Research Article

Video Analysis of Belt and Road Sports Events Based on Wireless Network and Artificial Intelligence Technology

Linlin Zhao and Delun Pan

Beijing Sport University, Beijing, China

Correspondence should be addressed to Linlin Zhao; zlinlin@bsu.edu.cn

Received 11 November 2021; Revised 19 December 2021; Accepted 28 December 2021; Published 17 February 2022

Academic Editor: Narasimhan Venkateswaran

Copyright © 2022 Linlin Zhao and Delun Pan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonprofessionals pay more and more attention to sports events, but in the process of live sports events, the semantics of professional terms are rich. And it is difficult for nonprofessionals to understand the meaning of the game. In order to improve the viewing experience of nonprofessionals, by intelligent target tracking technology, a system to capture athletes' movement process is proposed. It is possible to analyze the types of athletes' movements. In competitions or ball games, the system can predict the location of the ball or the player's drop point. The four different tennis courts, the player's motion capture, and the accuracy of the tennis landing prediction are further analyzed. The results show that when the athlete is facing the capture camera, the accuracy of motion recognition is high. In the absence of data on the playing field, the accuracy of the system for determining the position of the tennis ball fluctuates greatly. Therefore, this kind of system needs to have enough data to support it before it can be applied to fall point prediction. The proposed system provides a certain reference basis for the design of intelligent sports control facilities.

1. Introduction

With the normal holding of the 32nd Olympic Games and Paralympics in Tokyo, Japan in 2021, and the grand opening of the 14th National Games of People's Republic of China, the 11th National Paralympic Games and the 8th Special Olympics in Xi'an, Shaanxi Province in 2021, the Chinese people are very concerned about sports. The attention of the movement has reached an unprecedented grand occasion [1]. Among them, athletics, diving, usefulness, table tennis, and tennis have received very high ratings. However, because of this kind of large-scale sports competition, the explanation uses more professional terminology [2]. For most nonprofessionals who can only watch the game in front of the TV screen, it is difficult to understand certain professional terms and professional movements. The traditional form of watching games is gradually unable to meet people's demand for attention to sports programs. With the advent of the era of artificial intelligence (AI) [3], it is a hot research direction how to apply this technology to the intelligent upgrade of the broadcast form of sports games.

In recent years, when some large-scale sports events are held, some AI devices that can be worn by athletes have appeared [4]. Aithal used the random forest algorithm; different smart devices for measuring physical activity have been proposed. It can collect the athlete's training habits, infer the athlete's physical activity by capturing, storing, analyzing, and predicting four functions, and get the factors that affect the athlete's performance [5]. He designed a training program suitable for athletes according to their mental state, mental activity, and daily life by the deep learning model. This deep learning model can improve the accuracy and efficiency of recognition and greatly improve the utilization of sensor data [6]. Subsequently, Li et al. used fog-assisted computing to create an efficient wearable sensor network for continuous real-time monitoring of heart rate, breathing rate, and exercise rhythm of athletes during sports activities. This device can upload data to the Ethernet module of the Internet of Things (IoT) connection system and authorize individuals to access the data through the Internet to track the health of athletes [7]. In addition to wearable smart devices, there are also some smart devices

deployed in sports fields. Sarvestan and Khalafi proposed a new method of judging whether there is a ball in the volleyball court in the shortest possible time by manipulating the line and ball layer structure around the court. It can provide accurate decision-making for referees and provide faster competition process for athletes [8]. Fialho et al. used different AI technologies to propose a framework for developing an AI-based football match result prediction system [9]. Liu et al. learned that it is difficult for the system to obtain enough information to predefine multiple models before tracking. The system is also unable to accurately model the changeable and uncertain maneuvering movement in time. A deep learning maneuvering target tracking algorithm by the DeepMTT network is proposed. It verifies that the deep learning maneuvering target tracking algorithm is superior to other state-of-the-art maneuvering target tracking algorithms [10]. Wan et al. used sparse representation theory to propose a target tracking algorithm for locating pedestrians and vehicles captured by drones. Compared with other advanced trackers, the proposed tracker can achieve better performance in terms of accuracy and success rate [11]. The above studies have shown that various new technologies under AI can be applied to sports competitions, whether for individual athletes or competition venue management.

Taking the tennis match as the research goal, the target tracking and detection technology in the AI field and the wireless lightweight convolutional network model are used. The proposed model is used for real-time analysis of athletes' movements in the game and prediction of the landing area of tennis trajectory. Firstly, starting from the athlete itself, it can predict the athlete's real-time position, track the target, and obtain the athlete's movement data to analyze the action by the human body information. Secondly, the video frame is analyzed for the trajectory of the tennis ball, and the motion vector is obtained. Combined with relevant knowledge, its location is analyzed. The innovation lies in combining the wireless lightweight convolutional network with target detection technology to analyze and predict the real-time position of athletes and tennis balls. It provides a new way for the audience to watch the game.

2. Materials and Methods

2.1. Introduction to Commonly Used Target Detection Technologies. With the development and maturity of AI and deep learning technology, many detection systems combined with AI or deep learning have been proposed. This includes single-shot detection technology, regional neural networks, and unified real-time object detection algorithms. These algorithms are mainly divided into two detection algorithms by the candidate area and by the regression algorithm [12].

You Only Look Once (YOLO) [13] is divided into two levels, v2 and v3, and is the most widely used detection algorithm in recent years. It has the characteristics of high efficiency and accuracy. Table 1 shows the performance comparison of some commonly used detection algorithms.

The essence of target tracking [10] is to continuously track the observed target. It is one of the frequently used techniques in all sports. The tracking of the sphere in football, basketball, table tennis, and other ball sports and the tracking of the athletes themselves in track and field, diving, and swimming all belong to the category of target tracking.

With the rise of AI technology, the target motion data detected by the tracking algorithm is used as a basis for continuous tracking of the target. Currently, the most widely used tracking algorithm is called Multitarget Tracking Algorithm [14].

Target tracking involves the processing and matching of the state and motion trajectory of the target object. The state of an object refers to the use of an 8-dimensional space X to describe the state of the motion trajectory at a certain moment. The matching of the motion trajectory is to use the Mahalanobis distance between the detection result and the tracking result predicted in the filter to express its degree, as shown in the following:

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i), \quad (1)$$

where y_i is the predicted observation of the trajectory, d_j is the j -th result state, and S_i is the covariance matrix obtained through the filter.

The apparent degree of matching is the phenomenon that only uses Mahalanobis distance [15] to predict the data distortion. If the camera moves, it may cause the Mahalanobis distance prediction to fail, which needs to be remedied by the apparent degree. In the system, the cosine distance calculation is performed on the results within the latest $L=100$, as shown in (2) and (3):

$$d_{i,j}^{(2)} = \min\{1 - r_j^T r_k^{(i)} | r_k^{(i)} \in R_i\}, \quad (2)$$

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}. \quad (3)$$

Among them, λ is a kind of hyperparameter, whose function is to adjust different weights. d is the result state of the node. r is one of the observed variables. Equation (3) is the fusion average weight of these two measures.

2.2. Two-Dimensional Human Body Recognition. Human body gesture recognition [16] is a two-dimensional algorithm using affinity domain [17]. In Figure 1, it consists of two parts. The upper part is the training skeleton and the lower part has two branches. They are the direction of the regression skeleton and the corresponding key point calculations. The results of each stage are combined to determine the specific joint points of the human body.

In Figure 2, it is a complete human figure of the human body's 18-joint point recognition structure for human body gesture recognition. Through the human body gesture recognition, the specific position of the human body at different distances can be identified. Different points correspond to different positions: 0 (nose), 1 (neck), 2 (right shoulder), 3 (right elbow), 4 (right wrist), 5 (left shoulder), 6 (left elbow), 7 (left wrist), 8 (right hip), 9 (right knee), 10

TABLE 1: Performance comparison of commonly used detection algorithms.

Algorithm name	Feature	Frame rate	Data set used	Algorithm average accuracy
Local area network real-time object detection	High accuracy and low efficiency	7	COCO ^[66]	47.2
Regional neural network	The calculation speed is higher than YOLO, and the accuracy is similar to the above	19	COCO ^[66]	46.5
YOLO v2	High efficiency, but the detection of small targets is not perfect	40	COCO ^[66]	48.1
YOLO v3	On the high efficiency of v2, the detection performance of small targets is improved	30	COCO ^[66]	57.9

FIGURE 1: Structure diagram of human body gesture recognition.

(right ankle), 11 (left hip), 12 (left knee), 13 (left ankle), 14 (right eye), 15 (left eye), 16 (left ear), and 17 (right ear).

2.3. Wireless Lightweight Network. Wireless lightweight network [18] model has been favored by research experts in recent years. It has outstanding performance in the fields of

media classification, detection target, and segmentation graphics. In order to optimize and reduce the amount of calculation, the deep separable convolution is divided into two steps. The first is the deep convolution, and the second is the pointwise convolution.

Standard convolution is to change the layer of $M \cdot D_F \cdot D_F$ into $M \cdot D_G \cdot D_G$. D_F is the length and width of the input image.

FIGURE 2: Schematic diagram of human joint points.

M is the number of input channels. D_G is the length and width of the output graph. N is the number of output channels.

$$D_k \times D_k = M \times N \times D_k \times D_F \times D_k \times D_F. \quad (4)$$

Figure 3 is the pointwise convolution in the depth separable convolution, that is, to combine the feature maps. If the size of the convolution kernel is $1 \times 1 \times M \times N$, then the total calculation is $M \times N \times D_F \times D_F$.

Compared with the traditional volume measurement calculation, this method is shown in the following equation:

$$\frac{M \times N \times D_k \times D_F \times D_k \times D_F + M \times N \times D_F \times D_F}{M \times N \times D_k \times D_k \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_k^2}. \quad (5)$$

For example, the 3-channel given by the system is a picture or video with a size of 224×224 . It is the third convolutional layer in the network. Then, the size of the newly entered feature image is 112. The number of channels is 64. The number and size of the convolution kernel are $128 \times 3 \times 3$. Therefore, the traditional convolution operation is $3 \times 3 \times 128 \times 64 \times 112 \times 112 = 924844032$. If this traditional convolution operation is replaced with a deep convolution operation, the amount of calculation is $3 \times 3 \times 64 \times 112 \times 112 + 128 \times 64 \times 112 \times 112 = 109985792$. The ratio of calculation amount between the new deep convolution and the traditional convolution is $109985792/924844032 = 0.1189$. Therefore, the amount of calculation on the network is greatly reduced.

2.4. Structure Comparison between Deep Convolution and Ordinary Convolution. The structure of ordinary 3D convolution is shown in the upper part of Figure 4. The new type of deep convolution [19] is shown in the following half. This makes it possible to use deep convolution when it is convenient and at the same time add batching criteria and activation units to the result. Because this method uses a lot of 1×1 convolution operations, it can efficiently use a highly optimized mathematical inventory to complete.

2.5. Design of Semantic Analysis System for Tennis Matches. In Figure 5, the processing mode of the system is to perform video processing in high-resolution mode, which requires a lot of calculations. Therefore, for service facilities, Graphics Processing Unit (GPU) [20] is used to complete. In view of the low coupling and practicability, the user's convenient interaction must also be achieved, so the structural design with the front and rear parts separated is used.

The function of the camera or lens is to record and capture the video during the game. The service equipment relates to the lens and is responsible for solving the problems caused by the different models of the lens and preprocessing the media video of the mobile phone, unifying the format, and sending it to the postprocessing client. The work of the GPU is the back-end algorithm. The algorithm is divided into two parts: the first part is the processing of athletes' movements. Firstly, the athletes are monitored. According to the monitoring results, the location data is divided. Then, the athlete's active area is locked and tracked. Finally, the skeleton of the human body is extracted to obtain the

FIGURE 3: Schematic diagram of point-by-point convolution.

FIGURE 4: Comparison of ordinary convolution and deep convolution.

FIGURE 5: System structure diagram.

movement data of specific joint points. The second part is to deal with the tennis used in the game. In the middle of the game, tennis balls are smaller in size than football and basketball and move faster. The key frames of the middle part of the tennis balls [21] are extracted to predict the position of each tennis ball. In this way, a complete tennis ball prediction system is formed.

Figure 6 is a flowchart of the motion tracking for athletes. This part of the module completes the prediction of athletes'

actions by inputting video frames and outputting semantic information. Among them, the action elements include movement type, movement distance, speed, time, and system frame number. Starting from the input of the tennis match video, the processing flow is a process from rough to fine. Firstly, filter out the athletes from different objects in the video, and then refine them. Then, the athlete's body skeleton is extracted, and 18 skeletons are found to be tracked. Finally, use the above algorithm to make a judgment.

FIGURE 6: Athlete action tracking analysis flow chart.

Like the tracking process of athletes, they both complete the prediction of the position of the tennis ball by inputting video frames. Firstly, according to the horizontal speed of the tennis ball, theoretically, it should move at a constant speed. Secondly, through the calculation of the falling speed, the location of the falling point is finally obtained. Finally, output the data result and transmit it to the user client through the wireless network.

Figure 7 is a diagram of the semantic display process of a tennis video. In addition to the video sent back by the display

processor through the wireless network, the resulting data also needs to be drawn. Among them, it includes the human and tennis monitoring frame, the athlete's activity area, the athlete's skeleton structure, and 18 bone points.

In the analysis module of action semantics, motion analysis module plays a third of the role of the entire system module, mainly to analyze the meaning of the action of the detection target. It is mainly compiled in Python [22] and has two input function parameters. Among them, one is the key bone points of the athlete's movement in the previous

FIGURE 7: Video semantic display process.

frame, and the next is the new movement composed of the athlete's bone points in the new frame. Assign the coordinates of the four bone points to variables, using 0 for empty, 1 for left, and 2 for right. Assuming that the influence of the four skeletal points on the action is equal, that is, they are all 1, the calculation of the direction and distance four times is shown in (6) and (7):

$$k = \frac{y_1 - y_2}{x_1 - x_2}, \quad (6)$$

$$\text{distance} = \sqrt{(y_1 - y_2)^2 + (x_1 - x_2)^2}. \quad (7)$$

Through (6) and (7), the direction k and distance of the four bone points can be obtained. If $k > 0$ and distance > 5 , it means moving to the left; if $k < 0$ and distance > 5 , it means moving to the right.

Through the athlete's movement data and semantic information, various factors of the movement can be solved. The movement time is generated by the function library that comes with the system. The input parameters of the function are the center point of the bone point in the previous frame and the center point of the next frame. The actual stadium is reduced by a certain ratio in the video, so the values W_1 and W_2 of the length and width of the ratio are solved. The distance is $w_distance$ in width and $h_distance$ in length. The specific calculation is as (8) and (9):

$$\text{all_distance} = \sqrt{w_distance^2 + h_distance^2}, \quad (8)$$

$$w_distance = s - k * (y_1 - 336) * \frac{s}{336} * w_1. \quad (9)$$

Among them, $k = 1.17$ and $w_1 = 1828/663$. According to (8) and (9), the speed value can be obtained, as

$$\text{speed} = \frac{\text{all_distance}}{t}. \quad (10)$$

t is time.

2.6. The Impact of Tennis Impact Analysis on the Game Viewing Experience. The way of online sports events is different from that of watching them on-site. Online watching will be affected by network signals, cameras, or ball tracks, causing viewers to not receive real-time information about the game. The game is a real-time three-dimensional scene, and the way of watching TV and other network devices is two-dimensional. During watching the game, the device cannot predict some special positions. Therefore, the technique of predicting and analyzing the movement of the athletes and the trajectory of the ball used in the game can help the audience better understand the content of the game and analyze the game process.

3. Results and Discussion

3.1. Real-Time Analysis of the System. The real-time accuracy of this analysis system is tested. During the actual game, the venue environment, whether the server is stable, and the environment in which the system is running will all affect the accuracy of the tracking system. Therefore, this round of testing will analyze the speed of the system in four environments. The experiment is carried out four times, and a short segment is selected from the four kinds of competition videos. The results of the experiment are shown in Figure 8.

In Figure 8, the operating speed of the same site will fluctuate as the experiment increases. The fluctuation of the site's own data of the same material is not very large. The minimum frame rate of the Australian Open hard court is

FIGURE 8: The operating speed of the system under four venues (Australia is the Australian Open hard court, French is the French Open red court, Wimbledon is the Wimbledon grass court, U.S. is the US Open hard court).

6.83 frames, and the maximum frame rate is 7.3 frames, a difference of 0.47 frames. The lowest frame rate of the French Open red court is 6.78 frames, and the highest frame rate is 7.49 frames, a difference of 0.71 frames. The minimum frame rate of Wimbledon grass court is 7 frames, and the maximum frame rate is 7.96 frames, a difference of 0.96 frames. The minimum frame rate of the US Open hard field is 6.9 frames, and the maximum frame rate is 7.73 frames, a difference of 0.83 frames. The frame rate of each venue does not differ by more than 1 frame, and the average frame rate of the four venues is 7.04 frames on the Australian Open hard court. The French Open red court is 7.03 frames, the Wimbledon grass court is 7.4 frames, and the US Open hard court is 7.4 frames. The difference is no more than 0.4 frames, indicating that the operating speed of the system will not be affected by the different courts.

3.2. System Accuracy Analysis. For the accuracy of the tracking and capturing system, the same five experiments are carried out. The experiment is divided into two groups. The first group is the accuracy of the athlete's motion capture; the result is shown in Figure 9; the second group is the accuracy of the tennis capture and the prediction of the position of the landing; the result is shown in Figure 10. Experiments are still being carried out on four venues: the Australian Open hard court, the French Open red court, the Wimbledon grass court, and the US Open hard court.

In Figure 9, the system's judgment rate for the athlete's own action has a large range of fluctuations in the range of 60%–100% in the same field, and the fluctuation range of the first three different fields is large. In the fourth set of experiments, the experimental data of the four venues are

relatively similar. According to the analysis of the results after the experiment, because the athletes had their back to the camera in the first three groups of experiments, some bone points in the body are blocked. After the adjustment, the fourth group of experimental results are similar, and the venue has little effect on the accuracy of the system's action judgment.

In Figure 10, in the same court, the accuracy of the judgment of the tennis ball is very small. The Australian Open hard court has the lowest accuracy, between 47% and 78%, followed by the French Open clay court, between 65%–76%. Then came the US Open hard court, between 82%–96%. The highest is the Wimbledon grass field, which is between 85% and 98%. From this analysis, the different accuracy of the judgment of the landing point without the venue is due to the nature of the ground, and the physical phenomenon of rebounding is when the tennis ball falls on the ground. The system lacks references to physical quantities such as the elastic coefficient of the site, leading to the phenomenon of different accuracy.

4. Discussion

AI target tracking algorithms and wireless lightweight networks are used to build tennis impact prediction technology. The obtained results are compared and analyzed with previous studies. Among them, Jian et al. used computer vision, image processing, and software teaching technology to create a basketball sports monitoring data system for data collection. Traditional regression techniques are used to determine the location of individuals. The feasibility, performance, and efficiency of the proposed framework prove the reliability [23]. The method of this

FIGURE 9: Accuracy of athlete's action judgment.

FIGURE 10: The accuracy of the judgment of the area where the tennis ball falls.

literature research is similar to this paper. Image processing and target tracking are used to establish predictions for the rebound of the trajectory of ball sports in sports events. Therefore, this result further proves that the research in this direction is universal and can optimize the viewing experience of sports events.

5. Conclusions

In view of the two major sports competitions of the 32nd Olympic Games and Paralympics in Tokyo, Japan in 2021, and the grand opening of the 14th National Games of People's Republic of China, the 11th National Paralympic Games and the 8th Special Olympics in Xi'an, Shaanxi Province in 2021, people are paying more and more attention to sports competitions. The traditional way of

watching games cannot satisfy people's needs. In order to improve the viewing experience of nonprofessionals, an AI-based human motion capture analysis system and a tennis impact prediction and judgment system have been proposed. The prediction result is transmitted to the on-site central processing unit and the viewer's program watching equipment through the wireless network. By the athlete's motion tracking algorithm, the tennis real-time semantic analysis algorithm for posture estimation, and the motion landing area prediction algorithm, a set of semantics used in the live tennis match video is established through the wireless mobile network, two-dimensional human body recognition, and wireless lightweight network. It realizes the recognition output of the meaning of the sports data of the athletes and the types of actions. The established model is tested, and the system operation speed, the accuracy of the

athlete's action recognition, and the tennis landing are carried out on the four competition venues of the Australian Open hard court, French Open red court, US Open hard court, and Wimbledon grass court. Through experiments, the field will not affect the accuracy of the athlete's action recognition. The position of the athlete relative to the camera will affect the accuracy of motion capture. Therefore, when performing motion recognition, athletes should try to face the camera as much as possible so that the camera can capture a complete image of the human body. It did not test the degree of influence of other external factors on accuracy, such as weather, the physical fitness of the athletes, etc. Therefore, in subsequent experiments, external factors that affect the accuracy of the system will be tested and analyzed.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Tian and V. W. S. Tung, "Investigating non-host residents' perspectives of the impacts of a mega sports event: the Beijing 2022 winter olympic games," *Journal of China Tourism Research*, pp. 1–16, 2021.
- [2] A. LeCouteur and A. Yong, "Television commentary on women's and men's Australian rules football: a thematic analysis," *International Review For the Sociology of Sport*, Article ID 10126902211033221, 2021.
- [3] J. Chang and D. Lee, "Changes in user experience and satisfaction as media technology evolves: the reciprocal relationship between video games and video game-related media," *Technological Forecasting and Social Change*, vol. 174, Article ID 121219, 2022.
- [4] N. N. Xiao, W. Yu, and X. Han, "Wearable heart rate monitoring intelligent sports bracelet by Internet of things," *Measurement*, vol. 164, Article ID 108102, 2020.
- [5] P. S. Aithal, "Tracking and monitoring fitness of athletes using IoT enabled wearables for activity recognition and random forest algorithm for performance prediction," *International Journal of Health Sciences and Pharmacy (IJHSP)*, vol. 5, no. 1, pp. 72–86, 2021.
- [6] Y. He, "Athlete human behavior recognition by continuous image deep learning and sensors," *Wireless Networks*, pp. 1–12, 2021.
- [7] S. S. Li, B. X. Zhang, P. F. Fei, P. Mohamed Shakeel, and R. Dinesh Jackson Samueld, "Computational efficient wearable sensor network health monitoring system for sports athletics using IoT," *Aggression and Violent Behavior*, Article ID 101541, 2020.
- [8] J. Sarvestan and M. Khalafi, "Smart line judgement system: a novel technology in vol-leyball arbitration," *Insight—Material Science*, vol. 2, no. 1, pp. 1–7, 2019.
- [9] G. Fialho, A. Manhães, and J. P. Teixeira, "Predicting sports results with artificial intelligence—a proposal framework for soccer games," *Procedia Computer Science*, vol. 164, pp. 131–136, 2019.
- [10] J. Liu, Z. Wang, and M. Xu, "DeepMTT: a deep learning maneuvering target-tracking algorithm based on bidirectional LSTM network," *Information Fusion*, vol. 53, pp. 289–304, 2020.
- [11] M. Wan, G. Gu, W. Qian, K. Ren, X. Maldague, and Q. Chen, "Unmanned aerial vehicle video-based target tracking algorithm using sparse representation," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9689–9706, 2019.
- [12] R. Corizzo, G. Pio, M. Ceci, and D. Malerba, "DENCAST: Distributed density-based clustering for multi-target regression," *Journal of Big Data*, vol. 6, no. 1, pp. 1–27, 2019.
- [13] Y. Q. Huang, J. C. Zheng, S. D. Sun, and C. Y. Chen, "Optimized YOLOv3 algorithm and its application in traffic flow detections," *Applied Sciences*, vol. 10, no. 9, Article ID 3079, 2020.
- [14] I. Ahmed, M. Ahmad, and A. Ahmad, "Top view multiple people tracking by detection using deep SORT and YOLOv3 with transfer learning: within 5G infrastructure," *International Journal of Machine Learning and Cybernetics*, vol. 12, pp. 1–15, 2020.
- [15] M. Imani, "Difference-based target detection using Mahalanobis distance and spectral angle," *International Journal of Remote Sensing*, vol. 40, no. 3, pp. 811–831, 2019.
- [16] Z. Cao, G. Hidalgo, and T. Simon, "OpenPose: realtime multi-person 2D pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2019.
- [17] P. T. Chinimilli, S. Redkar, and T. Sugar, "A two-dimensional feature space-based approach for human locomotion recognition," *IEEE Sensors Journal*, vol. 19, no. 11, pp. 4271–4282, 2019.
- [18] S. B. Thigale, R. K. Pandey, P. R. Gaddekar, V. A. Dhotre, and A. A. Junnarkar, "Lightweight novel trust based framework for IoT enabled wireless network communications," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 7, no. 3, pp. 1126–1137, 2019.
- [19] R. Sun, X. Sun, F. Chen, H. Pan, and Q. Song, "An artificial target detection method combining a polarimetric feature extractor with deep convolutional neural networks," *International Journal of Remote Sensing*, vol. 41, no. 13, pp. 4995–5009, 2020.
- [20] J. Sunseri and D. R. Koes, "Libmolgrid: graphics processing unit accelerated molecular gridding for deep learning applications," *Journal of Chemical Information and Modeling*, vol. 60, no. 3, pp. 1079–1084, 2020.
- [21] M. J. Chen, X. F. Han, and H. Zhang, "Quality-guided key frames selection from video stream by object detection," *Journal of Visual Communication and Image Representation*, vol. 65, Article ID 102678, 2019.
- [22] S. T. Chopparapu and B. Seventline, "Object detection using matlab, scilab and Python," *Technology*, vol. 11, no. 6, pp. 101–108, 2020.
- [23] H. Jiang, T. Qiu, and T. K. Deepa, "Application of deep learning method in automatic collection and processing of video surveillance data for basketball sports prediction," *Arabian Journal for Science and Engineering*, pp. 1–11, 2021.