Intelligent Athletic Performance Monitoring System for Player Movement Detection and Step Analysis using Yolo

Sathish G
Dept. of Artificial Intelligence and Data Science
St. Joseph's College of Engineering
Chennai, India
sadie26032005@gmail.com

Sujan P
Dept. of Artificial Intelligence and Data Science
St. Joseph's College of Engineering
Chennai, India
sujanprabakaran2004@gmail.com

Arulprasanth S
Dept. of Artificial Intelligence and Data Science
St. Joseph's College of Engineering
Chennai, India
s.arulprasanth2004@gmail.com

Someshwar V
Dept. of Artificial Intelligence and Data Science
St. Joseph's College of Engineering
Chennai, India
somesh1812004@gmail.com

Sam Varghese George
Dept. of Artificial Intelligence and Data Science
St. Joseph's College of Engineering
Chennai, India
samvarghese936@gmail.com

***Abstract*—Sports analytics has revolutionized the understanding of player performance, game strategies, and fan engagement. Despite its benefits, implementing real-time sports analytics remains challenging due to the high computational demands and extensive setup requirements of existing models. The limitations of these models present serious obstacles to their implementation on resource-constrained devices, including edge or mobile systems. Using the YOLO (You Only Look Once) object detection framework, Which realize Precision, Recall, The work focuses on detecting and bounding football players the ball in real-time match scenarios. A meticulously annotated dataset was created, and implementation was carried out. When the IoU threshold was set at 0.5 (mAP 0.5), the proposed system achieved a mean Average Precision (mAP) of 60.7%; when the IoU threshold was set at 0.95 (mAP 0.5:0.95), the mAP was 38.3%. The results support the model's capability of accurately detecting and tracking objects in real time. This research has shown that the system is effective in solving the standard problem in sports analytics through an easy option in tracking players and balls on devices with limited resources.**

Keywords— Sports Analytics, Real-Time Detection, Football, Object Tracking, YOLO.

I. INTRODUCTION

High-quality images are of utmost importance in football detection, as they allow models to detect the ball accurately in a variety of real-world conditions. These include angle and perspective, lighting conditions, image resolution, situational diversity (fast gameplay or crowd important for training robust detection systems. These attributes have been crucial in football analytics as they advance the detection of players and football in dynamic environments. Traditional techniques of football detection include computationally less demanding techniques such as color-based detection, HSV color segmentation [1], and shape detection using approaches such as Hough Circle Transform [2] to identify spherical objects. Motion detection techniques such as background subtraction [3] and optical flow [4] are used for tracking. However, the traditional methods are not robust for real- world conditions, being weak with complex backgrounds, changing illumination, or occlusion.

Non-conventional approaches emphasize deep learning algorithms, they are excellent at directly extracting valuable features from unprocessed data. Hierarchical representation learning is a component of the deep learning concept. Low-level characteristics, such colors and edges. develop into high-level abstractions, including the shapes of objects, contextual patterns [5]. This ability to learn features automatically and adaptively is why deep learning is increasingly significant in football detection, offering scalability, precision, and robustness. Deep learning has been widely used in the domains of object detection and tracking during the last few decades. Mask RCNN [6], YOLO (You Only Look Once) [7] are notable architectures that segment images, The method involves predicting borders and likelihoods of classes while dividing the data into grids. Its effectiveness permits competitors and footballs to be detected and tracked in actual time in high-speed wearing situations,even on devices with limited processing power. However, even with such advancements, sometimes the non- conventional methods may face particular real-world challenges like occlusions and highly dynamic environments.

This study uses the YOLO model for real-time football player detection and tracking, leveraging deep learning to handle changes in object position, scale, and appearance. The system overcomes limitations of traditional methods, ensuring precision and robustness in sports analytics. Section 1 reviews prior research on football detection and YOLO applications. Section 2 outlines the methodology for player and ball recognition using YOLO. Section 3 presents results, comparing the model's effectiveness with conventional techniques.

II. LITERATURE REVIEW

Convolutional Neural Networks (CNNs) [9] revolutionize computer vision by excelling in image classification and visual data processing Convolutional, pooling, and fully connected layers which makes the architecture of neural networks using convolution (CNNs), which makes it easier to acquire hierarchical features. By learning directly from raw pixel data, CNNs demonstrate superior performance

979-8-3315-1485-3/25/$31.00©2025 IEEE

compared to conventional image processing techniques. They have become foundational in deep learning applications.

This achievement sent CNNs straight into the forefront of the researches conducted in machine learning. Follow-up architectures build and enlarged upon this framework, that might be understood by the models which constitute state of the art now- AlexNet [10], VGGNet [11], GoogleNet [12], ImageNet [13], ResNet [14], and man more. Those achievements had catalyzed innovation within various domains - such as analysis of satellite images [15,16], diagnostics of medical conditions [17], and advanced processing of images [18].

Deep learning methods counter the challenges of a real world, such as motion blur, varying lightings, and occlusions, to best suit complex tasks like live sport detection and tracking. Nonetheless, issues like unpredictable movement by players, overlaps, and cluttered backgrounds require innovative methods with both data-driven models and contextual understanding. The advance of CNN architectures has supported moves beyond classification, toward tasks like image segmentation which, in turn, asks for pixel-level accuracy about the delineation of an object. Semantic and instance segmentation have become vital in fields like autonomous driving and medical imaging, where accurate object identification is crucial. Instance Segmentation Advancements.

Mask R-CNN [6] was a landmark paper for instance segmentation, The Faster R-CNN [19] framework was enhanced by incorporating a parallel mask prediction feature branch. This advancement facilitated the concurrent forecasting of object bounding boxes, class labels, and pixellevel object masks. Before the introduction of Mask R-CNN, Faster R-CNN laid a strong groundwork through the implementation of Region Proposal Networks (RPNs). However, its limitations in pixel-level segmentation prompted the development of more sophisticated architectures.

The YOLO series emphasizes real-time object detection, with YOLOv3 enhancing accuracy using multi-scale features and residual blocks. YOLOv5 optimized the concept, while YOLOv8 introduced advanced techniques and modular design. Despite prioritizing speed, YOLO models remain vital for efficiency-critical tasks.

The football player tracking system exemplifies the practical application of advanced CNN technologies in sports analytics. In order to provide thorough player movement and performance analysis, the suggested system makes use of instance segmentation and object identification algorithms. The implementation makes use of the architectural advancements previously covered, especially the object detection and segmentation techniques used by the YOLO series and the Mask R- CNN model[20-22].

The objective of the system employs a multi-stage approach to player tracking:

- Initial player detection using pre-trained object detection models
- Precise player segmentation to isolate individual players
- Tracking player movements across video frames Extracting performance-related metrics such as distance covered, speed, and positional data.

III. Methodology

This section outlines the methods used in this work is described in this section, with a focus on the application of the YOLO model for real-time identification of football players. As shown in Fig 1 football match analysis, detecting players, balls, referees, and tracking relevant statistics presents several challenges. These challenges include accurately identifying objects in dynamic and complex environments, tracking objects across multiple frames, and calculating player statistics in real time.

To address these issues, a comprehensive methodology is proposed, encompassing video preprocessing, object detection, bounding box creation, tracking, and statistical analysis. The primary objectives are

1) Real-Time Detection and Tracking: Accurately detect and track players, the ball, and referees, with bounding boxes for analysis.

2) Player Statistics: Track player stats like steps, possession time, and successful passes.

3) Model Comparison: Compare YOLOv5 and YOLOv8 performance to select the best model for real- time analysis.

A. Video Preprocessing and Annotation

Video preprocessing segments match videos into frames at consistent intervals (e.g., every 30 frames) to capture key moments. Using Roboflow, frames are annotated with bounding boxes for players, referees, and the ball. This labeled dataset, with 1007 training and 293 validation images, forms the basis for training an object detection model to accurately identify and locate objects in the frames.

B. Object Detection Model

After annotating the frames, the next phase is to training an object detection model to recognize players, the ball, and referees throughout the game comes after the frames have been annotated. We compared the YOLOv5 and YOLOv8 models, two well-known models, in this study. YOLOv5, utilizing CSP-Darknet53 as its backbone, SPPF and CSP-PAN as its neck, and a YOLOv3 head for bounding box prediction, is known for its speed and efficiency. However, it struggles with detecting small objects like balls in complex scenes. YOLOv8 addresses these limitations with an optimized backbone for enhanced feature extraction, an improved neck for better feature fusion, and a refined head for higher precision in bounding box prediction. These advancements make YOLOv8 more robust and accurate, particularly for detecting small objects, marking a significant improvement over YOLOv5

C. Bounding Box Creation

Tracking and analyzing the locations of players, referees, and the ball are done based on the generation of bounding boxes. The bounding boxes are generated after the above-mentioned object detection model predicts the frames [22]. These bounding boxes capture their positions and sizes, as well as important references for further tracking. They are refined during training in order to achieve higher precision and accuracy that will provide the system with object-tracking ability.

D. Tracking System

The real-time tracking of players and the ball is achieved by associating the bounding boxes of the frames. This makes it possible to analyze players' interactions and game dynamics very accurately without losing but gaining insights at every pace. It further computes a host of player statistics – steps taken, distance covered, possession time, successful passes, etc. These figures give a distorted view of individual performance and team dynamics, allowing coaches and analysts to adjust their strategies without losing time.

IV. Results And Discussions

The results and discussion of the YOLO model's effectiveness for identifying football players are presented in this section. The recall, map, and precision of the suggested system's output are contrasted with those of the YOLOv5 model. The simulation makes advantage of Google Collab, an online cloud environment. For accurate and efficient annotation, Robo flow is used for the training process.

The object detection model is trained using these annotations as a foundational dataset. The dataset offers a strong collection of labelled data to train and evaluate the object detection model, including 1007 images for training and 293 validating images.

We used YOLOv8 to detect football players and the ball in match footage. As shown in Fig 2, the model successfully identified multiple players and the ball, providing confidence scores ranging from 0.49 to 0.94 for players and 0.77 for the ballYOLOv8 demonstrated precise bounding boxes, even with closely grouped players, and effectively handled scale, lighting, and motion variations. Despite some lower confidence in overlapping detections, the model proved reliable for real-time player and ball tracking in sports analytics.

A. Performance Metrics tracking system

A confusion matrix is used to assess the performance of the suggested model, as indicated in Table I. Better performance is typically indicated by a model with True Positives (TP) and True Negatives (TN) as well as False Positives (FP) and False Negatives (FN).

As presented in Table I, YOLOv8 demonstrates superior performance over YOLOv5 across all key metrics, highlighting notable advancements in object detection capabilities. The precision of YOLOv8 is 89%, compared to YOLOv5's 83%, showing a 6% improvement, which signifies better accuracy in identifying objects without false positives. Additionally, YOLOv8 achieves a recall of 76%, surpassing YOLOv5's 70% by another 6%, indicating its enhanced ability to detect most objects in the images without omissions.

The model's accuracy in identifying items at a particular Intersection over Union (IoU) level is measured by the mAP50 metric, which is 80% for YOLOv8, which is 5% better than YOLOv5's 75%.

Additionally, YOLOv8's mAP50-95 score, which offers a thorough assessment of performance across many IoU thresholds, is 51%, a 4% gain over YOLOv5's 47%. Summed, these aspects show that the use of YOLOv8 exhibits great accuracy in object detection tasks besides reliability as well as object localization to be bettered than previously.

B. Tracking system

The overall number of steps taken by each player in the team was determined through spatial distance analysis. Player locations were obtained via a tracking algorithm, and their movements were measured by calculating the Euclidean distance between successive positions on the field. As depicted in Fig. 3, the step count was obtained by converting the total distance covered into steps, utilizing an average stride length. The graph indicates that Player 8 achieved the highest step count followed closely by Player 10 and Player 3, suggesting elevated activity levels. Conversely, Player 11 exhibited the lowest step count, indicating limited movement. This visualization highlights player activity patterns, providing valuable insights for performance analysis and training optimization.

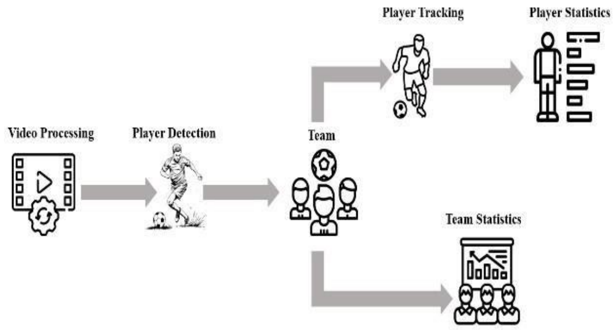

Fig. 1. Workflow

Fig.2. Player and Ball Detection

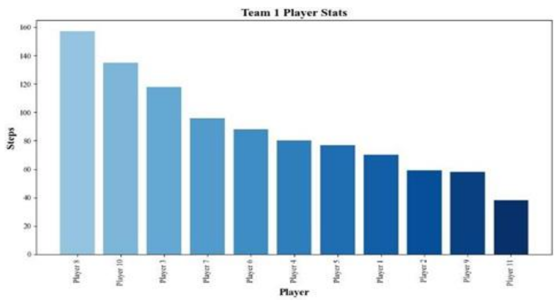

Fig.3. Team 1 Player Stats

TABLE I. MODEL COMPARISON

Model	Yolov5	Yolov8
Precision	0.8331	0.8881
Recall	0.7025	0.7625
mAP50	0.7476	0.7997
mAP50-95	0.4656	0.5103

V. Conclusion

This work has illustrated the feasibility of the effectiveness of YOLOv8 on real-time detection and tracking football players and balls within the game with notable accuracy and efficiency by achieving 60.7% in at an Intersection over Union (IoU) threshold of 0.5 and across a range of thresholds between 0.5 and 0.95, the proposed method achieves a mean Average Precision (mAP) of 38.3%. This approach effectively addresses various challenges in sports analytics, offering a versatile and lightweight solution suitable for implementation on low-resource devices. Future work will focus on enhancing ByteTrack's performance and robustness by improving occlusion handling, optimizing code for low- power devices, adding multi-camera inputs, and refining player/game metrics. These improvements will increase system efficiency and versatility.

ACKNOWLEDGMENT

We thank Data Science & Analytics domain, under the Centre for Innovations (CFI) and School of Computing, St. Joseph's College of Engineering, OMR, Chennai, for their valuable support in enabling this research.

REFERENCES

[1] Sural, S., Qian, G., & Pramanik, S. (2002, September). Utilizing HSV color space for image segmentation and histogram analysis in image retrieval applications. In *IEEE Int. Conf. on Image Processing* (Vol. 2, pp. II-II).

[2] Ni, J., Khan, Z., Wang, S., Wang, K., & Haider, S. K. (2016, June). Detecting and quantifying overlapping circular shapes using Hough Transform and contour techniques. In *12th World Congress on Intelligent Control and Automation (WCICA)* (pp. 2902–2906). IEEE.

[3] Lee, H., Kim, H., & Kim, J. I. (2016). Subtraction of background using reduced image and color space with multiple background references. *IEEE Trans. on Multimedia*, 18(10), 2093–2103.

[4] Kale, K., Pawar, S., & Dhulekar, P. (2015, September). Object tracking based on motion vector estimation and optical flow techniques. In *4th Int. Conf. on Reliability, Infocom Technologies and Optimization (ICRITO)* (pp. 1–6). IEEE.

[5] Acheli, M., Grigori, D., & Weidlich, M. (2021). Extracting and analyzing contextual user behavior from event records. *IEEE Trans. on Knowledge and Data Engineering*, 34(12), 5708–5721.

[6] Shi, J., Zhou, Y., & Zhang, W. X. Q. (2019, July). Enhanced Mask R-CNN for service robots in object detection tasks. In *Chinese Control Conference (CCC)* (pp. 8519–8524). IEEE.

[7] Liu, C., Tao, Y., Liang, J., Li, K., & Chen, Y. (2018, December). Real-time object recognition using a YOLO-based deep learning framework. In *IEEE ITOEC 2018* (pp. 799–803).

[8] Saleh, K., Szénási, S., & Vámossy, Z. (2021, January). Review of occlusion strategies in object detection systems. In *IEEE SAMI 2021* (pp. 000477–000484).

[9] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). Comprehensive survey on CNNs: methodologies, implementations, and future prospects. *IEEE Trans. on Neural Networks and Learning Systems*, 33(12), 6999–7019.

[10] Arya, S., & Singh, R. (2019, September). Evaluating CNN and AlexNet in identifying plant leaf diseases. In *ICICT 2019* (Vol. 1, pp. 1–6). IEEE.

[11] Jun, H., Shuai, L., Jinming, S., Yue, L., Jingwei, W., & Peng, J. (2018, November). Emotion recognition using a VGGNet-based CNN. In *Chinese Automation Congress (CAC)* (pp. 4146–4151). IEEE.

[12] Yao, X., Wang, X., Karaca, Y., Xie, J., & Wang, S. (2020). Classification of glomeruli using an optimized GoogLeNet model. *IEEE Access*, 8, 176916–176923.

[13] Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2015). Employing pre-trained ImageNet CNNs for Earth observation image classification. *IEEE Geoscience and Remote Sensing Letters*, 13(1), 105–109.

[14] Wen, L., Li, X., & Gao, L. (2020). ResNet-50-based transfer learning approach for diagnosing faults. *Neural Computing and Applications*, 32(10), 6111–6124.

[15] Ramalingam, A., George, S. V., Srivastava, V., Alagala, S., & Manickam, J. M. L. (2024). Efficient memory-based residual dilated CNN for extracting urban buildings via semantic segmentation. *Arabian Journal for Science and Engineering*, 1–20.

[16] Srivastava, V., Avudaiammal, R., & George, S. V. (2024). Deep learning-driven techniques for identifying urban structures in satellite imagery. *International Journal of Remote Sensing*, 45(1), 68–100.

[17] Jaspin, K., & Selvan, S. (2023). Multi-class CNN approach for brain tumor detection in MRI scans. *Biomedical Signal Processing and Control*, 82, 104542.

[18] Janaki, R., & Lakshmi, D. (2024). An integrated CNN and logistic regression model for early detection of esophageal conditions. *EURASIP Journal on Image and Video Processing*, 2024(1), 19.

[19] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Real-time object detection using RPN with Faster R-CNN architecture. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.

[20] Zou, S., Zhang, B., & Zhang, B. (2022, July). Darknet53-based image classification algorithm with improvements. In *Comm., Signal Processing, and Systems Conference* (pp. 73–80). Springer Nature.

[21] Wu, F., Jin, G., Gao, M., Zhiwei, H. E., & Yang, Y. (2019, May). Enhanced YOLOv3 network for helmet detection. In *IEEE ICNSC 2019* (pp. 363–368).

[22] Agarwal, S., Awan, A., & Roth, D. (2004). Object detection using sparse part-based models and structured representations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(11), 1475–1490.