Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 10
ISET 2025

A Review of Object Detection Empowering Sports: Key Technologies, Application Scenarios, and Future Outlook

Siyi Zhao

College of Computer Software, Southwest Petroleum University, Chengdu, China

Abstract. Object detection is now a cornerstone of 'Smart Sports,' yet the direct application of general-purpose models to the dynamic and often chaotic sports environment is fraught with challenges. This paper systematically reviews the core technologies of object detection in sports, including the adaptability and limitations of mainstream detectors (e.g., the YOLO series, Transformer-based models) in sports scenarios. It also examines the role of optimization strategies such as model pruning, quantization, and knowledge distillation in balancing performance and resource consumption, as well as specialized techniques for small object detection, motion blur processing, and occlusion robustness enhancement. Based on this, the paper provides an in-depth analysis of the diverse applications of object detection in professional sports training (e.g., motion capture and biomechanical analysis), competitive game analysis (e.g., tactical minimap reconstruction from match videos), intelligent officiating (e.g., foul recognition assistance), athlete performance evaluation, interactive sports broadcasting, and public fitness. Finally, the paper summarizes current challenges, including data bottlenecks, algorithm generalization, the complexity of multi-modal fusion, and the leap from perception to cognition. It also provides an outlook on future directions, including constructing sports-specific vision foundation models, deepening multi-modal intelligent fusion, enhancing dynamic scene understanding capabilities, and improving sports datasets and evaluation systems to promote the development of sports analytics toward intelligence, personalization, and accessibility.

Keywords: Object Detection; Sports Analytics; Computer Vision; Deep Learning; Model Optimization; AI in Sports; Sports Big Data

1. Introduction

With the deepening concept of "Smart Sports," computer vision has become increasingly critical in the digital transformation of sports due to its non-contact and information-rich advantages [1, 2], particularly in enhancing the "perception, comprehension, and decision-making" of athletic processes. This paper aims to systematically elaborate on how object detection technology empowers modern sports, focusing on its key technological advancements, diverse application scenarios, and future development trends, while also providing a deep analysis of the core challenges it currently faces. As a cornerstone of computer vision, object detection provides fundamental spatiotemporal data for subsequent high-level analyses (e.g., action recognition, tactical analysis) by accurately identifying and localizing key sports elements in images/videos, such as athletes, balls, and field markings. Its application has significantly improved the scientific basis of training (e.g., quantitative feedback on movements), deepened competitive insights (e.g., tactical minimap reconstruction [3]), revolutionized the viewing experience (e.g., automated highlights, data visualization), and created new commercial value. Specifically, object detection in sports aims to identify athletes (differentiating teams, numbers, and poses), various types of balls (often small and fast [4]), referees, field markings, and equipment. Its output is a prerequisite for multi-object tracking, action recognition, tactical analysis, and real-time officiating.

Deep learning has greatly advanced the development of object detection [5]. Mainstream algorithms include: *two-stage detectors* (e.g., the R-CNN series [6,7,8], Mask R-CNN [9]), *one-stage detectors* (e.g., the YOLO series [10,11,12,13,14, 15], SSD [16], RetinaNet [17]), and *Transformer-based detectors* (e.g., DETR [18], Deformable DETR [19], DINO [20]).

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

However, directly applying general-purpose models to sports scenarios presents challenges: large models are difficult to deploy on embedded or real-time systems; sports-specific problems such as small objects [4,21], fast motion and blur, frequent occlusions, and complex backgrounds and lighting [22] can all lead to performance degradation. Furthermore, the diversity of athlete poses and appearances, the need to balance speed and accuracy, the scarcity of high-quality specialized sports datasets (e.g., the limited coverage and high annotation cost of datasets like SoccerNet [23] and BasketballDBC [23]), and the gap between pixel-level perception and semantic understanding [1] are all issues that need to be addressed. The primary approaches to tackling these issues involve model-level optimization (lightweight networks, model compression, targeted detection capability enhancement) and innovation at the data and task levels (multi-modal fusion, weak/self-supervised learning [24], establishing evaluation benchmarks) [22]. Model optimization (pruning [25], quantization, knowledge distillation [26], lightweight network design [27,28,29]) is crucial for resource-constrained sports applications (e.g., portable analysis systems, smart sensors), aiming to balance accuracy and efficiency to enable the practical deployment of advanced algorithms.

Table 1. Mainstream Algorithm Comparison

Category	Typical Model	Core Idea	Accuracy in Sports	Speed in Sports	Small Object Handling	Computational Resource Req.	Main Pros & Cons
Two-Stage Detectors	R-CNN Series (R-CNN, Fast R-CNN, Faster R-CNN)	First generate region proposals, then classify and refine locations for each region.	High	Slow	Medium	High	Pros: High accuracy, suitable for complex scenes (e.g., fine-grained analysis of athlete movements). Cons: Slow, difficult to meet real-time requirements.
One-Stage Detectors	YOLO Series (YOLOv1 to YOLOX), SSD	Directly predict object classes and bounding boxes on the image, skipping the region proposal step.	Medium	Fast	Low (Improved in YOLOv3+)	Medium	Pros: Fast, suitable for real-time scenarios. Cons: Lower accuracy for small objects.
Transformer-based Detectors	DETR and its derivatives	Frame object detection as a set prediction problem, using the Transformer's self-attention mechanism to directly output a set of objects.	Medium-High	Medium	Medium-High (relies on attention mechanism)	High	Pros: Reduces hand-designed components (e.g., NMS), suitable for modeling long-range dependencies (e.g., team tactical analysis). Cons: High computational cost, long training cycles.

2. Methods

2.1. Model Optimization Techniques for Sports Object Detection

To effectively apply advanced object detection algorithms to resource-constrained sports analytics scenarios, such as deployment on edge devices, mobile terminals, and wearables, model optimization techniques play a crucial role. These techniques aim to achieve an optimal balance between real-time performance, accuracy, and energy efficiency. Key methods include pruning, quantization, and knowledge distillation. *Model Pruning* is a method aimed at reducing model size and computational complexity by removing relatively "unimportant" parameters, connections, channels, or even entire layers from a neural network. The work by Han et al. (2015) was pioneering in learning both weights and connections for efficient neural networks. Model pruning is mainly divided into two categories: *unstructured pruning*, which removes individual weights, leading to sparse weight matrices that often require specific hardware or libraries for effective acceleration; and *structured pruning*, which removes entire channels, filters, or layers, resulting in a more regular model structure that is easier to accelerate on general-purpose hardware. In sports object detection tasks, pruning can be used to compress large pre-trained models like ResNet, proposed by He et al. (2016), to enable more efficient application in tasks such as real-time athlete tracking. *Low-Precision Quantization* replaces the commonly used 32-bit floating-point (FP32) weights and/or activations in a model with lower-bit representations (e.g., FP16, INT8) for storage and computation. This significantly reduces model size, lowers memory bandwidth requirements, and leverages integer arithmetic units in hardware for accelerated inference. The main approaches are *Post-Training Quantization (PTQ)* and *Quantization-Aware Training (QAT)*. Although your provided reference list does not contain the original papers for these quantization techniques, they are widely recognized as effective optimization methods. For example, in a lightweight network framework like EITNet, described by Liu J et al. (2025) for real-time basketball action recognition, low-precision quantization is a key enabling technology for its efficient deployment on resource-constrained hardware. *Knowledge Distillation (KD)*, with its core idea originating from Hinton et al. (2015), involves a smaller, simpler "student model" learning the "knowledge" contained within a larger, more powerful "teacher model" to achieve model compression and performance improvement. Knowledge can be transferred in various ways, such as by mimicking the teacher model's outputs or intermediate feature representations. In the field of sports analytics, KD is widely applicable. For instance, in Liu J et al. (2025), to ensure the lightweight EITNet maintains good detection performance, knowledge distillation can be used to transfer capabilities from a more complex teacher model to the student model.

There is also a cutting-edge technology for automated design of neural network structures, *Neural Architecture Search, NAS)*, offering new solutions for model design in specific application scenarios. Elsken et al. (2019) provided a comprehensive survey on the principles and development of NAS, laying a foundation for understanding its potential. Drawing on this automated design concept, NAS shows significant promise for customizing efficient models for sports object detection, such as optimizing player detection networks for specific sports camera angles or automatically searching for the optimal architecture for resource-constrained embedded AI chips. This potential is also highlighted as an important future research direction in the sports computer vision survey by Naik B T et al. (2022), aligning with this paper's goal of enhancing the intelligence of sports analytics.

2.2. Specific Detection Enhancement Techniques for Sports Scenarios

The inherent complexity and uniqueness of sports scenarios place higher demands on the robustness and specificity of object detection techniques.

The challenge of *small object detection* is emphasized in the literature for its importance in sports, such as for fast-moving small balls (Hiemann et al., 2021) and in a general survey on small object detection (Du L et al., 2022). To address this issue, researchers have adopted several strategies:

- Using multi-scale feature fusion with Feature Pyramid Networks (FPN), first proposed by Lin T. Y. et al. (2017). *Variants of FPN* have also been developed, such as *PANet*, introduced by Liu S. et al. (2018), and *BiFPN*, a key component in EfficientNet proposed by Tan M. & Le Q. (2019).
- Introducing attention mechanisms to focus on key regions, such as the *Squeeze-and-Excitation (SE) Block* proposed by Hu J. et al. (2018).
- Employing super-resolution techniques to enrich the details of small objects, such as methods based on Generative Adversarial Networks (GANs) described by Li J. et al. (2017).
- Using specific inference techniques like *Slicing Aided Hyper Inference (SAHI)*, proposed by Akyon F. C. et al. (2022).

For *motion blur and fast-moving objects*, integrating temporal information is key. Researchers have explored various spatiotemporal feature learning methods. 3D convolutional networks are a common approach, with representative works including the *C3D (Convolutional 3D)* model by Tran D. et al. (2015) and the *I3D (Inflated 3D ConvNet)* model by Carreira J. & Zisserman A. (2017), which achieved significant success in action recognition.

In *handling occlusion*, integrating other information sources or technologies can be beneficial. For instance, using human pose estimation techniques, such as the real-time multi-person 2D pose estimation algorithm *OpenPose* proposed by Cao Z. et al. (2019), can provide clues for locating partially occluded athletes.

Finally, *multi-modal data fusion* offers a new avenue for improving detection robustness and scene understanding. A concrete example is the *EITNet framework* proposed by Liu J et al. (2025), which demonstrates how to effectively fuse visual information with kinematic and physiological data from IoT-enhanced devices (like wearable sensors) to enhance the accuracy and robustness of basketball action recognition.

2.3. Typical Application Scenarios of Object Detection in Sports

The applications of object detection in sports can be broadly categorized into three levels: *perception, understanding, and decision-making*.

At the fundamental perception level, object detection directly provides raw data for subsequent analysis. This includes real-time detection and localization of key elements: accurately identifying the positions of athletes on the field (positional accuracy typically required to be within 10-20cm), their team affiliation, jersey number, and even their general pose; rapidly locating various game balls such as footballs, basketballs, and table tennis balls (for example, in the 2024 European Championship, a chip-embedded football combined with high-frame-rate cameras and object detection achieved centimeter-level positioning), which is crucial for trajectory prediction and ball possession analysis; identifying field markings and facilities like boundary lines, functional area lines, goals, and baskets (e.g., goal-line technology requires a detection error of <1cm) to assist in judging key events; and detecting other relevant objects like referees and coaching staff. Building on this, Multi-Object Tracking (MOT) technology, as pursued by projects like SoccerNet [23], enables continuous and stable identity-preserving tracking of athletes and the ball in video sequences. The resulting motion trajectories are fundamental for calculating Key Performance Indicators (KPIs) such as running distance (error <5%), speed, activity heatmaps, and time of possession.

Moving to the understanding level, the system performs a deeper analysis of athlete behavior and game state by combining the object position and trajectory data from the perception layer with domain knowledge of the sport. This includes athlete performance analysis and tactical diagnosis, which can be broken down into: individual performance analysis (e.g., quantifying physiological metrics through running heatmaps, identifying and evaluating technical actions like shooting or kicking by combining with pose estimation, and even estimating biomechanical parameters [30] to analyze movement efficiency and injury risk) and team tactical analysis (e.g., real-time identification and tracking of formation changes, analyzing player positioning and space utilization during offensive

and defensive transitions, and understanding team coordination through player relative positions and passing routes). For instance, SoccerNet [3] can reconstruct a 2D tactical minimap from broadcast video with an error of less than 30cm, visualizing player positions and ball trajectories to help coaches with post-game reviews and tactical planning. Technologies like Graph Neural Networks (GNNs) [31,32] can analyze the interaction network between players to identify core players and key connections, such as finding the duo with the highest pass completion rate. Furthermore, object detection empowers intelligent officiating and assisted refereeing by enabling automatic detection of key events like offsides, ball out-of-bounds/goal decisions, and preliminary foul recognition. For example, in football, real-time detection of the positions of the attacking player, the second-to-last defender, and the ball helps in judging potential offsides; the advanced Semi-Automated Offside Technology (SAOT) can reduce the decision time from an average of 70 seconds to 25 seconds. Precise detection of the ball's position relative to boundary lines or the goal line helps determine if the ball has entirely crossed the line. By detecting abnormal contact and dangerous actions between athletes (e.g., a hand-checking foul in basketball, with model recognition accuracy reaching 85%), the system can provide preliminary foul alerts to the referee. This typically requires a combination of fine-grained action recognition and contact detection. The Video Assistant Referee (VAR) system also benefits, as object detection can provide more efficient footage processing, for example, by automatically filtering video clips containing controversial incidents (like potential penalties), improving processing speed by over 30%, highlighting the relevant players and the ball, and overlaying virtual auxiliary lines (like the offside line), thus helping the VAR make faster and more accurate decisions.

At the decision-making level, based on the outcomes of perception and understanding, object detection provides decision support and novel services for coaches, athletes, event organizers, and even casual fans. This benefits sports training and rehabilitation, allowing for the creation of personalized training plans, providing movement correction through pose evaluation (e.g., using AI to analyze a swimmer's stroke, identifying inefficient phases, and providing improvement suggestions, which can reportedly improve performance by 0.5-1%), monitoring training load and fatigue, and quantifying rehabilitation progress. Sports broadcasting and live streaming are enhanced with automated highlight generation (e.g., automatically identifying key events like goals or dunks with >90% accuracy), real-time overlay of immersive data visualizations (e.g., player names, running speed (accuracy ±0.5 km/h), possession stats, shot charts, tactical diagrams), and transforming traditional broadcast views into a top-down (bird's-eye) tactical view similar to what coaches use [3], providing deeper tactical insights for professionals and avid fans. Virtual advertising insertion and content enhancement also become possible, with intelligent placement of virtual ads in specific areas of the field, which can increase click-through rates by 5-10%. Personalized fitness and sports guidance are also developing; for example, an AI fitness coach via a mobile app or smart fitness mirror uses object detection and pose estimation to monitor a user's exercises (like squats, push-ups, or yoga poses) in real-time, judging if the form is correct (e.g., if knees go past the toes during a squat, with detection accuracy >95%) and complete, and providing voice or text feedback for guidance. Exercise data logging and analysis functions can automatically record the user's repetitions, duration, calories burned, and analyze trends to help users better manage their fitness plans. In terms of smart venue operations and public sports services, applications include crowd flow analysis and security monitoring (e.g., automatically issuing an alert if crowd density exceeds a preset threshold, with >98% accuracy), contactless timing and automatic results recording (e.g., at a marathon finish line, a detection system identifies athlete bib numbers and records their times with an error rate of <0.1%), and intelligent management of sports equipment and facilities.

3. Performance Comparison

A persistent core challenge is comparing the performance, especially the trade-off between accuracy and speed, required to realize the diverse sports applications mentioned above. The following table summarizes the performance requirements of different sports application scenarios and the suitability of various algorithm types.

Table 2. Analysis of Performance Trade-offs and Algorithm Suitability in Different Sports Application Scenarios

Application Scenario	Core Performance Demand	Key Performance Indicators (KPIs) & Requirements	Representative Algorithm Performance
Real-time Fast Ball Detection (e.g., Table Tennis, Badminton)	Speed (Low Latency)	- Latency: < 10-20ms - Metrics: FPS, Precision, Recall	- YOLO Series: Fast (e.g., YOLOv8 can reach 60-80 FPS), making it the primary choice for such scenarios. - DETR: Latency is too high (can be up to 120ms), making it unsuitable for real-time ball detection.
Detailed Tactical Analysis & Review (Offline Analysis)	Accuracy & Stability	- Localization Accuracy: < 5cm - Pose Keypoint Error: < 3 pixels - ID Switch Rate: < 1% - Metrics: mAP, MOTA, IDF1	- Faster R-CNN Series: High localization accuracy, especially adept at identifying small regions like athlete faces or numbers, making it a strong candidate. The slow speed is acceptable in offline analysis. - YOLO Series: May compromise on accuracy and ID stability.
Real-time Multi-Person Tracking & Analysis (e.g., Football, Basketball)	Balance of Speed and Accuracy (esp. under heavy occlusion)	- Processing Speed: > 30 FPS - ID Switch Rate: Must remain low in crowded scenes (>10 overlapping people) - Metrics: GAMESTATE-Acc, mAP, FPS	- YOLO Series (e.g., YOLOv8): Has a clear speed advantage. However, performance drops significantly under heavy occlusion (ID switch rate can increase from 3% to over 12%, mAP can drop by 10-15%). - DETR: Theoretically suited for modeling player interactions, but high computational cost and latency are major deployment hurdles. - EITNet: Can achieve 30 FPS on an edge device (Jetson AGX Xavier), demonstrating a good balance.

Algorithm performance is typically evaluated on public sports-related benchmark datasets. For instance, SoccerNet uses metrics like mAP, MOTA/IDF1, and GAMESTATE-Acc [3]. Basketball action recognition tasks [27] focus on accuracy and FPS (e.g., EITNet achieves real-time performance of 30 FPS on an NVIDIA Jetson AGX Xavier). Ball detection research focuses on precision, recall, and FPS (e.g., for high-speed small ball detection, some algorithms can boost precision to over 95% at the cost of some recall).

The adaptability of general-purpose algorithms to sports datasets varies. The YOLO series (e.g., YOLOv8) typically demonstrates a significant speed advantage on sports datasets (e.g., 60-80 FPS on a 2080Ti for 1080p video), making it suitable for scenarios requiring rapid response. However, in heavily occluded scenes, such as defending a corner kick in the penalty area where more than 10 players are overlapping, YOLOv8's ID switch rate can surge from a typical 3% to over 12%, and its

mAP may drop by 10-15%. This is mainly due to the lack of robustness in its anchor-matching or label-assignment strategy in high-overlap scenarios. Faster R-CNN and its variants have advantages in localization accuracy and handling complex scenes (e.g., higher accuracy in recognizing small areas like athlete faces or numbers), but the original models are slow. By using lightweight backbones, model compression, and hardware acceleration, their speed disadvantage can be partially mitigated, making them suitable for offline analysis tasks that demand higher accuracy, such as generating detailed post-match tactical reports. Transformer-based detectors (e.g., DETR) have potential in modeling global relationships, theoretically making them better suited for understanding complex interactions between players. However, in practical sports applications, such as detecting high-speed small objects like a shuttlecock in a badminton match, DETR's detection latency can be as high as 120ms, far exceeding the 30-50ms threshold typically required by real-time broadcasting or analysis systems. This is mainly limited by the computational complexity of its global self-attention mechanism. Therefore, for sports scenarios that require handling small samples and meeting high real-time demands, their training difficulty and inference efficiency remain challenges. Developing lightweight Transformer variants specifically for sports (e.g., MobileDETR) is a future research direction.

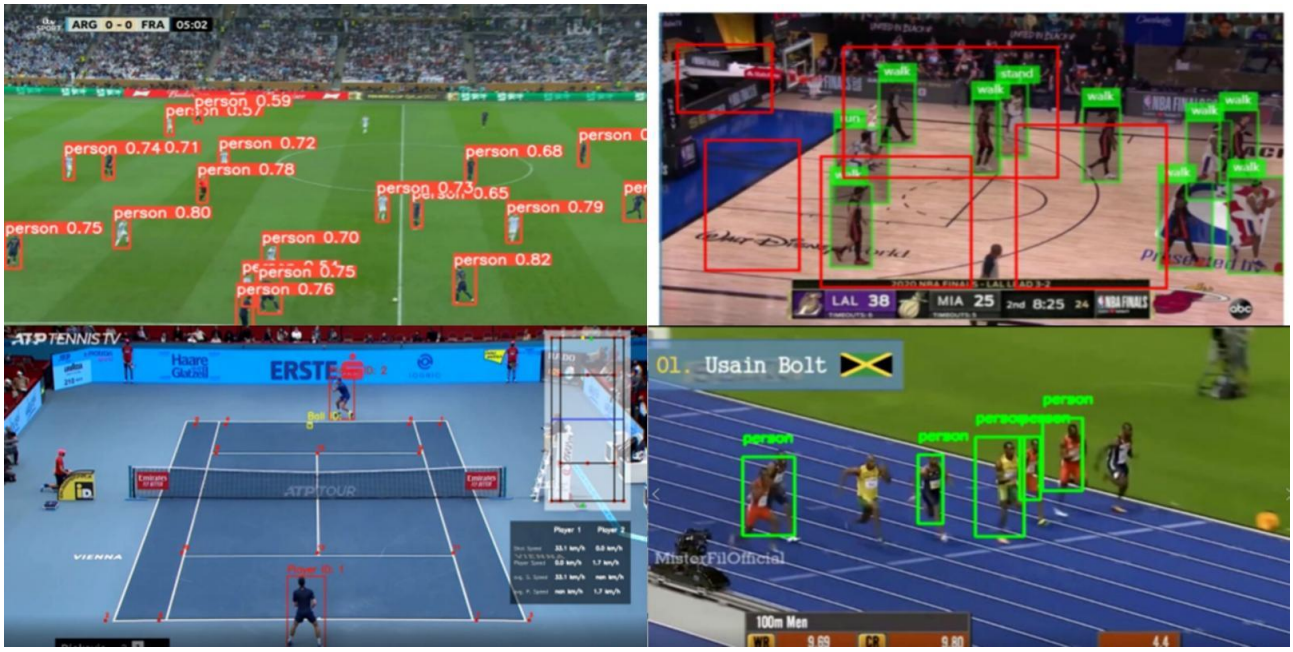

Figure 1. Application of Object Detection in Sports

4. Conclusion and Future Outlook

4.1. Summary of this Paper

This paper has systematically reviewed and analyzed the core role and profound impact of object detection technology in the modern sports domain. As the critical foundation of the perception layer in intelligent sports analytics, object detection provides indispensable raw data for higher-level motion understanding, behavior analysis, and intelligent decision-making. We have explored the adaptability, advantages, and inherent limitations of mainstream object detection algorithms (e.g., YOLO series [10-14], Faster R-CNN, DETR) in sports-specific scenarios. We detailed the key role of model optimization techniques (including model pruning, low-precision quantization, knowledge distillation, and neural architecture search) in balancing performance and resource consumption in sports applications. Concurrently, we conducted an in-depth analysis of specific detection enhancement strategies targeting the unique characteristics of sports scenes (such as small objects, fast motion and motion blur, and frequent occlusions). Building on this, the paper has broadly elaborated on the concrete applications and value of object detection technology in diverse scenarios,

including professional sports training and competitive analysis (e.g., reconstructing tactical minimaps from broadcast videos), intelligent assistant refereeing, interactive sports broadcasting, and public scientific fitness. Finally, we summarized and highlighted the severe challenges still facing sports object detection, which are mainly concentrated in the scarcity and difficulty of building high-quality specialized datasets, the insufficient robustness and generalization ability of algorithms in complex and dynamic real-world environments, the extreme demands for real-time performance and efficiency in edge deployment, the complexity of multi-modal information fusion, and the barrier of making the intelligent leap from low-level perception to high-level cognition.

4.2. Future Directions

Looking ahead, object detection technology in sports will become more tightly integrated with cutting-edge technologies like artificial intelligence, the Internet of Things, and big data, evolving in a direction that is more intelligent, precise, and accessible. This will primarily be reflected in the following aspects:

1. Advanced Model R&D and Foundational Capability Building: Continuously optimize and innovate sports-specific detection models, particularly lightweight architectures that are robust to disturbances like lighting variations, blur, and occlusion. Concurrently, explore the potential of foundation models like the Vision Transformer (ViT) [33] and Masked Autoencoders (MAE) [34]. By leveraging large-scale self-supervised learning [24] or domain adaptation for pre-training [35], we can build a sports-vision foundation model to enhance generalization capabilities in few-shot or novel scenarios.

2. Deepening Understanding and Multi-modal Fusion: Develop context-aware models that can more deeply understand the game situation, integrating temporal information, multi-agent interactions, and field semantics to comprehend player roles, intentions, tactics, and critical moments. This can be achieved using technologies like GNNs [31] and spatiotemporal attention [22,32]. Simultaneously, more intelligently integrate multi-modal information from visual data, wearable sensors, coaching instructions, and audio-video streams. By optimizing alignment, fusion, and collaborative reasoning, we can provide comprehensive insights.

3. Enhancing Intelligent Systems and Automation: Promote the distributed deployment and collaborative inference of models across edge, cloud, and end devices, while developing privacy-preserving techniques like federated learning [36]. Push sports analytics from mere data collection and basic statistics toward advanced applications such as automated tactical mining, intelligent report generation, personalized training recommendations, and injury prediction, building AI systems capable of long-term learning and self-evolution.

4. Data Ecosystem Construction and Trustworthy AI Development: Encourage the creation of high-quality, open-source sports datasets and evaluation [23,37]benchmarks with broader coverage and more detailed annotations (e.g., 3D pose, interactions, intent, multi-modal synchronization). At the same time, as AI plays a growing role in sports decision-making, it is crucial to enhance system transparency, explainability [38], and fairness [39] to ensure the technology is applied justly and gains user trust [22].

References

[1] Zhao Z, Chai W, Hao S, et al. A survey of deep learning in sports applications: Perception, comprehension, and decision[J]. IEEE Transactions on Visualization and Computer Graphics, 2025.

[2] Mendes-Neves T, Meireles L, Mendes-Moreira J. A survey of advanced computer vision techniques for sports[J]. arXiv preprint arXiv:2301.07583, 2023.

[3] Golovkin V, Nemtsev N, Shandyba V, et al. From Broadcast to Minimap: Achieving State-of-the-Art SoccerNet Game State Reconstruction[C]//Proceedings of the Computer Vision and Pattern Recognition Conference. 2025: 6028-6038.

[4] Hiemann A, Kautz T, Zottmann T, et al. Enhancement of speed and accuracy trade-off for sports ball detection in videos—finding fast moving, small objects in real time[J]. Sensors, 2021, 21(9): 3214.

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems (NIPS).

[6] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[7] Girshick, R. (2015). Fast R-CNN. Proceedings of the IEEE international conference on computer vision (ICCV).

[8] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in neural information processing systems (NIPS).

[9] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE international conference on computer vision (ICCV)*.

[10] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[11] Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[12] Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[13] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

[14] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. arXiv preprint arXiv:2107.08430.

[15] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. European conference on computer vision (ECCV).

[17] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision (ICCV)*.

[18] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. European conference on computer vision (ECCV).

[19] Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2020). Deformable DETR: Deformable transformers for end-to-end object detection. *International Conference on Learning Representations (ICLR)*.

[20] Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., ... & Wang, L. (2022). DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection. *International Conference on Learning Representations (ICLR)*.

[21] Du, L., Zhang, Z., Zhang, Y., Liu, J., & Wen, L. (2022). A survey on small object detection: Progress, challenges, and prospects. *Pattern Recognition*.

[22] Naik B T, Hashmi M F, Bokde N D. A comprehensive review of computer vision in sports: Open issues, future trends and research directions[J]. Applied Sciences, 2022, 12(9): 4429.

[23] Deliege, A., Cioppa, A., Giancola, S., Seikavandi, M. J., Dueholm, J. V., Nasrollahi, K., ... & Van Droogenbroeck, M. (2021). SoccerNet-v2: A Large-Scale Benchmark for Video Understanding in Football. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[24] Jing, L., & Tian, Y. (2021). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[25] Han, S., Pool, J., Tran, J., & Dally, W. (2015). Learning both weights and connections for efficient neural network. Advances in neural information processing systems (NIPS).

[26] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.

[27] Liu J, Liu X, Qu M, et al. EITNet: An IoT-enhanced framework for real-time basketball action recognition[J]. Alexandria Engineering Journal, 2025, 110: 567-578.

[28] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[29] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

[30] Zhang, Z. Q. (2020). Computer vision in sports biomechanics: A systematic review. *International Journal of Sports Science & Coaching*.

[31] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems.

[32] Ibrahim, M. S., & Mori, G. (2018). Hierarchical graph-based activity parsing and recognition in team sports videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[33] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR).

[34] He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[35] Wilson, G., & Cook, D. J. (2020). A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*.

[36] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*.

[37] Cabaset, S., Giancola, S., & Ghanem, B. (2019). ActivityNet-Sports: A Novel Dataset for Fine-Grained Activity Recognition in the Sports Domain. *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*.

[38] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE access*.

[39] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*.