

Research on the Application of Deep Learning in Ball Games

Hao Chen*

Department of Electronic Engineering, Tsinghua University, Beijing, China

*Corresponding author: hao-chen21@mails.tsinghua.edu.cn

Abstract. Deep learning algorithms are widely used in image detection. Ball games, popular in sports, are amenable to data fiction due to their straightforward interaction models. With AI advancements, deep learning (DL) techniques like AlexNet, SSD, and DNN are increasingly applied in ball sports. These models have excelled in image and speech recognition, and their adaptation to ball sports research has led to advancements in ball classification/localization, trajectory tracking, game video analysis, and decision-making in multiplayer ball games. Deep learning models are able to learn complex features and patterns to detect and locate balls more accurately in images. This is very critical for ball detection in sports games, monitoring systems, and other fields, which can improve the accuracy and reliability of detection. Deep learning models can apply knowledge learned from one task to a related task through transfer learning. This makes it easier for the model to adapt to different types of ball detection problems, even with limited training data. The future of DL in ball sports is bright, promising significant changes and improvements in training, competition, and spectatorship.

Keywords: Deep learning; neural networks; ball sports; video analysis; tactical decision-making.

1. Introduction

As an important part of sports, ball games are sought after all over the world, while most ball games can be simplified into a simple model of the interaction between players and balls and between players and players, which is conducive to data mining. With the rapid development of computer technology and artificial intelligence, deep learning, as an important branch of artificial intelligence, has achieved remarkable results in the application of image recognition, speech recognition, natural language processing, and other fields. Among the common deep learning models are AlexNet, SSD, DNN, each of which has many application scenarios. Researchers have also gradually tried to use deep learning technology in ball games, mainly in ball recognition and analysis, game video analysis, ball trajectory tracking and prediction, and multi-player ball decision-making problems. This paper analyzes the research and use of deep learning in ball sports and studies the application of learning models for ball sports.

2. The Basic Application of Deep Learning in Ball Image

There are many common traditional ball recognition and analysis methods, first of all, the colors and shapes of different balls are more obviously different. A color model called hue saturation intensity can select different colors based on the color threshold [1]. Edge detection is used in image analysis and processing such as feature description, image segmentation, enhancement, and pattern recognition. It can recognize objects based on shape [2].

However, these traditional methods detect a very narrow range of objects, whereas when deep learning is introduced into visual recognition and analysis, the extracted features are differentiated by detecting and fusing features through multiple convolutional and pooling layers. by integrating low-level features to create more abstract high-level features and by reducing the number of parameters that the computer needs to detect to discover distributed features of the data [3]. Then the recognition and analysis of ball images can be done through pre-training trials.

2.1. Deep Neural Networks for Sphere Classification

Tensorflow is an open-source tool with Python as the programming language, as shown in Fig 1, the ball image input into the DNN model that has been trained on a large number of ball image datasets, and after validation is complete, determine whether it is the declared classification of the ball, and if it is not, re-input the input into the DNN model until the output matches with the declared classification, and then the process of classification is completed at one time.



Fig.1 Image classification flowchart [4].

The researcher trained the model with 3495 ball images for classification and finally obtained the results as in Table 1.

Table 1. Training images [4]

Types of balls	Number of images
Baseball	112
Basketball	116
Beachballs	92
Billiard ball	143
Bowling ball	145
Brass	113
Buckey balls	83
Cannon ball	183
Chrochet ball	128
Cricket ball	150
Eye balls	139
Football	112
Golf ball	106
Marble	11
Meat ball	137
Medicine ball	77
Paint balls	32
Puff balls	106
Pokemon balls	86
Rubberband ball	134
Screwballs	140
Sepak takraw ball	141
Soccer ball	167
Tennis ball	139
Tetherball	105
Volleyball	131
Water polo ball	139
Wiffle ball	99
Wrecking ball	69
Crystal ball	161
Total	3495

The original input dataset is compared with this classification result for validation, and finally the result of training and validation accuracy is obtained as in Fig 2.

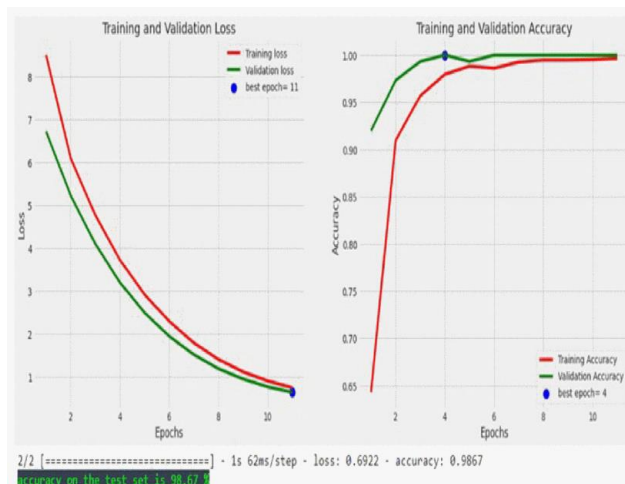


Fig.2 Accuracy results of training and validation [4]

As an example, the basketball in the test is shown in Fig.3

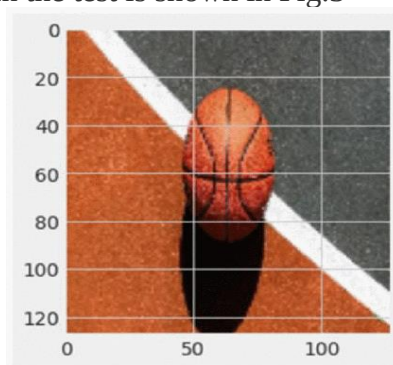


Fig.3 An image representing a basketball [4]

Five randomly selected basketball pictures used for testing were tested and the classification and probability results were obtained as shown in Table 2.

Table 2. Finding out the outcome and probability of classification as basketball [4]

File	Class	Probability
1.jpg	Basketball	95.36%
2.jpg	Basketball	95.93%
3.jpg	Basketball	86.65%
4.jpg	Basketball	92.11%
5.jpg	Basketball	97.97%

In summary, the study shows that the classification accuracy of this DNN model for 30 kinds of ball images is more than 90%, which is much higher than the traditional recognition method based on color shape, and the accuracy of the deep neural network increases with the increase of the dataset used for training, so the use of deep learning can be used to achieve accurate ball classification in ball sports.

2.2. SSD Model Implementation for Localization of Image Spheres

SSD was proposed by Wei Liu in 2016 [5]. SSD focuses on target detection problems including location and classification. SSD is faster and has comparable accuracy to the methods of additional region proposal. Compared to other single-stage methods, SSD is more advantageous in terms of accuracy, even when the input image size is small [6].

In the training phase, SSD uses images of objects and corresponding real labeled boxes. It evaluates a series of default frames with different aspect ratios for each location on feature maps at different scales and learns to predict the shape offsets and confidence levels of the ball categories

corresponding to these default frames. These default frames are then matched with the real labeled frames in the training data.

The SSD model is built on top of a feed-forward convolutional network that generates a fixed-size set of bounding boxes and corresponding category scores and subsequently refines the final detection results through a non-maximization suppression (NMS) step. Within the underlying network structure, auxiliary structures are introduced which are designed based on high-accuracy image classification. Key features of the model include multi-scale feature map detection, convolutional predictor for detection, default frame, and aspect ratio. The default frame design of SSD is similar to the anchor frame in Faster R-CNN [6].

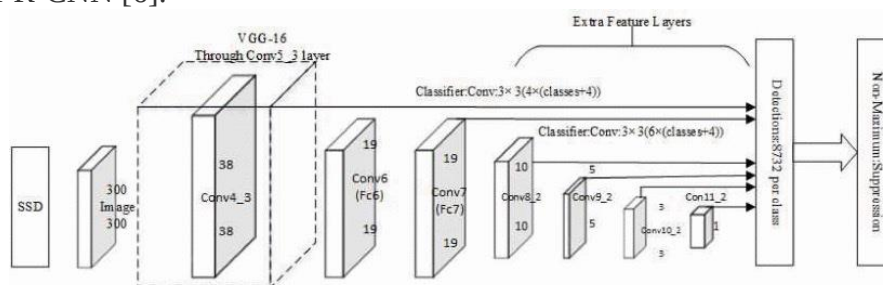


Fig. 4 Structure of the SSD model [6].

The process of the SSD model for ball localization recognition is as shown in Fig 4.

Before starting the deep learning task, it is crucial to prepare the ball dataset. First, a JPEGImages folder is created to store the field ball game images, and LabelImg software is used to generate label files in XML format for each image, which are stored in the Annotations folder. Meanwhile, four TXT files were generated in the ImageSets folder to distinguish the training, testing, and validation sets. Next, the data conversion script is run to generate the dataset in LMDB format. After that, the SSD model was trained using the Caffe framework. After the training was completed, the model was used to predict the test images, and the results are shown in Fig 5, which shows that the balls can be recognized and localized correctly [6].

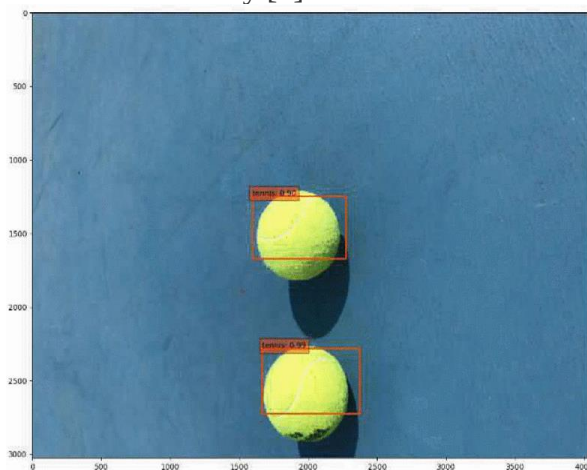


Fig. 5 The result of SSD model to identify the localized tennis ball in the image [6]

Through the application of the above SSD model, the recognition and localization of balls in the image can be realized, and this function can be further extended to realize the recognition and localization of various types of objects in the image. This kind of model also provides the basis for subsequent more practical applications such as tracking of ball trajectories, and video analysis of ball games.

3. Deep Learning Based Trajectory Detection for Small Scale Ball

The realization of ball positioning and movement trajectory tracking is most beneficial for small ball sports such as table tennis and badminton. Simple tracking prediction can not only facilitate the

player's training, but more importantly, it is a mode of detecting small targets in a large context, and after accurately realizing this aspect of the research and application, it can radiate the life, entertainment, and even military industry and other aspects of the research. At the same time, due to the small ball game being more predictable than the larger ball game, perhaps it can become the first station of the future robot to step into the field of a ball game.

With the popularity of deep learning, more and more neural network models are applied. Zhao Hongtu et al proposed a neural network that is based on a proportional conjugate gradient for table tennis target trajectory tracking. By setting a threshold and saving the implicit layer weight matrix, they improved their method and obtained better experimental results [10]. Sheng Xinjun et al. used a YOLOv3 convolutional neural network for ping pong ball detection and achieved 97.1% accuracy [11]. Lin Hsien'i et al. proposed a neural network based on an inverse propagation neural network-based table tennis ball trajectory detection scheme with 88% accuracy [12]. These methods have reached a relatively high level of accuracy, but are easily affected by the complex environment in the background.

3.1. Positioning the Table Tennis Ball

Focusing on the real-time requirements of table tennis ball detection, YOLOv5s was improved by adding a convolutional block attention module after the initial convolutional layer of the backbone network to create a YOLOv5s network based on the attention mechanism to enhance the detection of table tennis ball trajectories. [13]. For the detection of table tennis balls in motion, the background has a significant effect on the results due to the small target in the background.

To enhance the table tennis ball features, a convolutional block attention module is introduced to attenuate the background interference. The attention mechanism simulates the human eye to focus and filter the key information. The convolutional block attention module combines the advantages of both to improve performance and occupies little space[14].

3.2. Tracking of Table Tennis Balls

Tracking using positional information alone may result in insufficient accuracy. To improve the accuracy, the hyper pixel technique was used in tracking to optimize the results. The ping-pong image is segmented by a simple linear iterative clustering algorithm to extract the texture and combined with a K-mean clustering algorithm for hyperpixel clustering to accurately extract the target region [15][16]. After segmentation into multiple hyperpixels, the confidence of the frame can be obtained by calculating the confidence of individual hyperpixels, and the appearance model of the front and back frames is similar due to the short time between frames. Using the confidence map of the previous frame, then sampled and weighted by particle filtering, the maximum a posteriori estimation of the current frame is performed to find the best target. After obtaining the confidence map, the candidate region of the current frame is normalized and the super pixel confidence of each sampled particle is accumulated to compute the most probable particle to determine the target region and complete the super pixel-based tracking process [13], as shown in Fig 6.

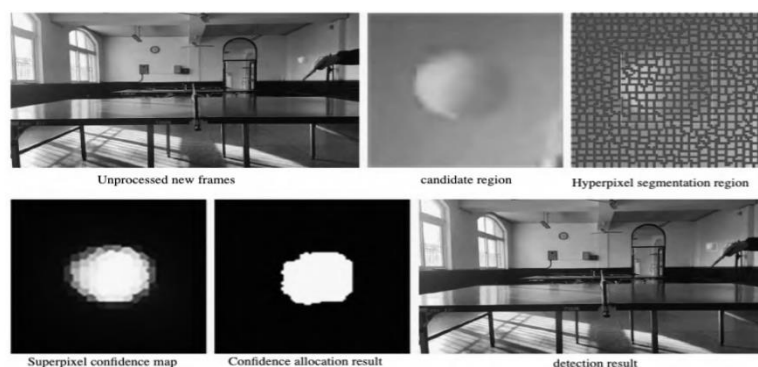


Fig. 6 Tracking process based on hyperpixel [13]

In table tennis ball trajectory detection, the occlusion problem is determined by setting a threshold. If the occlusion exceeds the threshold, the confidence of the previous frame is used as the current frame estimate.

3.3. Validating Deep Learning and Hyperpixel-Based Detection Methods

Firstly, to verify the localization function of the detection method based on deep learning with hyper pixel, combined with the table tennis game video production dataset, 1000 table tennis images were selected, and the experimental comparison of the detection methods based on color detection, based on contour detection, YOLOv4 algorithm, YOLOv5s algorithm, and Faster region convolutional neural network with the detection method based on deep learning with hyper pixel were obtained as follows as the results in Table 3.

Table 3. Comparison of table tennis ball positioning methods [13]

Method	Accuracy	Single frame detection Time/ms	Model size /MiByte
Based on color detection	55.50%	8.3	0.2
Based on contour detection	64.50%	9.5	0.2
YOLOv4 algorithm	97.20%	27.4	244
YOLOv5s algorithm	96.00%	15.2	13.7
Faster regional convolutional neural network	75.10%	195	113.8
Detection method based on deep learning and superpixels	98.10%	15.5	13.8

To verify the ability of the deep learning and hyper pixel-based detection methods to deal with occlusion and for ping pong ball tracking in complex environments, the YOLOv4 algorithm, the YOLOv5s algorithm, the Faster region convolutional neural network, and the deep learning and hyper pixel-based detection methods are tested in a comparative manner using the dataset from the previous experiment, and the results are obtained as shown in Table 4.

Table 4. Comparison of ping pong ball tracking methods [13]

Method	Center position Pixel error	Detection speed /(frame·s ⁻¹)	model size /MiByte
YOLOv4 algorithm	15 pixel	42	244
YOLOv5s algorithm	35 pixel	70	13.7
Faster Regional Convolutional Neural Network	—	5	113.8
Detection method based on deep learning and superpixels	12 pixel	65	1.3

Through the above two experiments, we can get that the detection method based on deep learning and superpixel has the highest accuracy rate and the detection time is only slightly higher than the YOLOv5s algorithm in the localization function compared with other methods, which meets the demand of fast and accurate localization of ping pong balls. In terms of trajectory tracking, the detection method based on deep learning and super pixel has the smallest model size, the detection rate is slightly slower than the YOLOv5s algorithm, but the pixel error at the center position is the lowest, which can better deal with the occlusion and all kinds of interfering factors, and adapt to different application scenarios, and it can effectively realize the fast and accurate detection of table tennis ball's trajectory.

The method of deep learning to detect the movement trajectory of the ball in a small range is also essentially a kind of small target detection in a large context, which is very helpful for the

development of the level of small ball players. Future research in the direction of more intelligent robots can be realized by combining the development of hardware.

4. Deep Neural Network Based Ball Video Analysis

4.1. ROI-based Video Behavior Recognition Network

The method constructs a video behavior recognition network for regions of interest (ROIs) from time-series data of individual actions. The core of the network employs an optimized multilevel fully convolutional network (MFCN) to extract multilevel features from video frames, supplemented by a deep fully convolutional network (DFCN) to initially identify athlete ROIs in the video. Further, the detected ROIs are pinpointed by a Markov Random Field (MRF) to obtain the complete set of athlete ROIs. Based on this, the method turns to analyzing the time series data of individual actions and performs inferential analysis on the ROI time series of the target athletes in order to predict the specific behaviors of the individuals. Ultimately, the network utilizes a specially designed ROI-matching recurrent neural network (ROI-match RNN) to model the time series of group behaviors and output predictions of group behaviors [17]. The general flow is shown in Fig 7.

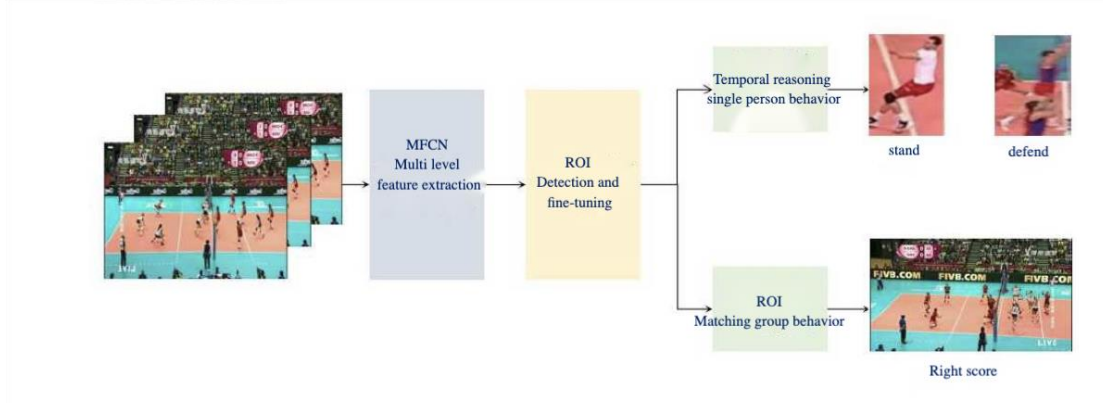


Fig. 7 ROI-based recognition network to analyze the video process [17]

4.2. Recognition Network Based on Dual Attention Mechanism

In order to enhance the discriminative and expressive power of features, a group behavior recognition network applied to video is proposed, which integrates two attention mechanisms, spatial attention and channel attention. The spatial attention model is designed to capture the interactions between different pixel points in a video frame, while the channel attention model is used to model the interdependencies between different channels in a feature map. In this way, the network can focus on key features while suppressing unimportant features, thus extracting key behavioral information in the video more effectively [17]. The specific process is as follows in Fig.8

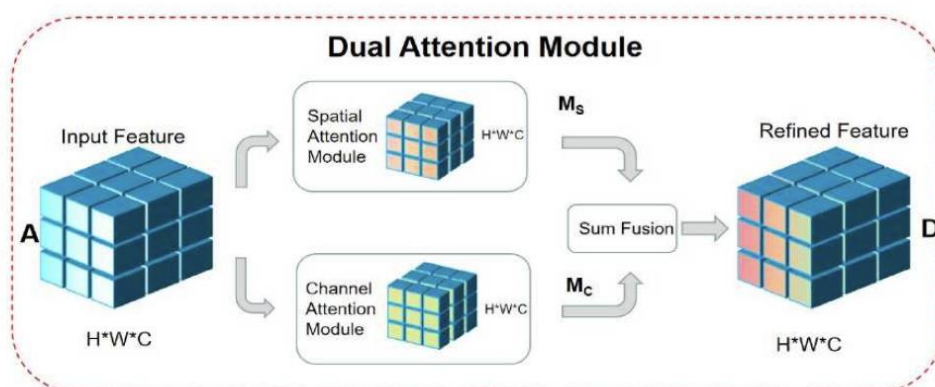


Fig.8 Recognition network based on dual attention mechanism [17]

4.3. Fusion of ROI-based Recognition Network

The dual attention-based video group behavior recognition network is introduced into the ROI-based video behavior recognition network to get the fusion network. After conducting comparison experiments of the attention model placed after the input image, after the input network; after global features; and after local features. When the dual attention model is applied after the input image, the group recognition accuracy is improved by 1.1% and 0.5%, while the individual recognition accuracy reaches 81.1% and 74.3%, respectively. However, when the dual attention model was located after global feature extraction, the group recognition accuracy reached 79.5% and 85.1%, while the individual recognition accuracy decreased to 74.6% and 74.5%, respectively. when the dual attention model was applied to local features, the group recognition accuracy increased by 0.3% and 2%, respectively, and the individual recognition accuracy increased by 0.1% and 2%, respectively [17]. In summary, placing the dual-attention model after local features and before ROI-based temporal inference of individual behaviors, and fusing it to form the final dual-attention-based video group behavior recognition network can be optimal.

Comparing this network with some advanced algorithms for video data analysis, as follows in Table 5.

Table 5. Comparison of the recognition network approach [17]

Method	Group Accuracy	Individual Accuracy	Time
Social Scene (2 groups)[Baseline]	83.8%	0.811	2016
Two-stage Hierarchical Model(2 groups)	81.9%		2016
CERN (1 group)	73.5%	690%	2017
CERN(2 groups)	83.3%	69%	2017
Social Scene(2 groups)	83.8%	81.10%	2017
Hierarchical LSTM (1 group)	70.3%		2018
Hierarchical LSTM V3 (1 group)	74.01%	75.96%	2018
SRNN-MaxNode (1 group)	74.39%	76.65%	2018
SRNN-MaxEdge (1 group)	68.39%	76.03%	2018
Hierarchical LSTM[12](2 groups)	81.9%		2018
Hierarchical LSTM V3 (2 groups)	83.12%	75.96%	2018
SRNN-MaxNode (2 groups)	83.47%	76.65%	2018
SRNN-MaxEdge 2(2 groups)	79.86%	76.03%	2018
B1(w/o HANs)(2 groups)	84.1%		2018
B2(w/o HCNs)(2 groups)	82.5%		2018
HANs+HCNs(2 groups)	85.1%		2018
Hybrid fusion (HF)(S+T)(HF)			
	70.2%		2019
+tRCNN(LF sum)			
Ours DA+SPTI(2 groups)	87.800		2020
Ours DA+SPTI(2 groups)	85.89%	83.800	2020

The experimental results show that the dual-attention model can effectively improve the accuracy of group and single behavior recognition in volleyball videos. In the group behavior recognition task, the model reaches 87.8% accuracy, showing strong competitiveness. When evaluating the two tasks simultaneously, the model reaches 85.8% for group recognition and 83.8% for single-person recognition, which are 5.0% and 2.5% higher than the baseline, respectively, validating the effectiveness of the algorithm. There is also a significant improvement compared with models such as Two-stage Hierarchical Model, CERN, Social Scene, and Hierarchical LSTM. Compared to the baseline model, the Dual Attention Model improved the accuracy by 5.0% for group behavior recognition and 2.5% for single-person behavior recognition. These results confirm that utilizing the

attention mechanism and individual temporal information can effectively enhance the feature representation and thus significantly improve the accuracy of recognition. [17]

The above algorithms can fulfill the role of data analysis for ball sports videos. They can assist professionals in the sports world to complete the live annotation and a series of data analyses of players and teams. In terms of technology, it lays the foundation for maximizing the profitability of decision-making by analyzing game videos and assists the coaching team in making tactical arrangements for players. If the technology is more mature, it can be used on more important occasions with large crowds to complete more complex analyses.

5. Conclusion

This paper explores deep learning applications in ball sports, covering neural networks, convolutional neural networks, and recurrent neural networks. DL models for ball image recognition, classification/localization, trajectory tracking, video analysis, and tactical guidance. Experiments validate the high accuracy and practicality of DL in ball games. Despite significant achievements, DL's full potential in ball sports remains untapped, with the potential for more accurate action recognition, analysis, game data prediction, player evaluation, and intelligent strategy planning. The integration of DL with virtual/augmented reality and hardware research can enhance training, gaming experiences, and human-machine interaction, potentially leading to the development of intelligent robots. The future of DL in ball sports is promising, with the potential to revolutionize training, competition, and spectatorship.

References

- [1] L. Fajie and C. Wanmi, "Research and design of basketball robot vision system on labview", *China Academic Journal Electronic Publishing House*, 2014,8:10-11.
- [2] C. Yuyun, *Research on Edge Detection of Gray Image*, 2009.
- [3] C. Zhen, X. Jingbo, B. Jun and X. Min, "Feature extraction algorithm based on evolutionary depth learning", *Computer Science*, 2015,42(11): 288-292.
- [4] S. L. Srikar, P. S. Gowtham, D. S. Swetha, M. H. Sri, B. B. Kumar and S. Bulla, "Deep Neural Networks (DNN) based Sports Balls Classification," *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2023.
- [5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, et al., *Ssd: Single shot multibox detector*, 2015.
- [6] S. Gu, L. Ding, Y. Yang and X. Chen, "A new deep learning method based on AlexNet model and SSD model for tennis ball recognition," *2017 IEEE 10th International Workshop on Computational Intelligence and Applications (IWCIA)*, Hiroshima, Japan, 2017.
- [7] Zhang Xuefeng, Dai Liangliang, Zhou Jianqin, et al. Fast Detection and Real-Time Tracking of Table Tennis Balls under Multi-color Models[J]. *Computer system applications*, 2018,27(5):232-237.
- [8] TEBBE J,GAO Y P,SASTRE- RIENIETZ M,et al.A Table Tennis Robot System Using an Industrial KUKA Robot Arm[C]//GCPR 2018:German Conference on Pattern Recognition,Stuttgart,2018.
- [9] Wang Binsheng, Liu Jianan, Liu Ji, et al.High-speed vision based table tennis robot[J].*military-industrial automation*,2018,37(3):60-63.
- [10] ZHAO H T,HAO F.Target Tracking Algorithm for Table Tennis Using Machine Vision[J/OL].*Journal of Healthcare Engineering*,2021.
- [11] SHENG X J,ZHU X Y,YANG L,et al.Feature Fusion Based Efficient Convolution Network for Real-time Table Tennis Ball Detection[C].2020 13th International Symposium on Computational Intelligence and Design (ISCID),Hangzhou,2020.
- [12] LIN H I,HUANG Y C.Ball Trajectory Tracking and Prediction for a Ping- Pong Robot[C]//2019 9th International Conference on Information Science and Technology (ICIST),Hulunbuir,2019.

- [13] Hu Xiaopeng,Zhao Yongjie,Huang Kexin,et al.Deep Learning and Hyperpixel Based Table Tennis Ball Motion Trajectory Detection[J].machinery manufacturing,2022,60(08):81-86.
- [14] WOO S H,PARK J C,LEE J Y,et al.CBAM:Convolutional Block Attention Module[C]//ECCV 2018:European Conference on Computer Vision,Munich,2018.
- [15] YIN J J,WANG T,DU Y L,et al.SLIC Superpixel Segmentation for Polarimetric SAR Images[J/OL].IEEE Transactions on Geoscience and Remote Sensing,2021.
- [16] Yang Junchuang, Zhao Chao. Review of Research on Clustering Algorithms[J].Computer Engineering and Applications, 2019,55(23):7-14,63.
- [17] Liu Yan. Research on Deep Learning Based Behavior Recognition Algorithm for Volleyball Videos[D].Shenzhen University,2021.
- [18] Fernández J, Bornn L, Cervone D. Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer[C]//13th MIT Sloan Sports Analytics Conference. 2019.