AI-Powered Football Match Analysis: From Object Detection to Tactical Insights

Asma A. Mahdi
KAUST
asmaa.mahdiii9@gmail.com

Haifaa M. Althobaiti
KAUST
haifaalthobaiti15@gmail.com

Norah W. Bin Mannie
KAUST
norahbinmannie@gmail.com

***Abstract*—Recent advancements in computer vision have enabled more precise, automated analysis of football matches. This research presents a comprehensive analytical framework that integrates YOLOv8 for real-time object detection of players, referees, goalkeepers, and the ball, and YOLOv8-Pose for keypoint detection on the football pitch. These models serve as the foundation for downstream tasks such as team classification, homography transformation, player tracking, and tactical analysis. YOLOv8 demonstrated superior performance in recall, mAP@50, mAP@50–95, and inference speed, while YOLOv8-Pose enabled accurate pitch landmark detection critical for spatial alignment and movement analysis. The system further supports advanced analytics by estimating possession, visualizing match intensity, and predicting goal threat. Experimental results confirm the effectiveness of task-specific training and dataset expansion, validating the framework's ability to deliver accurate, real-time insights for tactical evaluation and key event identification.**

I. Introduction

In recent years, the field of computer vision has seen substantial progress, particularly in the development of real-time object detection models capable of understanding and interpreting complex visual environments. These advancements have opened new opportunities for automated sports analytics, where accurate detection and analysis of dynamic events are essential for tactical evaluation and performance insights. Football, in particular, presents a challenging yet rich domain for such applications due to the fast-paced nature of the game, the small size of key objects like the ball, and the need for contextual understanding of player interactions. This research introduces a comprehensive analytical framework for football match analysis, integrating match intensity visualization, goal likelihood prediction, and detailed player positioning analysis. The framework leverages advanced object detection techniques, utilizing YOLOv8 models with various input resolutions to enhance the detection of small objects. These models are employed to identify key elements on the pitch and to automatically assign players to their respective teams. The primary objectives of this research are to develop a robust system capable of accurate object detection and team classification, to support tactical understanding, and to enable the identification of key moments within a match. By combining high-performance detection models with contextual match analysis, this work contributes to the growing field of AI-driven sports analytics and demonstrates the potential of computer vision in enhancing football performance evaluation.

II. Project objectives and goals

The primary goal of this project is to design an AI-powered framework for football match analysis with a strong focus on ball possession estimation and match intensity visualization. Ball possession is tracked in real time by combining object detection, tracking, and homography transformation, providing insights into which team controls the game. Match intensity is quantified by analyzing player movements and ball dynamics, offering a clear view of high-pressure moments and tactical shifts during the match. These analytics, alongside object detection and team classification, enable coaches and analysts to better evaluate team performance and key match events.

III. Related work

Recent research has increasingly applied computer vision techniques to football analysis, focusing on areas such as object detection, pose estimation, and team classification.

YOLOv8 represents the most recent generation of the YOLO family of single-shot object detectors, introducing several architectural innovations that improve both accuracy and inference speed. The model incorporates an anchor-free detection mechanism, a CSPDarknet backbone enhanced with c2f modules, and a combination of Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN) to enable robust multi-scale feature extraction. These design choices make YOLOv8 particularly effective at detecting small, fast-moving, and partially occluded objects. In their work, Reis et al. [1] demonstrated the effectiveness of YOLOv8 in real-time object detection by applying it to flying objects, where the model achieved an impressive 99.1% mAP@50 while running at 50 FPS on 1080p video. This balance between speed and accuracy highlights the suitability of YOLOv8 for time-critical applications such as sports analytics, where detecting players, referees, and the ball in real-time is essential for downstream tasks like homography transformation, team classification, and match intensity analysis.

YOLOv8-Pose is an extension of the YOLO architecture designed for real-time multi-person keypoint detection. Unlike traditional approaches that rely on heatmaps or multi-stage processing, YOLOv8-Pose directly predicts keypoints for each detected object in a single pass. This design reduces computational cost and increases efficiency, making it more practical for real-time applications. The model is trained with the Object Keypoint Similarity (OKS) loss, which improves the

accuracy of keypoint localization. Maji et al.[2] demonstrated that YOLO-Pose achieved state-of-the-art performance on the COCO dataset, surpassing 90% AP50, which confirms its effectiveness for accurate and efficient keypoint estimation. In this project, YOLOv8-Pose was utilized to detect football pitch keypoints, serving as a reference for homography transformation.

Pan et al. [3] proposed an unsupervised clustering framework for football analysis that relies on color segmentation and illumination equalization to distinguish between players, referees, and goalkeepers. Their method introduces shadow correction and Retinex-based enhancement to address the challenges of uneven lighting in outdoor matches, significantly improving clustering accuracy. By operating solely on color features, the approach avoids the need for large labeled datasets and achieves state-of-the-art results on benchmarks such as SoccerNet-Tracing and Sn-gamestate. This demonstrates the effectiveness of K-means and related clustering techniques for team affiliation tasks under realistic match conditions.

Claasen and de Villiers [4] proposed a video based sequential Bayesian framework for homography estimation in soccer field registration. Their method explicitly models temporal dependencies between consecutive frames using a two-stage Kalman filter and integrates keypoint uncertainty into the transformation estimation. This leads to more stable and accurate homography alignment in broadcast video sequences, especially when field markings are not consistently visible due to camera movement or occlusions. The refined annotated datasets (CARWC) they released further contribute to practical applications in sports analytics.

IV. Problem Statement

Automated football match analysis has attracted considerable interest due to the increasing demand for detailed, real-time insights that support coaching, broadcasting, and fan engagement. Football presents unique challenges for computer vision systems, as the sport involves fast-paced actions, complex player interactions, and the presence of small, fast-moving objects such as the ball. Accurate detection and tracking of players, referees, goalkeepers, and the ball are essential for meaningful tactical and performance analysis. Furthermore, the variability of camera angles, occlusions, and frequent player clustering on the pitch complicate the detection and classification tasks. Separating players into their respective teams automatically adds another layer of complexity, requiring robust algorithms capable of understanding contextual information beyond simple object recognition. This research aims to develop a comprehensive analytical framework that leverages advanced object detection methods, such as YOLOv8 with optimized configurations, to deliver accurate, real-time football analysis in support of tactical evaluation and key event identification.

V. Methodology

The overall methodology of the proposed system is illustrated in Figure 1. The pipeline comprises six stages: (1) video-frame ingestion, (2) object detection, (3) tracking, (4) homography transformation, (5) team classification, and (6) match analysis. Object detection uses YOLOv8 models specialized for three streams: a field line/keypoint detector, a dedicated ball detector, and a player/referee/goalkeeper detector. A ByteTrack then assigns persistent IDs across frames. Homography maps image pixels to field coordinates to obtain geometry-aware positions. For team classification, cropped player images are embedded (SigLIP) and clustered with K-Means into two squads. Finally, the match analysis stage computes live ball possession and goal-threat intensity on the unified field view.

A. Data Acquisition

The datasets utilized in this research were sourced from Roboflow [5], which provides publicly available football image datasets that had been previously annotated for object detection tasks. Two annotated datasets were employed: one for detecting players, referees, and goalkeepers (663 annotated images), another for pitch keypoints (813 annotated images), and a third for detecting the ball (2226 annotated images). Figure 3 shows an example from the pitch keypoints dataset, Figure 2 illustrates a sample from the player detection dataset, and Figure 4 presents an annotated example from the ball detection dataset.



Fig. 2. Sample image from the player and ball detection dataset, showing annotated bounding boxes for players, referees, and the ball.



Fig. 3. Sample image from the pitch keypoints dataset, illustrating annotated field landmarks.



Fig. 4. Sample image from the ball detection dataset, illustrating annotated ball positions during a football match.

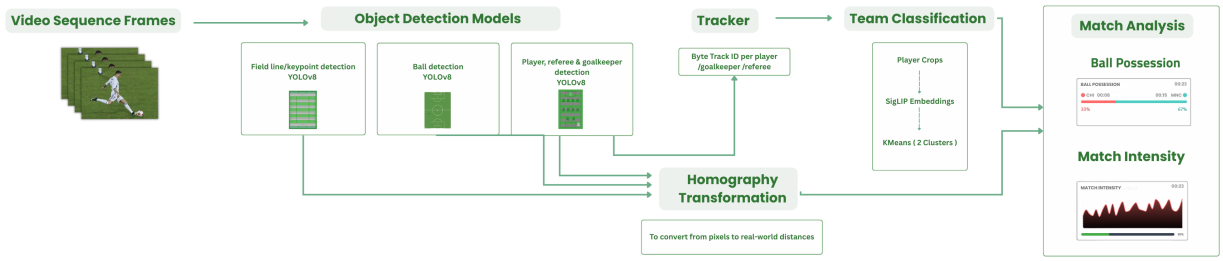


Fig. 1. Overall system methodology

Since the datasets were already labeled, no additional manual annotation was required. Each dataset was divided into training and validation subsets to facilitate model learning and ensure reliable evaluation. This preprocessing step provided a clean and structured input for the subsequent model training phase.

B. Object Detection Models

- **YOLOv8**
  For object detection, we utilized YOLOv8, the most recent generation of the YOLO family. The model was employed to detect players, referees, goalkeepers, and the ball in real time. These detections provided the essential inputs for subsequent tasks such as team classification, homography transformation, and match analysis. By leveraging YOLOv8's accuracy and efficiency, the system ensured reliable detection performance in football scenarios [1].
  Figure 5 illustrates the general YOLO architecture, where convolutional layers are stacked to extract multi-scale features and followed by fully connected layers for final predictions. This design principle has evolved across YOLO versions, with YOLOv8 adopting modern architectural improvements that enhance both detection accuracy and inference speed.

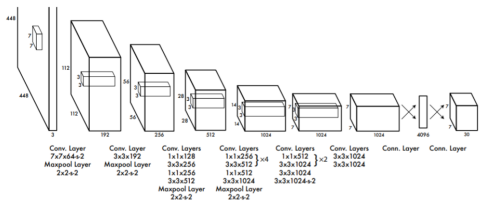


Fig. 5. YOLO architecture structure used for object detection.

- **YOLOv8-Pose** For the detection of pitch keypoints, we employed YOLOv8-Pose, an extension of the YOLO architecture designed for the estimation of real-time keypoints. In this project, YOLOv8-Pose was used to detect landmarks on the football pitch, and these keypoints served as references for the homography transformation that maps the coordinates of pixels to the real world positions [2].

C. Tracker

In our system, a tracking is integrated to keep consistent identities for players, goalkeepers, and referees across video frames. We employ ByteTrack, a high-performance multi-object tracker that associates detections frame-to-frame to maintain stable track IDs. This preserves identity continuity under occlusion and rapid motion, improving the reliability of downstream analyses such as team assignment, possession estimation, and goal-threat intensity.

D. Homography Transformation

For geometric mapping, we applied a homography transformation to convert image-based detections into real-world pitch coordinates. Inspired by Claasen and de Villiers [4], we treated homography as a temporal estimation task, incorporating frame-to-frame consistency to handle camera motion and irregular field marking visibility.

E. Team Classification

To represent players in a discriminative feature space, we employed SigLIP embeddings extracted from cropped player images. SigLIP generates robust visual representations by aligning image features with semantic concepts, allowing the model to capture distinctive cues such as jersey colors and textures. These embeddings were then used as input to the KMeans clustering algorithm, which grouped players into two clusters corresponding to team affiliations. By leveraging the uniformity of kit colors, the method eliminates the need for manual labeling and provides an efficient solution for real-time team classification, inspired by Pan et al. [3].

F. Match Analysis

1) Possession Estimation: We estimate possession by assigning ball ownership each frame to the nearest player or goalkeeper whose footpoint lies within an adaptive radius tied to scene scale. To avoid flicker, a change of owner is accepted only after short, sustained evidence and a minimum ball displacement; otherwise the previous owner is held briefly through dropouts. Referees are excluded, goalkeepers are first mapped to teams, and per-frame ownership is integrated over time to yield cumulative percentages.

2) Attack Intensity (Goal-Treat): We define a per-team, real-time measure of goal threat that combines how close the ball is to the opponent's goal with how strongly it is moving toward that goal. After mapping the ball to field coordinates via homography, we compute a normalized proximity term (closer is higher) and a directional velocity term (forward motion contributes, backward does not). The score can be gated by possession so only the team currently in control accumulates threat. To avoid jitter, we smooth the signal with an exponential moving average and clamp values to a fixed range. This yields a stable, interpretable curve that rises during sustained attacks and falls as the ball moves away or changes direction.

G. Evaluation Metrics

To evaluate the performance of the object detection models, we employed standard metrics commonly used in computer vision benchmarks such as COCO, including Recall, mAP@50, and mAP@50–95.

1) Recall: Recall measures the model's ability to correctly identify all relevant objects. It is defined as the ratio of true positive detections to the total number of actual objects (true positives + false negatives):

$$Recall = \frac{TP}{TP + FN}$$

A higher recall indicates that the model is successfully detecting most of the objects present in the image, even if some of the bounding boxes may be less precise.

2) mAP@50: This metric computes the mean Average Precision at a fixed Intersection over Union (IoU) threshold of 0.5. A detection is considered correct if the IoU between the predicted bounding box and the ground truth exceeds 0.5. It is formally defined as:

$$mAP@50 = \frac{1}{N} \sum_{i=1}^{N} AP_i(IoU = 0.5)$$

3) mAP@50–95: This is a more comprehensive evaluation that averages the AP across multiple IoU thresholds, ranging from 0.5 to 0.95 with a step size of 0.05. This metric provides a stricter assessment of localization accuracy and robustness.

$$mAP@50-95 = \frac{1}{10N} \sum_{t=0.5}^{0.95} \sum_{i=1}^{N} AP_i(IoU = t)$$

While $mAP@50$ emphasizes detection performance at a moderate threshold and is more tolerant to bounding box alignment, $mAP@50$–95 serves as a stricter and more robust measure of detection precision and localization accuracy. Recall complements these by indicating the completeness of the model's detections.

VI. Results and Discussion

The following tables summarize the performance of our models on three key detection tasks: field line/keypoints, players/referees/goalkeepers, and ball detection.

TABLE I
Field Line / Keypoint Detection Performance

Model Type	Recall	mAP@50	mAP@50–95
Benchmark (317 images)	94%	97%	47%
Improved (813 images)	96%	98%	87%

TABLE II
Player, Goalkeeper, Referee Detection Performance

Model Type	Recall	mAP@50	mAP@50–95
Benchmark (All Classes)	82%	88%	67%
Improved (Without Ball)	98%	98%	81%

TABLE III
Ball Detection Performance

Annotated Images	Recall	mAP@50	mAP@50–95
2,226	83%	93%	61%

The results presented in Tables I–III demonstrate the effectiveness of task-specific training and increased data annotation in improving detection performance for football analysis.

Table I (Field line/keypoint detection) shows that recall improved from 94% to 96%, while mAP@50–95 increased markedly from 47% to 87%, indicating stronger generalization across diverse match conditions.

Table II (Player, referee, and goalkeeper detection) highlights that excluding the ball class and augmenting the training data substantially boosted performance, with recall increasing from 82% to 98% and mAP@50–95 improving from 67% to 81%.

Table III (Ball detection) reports results from over 2,226 annotated images, achieving 83% recall and 93% mAP@50. However, the lower mAP@50–95 of 61% underscores the inherent difficulty of accurately tracking such a small, fast-moving object.

Overall, these findings validate the proposed framework's ability to deliver accurate, real-time insights for tactical analysis and match event detection.

- **Improved vs. Benchmark Models**
  Across all detection tasks, the improved models trained with more annotated data significantly outperformed the benchmark versions. For example, in field line/keypoint detection, the mAP@50–95 increased from 47% to 87%, and recall improved from 94% to 96%. Similarly, for player, referee, and goalkeeper detection, the recall rose from 82% to 96% and mAP@50–95 from 67% to 87%. These improvements highlight the value of task-specific data expansion and fine-grained labeling in boosting model performance.

The improved models also retained high mAP@50 values (98%) across detection tasks, indicating strong performance even with looser overlap thresholds. The combination of increased data and optimized model configurations (e.g., using YOLOv8 with 1280 input size) contributed to these gains.

- **Effectiveness of Dataset Separation**
  Separating the ball class during training for player-related detection tasks led to significant improvements. The reduced class ambiguity and focused learning helped the model better distinguish between players, referees, and goalkeepers. This strategy contributed to the observed recall and mAP@50–95 improvements in those categories, supporting the importance of clear task separation when designing multi-class detection pipelines.

VII. Challenges

The main challenges faced in this project were the detection of small fast-moving objects like the ball, player occlusion and clustering, and variations in camera angles and lighting that affected tracking and homography. Ensuring real-time performance while processing large video data was also a key obstacle.

VIII. Resources and tools

The project utilized Python, PyTorch, and OpenCV as core development tools. For detection and analysis, YOLOv8, YOLOv8-Pose, ByteTrack, and K-Means with SigLIP embeddings were employed. Datasets were obtained from Roboflow, and experiments were conducted on GPU-enabled environments to ensure real-time performance.

IX. Conclusion and Future Work

This study demonstrates the effectiveness of using advanced object detection models, particularly YOLOv8 with high-resolution inputs, for automated football analytics. By leveraging task-specific training and increased annotation, the system achieved high accuracy in detecting players, referees, goalkeepers, the ball, and pitch keypoints. The improved models significantly outperformed benchmark baselines, confirming the importance of targeted data preparation and model tuning.

While overall detection performance was strong, small object detection remains a challenge under stricter evaluation thresholds, highlighting areas for further refinement.

For future work, we aim to extend the system's capabilities by incorporating offside detection using spatial positioning and homography mapping. Additionally, we plan to analyze player velocity over time to support sprint analysis and movement-based tactical insights. These enhancements will further contribute to the development of a robust, real-time football analytics framework.

References

[1] Dillon Reis et al. "Real-Time Flying Object Detection with YOLOv8". In: *arXiv preprint arXiv:2305.09972* (2024). URL: https://arxiv.org/abs/2305.09972.

[2] Debapriya Maji et al. *YOLO-Pose: Enhancing YOLO for Multi-Person Pose Estimation Using Object Keypoint Similarity Loss*. arXiv preprint arXiv:2204.06806. 2022. URL: https://arxiv.org/abs/2204.06806.

[3] Weiwei Pan et al. "Unsupervised Clustering in Football Analysis: A Color-Segmentation and Lighting Adaptation Approach". In: *IEEE Access* 12 (2024), pp. 178127–178140. DOI: 10.1109/ACCESS.2024.3506827. URL: https://doi.org/10.1109/ACCESS.2024.3506827.

[4] Paul J. Claasen and J. P. de Villiers. "Video-based Sequential Bayesian Homography Estimation for Soccer Field Registration". In: (2023). URL: https://arxiv.org/abs/2311.10361.

[5] Roboflow. *Give your software the sense of sight*. https://roboflow.com. Accessed: 2025-08-22. 2024.