Research Article

Video Analysis in Sports by Lightweight Object Detection Network under the Background of Sports Industry Development

Yifei Zheng [iD][1] and Hongling Zhang [iD][2]

[1]*Physical Department, Chang'an University, Xi'an 710064, Shaanxi, China*
[2]*Physical Institute, Yan'an University, Yan'an 716000, Shaanxi, China*

Correspondence should be addressed to Yifei Zheng; zhengyifei@chd.edu.cn

This study uses the video image information in sports video image analysis to realize scientific sports training. In recent years, game video image analysis has referenced athletes' sports training. The sports video analysis is a widely used and effective method. First, the you only look once (YOLO) method is explored in lightweight object detection. Second, a sports motion analysis system based on the YOLO-OSA (you only look once-one-shot aggregation) target detection network is built based on the dense convolutional network (DenseNet) target detection network established by the one-shot aggregation (OSA) connection. Finally, object detection evaluation principles are used to analyze network performance and object detection in sports video. The results show that the more obvious the target feature, the larger the size, and the more motion information contained in the sports category feature, the more obvious the effect of the detected target. The higher the resolution of the sports video image, the higher the model detection accuracy of the YOLO-OSA target detection network, and the richer the visual video information. In sports video analysis, video images of the appropriate resolution are fed into the system. The YOLO-OSA network achieved 21.70% precision and 54.90% recall. In general, the YOLO-OSA network has certain pertinence for sports video image analysis, and it improves the detection speed of video analysis. The research and analysis of video in sports under the lightweight target detection network have certain reference significance.

1. Introduction

With the continuous maturity of "Internet +" technology, the sports industry has gradually developed in the direction of wisdom, inclusiveness, and services. Smart technology is used to collect and distinguish the needs of mass sports accurately. The concept of sharing integrates sports resources and realizes the online and offline integration of the sports industry by enriching sports service projects and innovating development concepts. Based on meeting the needs of mass sports, it realizes the multiple values of the sports industry [1–3]. With the help of social media, the public has become an important force in promoting the high-quality development of the sports industry. The public uses pedometer applications (APP) and social media platforms to share and exchange sports experiences. This behavior has formed a new sports social model, helping sports become a way of life for the public to share and communicate [4, 5]. However, when videos in sports are analyzed, there are certain drawbacks. Most sports videos require the experience of professional scouts [6]. With the gradual penetration of intelligent video analysis applications in various industries, from sports to climate science, intelligent video analysis applications in various industries are changing the future [7]. Intelligent video analytics is also known as visual artificial intelligence (AI), or visual AI for short. This technology can turn complex video data into actionable insights to help humans manage urban traffic, improve urban planning, increase safety in public spaces, and even curate theatrical performance content [8, 9]. Visual AI can improve the operational efficiency of sports venues, improve the quality of live streaming, and evaluate athlete

performance for athletes and fans of the sports world. It also allows users to exercise at home to adjust their workouts under the guidance of a fitness trainer [10].

The Swiss "Dartfish" is the most advanced and widely used professional motion video analysis system in the world. It is powerful and comprehensive, with professional, project tactical analysis, training, and other functions, especially multiscreen comparison, video overlap analysis, key action decomposition, video marking, and other functions, which is particularly suitable for the sports field [11]. Therefore, Dartfish not only uses diagrams for analysis and time stamping but also enables real-time actions to be compared with typical characteristics, allowing experts and athletes to see the most important elements immediately [12]. Bilai and Hanif used histogram of oriented gradient (HOG) for feature extraction and support vector machine (SVM) as a classifier for pedestrian detection. After extensive testing, HOG hybrid SVM is a pedestrian detection method with a better balance between speed and effect [13]. In 2019, fully convolutional one-stage object detection (FCOS) in object detection algorithms prevailed in the society. The algorithm uses detection heads with shared weights, that is, using the same set of convolutions to predict detection boxes for the multiscale feature map of feature pyramid network (FPN). Then, each layer uses a learnable scale value to solve the object detection problem in a pixel-by-pixel prediction manner, like semantic segmentation [14]. The FCOS algorithm is not the first to propose the use of a full-scroller network for object detection. DenseBox is also an algorithm based on fully convolutional network (FCN) [15]. FCOS will predict a 4-dimensional vector plus a piece of category information for each point on the feature map, that is, each point on the feature map of width by height ($W \times H$). Channels are set to 5. In fact, this is like semantic segmentation, except that there is four more bounding box offset information [16]. When FCOS is lightweight, the model effect is not as good as expected because the centerness branch of FCOS is difficult to converge on lightweight models [17]. Chen and Qin proposed the generalized focal loss function. This function can remove the centerness branch of FCOS and save a lot of convolutions on this branch, thereby reducing the computational overhead of the detection head, which is very suitable for lightweight deployment on mobile terminals [18]. Koo et al. used you only look once (YOLO) to achieve real-time object detection used in cutting-edge technologies such as self-driving cars [19].

Target detection and recognition methods based on deep learning have become mainstream. The study deeply discusses the regression-based target detection and recognition algorithm and optimizes the YOLO method. The lightweight target detection method is applied to sports, which not only combines the current technical methods with practice but also highlights the characteristics of the era of sports operation. These provide a video analysis research method for sports and a new path for the economic development of the sports industry.

This study aims to use the video image information in sports video image analysis to realize scientific sports training. First, the YOLO method in lightweight object detection is explored. Second, based on the one-shot aggregation (OSA) connection, the dense convolutional network (DenseNet) is established, and the sports analysis system is established based on the you only look once-one-shot aggregation (YOLO-OSA) target detection network. Finally, object detection evaluation principles are used to analyze network performance and object detection in sports video. This study has certain reference significance for the research and analysis of video in sports under the lightweight target detection network.

2. Object Detection Network Method

2.1. YOLO Target Detection. Deep learning target detection has been developed for many years, from two-stage to one-stage, from anchor-base to anchor-free, and then to a transformer for target detection. Various methods are blooming. However, in mobile target detection algorithms, anchor-base models such as the YOLO series and single shot multiBox detector (SSD) have always dominated [20]. YOLO can input a picture and directly output the result, including the name and score of the box and objects in the box [21]. The principle of YOLO target detection is shown in Figure 1.

In Figure 1, first, the size of the sports picture is fixed. The input image is divided into an $n * n$ grid, and classification and localization are performed on each grid. Compared to sliding windows, this greatly reduces the amount of computation. Before YOLO performs detection, the image has meshed. When an object's center falls within a grid, that grid is responsible for predicting the object. Each grid needs to predict a bounding box, and each bounding box needs to predict a confidence value in addition to its own position. This confidence value represents the confidence of the object in the predicted box (what object is contained in the grid and the accuracy of predicting this object) and the multiple information predicted by this box. By dividing the image into multiple blocks, a tensor with the same number of blocks is directly output. The prediction vector of each block determines whether it has a target and what type it is, and the correct calculation of the specific position is shown in the following equation:

$$\Pr(\text{Object}) * IOU^{\text{truth}}_{\text{pred}}. \tag{1}$$

In the equation, if an object falls in a grid unit, then the first item takes 1; otherwise, it takes 0. The second term is the intersection over union (IoU) value between the predicted bounding box and the actual bounding box. When there is no object in the box, the entire confidence will become 0. However, with the evolution of time, there is a lack of lightness in the application, and the practical application is not strong.

2.2. Lightweight Object Detection Network Based on OSA Connection. Different networks are introduced to optimize the object detection model. The inception series is different from the residual network (ResNet). The inception network mainly improves the network structure from the width of the

FIGURE 1: The principle of YOLO target detection.

network to improve the expressive ability of the network. ResNet mainly improves the network structure from the depth of the network to improve the expressive ability. DenseNet explores the potential of the network through feature map reuse. DenseNet can make the input of each layer of the network become the superposition of all previous layers. Then, its feature map is passed to all subsequent network layers [22]. The 5-layer dense block structure of the DenseNet network with a growth rate of 4 is shown in Figure 2.

In Figure 2, a growth rate of 4 means that the output feature map dimension of each dense layer is 4. A 5-layer dense block represents 5 BN-ReLU-Conv (3 ∗ 3) layers. All layers are directly connected to ensure the maximum information flow between layers. In order to maintain the feed-forward feature, the input of each layer is the output of all previous layers, and its own feature map results are also used as the input of the subsequent layers. DenseNet consists of dense blocks. In these blocks, the layers are closely connected, and each layer takes its input from the output feature map of the previous layer. This extreme reusability of residuals results in deep supervision, with each layer receiving more supervision from the previous layer, so the loss function will react accordingly. In dense blocks, a single block consists of batch normalization, ReLU activation, and 3 × 3 convolution. *Transition Layer*. In ResNet, the summation of residuals is performed, while DenseNet concatenates all feature maps together. Batch

normalization, convolutional layer, and average pooling constitute the transition layer. DenseNet's convolutions generate fewer feature maps. DenseNet has lower requirements for wide layers. With tight connections between layers, it learns features with little redundancy. The number of output feature maps of a layer is defined as the growth rate. Ultimately, the growth rate controls how much new information each layer contributes to the global [23]. The input and output transformation graphs of the first convolutional layer in the dense block are shown in the following equation:

$$x_l = H_l\left(\left[x_0, x_{1,\dots,x_{l-1}}\right]\right). \tag{2}$$

In the equation, $[x_0, x_{1,\dots,x_{l-1}}]$ is the output feature of the 0th layer to the $l-1$th layer in the module. H_l completes nonlinear transformation for the lth convolutional layer. The residual connection module of the input and output transformation of the first convolutional layer is shown in the following equation:

$$x_l = H_l\left(x_{l-1}\right) + x_{l-1}. \tag{3}$$

In the equation, l is the representation layer, x_l is the output of the l layer, x_{l-1} is the output of the $l-1$ layer, and H_l is a nonlinear transformation. Therefore, for ResNet, the output of layer l is the output of layer $l-1$ plus a nonlinear transformation of the output of layer $l-1$. The core module of DenseNet is the dense block. This dense connection
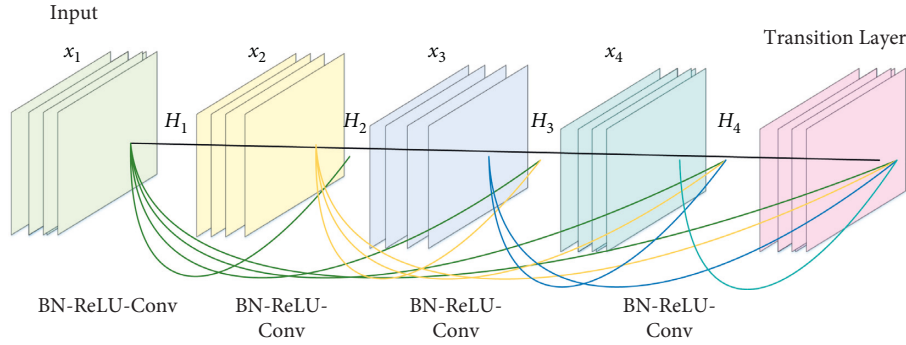
FIGURE 2: DenseNet 5-layer dense block structure with a net growth rate of 4.

aggregates all previous layers, resulting in a linear increase in the input of each layer. The output of each layer is fixed size, limited by floating-point operands and model parameters. The problem with it is that the input and output are inconsistent. Currently, the network model is not optimal. In addition, due to the large input, DenseNet uses a 1×1 convolutional layer to compress the features first. The introduction of this extra layer is detrimental to the efficient computation of the graphics processing unit (GPU). Therefore, although DenseNet's floating-point operations and model parameters are not large, the inference is not efficient. When the input is large, more memory and inference time are often required. Therefore, although the floating point operations (FLOPs) and model parameters of DenseNet are not large, the inference is not efficient. It tends to require more memory and inference time when the input is larger. The OSA connection is proposed, as shown in Figure 3.

In Figure 3, F is a floating-point operand. X is the output value of all layers. That is, all previous layers are aggregated only at the last time. This connection change will solve the problems described earlier in DenseNet. The number of input channels of each layer is fixed, assuming that the number of output channels is consistent with the input to obtain the smallest memory access cost and no longer needs 1×1 convolutional layers to compress features. The OSA module is computationally efficient [24]. The OSA module is shown in Figure 4.

In Figure 4, the module contains m convolutional layers. f is the output of each convolutional layer along with the input channel. g is the channel number of the input module. At the end of the OSA module, the outputs of the m convolutional layers are superimposed with the input channels of the OSA module. If the number of superimposed is too large, a $1 * 1$ convolutional layer is added at the end of the OSA module for the compression module. Finally, each stage uses a max-pooling layer for downsampling. Like other networks, the number of feature channels is increased after each downsampling.

2.3. Sports Video Analysis System. Sports is a social activity, in which people follow the growth and development laws of the human body and the laws of physical activity through physical exercise, technology, training, competitive competitions, and other methods to enhance physical fitness, improve the level of sports technology, and enrich cultural life. It is a general term for all competitive physical activities that provide entertainment to participants through the use, maintenance, or improvement of physical fitness, and organized or unorganized participation. There are hundreds of sports in the world, and their forms vary widely. Some only require two people to complete, while others require multiple teams to complete [25]. The main categories of sports are listed in Table 1.

In Table 1, different sports models have the characteristics of fast target movement and fast change of target speed in sports videos. Therefore, four motion models such as uniform velocity, uniform acceleration, rest, and collision are proposed. In sports events, single data statistics and unstructured video recordings have certain limitations and do not produce much value. Combining statistics and video to structure and refine video data can assist sports events and sports training and provide valuable solutions. The 3D panoramic video surveillance and video processing flow are shown in Figure 5.

In Figure 5, the video image is preprocessed and noise removed. The geometrically distorted images are corrected. Additionally, a histogram equalization image enhancement algorithm is incorporated. The images with lower definition are image enhanced, and the frame image registration algorithm is used to complete the image feature point detection, feature description, and feature matching. The overlapping parts between the images are aligned. The image to be stitched is transformed into the coordinate system of the reference image to form a complete image. The frame image fusion algorithm is used to fuse the overlapping areas of the splicing so that the color gradually transitions, ensures the smooth transition of the splicing boundary, eliminates the splicing line, and generates a smooth and seamless panoramic video image. Multiple monitoring probes are equipped with an embedded processor, a video acquisition module connected with the embedded processor for video image acquisition in the monitoring area, and a video processing module for video processing of the collected video images and positioning of the probe position [26]. In addition, for moving video, the principle diagram of the background difference method in the process of target detection is shown in Figure 6.
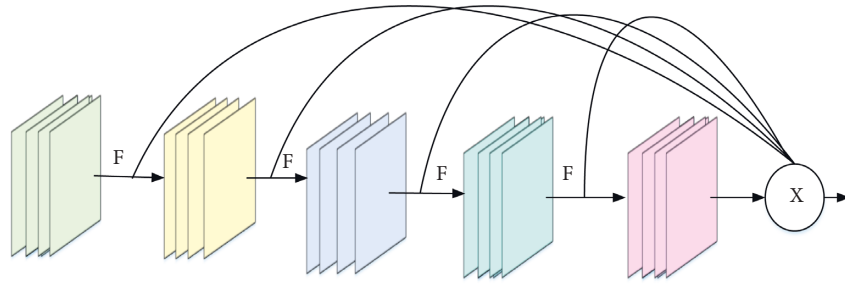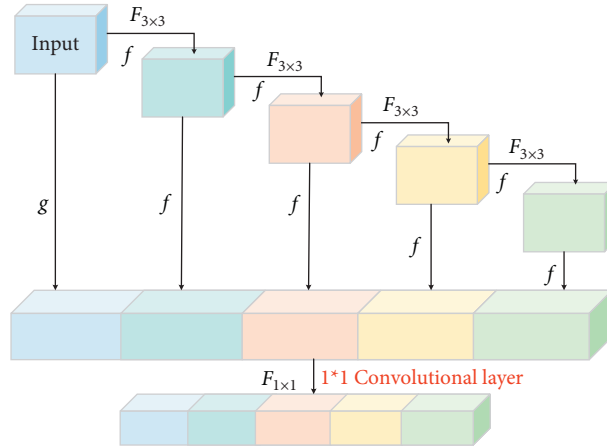
FIGURE 3: OSA connection.

FIGURE 4: OSA module.

TABLE 1: Main categories of sports.

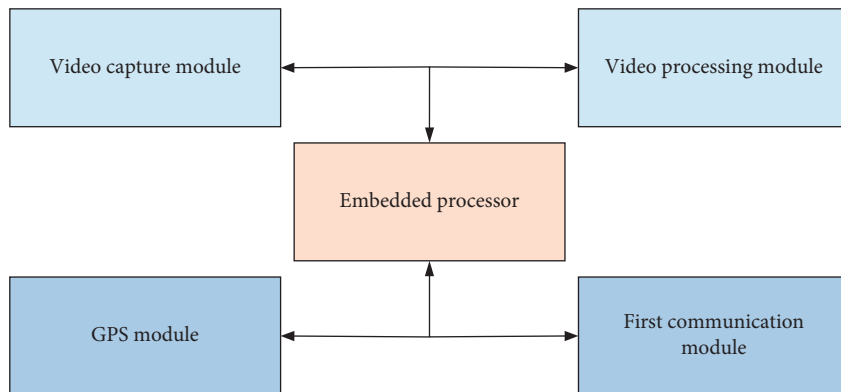| Sports category	Features
Athletics	Adversarial, mobilizing, technical, international, and recognized
Entertainment sports	Amateur, recreational, and recreational
Popular sports	Extensive
Medical sports	Active, restorative, and therapeutic

FIGURE 5: 3D panoramic video surveillance and video processing flow.

In Figure 6, the basic idea of the background difference method is to compare the input image with the background model. The moving objects are segmented by judging changes in features such as grayscale or using changes in statistical information such as histograms. First, the background image frame B and the stored background image are established [27]. The difference in image calculation is shown in the following equation:

$$D_n(x, y) = \left| f_n(x, y) - B(x, y) \right|, \quad (4)$$

‖

+

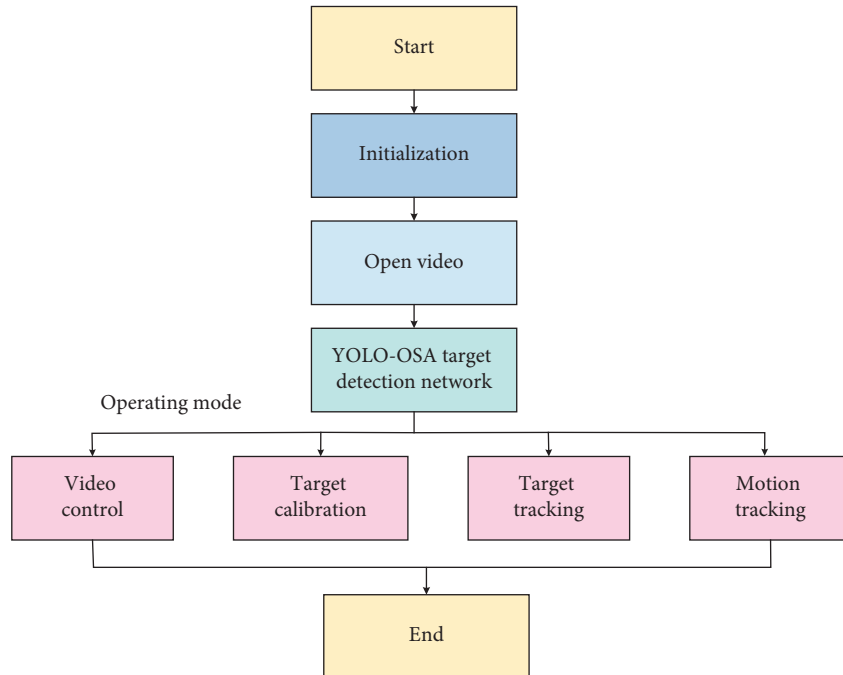FIGURE 6: Schematic diagram of background difference method.

FIGURE 7: Sports analysis system based on YOLO-OSA target detection network.

where $B(x, y)$ and $f_n(x, y)$ are the grayscale values of the pixels corresponding to the background frame and the current frame. $D_n(x, y)$ is the grayscale value of the differential image. The binarization processing calculation is shown in the following equation:

$$R_n'(x, y) = \begin{cases} 255, & D_n(x, y) > 1, \\ 0, & \text{else}. \end{cases} \quad (5)$$

In the equation, $R_n'(x, y)$ is a binarized image, in which the point with a gray value of 255 is the foreground (moving target) point, and the point with a gray value of 0 is the background point. When the subtraction of the current frame and the background image is greater than a certain threshold, the pixel is determined to be the foreground target: the input image is subtracted from the background image, and T is the threshold or difference. After
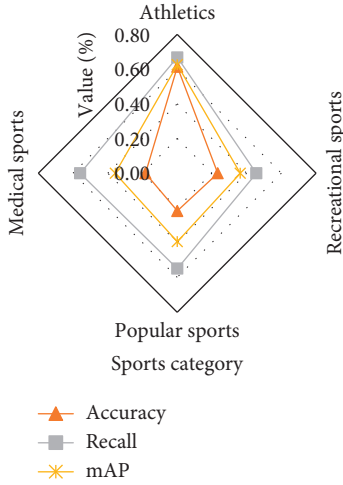
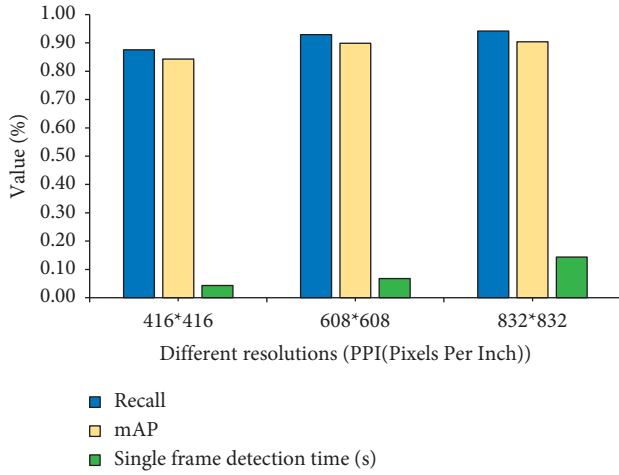FIGURE 8: Sports video target detection analysis.

FIGURE 9: Influence of sports video images at different resolutions on the YOLO-OSA target detection network.

thresholding, the point corresponding to the 0 value in the image is the background image, and the point corresponding to the 1 value is the moving pixel in the scene. The sports analysis system based on the YOLO-OSA target detection network is shown in Figure 7.

In Figure 7, target detection in sports videos is solved based on the YOLO-OSA target detection network. First, the video obtained through the camera is punched. Second, the target detection network enters the working mode to perform sports video control, target calibration, target tracking, and motion tracking analysis.

2.4. Principles of Model Evaluation. The IOU of the standard frame predicted by the system model and the real frame is greater than or equal to 0.5, and the sports target recognition is correct. The system adopts the accuracy rate, precision rate, recall rate, and average precision as the system evaluation principles [28]:

(1) The accuracy rate is the percentage of correctly predicted samples in the total samples. Although the accuracy rate can judge the correct overall rate, in the case of unbalanced positive and negative samples, the evaluation index of the accuracy rate has a great defect. For example, in the total sample, positive samples account for 90%, and negative samples account for 10%. In this case, it is only necessary to predict all samples as positive samples to get high accuracy of 90%. The calculation is shown in the following equation:

$$A\,(\text{accuracy}) = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{6}$$

True positive (TP) is the number of positive samples that are correctly identified. True negative (TN) is the number of correctly identified negative samples. False positive (FP) is the number of negative samples that are incorrectly identified as positive samples. False negative (FN) is the number of positive samples that are falsely identified as negative samples.

(2) *Accuracy Rate.* The accuracy rate is also called the precision rate. The percentage of positive samples that are correctly predicted to all detected (predicted as positive) samples is for the prediction results. The calculation is shown in equation (7):

$$P\,(\text{precision}) = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{7}$$

(3) *Recall Rate.* The recall rate is the proportion of positive samples that are correctly predicted to all positive samples (ground truth). It is for the original sample. The calculation is shown in the following equation:

$$R\,(\text{recall}) = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{8}$$

(4) *Average Precision.* Average precision is a performance metric for this class of algorithms that predict target locations as well as classes. The calculation is shown in the following equation:

$$AP\,(\text{average precision}) = \frac{\sum_{0,0.1,\ldots,1.0} P_{\text{smooth}}\,(i)}{11}. \tag{9}$$

In the equation, P is the value of different recall rates. $\sum_{0,0.1,\ldots,1.0} P_{\text{smooth}}\,(i)$ is the precision of the recall rate, Recall = {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1}. Mean average precision (MAP) is the average of multiple categories of AP. Mean means that the AP obtained under each category is averaged. The value range of MAP is [0, 1]. In this range, the larger the MAP value, the higher the average accuracy.

3. Results and Discussion

3.1. Analysis of Sports Video Target Detection under the YOLO-OSA Target Detection Network. Through the sports video analysis under different sports types under the YOLO-OSA
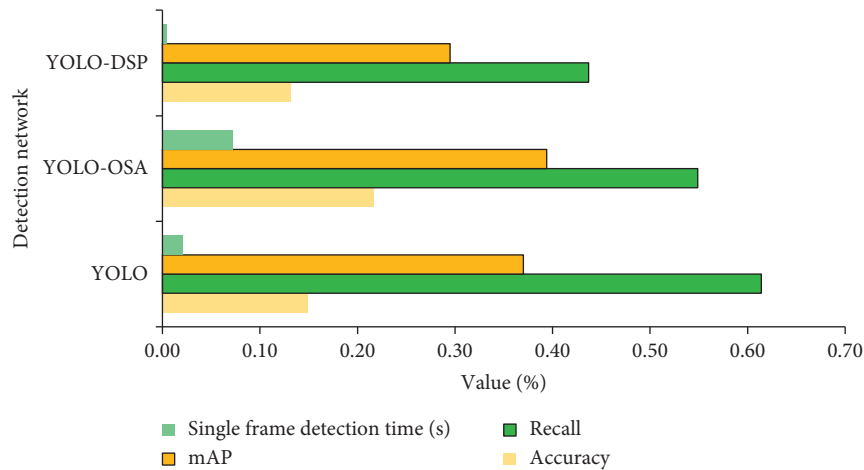
FIGURE 10: Performance comparison of sports video analysis under different networks.

target detection network, the sports video target detection analysis is shown in Figure 8.

Figure 8 shows the detection accuracy of video objects in each category of sports by YOLO-OSA. The target detection accuracy rate of competitive sports is 61.30%, the recall rate is 66.80%, and the MAP value is 62.50%. The accuracy rate of entertainment sports and target detection is 23.30%, the recall rate is 45.50%, and the MAP value is 36.30%. The accuracy rate of mass sports target detection is 21.70%, the recall rate is 54.90%, and the MAP value is 39.40%. The accuracy rate of medical sports target detection is 18.30%, the recall rate is 55.90%, and the MAP value is 35.70%. According to the sports category characteristics, the more obvious the target features, the larger the size, and the more motion information contained, the more obvious the effect of the detected target.

3.2. The Influence of Different Sports Video Input Resolutions on the YOLO-OSA Target Detection Network. Sports video images at different resolutions test the performance of the YOLO-OSA object detection network. The "horizontal pixel number * vertical pixel number" of the video image is used to express the video image, and the videos with different pixel sizes of $416 * 416$, $608 * 608$, and $832 * 832$ are detected, respectively. The influence of sports video images at different resolutions on the YOLO-OSA target detection network is shown in Figure 9.

In Figure 9, the higher the resolution of the sports video image, the higher the model detection accuracy of the YOLO-OSA target detection network, and the richer the visual video information. However, the higher the resolution of the sports video image, the longer the detection time of a single frame, which means the higher the resolution of the video image. The amount of computation and access to the network reduces the computation speed of the network. Therefore, in sports video analysis, video images of appropriate resolution size are selected for input into the system.

3.3. Performance Comparison of Sports Video Analysis under Different Networks. The core idea of YOLO is to use the entire image as the network's input. Multiple downsampling layers are employed at the output layer to regress the location of the bounding box and the class, to which it belongs. The object features learned by the network are not fine, so it also affects the detection effect. The OSA connection and the digital signal processing (DSP) method are used to make up for the shortcomings of the YOLO method. Each stage of the OSA connection uses a max-pooling layer for downsampling and no longer requires a 1×1 convolutional layer to compress features. Therefore, the OSA module is GPU computationally efficient. You only look once-digital signal processing (YOLO-DSP) algorithm is feasible in identifying medium or large-size objects in unrestricted natural scenes. The proposed lightweight target detection network uses the YOLO network, YOLO-OSA network, and you only look once-digital signal processing (YOLO-DSP) network for performance analysis under the sports video dataset. A dataset is obtained from the University of Florida Sports Behavior Database, which consists of a series of actions collected from various sports, containing 150 sequences. The resolution is $720 * 480$, and this set represents the natural action pool in various scenes and viewpoints. The resolution is uniformly set to 800, and the confidence is set to 0.01. The performance comparison of sports video analysis under different models is shown in Figure 10.

In Figure 10, the network structure and the multiscale improved YOLO-OSA network obtained 21.70% precision and 54.90% recall. This is an increase of 6.80% and 6.50%, respectively, compared to YOLO. The YOLO-DSP network achieves 13.20% precision and 43.90% recall. The average detection accuracy of the YOLO-OSA network is higher than other networks and 10.00% higher than that of the YOLO-DSP, indicating that the YOLO-OSA network has a certain pertinence for the analysis of sports video images and improves the detection speed.

4. Conclusions

At present, most of the target tracking methods are based on a static background. The frame difference between the target and the background is used to extract the contour information of the moving target to obtain the moving position of the target. However, in sports videos, the camera often moves violently, and the background changes every moment, so it is difficult to use the background frame difference to obtain the moving position of the target. Through the YOLO-OSA target detection network to conduct sports video analysis research, the results show that the average detection accuracy of the YOLO-OSA network is higher than that of other networks. Sports video image analysis has certain pertinence and improves the detection speed. Under the lightweight target detection network, video research and analysis in sports have certain reference significance. The development of the sports industry itself is very important for national economic development. In the future, the sports industry can use scientific definitions and suggestions in sports video analysis to make standardized and creative sports industry development. This industry will be able to regulate the marketization of the sports industry better, thus promoting the development of the national social economy. In terms of the design model method, this study is based on the YOLO implementation model. In view of the endless network structures of efficient feature extraction of lightweight target detection networks, it is hoped to improve the detection accuracy of the lightweight target while ensuring the detection speed. In sports videos, the increase in data amount reduces the processing speed. In the future, the study will be able to design a more efficient detection model for unified training of different targets.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Yang, J. Xu, and R. Yang, "Research on coordination and driving factors of sports industry and regional sustainable development-empirical research based on panel data of provinces and cities in eastern China," *Sustainability*, vol. 12, no. 3, p. 813, 2020.

[2] S. Liu and J. Tang, "Research on the development mode of sports industry in gannan prefecture," *Curriculum and Teaching Methodology*, vol. 3, no. 1, pp. 21–24, 2020.

[3] Y. Seonghyun and H. Song, "The endogenous logic and predictive model of converging development between sports industry and elderly care service industry in China," *International Journal of Human Movement Science*, vol. 14, no. 2, pp. 149–164, 2020.

[4] S. Na, T. Kunkel, and J. Doyle, "Exploring athlete brand image development on social media: the role of signalling through source credibility," *European Sport Management Quarterly*, vol. 20, no. 1, pp. 88–108, 2020.

[5] W. K. Chen, D. Riantama, and L. S. Chen, "Using a text mining approach to hear voices of customers from social media toward the fast-food restaurant industry," *Sustainability*, vol. 13, no. 1, p. 268, 2020.

[6] S. Hendricks, K. Till, Savage et al., "Consensus on a video analysis framework of descriptors and definitions by the Rugby Union Video Analysis Consensus group," *British Journal of Sports Medicine*, vol. 54, no. 10, pp. 566–572, 2020.

[7] K. Muhammad, T. Hussain, J. D. Ser, V. Palade, and V. H. C. de Albuquerque, "DeepReS: a deep learning-based video summarization strategy for resource-constrained industrial surveillance scenarios," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5938–5947, 2019.

[8] J. Jin, H. Guo, J. Xu, X. Wang, and F. Y. Wang, "An end-to-end recommendation system for urban traffic controls and management under a parallel learning framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1616–1626, 2020.

[9] F. Zhang, C. Li, and F. Yang, "Vehicle detection in urban traffic surveillance images based on convolutional neural networks with feature concatenation," *Sensors*, vol. 19, no. 3, p. 594, 2019.

[10] M. Herold, F. Goes, S. Nopp, Bauer, Thompson, and Meyer, "Machine learning in men's professional football: current applications and future directions for improving attacking play," *International Journal of Sports Science & Coaching*, vol. 14, no. 6, pp. 798–817, 2019.

[11] D. Son, W. Chun, and S. Park, "Comparison of craniovertebral angle and muscle properties after smartphone use in healthy individuals with and without forward head posture," *Journal of The Korean Society of Integrative Medicine*, vol. 9, no. 4, pp. 149–158, 2021.

[12] Z. Lu, G. Nazari, J. C. MacDermid, Modarresi, and Killip, "Measurement properties of a 2-dimensional movement analysis system: a systematic review and meta-analysis," *Archives of Physical Medicine and Rehabilitation*, vol. 101, no. 9, pp. 1603–1627, 2020.

[13] M. Bilal and M. S. Hanif, "Benchmark revision for HOG-SVM pedestrian detector through reinvigorated training and evaluation methodologies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1277–1287, 2019.

[14] H. Wang, Y. Yu, Y. Cai, X. Chen, L. Chen, and Y. Li, "Soft-weighted-average ensemble vehicle detection method based on single-stage and two-stage deep learning models," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 100–109, 2020.

[15] Z. Xia, J. Gu, K. Zhang, Wang, and Li, "Research on multi-scene electronic component detection algorithm with anchor assignment based on K-means," *Electronics*, vol. 11, no. 4, p. 514, 2022.

[16] N. Wang, Y. Gao, H. Chen et al., "NAS-FCOS: efficient search for object detection architectures," *International Journal of Computer Vision*, vol. 129, no. 12, pp. 3299–3312, 2021.

[17] X. He, S. Ma, L. He, Ru, and Wang, "Multi-sector oriented object detector for accurate localization in optical remote sensing images," *Remote Sensing*, vol. 13, no. 10, p. 1921, 2021.

[18] G. Chen and H. Qin, "Class-discriminative focal loss for extreme imbalanced multiclass object detection towards

autonomous driving," *The Visual Computer*, vol. 38, no. 3, pp. 1051–1063, 2022.

[19] Y. Koo, S. Kim, and Y. Ha, "OpenCL-Darknet: implementation and optimization of OpenCL-based deep learning object detection framework," *World Wide Web*, vol. 24, no. 4, pp. 1299–1319, 2021.

[20] Y. Ariji, Y. Yanashita, S. Kutsuna et al., "Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 128, no. 4, pp. 424–430, 2019.

[21] W. Kim, H. Cho, J. Kim, Kim, and Lee, "YOLO-based simultaneous target detection and classification in automotive FMCW radar systems," *Sensors*, vol. 20, no. 10, p. 2897, 2020.

[22] R. Yang, R. Wang, Y. Deng, Jia, and Zhang, "Rethinking the random cropping data augmentation method used in the training of CNN-based SAR image ship detector," *Remote Sensing*, vol. 13, no. 1, p. 34, 2020.

[23] D. O'Neill, B. Xue, and M. Zhang, "Evolutionary neural architecture search for high-dimensional skip-connection structures on densenet style networks," *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 6, pp. 1118–1132, 2021.

[24] F. Li, J. Li, A. Zhu, Xu, Yin, and Hua, "Enhanced spatial and extended temporal graph convolutional network for skeleton-based action recognition," *Sensors*, vol. 20, no. 18, p. 5260, 2020.

[25] J. Orlowski and P. Wicker, "Monetary valuation of non-market goods and services: a review of conceptual approaches and empirical applications in sports," *European Sport Management Quarterly*, vol. 19, no. 4, pp. 456–480, 2019.

[26] J. Benito-Picazo, E. Domínguez, E. J. Palomo, and López-Rubio, "Deep learning-based video surveillance system managed by low cost hardware and panoramic cameras," *Integrated Computer-Aided Engineering*, vol. 27, no. 4, pp. 373–387, 2020.

[27] B. Ghanbari, K. S. Nisar, and M. Aldhaifallah, "Abundant solitary wave solutions to an extended nonlinear Schrödinger's equation with conformable derivative using an efficient integration method," *Advances in Difference Equations*, vol. 2020, no. 1, pp. 1–25, 2020.

[28] A. J. Sabnis and T. G. Bivona, "Principles of resistance to targeted cancer therapy: lessons from basic and translational cancer biology," *Trends in Molecular Medicine*, vol. 25, no. 3, pp. 185–197, 2019.