cycle of a motion, which is shown as Fig. 4.

Because each key point extracted from HRNET corresponds to a specific joint, we group all the specific key points in time series as an analysis unit. Among the key points in the series, the average number of frames between adjacent points with the maximum

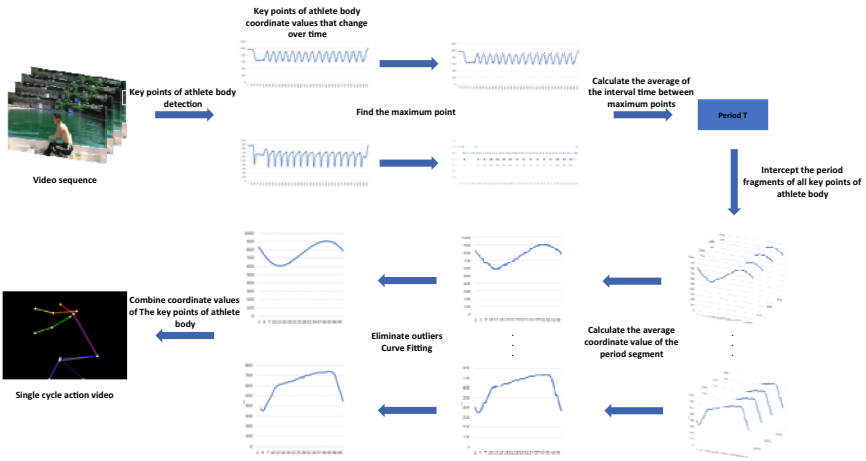


Fig. 4. Scheme flow of the motion information extraction

coordinates is calculated as the average period T of the periodic repetitive actions. By summing and averaging the coordinates of the corresponding points of all the periods, the average curve of the athlete’s joint points in a cycle can be obtained. Finally, the corresponding joint angle could be obtained from the cycle information of each key point.

The specific details are as follows:

First, it is necessary to find the maximum points through the periodic law of the coordinates of the key points, and the conditions of the maximum points satisfy Eq. (1):

$$(f_{k,i}, y_{f_{k,i}}^k) = \max\{y_{f_{k,i}-r}^k, y_{f_{k,i}-r+1}^k, \dots, y_{f_{k,i}+r}^k\} \quad (1)$$

where $f_{k,i}$ represents the number of frames corresponding to the maximum value of i^{th} coordinate of the key point k in the video sequence, y_f^k represents the coordinate value of the key point k in f^{th} frame, and r represents the search range of the number of frames (It is set to 30 in the experiment, that is, the search range is $-30-30$, which needs to be less than the estimated motion period).

After all the maximum points are counted, the average period of each repeated motion in the dragon boating can be obtained by taking the difference between two adjacent maximum points and taking the average, as shown in Eq. (2).

$$T = \frac{1}{K(N-1)} \sum_{k=1}^K \sum_{i=1}^{N-1} (f_{k,i+1} - f_{k,i}) \quad (2)$$

where T represents the period, K represents the number of the key points, N represents the number of maximum points, and $f_{k,i}$ represents the number of frames corresponding to the key point k at i^{th} coordinate maximum in the video sequence.

The Cycle Average Action can be obtained by adding the corresponding points in each cycle segment and taking the average, as shown in Eq. (3)

$$CAA_j^k = \frac{1}{N} \sum_{i=0}^{N-1} y_{f_0+T \cdot i+j}^k \quad (3)$$

where CAA_j^k represents the coordinate value of the key point k in j^{th} frame in a single cycle after averaging, y_f^k represents the coordinate value of the key point k in f^{th} frame, and f_0 represents the initial number of frames in the first cycle (Set to 207 in experiment).

After obtaining the required average coordinate value of each key point, the four kinds of angle information required by the paper can be obtained through some angle transformations.

3.3 Motion Transformation Evaluation

After obtaining the required four joint angles, we can evaluate the motion of dragon boat athlete by comparing it with the angle information of the standard action. The angles difference guides the final suggestion, as shown in Eq. (4):

$$\partial_i = \beta_i - \mu_i \quad (4)$$

where β is the angle information in the standard action, μ is the angle information of athlete, ∂ represents the angle difference and i represents the serial number of a certain angle ($i = 1-4$) respectively represent paddle, shoulder, elbow and the hip angles). When $\partial > 5^\circ$ indicates that a certain angle of the athlete is too large during exercise, attention should be paid to reduce the angle appropriately. When $\partial < -5^\circ$ indicates that a certain angle of the athlete is too small during exercise, and the range of this part of the movement needs to be increased, when $-5^\circ < \partial < 5^\circ$ means that a certain angle of the athlete's movement is quite standard.

4 Experiment

4.1 Experimental Results

This paper analyzes the specific actions of the characters through the input of the dragon boat rowing video sequence. First, by key point detection, the coordinates of the key points of athlete's body in each frame are obtained, and the periodic curve of the key points changing with time can be drawn as Fig. 5.

Since the rowing cycle of the dragon boat action is stable, the period fragments of all related joints can be obtained and further fit with polynomial curve after part of the abnormal values are removed, which are shown in Fig. 6.

By the obtained average coordinate values of the key points of human body, a video of the key points of human body motion in a single cycle is obtained, and the representative rowing motion is grouped as shown in Fig. 7.

The average angle change curves of the four angles are also finally drawn as shown in Fig. 8:

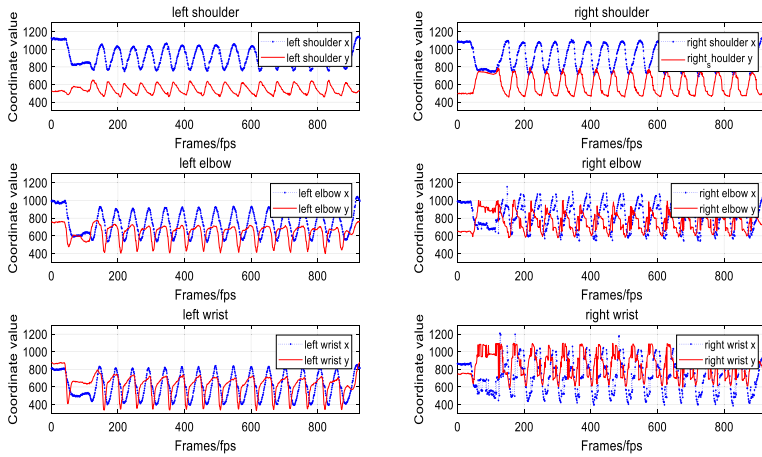


Fig. 5. Coordinate change curve of the key points of athlete

With the above analysis results, an example of suggestions can be drawn as follows:

- (1) The change range of the shoulder angle when the dragon boat athlete is in the beginning of paddle is $100^{\circ}\text{--}120^{\circ}$, and the normal range of shoulder joint angle change should be $120^{\circ}\text{--}130^{\circ}$. We can see that $\partial_2 = 10^{\circ}\text{--}20^{\circ} > 5^{\circ}$. It means that the athlete’s shoulders are not extended enough when rowing, resulting in the athlete’s failure to obtain a reasonable insertion point;
- (2) The dragon boat athlete’s elbow angle is 178° before entering the water. The normal range should be $165 \pm 5^{\circ}$. We can see that $\partial_3 = 13^{\circ} > 5^{\circ}$. It means that the elbow joint angle is too large and the arm is too stretched away and too stiff. Therefore, the elbow joint angle should be slightly reduced;
- (3) The hip angle is 51° when the athlete inserts the Paddle. The normal range should be 30° . We can see that $\partial_4 = 21^{\circ} > 5^{\circ}$. It means that the body is too stretched and the lift is too large. It should be controlled at about 30° to obtain a reasonable insertion point.

4.2 Comparison Experiment

The method used in this paper has significant advantages in accuracy and efficiency. Compared with the traditional method of analyzing human movements, this method can also obtain better and more accurate human motion information. We compared the results of the method proposed in [5]. For simplicity, only one cycle is measured during the experiment. The changes of the hip angle are compared, and the specific experimental results are shown in Fig. 9:

It can be seen from Fig. 9 that the hip angle change curve obtained by the analysis method in [5] and that of our proposed method are very similar, but our proposed method have two advantages of the method in [5]:

- (1) Time complexity is low. The method in [5] needs to analyze the input video frame by frame slowly, and the time cost is too high. However, the method proposed in this

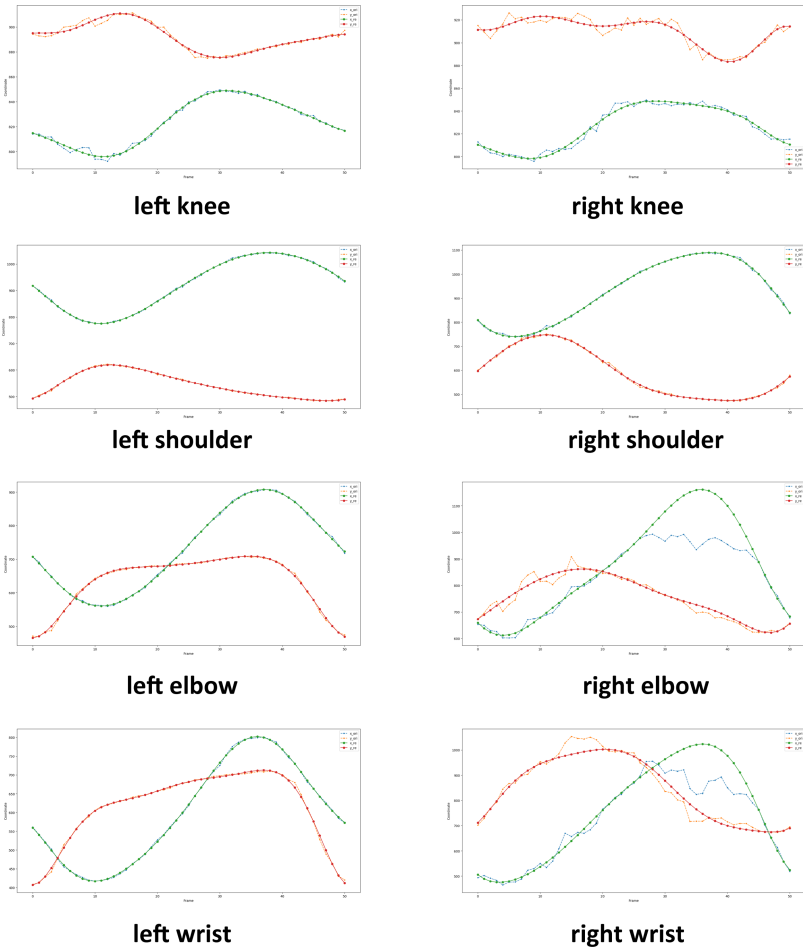


Fig. 6. Comparison of joint curves before and after fitting of each key point

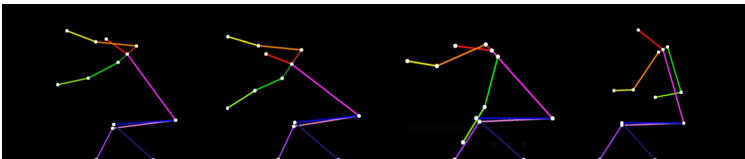


Fig. 7. Rowing sequence represented by the key points of athlete

paper relies purely on algorithm analysis, which can automatically and quickly obtain the required various angles. Generally speaking, using the method in [5] to process a video sequence of an athlete’s periodic actions will take nearly 40 min. However, using the method proposed in this paper will only take nearly 10 s.

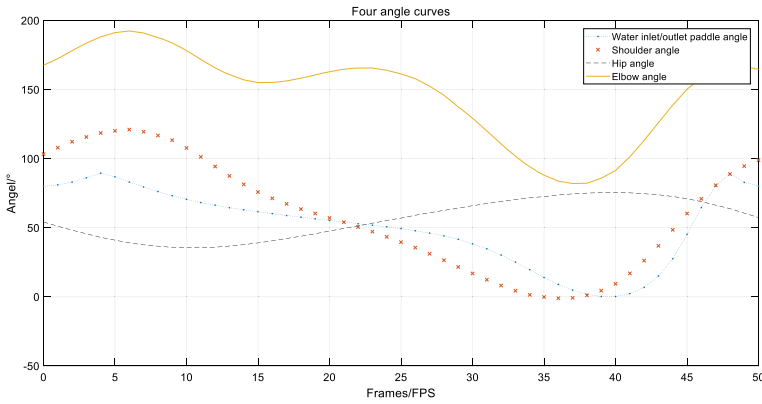


Fig. 8. Four-angle change curves

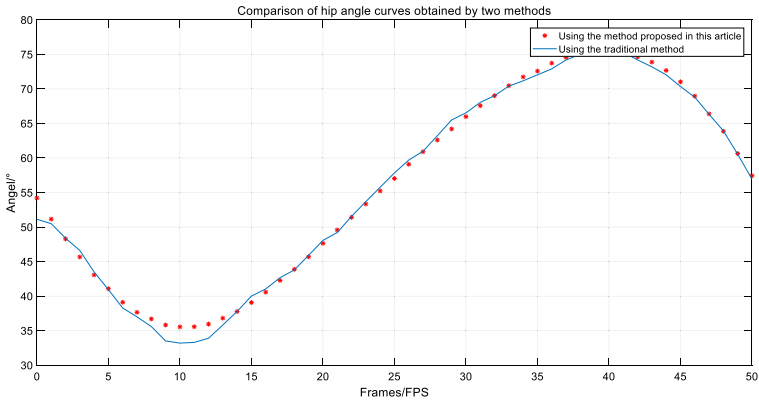


Fig. 9. Curves of hip joint angle changes extracted by two methods

(2) Accuracy is better. The method in [5] needs to manually detect the key angles of the characters in the video, and there are many errors in the measurement process. The method proposed in this paper first uses the posture estimation algorithm to obtain the reliable key points of human body, and then performs some coordinate transformations on the key points coordinates to obtain the required angle. There is no need to measure in the middle, and the accuracy of the experimental results is high.

Therefore, in general, the method of analyzing the rowing movement of dragon boat athletes proposed in this paper has obvious advantages in time cost and measurement accuracy compared with the traditional method of analyzing sports movements.

5 Conclusion

This paper mainly discusses the extraction of the key points of athlete's body and the analysis of dragon boat rowing movements. After obtaining the reliable key points of athlete's body, the average period of each repetitive movement is extracted, which is further

used to analyze the motion parameters represented as four joint angles. Finally, the joint parameters are compared with standard action to give some improvement suggestions. The experiments of the action analysis method proposed in this paper demonstrate that it can qualitatively evaluate the action of dragon boat athlete.

In future, there are two directions can be further explored:

First, the optimal design of the key points of human body detection network is very important. To reduce the complexity of training and testing, it is necessary to design lightweight neural networks.

Second, this study only studies motion parameters of paddle, elbow, shoulder and hip angles, so that more complicated motion description framework needs to be studied to further analyze the movement rhythm and the speed of each joint from multiple angles.

References

1. Rogez, G., Rihan, J., Ramalingam, S., et al.: Randomized trees for human pose detection. In: IEEE Conference on Computer Vision & Pattern Recognition, pp. 1–8 (2008)
2. Urtasun, R., Darrell, T.: Sparse probabilistic regression for activity-independent human pose inference. In: IEEE Conference on Computer Vision & Pattern Recognition, pp. 149–156 (2008)
3. Hu, G.: Research of Human Pose Estimation Based on Pictorial Structure Models. Wuhan University of Technology (2014)
4. Sun, K., Xiao, B., Liu, D., et al.: Deep high-resolution representation learning for human pose estimation. IEEE Conference on Computer Vision & Pattern Recognition, pp. 5686–5696 (2019)
5. Qiong-zhu, G., Feng, C.: Sports biomechanical analysis of landing stability of maneuvers in women's Wushu. *J. Wuhan Inst. Phys. Educ.* **44**(6), 48–52 (2010)
6. Sun, X., Xiao, B., Wei, F., et al.: Integral Human Pose Regression. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*, vol. 11210, pp. 536–553. Springer, Cham (2018)
7. Newell, A., Yang, K., Deng, J.: Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*, vol. 9912, pp. 483–499. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46484-8_29
8. Luvizon, D.C., Tabia, H., Picard, D.: Human pose regression by combining indirect part detection and contextual information. *Comput. Graph.* **85**, 15–22 (2017)
9. Li, S., Liu, Z.Q., Chan, A.B.: Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network. *Int. J. Comput. Vis.* **113**(1), 19–36 (2015)
10. Fang, H.S., Xie, S., Tai, Y.W., et al.: RMPE: regional multi-person pose estimation. In: IEEE International Conference on Computer Vision, pp. 2353–2362 (2017)
11. He, K., Georgia, G., Piotr, D., et al.: Mask R-CNN. In: IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
12. Chen, Y., Wang, Z., Peng, Y., et al.: Cascaded pyramid network for multi-person pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 7103–7112 (2017)
13. Cao, Z., Simon, T., Wei, S.E., et al.: Realtime multi-person 2D pose estimation using part affinity fields. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1302–1310 (2017)

14. Pishchulin, L., Insafutdinov, E. Tang, S., et al.: DeepCut: joint subset partition and labeling for multi person pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4929–4937 (2016)
15. Sheng-nan, J., En-qing, C., Ming-yao, Z., et al.: Human action recognition based on ResNeXt. *J. Graph.* **41**(2), 277–282 (2020)
16. Li, Y.: Abnormal Behavior Detection Based on Head Movement Analyze in Examination Room. Changchun University of Science and Technology (2018)
17. Zhu Jian-bao, X., Zhi-long, SY.-w., et al.: Detection of dangerous behaviors in power stations based on OpenPose multi-person attitude recognition. *Autom. Instrum.* **35**(2), 47–51 (2020)
18. Li, K.: Capture, Recognition and Analysis of Badminton Players Swing. University of Electronic Science and Technology of China (2020)
19. Ji, Y.: The Research on Golf Swing Action Comparison Based on Video Human Body Pose Estimation. Nanjing University of Posts and Telecommunications (2019)