Badminton Action Classification Based on PDDRNet

Xian-Wei Zhou(✉), Le Ruan, Song-Sen Yu, Jian Lai, Zheng-Feng LI, and Wei-Tao Chen

School of Software, South China Normal University, Foshan, China
{zhouxianwei,2020023863,yusongsen,2021024160,2020023841, 2021024129}@m.scnu.edu.cn

Abstract. Badminton is one of the most popular sports nowadays. To assist badminton teaching, a two-stage badminton movement classification method based on PDDRNet is proposed in this paper. In the first stage, the PDDRNet model for human pose estimation is trained using the knowledge distillation architecture of the teacher student network, the student network uses the lightweight model SECANet, while SimCC is simultaneously applied to replace the heatmap for representation. In the second stage, the estimated poses from the first stage are used for feature engineering, and XGBOOST is applied to classify the underlying badminton movements. In order to verify the performance of our proposed algorithm, we leverage the MPII datasets for human pose estimation experiments, and a proprietary badminton movement dataset for badminton movement classification. The results show that on the MPII dataset, it achieves a 3.1% improvement in PCKh when compared to lite-HRNet. In the second stage, the accuracy of the badminton movement classification algorithm using XGBOOST reaches 93.5% , which is 7.60% higher than the KNNbased badminton movement classification method.

Keywords: badminton action classification · human pose estimation · model lightweight · knowledge distillation

1 Introduction

In recent years, with the development of artificial intelligence technology, human pose estimation has begun to be applied to competitive sports video analysis and teaching, such as diving, Table tennis. Using artificial intelligence to recognize badminton movements is more difficult and has the value of in-depth research.

In order to assist badminton teaching, a lightweight two-stage badminton action classification method based on human body pose estimation is proposed in this paper. In the first stage, a lightweight human pose estimation model-poodenr is proposed; in the second stage, badminton actions are classified on the badminton sports data set based on PDDRNet (Pose Distillation model based on Disentangled Representation). We summarise our contributions in follows:

D. Kumar et al. (Eds.): IEIT 2023, AHSSEH 10, pp. 980–987, 2023.
https://doi.org/10.2991/978-94-6463-230-9_118

(1) We proposes a pose distillation model based on disentangled representation. The teacher network uses HRNET, and the student network adopts a lightweight network SECANet, which adds Sandglass and ECA modules, and uses SimCC [1] for characterization. Finally, comparative experiments and analysis were carried out on the MPII data set. The experimental results show that the parameters of this method are reduced by 14.9 M compared with HRNet-W32, and compared with Lite-HRNet, the average detection accuracy on MPII is increased by 3.1%.
(2) We presents a method for classifying badminton moves based on PDDRNet pose estimation results. Then calculate the relative distance, height difference, joint angle and other information, and use XGBOOST to classify badminton movements. The accuracy rate reached 93.5% on the self-made badminton action classification data set.

2 Related Work

There are two main methods for classification of badminton movements using computer methods. One is the sensor-based motion recognition algorithm, the algorithm first obtains the athlete's motion data through the acceleration sensor and gyroscope, and then use KNN [2], HMM [3], SVM [4], KNN [5] for action classification. This type of method only needs the data of sensor equipment, and the method is simple and reliable, but athletes need to wear special sensor equipment, which is often difficult to meet. The other is a vision-based badminton action classification algorithm. This type of method mostly uses CNN [6] or LSTMmodels to extract and classify image or video features, and the detection accuracy and model generalization ability are stronger than sensor-based methods.

We propose a badminton action classification method based on human pose estimation. Since the proposal of DeepPose [7], the research on 2D human pose estimation has developed rapidly. Networks such as CPM [8] and Multi-Context Attention [9] focus on solving the key point occlusion problem. Hourglass, HRNet [10] proposed a new skeleton network, which greatly improved the detection accuracy. Model such as FPD [11], lite-hrnet use lightweight networks or modules to reduce the amount of network parameters and make the model more lightweight.

3 Methods

3.1 PDDRNet Network Module

The Structure of PDDRNet. PDDRNet adopts the knowledge distillation teacher-student network structure, the teacher network uses HRNet, and the student network designs a lightweight attitude estimation method SECANet. At the same time, SIMCC is used instead of heat map for coordinate representation. The specific structure is shown in Fig. 1.

SECANet integrates Sandglass module and ECA module on the basis of HRNet to make the model more lightweight. SECANet changes the Bottleneck in HRNet to SECANeck, and changes the Basic module to the SECAblock module. The SECAneck module consists of two convolutions, one Sandglass, and one ECA module, as shown in Fig. 2-a. The SECAblock module consists of two Sandglass modules and one ECA module, as shown in Fig. 2-b. Compared with the Bottleneck and Basic modules, on the basis of retaining the original residual structure, Sandglass is used to replace the 3 × 3 convolution in the original model, and the feature extraction ability of the model is guaranteed through the ECA lightweight attention module.

Model Representation Part. HRNet uses a heatmap-based representation method, but the heatmap-based method requires additional post-processing, which will bring new errors while introducing new calculations. Therefore, this paper uses the space-aware decoupling representation method SimCC for coordinate encoding. For the input image, the and coordinates of its key points are represented by two independent one-dimensional

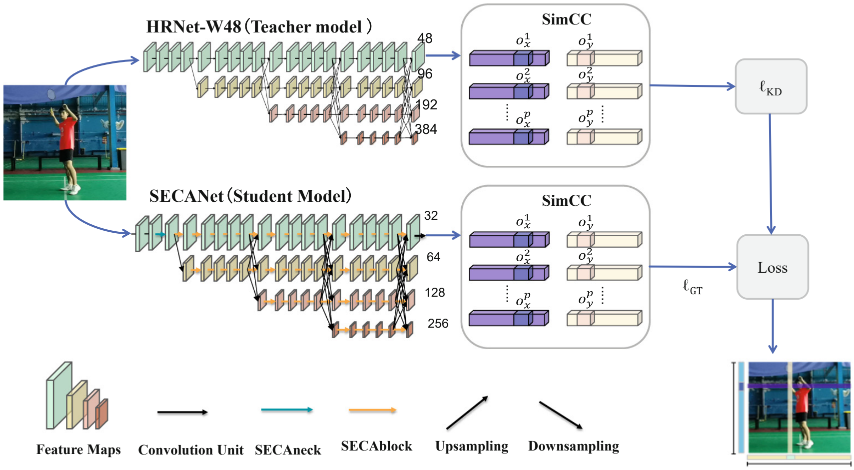

Fig. 1. PDDRNet Network framework

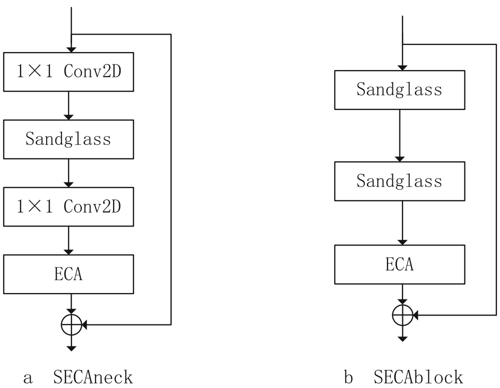

Fig. 2. SECAblock module and SECAneck module structure

vectors. For the th key point P, its coded coordinates will be expressed as Eq. (1):

$$P' = (x', y') = (round(x^p \times k), round(y^p \times k)) \tag{1}$$

Among them, the function of the scaling factor is to enhance the positioning accuracy to the sub-pixel level. During the coordinate decoding process, the model output is two one-dimensional vectors, and the coordinate calculation method of the predicted point is shown in Eq. (2) and Eq. (3):

$$\overline{o}_x = \frac{\arg\max_i(o_x(i))}{k} \tag{2}$$

$$\overline{o}_y = \frac{\arg\max_j(o_y(j))}{k} \tag{3}$$

Through this method, the response maximum point on the one-dimensional vector can be divided by the scaling factor to restore the original image scale.

3.2 Badminton Action Classification Module

In order to facilitate deployment on edge devices such as mobile terminals, this paper uses XGBOOST for the second stage of badminton movement classification. Specifically, after obtaining the coordinates of the key points of the human body in the first stage, the construction of feature engineering will be carried out in the second stage. Information such as the height difference, relative distance and included angle between each key point will be calculated, and the specific calculation methods are shown in Eq. (4), Eq. (5), and Eq. (6):

$$\mathrm{H(i, j)} = \mathrm{y}_j - y_i \tag{4}$$

$$\mathrm{L(i, j)} = \sqrt{(\mathrm{y}_j - y_i)^2 + (x_j - x_i)^2} \tag{5}$$

$$A(\vec{a}, \vec{b}) = \arccos \frac{\vec{a} \times \vec{b}}{|\vec{a}|\left|\vec{b}\right|} \tag{6}$$

Among them, $\vec{a}, \vec{b}$ is calculated by f Eq. (7) and Eq. (8).

$$\vec{a} = \left(x_j - x_i, y_j - y_i\right) \tag{7}$$

$$\vec{b} = \left(x_k - x_j, y_k - y_j\right) \tag{8}$$

Among them, $i, j, k = \{1, 2, ..., n\}$ represents the total number of joint points, which is 16 in this project. A total of 16 key points can be obtained through human body pose estimation, and (x_i, y_i) represents the position of the key point pose estimation.

4 Experimental Evaluation

4.1 Experimental Setup

Human Pose Estimation Datasets and Evaluation Metrics. We were used MPII dataset for human pose estimation. Which contains about 25k pictures, and there are about 40k human objects in total. Each human object contains the information of 16 joint points. The evaluation index of the MPII data set is PCKh.

Badminton Action Classification Database. We collect badminton action image data sets in the second stage. Taking the high and long forehand as an example, as shown in Fig. 3, the professional badminton coach's actions are divided into four parts. Image acquisition is performed for each small action, and 350 images of each small action are obtained. A total of 1400 pictures are divided into training set and test set according to the ratio of 7:3.

4.2 Human Pose Estimation Experiment

In order to verify the performance of PDDRNet, compare PDDRNet with the common models in recent years. The input pixel size of the experimental image is 256*256, and the corresponding experimental results on the MPII data set are shown in Table 1.

It can be seen from Table 3-5 that the PDDRNet network parameters proposed in this chapter are 13.6 M, and the average accuracy rate in PCKh is 89.8%. When the performance is similar to the mainstream model, the number of parameters is greatly reduced, and compared with the lightweight human pose estimation algorithm Lite-HRNet, the average accuracy rate is increased by 2.8%.

It can be seen from Table 2 that due to the use of SimCC for characterization, PDDRNet has higher superiority in small-size input images. Compared with 256*256 size images, PDDRNet performance is improved by 3.4% in 64*64 size images, The performance improvement is even greater.

Table 1. Comparison of MPII performance between PDDRNet and mainstream networks

Method	#Params	Head	Sho.	Elbo.	Wri.	Hip	Knee	Ank.	Mean
CPM	27.0 M	97.8	95.0	88.7	84.0	88.4	82.8	79.4	88.5
Multi-Context Attention	–	98.5	96.3	91.9	88.1	90.6	88.0	85.0	91.5
SimpleBaseline	34.0 M	98.5	96.6	91.9	87.6	91.1	88.1	84.1	91.5
HRNet	28.5 M	98.6	97.0	93.0	89.2	91.5	89.0	85.7	92.3
FPD	3 M	98.3	96.4	91.5	87.4	90.9	97.1	83.7	91.1
OKDHP	24.7 M	98.2	96.6	92.3	88.0	91.0	88.5	84.5	91.7
Lite-HRNet	1.8 M	–	–	–	–	–	–	–	87.0
PDDRNet	13.6 M	97.3	96.2	90.4	85.9	89.3	86.1	82.1	90.1

Table 2. Performance comparison under different image sizes

Size	Model	PCKh
256*256	HRNET	92.3
256*256	PDDRNET	90.1
64*64	HRNET	79.4
64*64	PDDRNET	79.8

Table 3. Ablation experiment configuration and results

SECANet	KD	SimCC	Head	Sho.	Elbo.	Wri.	Hip	Knee	Ank.	Mean
			93.6	89.6	77.9	68.3	81.2	72.8	66.7	79.4
√			90.8	86.8	74.8	65.8	77.6	69.2	62.9	76.4
√	√		93.3	89.3	78.1	67.4	79.8	70.7	65.3	78.6
√		√	93.3	89.3	78.1	67.4	79.8	70.7	65.3	79.6
√	√	√	93.6	89.6	78.7	69.3	80.8	72.7	68.1	79.8

Ablation Experiment. In order to evaluate the effectiveness of the main components of PDDRNet proposed in this paper in human pose estimation, an ablation experiment is carried out on the MPII dataset. This part contains a total of four comparative experiments, and the experimental results are shown in Table 3.

Through comparison, it is found that on the basis of SECANet, using SimCC or knowledge distillation methods, the PCKh detection index has been improved to varying degrees, but using the PDDRNet network can increase by 3.4%, which is higher than the improvement brought by only using any one of the methods, so it is reasonable to use the PDDRNet network.

4.3 Badminton Action Classification Experiment

On the badminton action classification data set, HRNet and PDDRNet are used to conduct action classification experiments respectively. The input image size is 64*64. The experimental results are shown in Table 4.

It can be seen that the algorithm based on PDDRNet is more accurate than the algorithm based on HRNet. And in the action classification based on the PDDRNet algorithm, compared with the KNN and random forest methods, XGBOOST has the highest classification accuracy rate, reaching 93.50%.

The experimental results show that the PDDRNet + XGBOOST badminton action classification method proposed in this paper has good performance and practical value.

Table 4. Performance comparison of machine learning algorithms for badminton action classification

classification method	KNN		Random Forest		XGBOOST	
Pose Estimation Model	HRNET	PDDRNET	HRNET	PDDRNET	HRNET	PDDRNET
Action Classification Accuracy	82.13	85.90	76.94	77.61	92.35	**93.50**

5 Conclusion

We proposes a two-stage badminton action classification method based on PDDRNet + XGBOOST in this paper. In the first stage, a lightweight human body pose estimation method PDDRNet is proposed, which is trained by knowledge distillation. The teacher network is HRNet, and the student network proposes SECANet. SECANet replaces the BASIC module in HRNET with the lightweight SECAblock, and replaces the Bottleneck module with the SECAneck module. SimCC is used for characterization in this paper. In the second stage, feature engineering was constructed based on the results of the first stage, and the XGBOOST method was used to classify badminton movements on the self-made badminton movement classification dataset. This method performed excellently, and the classification accuracy rate reached the best 93.5%.

However, it is also observed in the experiment that the performance of this model is lacking in rare badminton postures. We will increase the badminton data set and improve the detection accuracy in the future.

References

1. Li, Y., Yang, S., Liu, P., Zhang, S., Wang, Y., Wang, Z., ... & Xia, S. T. (2022, November). SimCC: A Simple Coordinate Classification Perspective for Human Pose Estimation. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI (pp. 89–106). Cham: Springer Nature Switzerland.
2. Steels, T., Van Herbruggen, B., Fontaine, J., De Pessemier, T., Plets, D., & De Poorter, E. (2020). Badminton activity recognition using accelerometer data. Sensors, 20(17), 4685. https://doi.org/10.3390/s20174685
3. Anik, M. A. I., Hassan, M., Mahmud, H., & Hasan, M. K. (2016, December). Activity recognition of a badminton game through accelerometer and gyroscope. In 2016 19th Inter-national Conference on Computer and Information Technology (ICCIT) (pp. 213–217). IEEE. https://doi.org/10.1109/ICCITECHN.2016.7860197
4. Wang, Z., Guo, M., & Zhao, C. (2016). Badminton stroke recognition based on body sensor networks. IEEE Transactions on Human-Machine Systems, 46(5), 769–775. https://doi.org/10.1109/THMS.2016.2571265
5. Bulat, A., Kossaifi, J., Tzimiropoulos, G., & Pantic, M. (2020, November). Toward fast and accurate human pose estimation via soft-gated skip connections. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (pp. 8–15). IEEE. https://doi.org/10.1109/FG47880.2020.00014

6. binti Rahmad, N. A., binti Sufri, N. A. J., bin As' ari, M. A., & binti Azaman, A. (2019). Recognition of badminton action using convolutional neural network. Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 7(4), 750–756. https://doi.org/10.52549/ijeei.v7i4.968
7. Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653–1660). https://doi.org/10.1109/CVPR.2014.214
8. Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 4724–4732). https://doi.org/10.1109/CVPR.2016.511
9. Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A. L., & Wang, X. (2017). Multi-context attention for human pose estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1831–1840). https://doi.org/10.1109/CVPR.2017.601
10. Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5693–5703). https://doi.org/10.1109/CVPR.2019.00584
11. Zhang, F., Zhu, X., & Ye, M. (2019). Fast human pose estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3517–3526). https://doi.org/10.1109/CVPR.2019.00363