Hindawi
Scientific Programming
Volume 2022, Article ID 4664388, 10 pages
https://doi.org/10.1155/2022/4664388

*Research Article*

# Research on Badminton Teaching Technology Based on Human Pose Estimation Algorithm

**Zhang Xipeng**,[1] **Zhao Peng**,[1] **and Cao Yecheng**[2]

[1]*Qingdao Huanghai University, Qingdao 266000, Shandong, China*
[2]*Shijiazhuang University, Shijiazhuang 050035, Hebei, China*

Correspondence should be addressed to Zhang Xipeng; zhangxp@qdhhc.edu.cn

Received 28 January 2022; Revised 15 February 2022; Accepted 28 February 2022; Published 24 March 2022

Academic Editor: Baiyuan Ding



Human pose estimation is an important task in physical education, which can provide a valuable reference for teachers and students. We propose a human pose estimation method based on part affinity field. Firstly, the correlation of position information and orientation information between limb regions is maintained by part affinity field. Then the key points of limb pose are localized by part confidence map, and finally, the part affinity field is integrated to correlate all the acquired feature key points to obtain the human pose estimation. With the aid of computer vision technology, the students' training movements can be compared with the standard movements. It enables the students to feel the standard movements and badminton hitting points more intuitively. In the experiment, we set up a comparison experiment to compare the teaching mode of the method in this paper with the traditional teaching mode. The experimental results prove that through the teaching mode of our method, students have more standard strokes, more smooth skill switching between badminton serves and strokes, and higher badminton stroke scores. At the same time, such a teaching system adds a lot of fun to the course and makes the students' participation higher.

## 1. Introduction

Badminton is a popular sport around the world and is within our reach in our lives. As a sport that requires quick reaction and moving to catch the ball [1]. It requires a synergistic stroke variation between both badminton participants. Therefore, badminton is also a sport that tests the agility of the human body, due to its fast hitting and strategic tracking and prediction of badminton landing points, which turns badminton into a high-spectator sports competition. On the other hand, from a professional point of view, badminton is an extremely complex sport [2] with high demands on the physical and mental strength of the players. According to statistical reports from badminton professional bodies, it is estimated that there are about 150 million badminton enthusiasts and about 10 million professional badminton players worldwide [3, 4].

The key performance indicators of badminton are mainly reflected in the use of court areas, stroke distribution, technical movements, and stroke effectiveness. Like tennis [5–7], squash [8–10], and table tennis [11–13], the key performance indicators are mainly reflected in tactics. A successful badminton matchup will hold the opponent in spatial pressure to hit the badminton to the opponent's most disadvantageous area. The athlete widens the gap with the opponent when hitting the ball, inducing the opponent to create a large distance between the shots, making it impossible to return the ball in the next round, and at the same time consuming the opponent's physical strength [14, 15]. A professional badminton player should learn to fully control the integration of space, time, and tactics. Coaches, on the other hand, can obtain training details in the data of each game, which can be of great reference in future training [16]. For badminton teaching, teaching professional badminton skills movements in a decomposed manner can achieve a good teaching effect. as shown in Figure 1. Then human posture estimation can better correct the wrong badminton strokes.

With the popularity of badminton, badminton has become a physical education course in colleges and universities. At present, most of the badminton teaching adopts the traditional physical education class mode. The physical

Figure 1: Badminton stroke breakdown.

education teacher will first explain the rules system, action points, tactical skills, and scoring techniques of badminton through theory. Then take a practical face-to-face teaching, mainly to explain some action specifications, bucketing techniques, and sports injury prevention. This traditional didactic course does not allow students to truly feel the details of the action, and students' blind imitation is likely to lead to problems of sports strains and muscle injuries. This leads to ineffective physical education [17, 18]. With this teacher-led learning style, students can only learn badminton by imitating to feel the badminton action and feel the badminton skills through repeated practice later. This makes the quality of badminton teaching half the effort. In the long run, students will be dependent on the physical education teacher's demonstration guidance, in their practice, often will not get the point. Through this traditional way of teaching badminton, students often do not feel the main points of badminton personally, which reduces their motivation and participation, and they do not learn the main points of badminton to think after class [19]. Relevant studies have proven that there is a direct link between students' motivation and their sense of learning experience. Using smart strategy instruction in physical education can better increase students' engagement and interest [20].

A researcher has conducted solid three-dimensional modeling of player positions and stroke trajectories in badminton and analyzed the relationship between stroke trajectories and scoring points. This is a meaningful and highly expressive study. The presentation model of badminton is transformed from a pure spectator perspective to a data perspective. Data visualization can better assist badminton teaching and training [21]. Some researchers also started with videos of past badminton events and dissected the details of badminton skills from a video analysis perspective. Through the action decomposition algorithm, each frame of action is perfectly presented, and the students' attention is transferred from the audience's point of view to the skill learning link, so that students have an intuitive understanding of serving and hitting skills [22]. Chu et al. documented the link between space and stroke types in badminton through video analysis and proposed a classification detection model to analyze stroke technique characteristics. Although this analysis method of badminton players' performance patterns is able to present scoring details in complex matches, spatial and opponent-related movement details are not analyzed in an integrated manner [23, 24]. Badminton, as a high-intensity sport, has high demands on the quick reaction ability of athletes. Therefore, the technical details of opponents in space should be combined with before and after frames for detailed reference in multivariate performance analysis, and it is more informative to integrate the analysis of the matchups between athletes and opponents [25].

With the rapid development of technology, the sports industry urgently needs the intervention of artificial intelligence technology. The traditional badminton teaching mode is difficult to mobilize students' enthusiasm and participation. It makes it impossible for students to quickly acquire badminton action essentials and the quality of badminton teaching is low. With the upgrade of computer vision technology, stance estimation technology has the ability to be applied in badminton teaching. The development of these technologies has directly promoted badminton teaching and improved students' interpretation and interaction with badminton movements. The introduction of new technical systems for badminton instruction and practical practice is directly related to the quality of badminton instruction. Understanding the technical movements of badminton is quite important, the number of movements that occur in badminton matches is very high, and it is difficult for badminton teaching to explain the movements in a decomposed manner, making students understand the technical movements and thus comprehend the key points.

In this paper, we analyzed the current situation of badminton and found that the quality of badminton teaching is not optimistic. In order to further improve the quality of badminton teaching and students' participation. We consider the introduction of artificial intelligence technology into physical education. With the assistance of AI technology, the quality of physical education will be greatly improved. So we analyzed the research results related to human posture estimation and considered integrating it into physical education. Finally, we constructed a human-computer interactive badminton teaching system integrating computer vision technology and neural network algorithm. We propose a human pose estimation algorithm to correlate all the key points of acquired human features based on part affinity fields. Thus, human pose estimation and prediction are obtained. Finally, we demonstrate through comparison experiments with the traditional teaching model that the teaching model through our method results in more standardized strokes, higher badminton stroke scores, and more smooth skill switching between badminton serves and strokes. In addition, through our method, badminton teaching is more intuitive and interesting, which can greatly motivate students to learn.

## 2. Related Work

Ong et al. take inspiration from the flower pollination algorithm (FPA) and use FPA in real-time tracking of video actions of athletes' events. Real-time tracking of the athlete's center-of-mass pixel coordinates is achieved, and the search window is set following the adaptive law. The detection accuracy of the method is verified by experimental tests to meet the accuracy requirements [26]. To address the problem that the predicted range of the athlete's pose estimation does not enable energy-efficient point-to-point motion, Wang developed a linear time-invariant system in 2012. The system is able to control the optimal time to complete the point-to-point motion control and reduce the cost due to equipment capacity loss by energy optimization strategy, achieving an accuracy of 85% in the final system attitude estimation experiment [27]. Rutten et al. embarked on the study of badminton interactive robots and proposed an evolutionary operational derivation approach. The optimal pattern of energy is ensured from the time level to ensure the optimal badminton stroke trajectory. And by simulating the badminton player's stroke, the opponent's stroke trajectory is predicted as a way to achieve a comeback [28].

The biggest challenge facing human posture is the variation of human appearance, which is challenging in predicting the coordinates of key points in human space [29, 30]. Human pose estimation is a necessary joint technology for many industries, and the subject has been invested in research and development in various industries in recent years. Human pose estimation has different cut-offs in terms of the number of people, and single-person pose estimation has better accuracy and stability than multi-person pose estimation. In single-person pose estimation, only one person in the image is specified and localized, and then its key points are obtained. In multi-person pose estimation, it is necessary to first specify the number of people to be detected, then extract the individual with the highest human recognition weight from the image, and then localize its key points to capture the spatial coordinates. Obviously, multi-person pose estimation is more challenging, especially involving the effects of unstructured environments, such as crowds, occlusions, and multi-person interactions. The main structure of the human pose estimation algorithm is a convolutional neural network. In this paper, to estimate the pose of badminton action using this method, we choose multi-person pose estimation. In multi-person pose estimation, there are two main approaches: top-down and bottom-up.

The top-down approach will first detect the human body through the target recognition CNN network and then label it in the form of a rectangular box. The rectangular border is then used as a boundary to locate the human center of mass points in the rectangular box. The same operation is then used to locate other people appearing in the image in the same number as the specified number of people. Thus, the computational cost of multi-person pose estimation is closely related to the number of people specified. This type of algorithm uses mainly human detector techniques for the annotation of rectangular boxes and then applies a pose estimator to each detected person. So the detection results of multi-person pose estimation algorithms depend heavily on the accuracy of human detectors [31]. The human pose estimation based on convolutional neural networks can all be presented in the form of a heat map, where a deeper presentation represents a richer nodal capture. A convolutional pose machine is added to the output stage of each network, which has arrived at a supervised training role for the network, thus solving the gradient disappearance problem [32]. Besides, the stacked hourglass network can effectively expand the perceptual field and is a necessary continuous structure for pooling and upsampling in human pose estimation networks [33]. Of course, some researchers have also adopted the structure of feature pyramids to deal with feature maps of different resolutions [34]. However, in this paper, a cascaded pyramid network [35] is adopted to pool all the feature maps together to achieve multi-scale detection and obtain multi-scale human features. The literature [36] proposed a combined algorithm for the human pose keypoint prediction method, which can accurately predict the offset of the pose heat map and feature points and then perform offset correction to obtain accurate human spatial keypoint coordinates. Other researchers have tried end-to-end algorithms for bounding box regression and keypoint estimation for the human pose, such as Mask RCNN [37] and Faster R-CNN [38].

The bottom-up approach is significantly different, as it first detects all keypoint coordinates of multiple people in the full map and then performs keypoint combination on an individual basis to obtain pose estimation results. The literature [39] takes inspiration from the integer linear programming (ILP) algorithm and divides the detection population into clusters. The clusters are then matched with the labeled individuals as a way to obtain the final pose

estimation coordinate information. The literature [40] incorporates ResNet [41] based on the former to perform pairwise matching operations on adaptive images through conditional constraints. Both of these methods are representatives of ILP methods, but the computational cost of this type of method is too high for many researchers to adopt. Later, Newell et al. proposed fractional map and pixel-by-pixel embedding methods to successfully match pose key points to the corresponding people in different clusters, thus obtaining pose estimation information [42]. The literature [31] profoundly investigated a balanced approach between performance and speed, which was effective in saving computational costs in later studies. Just because of its high stability and adaptability and low computational cost, this algorithm is chosen as the basis of our human pose estimation method in this paper.

## 3. Method

*3.1. Part Affinity Fields.* In the input of the pose estimation algorithm, the input image is assumed to be of size $w \times h$. The input is fed into a convolutional neural network for the prediction of body joints to obtain a confidence map $H$ and a part affinity field (PAF) $L$, $L$ for each limb. The confidence map $H = (H_1, \ldots, H_J)$ indicates the existence of $J$ confidence maps in the whole pose estimation, where $H_j \in \mathbb{R}^{w \times h \times 2}$, $j \in \{1, \ldots, J\}$. $H_j^{GT}$ denotes the average position of the key points of each human body in the image, in terms of the whole without distinguishing individuals, and its label information is generated by a Gaussian distribution. PAFs $L = (L_1, \ldots, L_c)$, $C$ denotes the number of vector fields, and each limb is assigned a vector field, where $L_c \in \mathbb{R}^{w \times h \times 2}$, $c \in \{1, \ldots, C\}$. $L_c^{GT}$ denotes the unit vector real information generated independently for each limb, and each unit vector corresponds to a key point pair $j_1$ and $j_2$, and all vector directions within the rectangular box boundaries are $j_1$ pointing to $j_2$. Given the true label $y_{GT} = (H^{GT}, L^{GT})$ and the model prediction $P = (H, L)$, the model is trained using the mean square error $E_{L2}(P, y_{GT})$, defined as follows:

$$E_{L2}(P, y_{GT}) = \sum_{j=1}^{J} \sum_{p} W(p) \left\| H_j(p) - H_j^{GT}(p) \right\|_2^2 + \sum_{c=1}^{C} \sum_{p} W(p) \left\| L_c(p) - L_c^{GT}(p) \right\|_2^2, \tag{1}$$

where $P$ denotes the two-dimensional coordinates of the key points, $W$ denotes the binary mask, mainly for regions with nonstructural factors (e.g., overlapping occlusion cases like crowds), and $W(p) = 0$. The purpose of the mask is mainly to prevent the true predictions from being deleted by mistake during the training process. In addition to this, supervised units are added in the middle of each stage to periodically compensate for the gradient and prevent the occurrence of gradient disappearance [43]. The overall objective is $f$.

$$f = \sum_{t=1}^{T_p} f_L^t + \sum_{t=T_p+1}^{T_p+T_c} f_S^t. \tag{2}$$

In the parsing process, we use a set of two-assignment methods in order to better match the body part candidate matches. Firstly, we set the detection candidate region of the key points to the maximum position of the confidence map. Then, we match the linear integrals between the key points and then obtain the confidence scores of the body joints by calculation. Finally, all confidence scores are combined, and the final result of pose estimation is obtained by greedy correlation law.

*3.2. Part Confidence Map.* In order to evaluate the $f$ in equation (2), we chose to start in two dimensions and generate confidence maps $S^*$ from the annotated key points. Each confidence map corresponds to a corresponding limb site, and each limb site has a corresponding pixel coordinate representation in 2D coordinates. Ideally, if one person is detected during the detection process, a peak appears in the confidence map for each of its limb parts. If more than one person is detected, a peak appears similarly for each visible part $j$ of each person $k$.

Suppose that each person $k$ generates a corresponding individual confidence map $S_{j,k}^*$, where $x_{j,k} \in \mathbb{R}^2$ denotes the true position of body part $j$ of each person $k$ in the image. Then the value of position $p \in \mathbb{R}^2$ in $S_{j,k}^*$ is defined as follows:

$$S_{j,k}^*(p) = \exp\left( - \frac{\left\| p - x_{j,k} \right\|_2^2}{\sigma^2} \right), \tag{3}$$

where $\sigma$ is the controller of the peak size; the convolutional neural network first predicts to obtain a single confidence map and then aggregates through multiple maximal operators to obtain the final ground truth confidence map as a single confidence map.

$$S_j^*(p) = \max_k S_{j,k}^*(p). \tag{4}$$

Throughout the network prediction process, we took the average value of the confidence map instead of the maximum value in order to maintain the same accuracy of the nearby peaks. As shown in Figure 2, in our experimental tests, when we predict confidence maps, we tend to choose the no-maximal suppression method to obtain independent body part candidates. In the comparative experiment, we selected Gaussian curves with different $P$ values, namely Gaussian 1 and Gaussian 2. From the experimental results, it can be seen that the intersection between Gaussian 1 and Gaussian 2 is the optimal range of nonmaximum

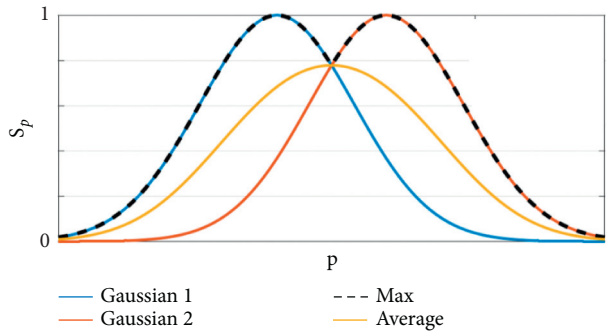


Figure 2: Confidence values for different peaks.

suppression because we take Gaussian 1 and Gaussian 2 to define the upper bound for nonmaxima suppression.

*3.3. Detection and Association.* The images are first fed into the convolutional neural network to generate the feature maps $F$ required in the first stage, and then the VGG-19 initial layers are adjusted and initialized. In the first stage, $L^1 = \phi^1(F)$, represents each set of partial affinity fields (PAFs), where $\phi^1$ represents the first inference result after network initialization. In the subsequent stages, each round of original image and feature $F$ predictions will be based on the results of the previous stage, and then the PAF logic will be computed to obtain the accurate predictions.

$$L^t = \phi^t\left(F, L^{t-1}\right), \forall 2 \le t \le T_p, \tag{5}$$

where $\phi^t$ denotes the $t$ stage of inference in the convolutional neural network and $T_p$ denotes the total number of PAF stages. After each round of $T_p$ iterations, the latest PAF predictions are passed through the confidence graph detection module.

$$\begin{aligned} S^{T_p} &= \rho^t\left(F, L^{T_p}\right), \forall t = T_p, \\ S^t &= \rho^t\left(F, L^{T_p}, S^{t-1}\right), \forall T_p \le t \le T_p + T_c, \end{aligned} \tag{6}$$

where $\rho^t$ is the CNN used for inference at stage $t$ and $T_c$ is the number of total confidence map stages.

In contrast to the approach mentioned in the literature [31], the PAF and confidence maps are refined in each convolutional neural network inference stage. Thanks to the refinement process, the number of parameters in each inference stage is halved. Our preliminary experimental validation shows that the affinity field prediction is proportional to the confidence result. More generally, the output fusion results of each channel of PAF can be used to infer the corresponding body part. However, if we are only given fragmentary information about the body parts, we instead obtain PAF channel information and the associated confidence map results.

The effect of the refinement of the affinity field at different stages is shown in Figure 3. All the confidence map results unfold the predictions within certain PAF bounds, and this also creates a problem that there is no significant variation in variability between all the stage confidence maps. In the iterative process of the neural network, in order to ensure the correct matching of body joint point features. We add the loss function L2 at the end of each stage, which are applied to the PAFs of the body parts in the first branch and the confidence maps in the second branch. Finally, the real graphs and fields are matched. In addition to this, to address the drawback of the dataset, we used a spatial loss function weighting.

*3.4. Human Pose Estimation Network.* The structure of our proposed human pose estimation network is shown in Figure 4. The coded partial prediction is performed by an iterative approach, which is mainly divided into two parts: part affinity field $\phi^t$ and detection confidence map $\rho^t$. Based on the network structure mentioned in the literature [32], we optimize the stage prediction by designing its confidence map as a continuous prediction based on this, where $t \in \{1, \ldots, T\}$; each stage contains supervised units.

Compared with the network structure in the literature [31], the human pose estimation network in this paper has more layers. The initial conv $7 \times 7$ is replaced by three consecutive conv $3 \times 3$, such an improvement is inspired by the inception structure proposed by Google, and after such a structural optimization, the number of parameters of the network can be greatly reduced, thus reducing the computational cost. In addition, we also refer to the method of DenseNet [44], which takes three convolutional kernels as groups and concatenates the output of each group. Such an operation can increase the number of nonlinear layers and also extract higher- and lower-level features from the maximum perceptual field.

## 4. Experiments

*4.1. Data Set.* Badminton is a sport, and there is no dedicated badminton stance data set in the world. In order to verify the performance of our method, we requested data from badminton matches from relevant authorities. Then we perform manual classification based on badminton teaching points, then use video editing software for badminton action segmentation, and then perform manual annotation for each classified action. Finally, all the collected data were separated into a training set and test set, and a total of 30,000 movements were produced for the training set and 5,000 movements for the test set. All the following experimental results analysis is based on this data set.

*4.2. Training.* All experiments in this article were performed on Ubuntu 16.04 with python 3.7 configured as the programming language environment version. The experimental hardware environment uses an RTX 3080 Ti GPU, an Intel i7-7700 CPU, and 50 GB of RAM. The normalization criterion is used during training. It can optimize the model and improve the overall robustness and stability of the model, and a method based on stochastic policy estimation is proposed as an alternative. To avoid the difficulty of overfitting during the training of the deep network, the early stopping method is used in the training. The relevant training parameters are shown in Table 1.

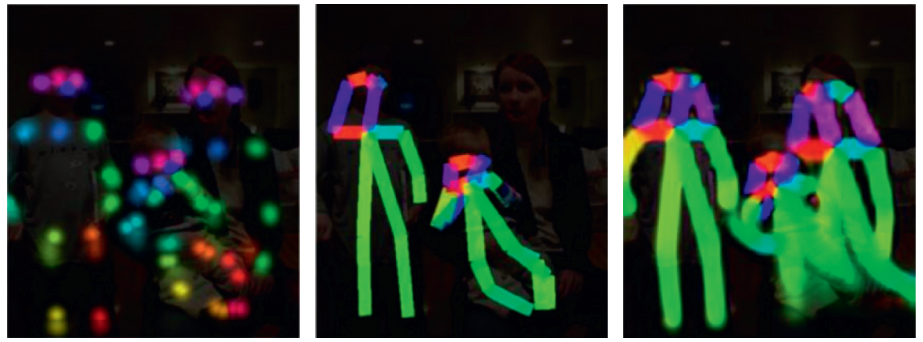


Figure 3: The effect of part affinity fields at different stages.

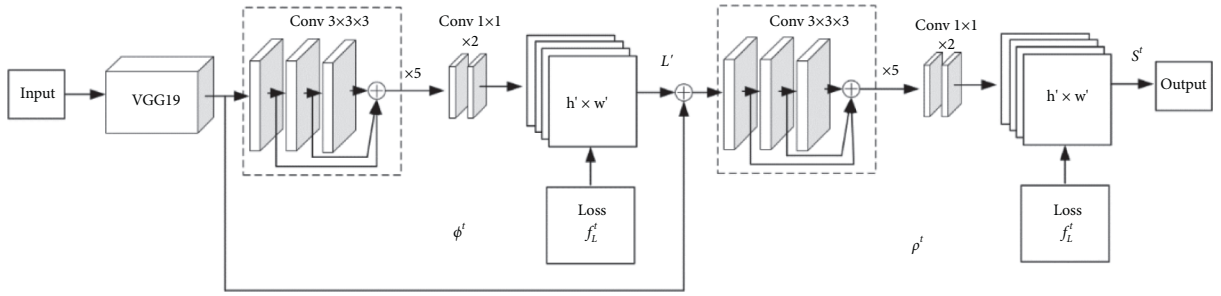


Figure 4: Human pose estimation network.

Table 1: Training parameter settings.

| Parameter | Value |
|---|---|
| Epoch | 30 |
| Regularization | 0.001 |
| Initial learning rate | 0.02 |
| Hidden unit number | 100 |
| Weight attenuation coefficient | 0.001 |
| Momentum | 0.9 |

*4.3. Experimental Procedure.* According to the physical education task set by the school, badminton class is held once a week for 90 minutes. For this reason, we set up a control group in the experiment: group A for the teaching method applying the human pose estimation method of this paper and group B for the traditional teaching method. For this purpose, we made detailed planning of the badminton teaching schedule as shown in Table 2.

The students in group A follow the human-machine interaction model. The class environment is in the machine vision arrangement of the field. We used the depth camera Kinect DK as a computer vision sensor to capture the badminton teaching situation through the depth camera and then feed it to the algorithm processing unit as shown in Figure 5.

Firstly, a human pose estimation algorithm is used to perform action decomposition of badminton strokes and then according to the main points explained. Then let the students in the machine vision field and capture the students comprehend the main points of badminton action; students can also visualize their own action details on the big screen and the standard action for comparison; students can correct their own action according to the standard action, so as to achieve a teaching purpose. At the end of the class, the teacher will comment on individual student movements with large differences and make comments. The video of the lesson with the human posture estimates is then given back to the students in the form of a video. Students can comprehend and think about badminton movements again after the class. The operation flow of the human-computer interaction system in group A teaching mode is shown in Figure 6.

The students in group B who follow the traditional teaching model learn badminton completely. The teacher will first verbally introduce the main points related to badminton, then demonstrate the main points of each badminton action through practice, and then let the students then practice in a cut and feel each badminton action. The teacher will give individual instruction to students with substandard posture. After the lesson, students are required to reflect on the instructional video and then write a summary.

*4.4. Badminton Serve Evaluation.* Regarding the assessment of badminton serve quality, this assessment guideline was followed for both groups A and B. Due to a large number of badminton skill movements, the assessment guidelines for each movement had subtle differences. In order to reflect the comparative performance of this study, we chose two skill movements, backhand and forehand, for the comparative experimental analysis. We assessed the pre- and post-test scores of the serve in four main areas: contact point, movement fluency, stroke trajectory, and stroke score. We

Table 2: Comparing the badminton teaching schedule of the experimental group.

| Schedule | A | B |
|---|---|---|
| Week 1–4 | Course introduction, human-computer interactive badminton action explanation, and learning | Course introduction and basic badminton skills learning |
| Week 5 | Automatic pre-testing of badminton serve quality and serve correctness with feedback under machine vision | The pre-test of badminton serve accuracy and badminton serve quality |
| Week 6–9 | Decompose difficult badminton skill movements through the human pose estimation algorithm and learn its key points | Learned difficult badminton skills, including ball control, footwork, swing, and stroke |
| Week 10 | Automatic post-test of badminton serve quality and correctness based on machine vision and give feedback report | The post-test of badminton serve accuracy and badminton serve quality, a self-reflection report |
| Week 11 | Badminton teaching effect test | Badminton teaching effect test |

Figure 5: Computer vision environment construction.

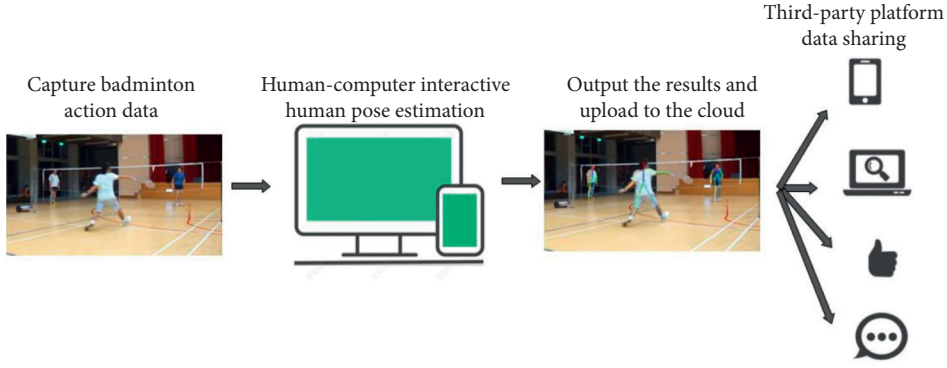


Figure 6: Human-computer interaction system operation flow for human posture estimation.

have summarized the scoring points of badminton strokes, and the specific scoring points are shown in Figure 7. The scoring rules of badminton are only divided according to the badminton court, and the extra points on the difficulty of the batter's actions are counted separately. We mainly study the scoring situation of badminton on the court.

We also assessed the variance of each weighted score, and the test of equality of variances implied satisfaction with the movement skills of the two groups assessed. To prevent differences in scores due to differences in student fitness, we also used ANCOVA analysis, which was primarily used to compare one variable in two or more aggregates, while considering other variables. The specific assessment results are shown in Table 3.

From the experimental results, it can be seen that group A was better than group B in serve quality assessment overall, the mean value of contact points in backhand action was higher in group A than group B by 0.35, the fluency of action in group A was more than group B by 0.31, the stroke trajectory in group A was better than group B by 0.25, and the average total score in group A was 1.09 ahead of group B. The experiment proves that in badminton teaching, the human-computer interaction teaching mode using the human posture assessment method, students' serving quality and scoring are excellent. Students' mastery of badminton movements was more precise. Students were more fluent in overall movement coherence for badminton sparring.

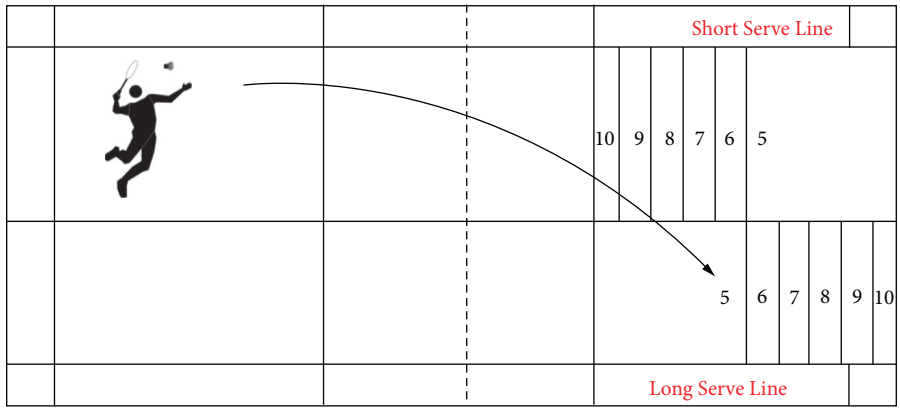


Figure 7: Badminton stroke scoring points.

Table 3: A and B group experimental group badminton stroke results.

| Badminton skill | Evaluation options | Mean | | SD | | $\eta^2$ | |
|---|---|---|---|---|---|---|---|
| | | A | B | A | B | A | B |
| Backhand | Contact point | 2.13 | 1.78 | 0.33 | 0.42 | 0.11 | |
| | Movement fluency | 2.16 | 1.85 | 0.41 | 0.36 | 0.12 | |
| | Stroke trajectory | 2.23 | 1.98 | 0.41 | 0.38 | 0.13 | |
| | Stroke score | 6.63 | 5.54 | 0.78 | 0.66 | 0.18 | |
| Forehand | Contact point | 2.23 | 1.66 | 0.43 | 0.54 | 0.24 | |
| | Movement fluency | 2.41 | 1.84 | 0.33 | 0.43 | 0.23 | |
| | Stroke trajectory | 1.34 | 1.85 | 0.61 | 0.61 | 0.15 | |
| | Stroke score | 6.97 | 5.36 | 0.98 | 0.99 | 0.31 | |

## 5. Conclusion

In this paper, we analyzed the current situation of badminton and found that the quality of badminton teaching is not optimistic. In order to further improve the quality of badminton teaching and students' participation. We integrate the human pose estimation algorithm into badminton teaching and construct a human-computer interactive badminton teaching system integrating computer vision technology and neural network algorithm. In this paper, we propose a human pose estimation algorithm, which firstly maintains the correlation of position information and orientation information between limb regions through part affinity fields. Then the center-of-mass key points are localized by part confidence map for the limb pose, and finally, all the feature key points obtained are correlated according to the part affinity field. Then the key points in each frame are integrated to obtain the features in the temporal dimension, and the pose estimation and prediction can be obtained. With our approach, real-time stance estimation can be achieved for badminton teaching, and students can observe the gap between their movements and the standard ones in the human-computer interface, thus feeling more deeply about the essentials of badminton strokes. In addition, such a teaching system makes badminton lessons more vivid and interesting and fully mobilizes the curiosity and motivation of students. Finally, we set up a comparison experiment to compare the teaching mode of this paper's method with the traditional teaching mode. The experimental results proved that through the teaching mode of our method, students' strokes were more standardized, the skill switching between badminton serves and strokes was more fluent, and badminton strokes scored higher.

Due to the wide range of badminton skill movements, the currently constructed dataset contains only two skill movements. For the neural network algorithm, the number of data sets determines the performance of the model. Therefore, in future research, we will further expand more professional badminton movements and improve the generalization ability of the model.

## Data Availability

The data set can be accessed upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## References

[1] W. Chen, T. Liao, Z. Li et al., "Using FTOC to track shuttlecock for the badminton robot," *Neurocomputing*, vol. 334, pp. 182–196, 2019.

[2] A. Fabisch, C. Petzoldt, M. Otto, and F. Kirchner, "A survey of behavior learning applications in robotics-state of the art and perspectives," 2019, https://arxiv.org/abs/1906.01868.

[3] S. Lupashin, M. Hehn, M. W. Mueller, A. P. Schoellig, M. Sherback, and R. D'Andrea, "A platform for aerial robotics research and demonstration: the flying machine arena," *Mechatronics*, vol. 24, no. 1, pp. 41–54, 2014.

[4] C. Z. Shan, *Sensor-Based Assessment Using Machine Learning for Predictive Model of Badminton skills*, Universiti Teknologi Malaysia, Johor Bahru, Malaysia, 2018.

[5] R. Martínez-Gallego, J. F. Guzmán, N. James, J. Pers, J. Ramón-Llin, and G. Vuckovic, "Movement characteristics of elite tennis players on hard courts with respect to the direction of ground strokes," *Journal of Sports Science & Medicine*, vol. 12, no. 2, p. 275, 2013.

[6] N. S. Kolman, T. Kramer, M. T. Elferink-Gemser, B. C. H. Huijgen, and C. Visscher, "Technical and tactical skills related to performance levels in tennis: A systematic review," *Journal of Sports Sciences*, vol. 37, no. 1, pp. 108–121, 2019.

[7] C. Martin, B. Bideau, P. Touzard, and R. Kulpa, "Identification of serve pacing strategies during five-set tennis matches," *International Journal of Sports Science & Coaching*, vol. 14, no. 1, pp. 32–42, 2019.

[8] S. Murray, N. James, J. Perš, R. Mandeljc, and G. Vučković, "Using a situation Awareness Approach to Identify differences in the performance Profiles of the World's top two squash players and their opponents," *Frontiers in Psychology*, vol. 10, p. 1036, 2019.

[9] G. Vučković, N. James, M. Hughes et al., "A new method for assessing squash tactics using 15 court areas for ball locations," *Human Movement Science*, vol. 34, pp. 81–90, 2014.

[10] G. Vučković, N. James, M. Hughes, S. Murray, Z. Milanović, and J. Perš, "The effect of court location and available time on the tactical shot selection of elite squash players," *Journal of Sports Science and Medicine*, vol. 12, no. 1, p. 66, 2013.

[11] M. Fuchs, R. Liu, I. Malagoli Lanzoni et al., "Table tennis match analysis: a review," *Journal of Sports Sciences*, vol. 36, no. 23, pp. 2653–2662, 2018.

[12] J. Wang, "Comparison of table tennis serve and return characteristics in the London and the Rio Olympics," *International Journal of Performance Analysis in Sport*, vol. 19, no. 5, pp. 683–697, 2019.

[13] G. Munivrana, G. Furjan-Mandić, and M. Kondrič, "Determining the structure and evaluating the role of technical-tactical Elements in basic table tennis playing systems," *International Journal of Sports Science & Coaching*, vol. 10, no. 1, pp. 111–132, 2015.

[14] P. Abián, A. Castanedo, X. Q. Feng, J. Sampedro, and J. Abian-Vicen, "Notational comparison of men's singles badminton matches between Olympic Games in Beijing and London," *International Journal of Performance Analysis in Sport*, vol. 14, no. 1, pp. 42–53, 2014.

[15] C. L. Ming, C. C. Keong, and A. K. Ghosh, "Time motion and notational analysis of 21 point and 15 point badminton match play," *International Journal of Sports Science and Engineering*, vol. 2, no. 4, pp. 216–222, 2008.

[16] M. Hughes, M. T. Hughes, and H. Behan, "The evolution of computerised notational analysis through the example of racket sports," *International Journal of Sports Science and Engineering*, vol. 1, no. 1, pp. 3–28, 2007.

[17] J. Zeller, "Reflective practice in the ballet class: bringing progressive pedagogy to the classical tradition," *Journal of Dance Education*, vol. 17, no. 3, pp. 99–105, 2017.

[18] M. Xie, "Design of a physical education training system based on an intelligent vision," *Computer Applications in Engineering Education*, vol. 29, no. 3, pp. 590–602, 2021.

[19] P. Petsilas, J. Leigh, N. Brown, and C. Blackburn, "Creative and embodied methods to teach reflections and support students' learning," *Research in Dance Education*, vol. 20, no. 1, pp. 19–35, 2019.

[20] T. H.-C. Chiang, S. J. H. Yang, and C. Yin, "Effect of gender differences on 3-on-3 basketball games taught in a mobile flipped classroom," *Interactive Learning Environments*, vol. 27, no. 8, pp. 1093–1105, 2019.

[21] W. T. Chu and S. Situmeang, "Badminton video analysis based on spatiotemporal and stroke features," in *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval*, pp. 448–451, New York, NY, USA, 2017.

[22] S. Vial, J. Cochrane, A. J. Blazevich, and J. L. Croft, "Using the trajectory of the shuttlecock as a measure of performance accuracy in the badminton short serve," *International Journal of Sports Science & Coaching*, vol. 14, no. 1, pp. 91–96, 2019.

[23] G. Torres-Luque, Á. I. Fernández-García, J. C. Blanca-Torres, M. Kondric, and D. Cabello-Manrique, "Statistical differences in set Analysis in badminton at the RIO 2016 Olympic Games," *Frontiers in Psychology*, vol. 10, p. 731, 2019.

[24] M. Phomsoupha and G. Laffaye, "The science of badminton: game characteristics, anthropometry, physiology, visual fitness and biomechanics," *Sports Medicine*, vol. 45, no. 4, pp. 473–495, 2015.

[25] M. D. Hughes and R. M. Bartlett, "The use of performance indicators in performance analysis," *Journal of Sports Sciences*, vol. 20, no. 10, pp. 739–754, 2002.

[26] P. Ong, T. K. Chong, K. M. Ong, and E. S. Low, "Tracking of moving athlete from video sequences using flower pollination algorithm," *The Visual Computer*, vol. 38, pp. 939–962, 2022.

[27] X. Wang, J. Swevers, J. Stoev, and G. Pinte, "Energy optimal point-to-point motion using model predictive control," *American Society of Mechanical Engineers*, vol. 45301, pp. 267–273, 2012.

[28] K. Rutten, J. De Baerdemaeker, J. Stoev, M. Witters, and B. De Ketelaere, "Constrained online optimization using evolutionary operation: a case study about energy-optimal robot control," *Quality and Reliability Engineering International*, vol. 31, no. 6, pp. 1079–1088, 2015.

[29] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," in *Proceedings of the British Machine Vision Conference*, Aberystwyth, UK, September 2010.

[30] B. Sapp and B. Taskar, "Modec: Multimodal decomposable models for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3674–3681, Portland, OR, USA, June 2013.

[31] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, Honolulu, HI, USA, 2017.

[32] S. E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 4724–4732, Las Vegas, NV, USA, June 2016.

[33] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proceedings of the European Conference on Computer Vision*, pp. 483–499, Springer, Amsterdam, Netherlands, 2016.

[34] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, 2017.

[35] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded pyramid network for multi-person pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7103–7112, Salt Lake City, UT, USA, June 2018.

[36] G. Papandreou, T. Zhu, N. Kanazawa et al., "Towards accurate multi-person pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4903–4911, Honolulu, HI, USA, 2017.

[37] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, Venice, Italy, October 2017.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.

[39] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, and B. Schiele, "DeepCut: joint subset partition and labeling for multi person pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4929–4937, Las Vegas, NV, USA, 2016.

[40] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: a deeper, stronger, and faster multi-person pose estimation model," in *Proceedings of the European Conference on Computer Vision*, pp. 34–50, Amsterdam, Netherlands, 2016.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[42] A. Newell, Z. Huang, and J. Deng, "Associative embedding: end-to-end learning for joint detection and grouping," 2016, https://arxiv.org/abs/1611.05424.

[43] I. Radosavovic, P. Dollár, R. Girshick, G. Gkioxari, and K. He, "Data distillation: towards omni-supervised learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4119–4128, Salt Lake City, UT, USA, 2018.

[44] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, Honolulu, HI, USA, July 2017.