Automated Football Analytics for the Jordanian Pro League Using Computer Vision

Prepared by:

Abdelrahman Abunasser (160825)
Hamza Sawafta (161395)
Hashem Elshaweesh (162374)
Raad Shraiedeh (161287)

Supervisor:
Rasha Obeidat

January 2026

Automated Football Analytics for the Jordanian Pro League Using Computer Vision

Abdelrahman Abunasser
Department of Computer Science
Jordan University of Science and Technology (JUST)
Irbid, Jordan
Email: arahmadabunsaer22@cit.just.edu.jo

Hamza Sawafta
Department of Computer Science
Jordan University of Science and Technology (JUST)
Irbid, Jordan
Email: hsalsawaftah22@cit.just.edu.jo

Hashem Elshaweesh
Department of Computer Science
Jordan University of Science and Technology (JUST)
Irbid, Jordan
Email: hkelshaweesh22@cit.just.edu.jo

Raad Shraiedeh
Department of Computer Science
Jordan University of Science and Technology (JUST)
Irbid, Jordan
Email: rmalshraiedeh22@cit.just.edu.jo

Dr. Rasha Obeidat
Department of Computer Science
Jordan University of Science and Technology (JUST)
Irbid, Jordan
Email: rmobeidat@just.edu.jo

***Abstract*—Advanced football analytics have historically been limited to well-resourced leagues with abundant tracking and event data, leaving regional competitions underrepresented. This study addresses that gap by developing a computer vision system for automated analysis of Jordanian professional football matches. A custom-labeled dataset is constructed from broadcast footage using annotation tools such as CVAT, given the absence of public data for this league. Several deep learning models for object detection (e.g., YOLO variants) and multi-object tracking (e.g., DeepSORT, ByteTrack) are trained and evaluated to detect players and the ball and to maintain their identities over time. A learning-based ball possession module is further introduced and integrated into the overall framework to infer the controlling player from visual and spatiotemporal cues and aggregate predictions to compute team possession percentages. The system produces automated metrics, including player trajectories, individual and team possession, and other derived indicators, directly from raw video. Initial experiments suggest that modern detectors and trackers can reliably operate on Jordanian league footage and, combined with the learned possession model, yield realistic, interpretable statistics suitable for low-resource sports analytics.**

***Index Terms*—football analytics, computer vision, object detection, multi-object tracking, ball possession, Jordanian Pro League**

I. PROJECT GOALS AND OBJECTIVES

This project aims to build an automated football analytics pipeline for the Jordanian Pro League using only broadcast video. The main objectives are:

- Construct a Jordanian Pro League dataset by annotating broadcast footage (players, goalkeepers, referees, ball) using a professional annotation workflow.
- Train and benchmark state-of-the-art object detectors for player and ball detection under Jordanian broadcast conditions.
- Apply multi-object tracking to generate stable player trajectories and maintain identities across frames.
- Assign teams based on jersey appearance cues and support match-wide team-level analytics.
- Develop a learning-based possession inference module to estimate individual and team ball possession.
- Produce interpretable automated statistics (trajectories, possession percentages, and derived indicators) from raw match footage.

II. INTRODUCTION

Football analytics has become an essential component of modern performance analysis, supporting objective evaluation of players and teams and informing tactical decision-making [1]. In top-tier competitions, clubs routinely employ commercial data providers and sensor-based tracking systems to obtain detailed positional and event data throughout each match [2]. By contrast, such infrastructures are rarely available in smaller or regional leagues, where financial and logistical constraints limit access to advanced technology.

The Jordanian Pro League exemplifies this disparity: there are no large-scale, publicly available annotated video datasets or in-stadium optical/GPS tracking comparable to those used in major European leagues. Most publicly accessible soccer video datasets—such as SoccerNet—focus on European competitions and specific broadcast styles, and are primarily designed for action spotting rather than league-specific tracking

and possession analysis [3]. This lack of regionally relevant benchmarks forces analysts in leagues like Jordan's to rely on manual video review or limited tabular statistics, highlighting the need for video-based analytics that do not assume big-data infrastructure.

Recent developments in computer vision and deep learning offer a way to bridge this gap. Single-stage object detectors such as YOLO can detect players and the ball in real time while maintaining strong accuracy [4]. Multi-object tracking (MOT) algorithms—including DeepSORT and subsequent variants— maintain identities across frames by combining motion and appearance information [5]. Together, these techniques can convert ordinary broadcast footage into rich spatiotemporal data without additional hardware. Among the many metrics that can be derived from such data, ball possession is particularly important, as it is frequently used to characterize tactical styles and match dominance [6]. However, possession is often still estimated manually or using simple heuristics.

This study uses the Jordanian Pro League as a representative setting for developing video-based football analytics in a low-resource environment. It involves (i) constructing a custom annotated dataset from Jordanian broadcast matches, (ii) adapting and evaluating state-of-the-art object detection and tracking methods for this domain, and (iii) developing a supervised model that infers ball possession from spatiotemporal features derived from tracking. This framework enables a systematic examination of both the technical feasibility and the practical value of computer-vision-based analytics in an underrepresented league.

This paper is organized as follows:

1) **Motivation:** The need for automated analytics in low-resource leagues.
2) **Previous Work and Research Gap:** Detection, tracking, and possession estimation literature and the gap for Jordanian league data.
3) **Methodology:** Dataset construction, detection and tracking pipelines, and possession inference.
4) **Expected Results:** Planned outputs and evaluation criteria.

III. LITERATURE REVIEW

A. Object Detection in Sports Footage

Object detection is a core component of automated sports video analysis. Deep convolutional neural networks dominate modern practice, with the "You Only Look Once" family especially influential due to real-time performance and strong accuracy [4], [7]. In football, such detectors have been applied to detect players and the ball in broadcast footage, with performance improving when models are fine-tuned on soccer-specific data [8].

Detecting the ball remains challenging due to its small size, rapid motion, and frequent occlusions. Heatmap-based approaches such as TrackNet have been used for high-speed tiny-object tracking in sports [9]. In practice, robust football ball detection typically requires a strong base detector plus training data that includes diverse ball appearances (rolling, airborne, partially occluded). Our study follows this strategy by creating a custom dataset tailored to Jordanian broadcast conditions.

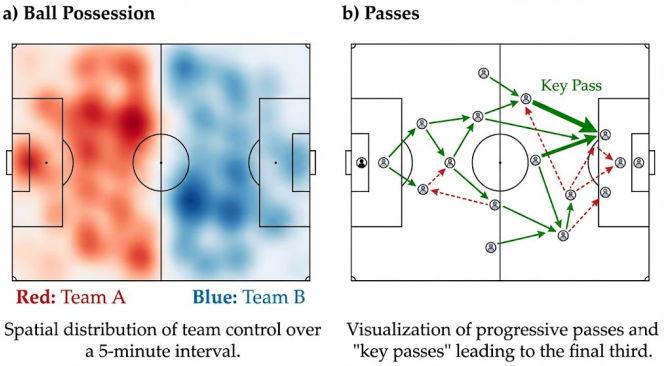

Fig. 1. Example visualizations of (a) team ball possession distribution and (b) passing patterns.

Beyond object detection, the proposed system is specifically designed to analyze two fundamental match events at the team level: ball possession and passing. These events are derived from spatiotemporal relationships between tracked players and the ball. Ball possession is inferred by combining sustained player–ball proximity with motion consistency over consecutive frames and is subsequently aggregated to estimate team possession phases. Passing is analyzed as a team-level event, identified through transitions of ball control within the same team, characterized by a distinct ball-velocity burst followed by a stable receiving phase, without explicitly attributing the pass to individual players. By restricting the event set to team possession and passing, the system focuses on reliable, interpretable analytics that can be robustly extracted from broadcast footage in a low-resource setting such as the Jordanian Pro League.

B. Multi-Object Tracking of Players

To derive player trajectories and enable higher-level movement and possession analysis, per-frame object detections must be associated consistently over time. In this work, we adopt a tracking-by-detection paradigm, in which players detected independently in each frame are linked into continuous trajectories using multi-object tracking (MOT) algorithms.

We evaluate two widely used MOT approaches that differ in how detections are associated across frames. SORT serves as a classical baseline, relying on a Kalman filter to predict object motion and the Hungarian algorithm to match detections between frames based on spatial overlap (Intersection-over-Union) [10]. While SORT is computationally efficient, it does not model visual appearance, making it sensitive to occlusions and identity switches in crowded football scenes.

DeepSORT extends SORT by incorporating deep appearance embeddings extracted from player crops [5]. These

Tracker	Motion Model	Appearance Features	Strengths	Observed Behavior in Our Dataset
SORT	Kalman Filter	No	Fast, simple	Frequent identity switches in crowded scenes
DeepSORT	Kalman Filter	Yes (ReID embeddings)	Improved identity consistency	More stable tracks, occasional fragmentation
ByteTrack	Kalman Filter	Optional	Robust under occlusion	Most complete and stable trajectories

TABLE I
COMPARISON OF TRACKING METHODS USED IN OUR PIPELINE.

embeddings enable the tracker to distinguish visually similar players and to recover identities after short-term occlusions. In our system, DeepSORT is used to maintain player identities across frames, producing more stable trajectories than motion-only tracking, particularly during close player interactions.

More recently, ByteTrack has been proposed to further improve tracking robustness by associating both high-confidence and low-confidence detections, rather than discarding detections below a confidence threshold [11]. This strategy is especially beneficial in football broadcasts, where players may be partially occluded or blurred due to camera motion. In our pipeline, ByteTrack operates directly on the outputs of the YOLO-based detector and emphasizes track continuity under challenging visual conditions.

All tracking methods are integrated downstream of the same detection backbone to ensure a fair comparison. Qualitative and quantitative evaluation on the Jordanian Pro League dataset indicates that DeepSORT substantially reduces identity switches compared to SORT, while ByteTrack produces the most complete and temporally stable tracks, particularly during occlusions and rapid camera movements. Based on this comparison, ByteTrack is selected as the primary tracker for downstream possession and passing analysis, while DeepSORT is retained as a secondary baseline for evaluation.

C. Ball Possession Estimation in Football

Ball possession is a key performance indicator linked to tactical style and match dominance [6], [12]. Simple proximity rules (assigning possession to the nearest player) can fail in contested situations. Learning-based approaches instead infer possession from spatiotemporal features derived from tracking. Prior work has shown that supervised models can better capture possession dynamics than static rules [6], [13], [14]. Our approach follows this direction by training a possession classifier using features extracted from our detection and tracking pipeline.

D. Sports Match Video Dataset

Existing sports video datasets span multiple tasks (action spotting, detection, tracking) and multiple sports, but most are collected under specific production standards and therefore do not always transfer well across leagues or broadcast settings. In football, widely used resources include SoccerNet for action spotting [3] and SoccerNet-Tracking for player/ball multi-object tracking [15]. Additional soccer datasets and benchmarks have also been used in the literature for detection/tracking and event analysis (e.g., multi-camera or broadcast-derived collections such as ISSIA-CNR and World Cup–based event datasets) [16], [17]. Beyond football, large-scale datasets exist for other sports and settings, such as SportsMOT (multi-sport multi-object tracking across basketball, football, and volleyball) [18], basketball-focused datasets including DeepSport [19] and multi-view datasets such as APIDIS [20], and recent action understanding benchmarks for sports such as MultiSports [21], FineGym [22], and FineSports [23]. At a broader scale, generic video action datasets that include many sports categories—such as Sports-1M [24] and Kinetics [25]—are often used for pretraining but still exhibit domain mismatch for league-specific broadcast analytics. Despite this breadth, multiple studies and surveys report that transferring models across leagues and broadcast environments remains difficult due to changes in camera viewpoints, resolution, lighting, field appearance, and team kits [26]. Consequently, constructing a league-specific dataset is essential for underrepresented competitions. We use the Computer Vision Annotation Tool (CVAT) to annotate Jordanian Pro League match footage and create a dataset tailored to local broadcast conditions [27].

Study (Authors + citation)	Dataset (and # classes)	Method summary	Main results
Redmon et al. [4]	**COCO** (80 classes)	Single-stage real-time object detection (YOLO)	Real-time detection with strong accuracy
Hartono et al. [8]	Football broadcast (players/ball; setup-dependent)	YOLO-based football detection + tracking pipeline	Fine-tuning improves soccer detection performance
Bewley et al. (SORT) [10]	MOT benchmarks (varies)	Kalman filter + Hungarian matching	Efficient baseline; weaker under occlusion
Wojke et al. (DeepSORT) [5]	MOT benchmarks (varies)	SORT + deep appearance embeddings (ReID)	Fewer identity switches compared to SORT
Zhang et al. (ByteTrack) [11]	MOT benchmarks (varies)	Associates high- and low-confidence detections	More complete and stable tracks under occlusion
Cioppa et al. (SoccerNet-Tracking) [15]	**SoccerNet-Tracking** (soccer; benchmark-defined classes)	Soccer-specific MOT benchmark and analysis	Shows camera cuts and occlusion remain challenging
Giancola et al. (SoccerNet) [3]	**SoccerNet** (action labels; class count varies by version)	Large-scale dataset for soccer video understanding	Enables standardized benchmarking in soccer video
Link & Hoernig [6]	Tracking/positional data (players+ball)	Supervised possession inference from spatiotemporal features	High agreement with expert possession annotations (reported)
Borghesi et al. [14]	Spatiotemporal/trajectory data (team focus)	Deep learning for team-level possession estimation	Demonstrates feasibility of team possession modeling
Sekachev et al. (CVAT) [27]	Tool (not a dataset)	Annotation platform supporting bounding boxes and tracks	Enables practical labeling and track annotation
Manafifard et al. (survey) [26]	Multiple soccer datasets	Survey emphasizing domain shift and open challenges	Highlights generalization issues across leagues and broadcasts

TABLE II
SUMMARY OF RELATED WORK RELEVANT TO DETECTION, TRACKING, DATASET CONSTRUCTION, AND POSSESSION INFERENCE.

E. Significance of Work

This project contributes a practical pipeline for producing football analytics in a low-resource setting using only broadcast video. By building a Jordanian Pro League dataset and validating detection, tracking, and possession inference under local conditions, the work helps close a regional benchmarking gap and supports scalable analytics without specialized in-stadium hardware.

IV. APPROACH AND METHODOLOGY

A. Pipeline Overview

The system follows a tracking-by-detection pipeline: (i) curate and annotate Jordanian broadcast footage, (ii) train and benchmark detection models for players and the ball, (iii) maintain identities through multi-object tracking, (iv) assign teams using jersey appearance cues, and (v) infer possession and passing from spatiotemporal signals.

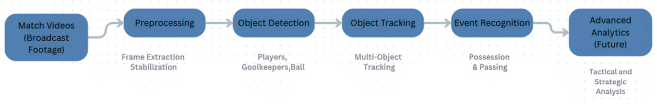

Fig. 2. Overview of the proposed football analytics pipeline from broadcast video to event recognition and downstream analytics.

B. Dataset Acquisition and Preparation

Source Data. Raw broadcast footage is collected from Jordanian Pro League matches. Compared to standardized top-tier broadcasts, this footage often exhibits variable camera quality, non-uniform angles, motion blur, and frequent shot changes, increasing the difficulty of small-object detection (ball) and long-horizon identity preservation (players).

Custom Annotation. Due to the absence of public benchmarks for this league, we establish a labeling pipeline using professional tools, primarily CVAT [27]. We annotate bounding boxes and (when applicable) consistent track identities across consecutive frames.

Classes. Annotated classes include *Team A Player*, *Team B Player*, *Goalkeeper*, *Referee*, and *Ball*.

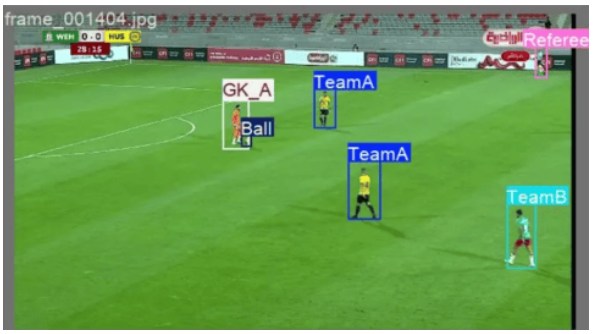

Fig. 3. Example annotations from our Jordanian Pro League dataset (Team A/Team B players, goalkeeper, referee, and ball).

Dataset Organization and Export Format. For reproducibility and efficient processing, the dataset is stored using a folder-per-match structure. Each match directory contains the original broadcast video (`video.mp4`) and a sequence of per-frame annotation files (`frame_000000.txt`, `frame_000001.txt`, ...), where each text file stores the labels corresponding to a single frame. This organization simplifies alignment between frames and labels, supports batch processing, and enables training/evaluation pipelines that operate on frame-indexed annotations.

Training Strategy. To mitigate limited labeled data, we apply transfer learning by initializing models from large-scale pretraining and then fine-tuning on the Jordanian dataset. Where appropriate, we leverage football-domain datasets for additional context [3].

C. Object Detection Framework

We benchmark multiple detectors to identify robust and efficient models for Jordanian broadcast footage:

- **YOLOv8** as a strong real-time baseline [28].
- **YOLOv9** for improved feature learning [29].
- **YOLOX** as an anchor-free baseline [30].
- **YOLOv10** as an end-to-end efficiency-oriented baseline [31].
- **RT-DETR** to test transformer-based detection for small targets [32].
- **FootAndBall** as a football-specialized detector [33].

D. Multi-Object Tracking (MOT)

Objective. The goal of this module is to convert per-frame detections (players and ball) into temporally consistent trajectories (track IDs). These trajectories are essential for downstream analytics, particularly team-level possession estimation and pass inference, which require stable identities over time rather than isolated frame detections.

Tracking-by-detection pipeline. We follow a standard tracking-by-detection approach. For each video frame, the detector produces bounding boxes; the tracker then links these boxes across time by predicting object motion and associating new detections to existing tracks.

Primary tracker (BoT-SORT). We use BoT-SORT as the main tracking method because it is designed for robust association in dynamic scenes. BoT-SORT combines: (i) Kalman filtering to predict the next position of each tracked object, (ii) camera-motion compensation to reduce errors caused by broadcast camera panning/zooming, and (iii) optional appearance (ReID) embeddings to improve identity preservation in crowded scenes [34].

Association rule. At each frame, we (1) run detection, (2) predict track states using the Kalman filter, and (3) match detections to tracks using geometric overlap and (when enabled) appearance similarity. Geometric overlap is measured by Intersection-over-Union (IoU).

Baselines for comparison. To verify robustness under Jordanian broadcast conditions (occlusions, motion blur, and camera motion), we compare BoT-SORT against two strong alternatives: (i) ByteTrack, which improves track continuity by also associating low-confidence detections, helping recovery under occlusion [11], and (ii) OC-SORT, which emphasizes improved handling of non-linear motion and occlusion recovery through observation-centric updates [35]. When appearance features are enabled, association follows the same principle as DeepSORT, using ReID embeddings to reduce identity switches among visually similar players [5].

Outcome. This module outputs track IDs and trajectories for detected objects, forming the spatiotemporal representation used by the possession and passing analysis stages.

E. Team Assignment

Each player track is assigned to a team using jersey-color cues. We extract dominant color features from the jersey region and cluster players into two groups using K-Means [36]. The clusters correspond to home/away team colors, enabling match-wide team labeling.

F. Event Detection and Possession Logic

Baseline heuristic. A player P_i is assigned possession if their distance to the ball B remains below a threshold δ for at least t consecutive frames [6]:

$$d(P_i, B) = \sqrt{(x_i - x_B)^2 + (y_i - y_B)^2} < \delta \qquad (1)$$

Learning-based possession inference. To handle contested situations, we train a supervised classifier using spatiotemporal features derived from tracking, following learning-based possession literature [13], [14]. Features include distance-to-ball, velocity and acceleration of nearest players, and alignment between player motion and ball displacement.

Pass detection. A pass is detected when possession transfers between two players on the same team after a ball-velocity burst and a consistent receiving phase. Pass length and speed can be computed when mapping image coordinates to pitch coordinates via camera calibration / homography methods [37], [38].

G. Evaluation Metrics

Overview. We evaluate the proposed framework at three levels: (i) object detection quality, (ii) multi-object tracking (trajectory consistency), and (iii) team-level event analytics (possession and passing). Intersection-over-Union (IoU) is used both as an overlap measure for bounding boxes and, separately, as a matching criterion inside the tracker; therefore, IoU is defined earlier in (**??**) in the context of evaluation, while tracking association uses it only for matching.

1) Object Detection Metrics: **IoU and mAP.** Detection quality is measured using IoU between a predicted box b_p and a ground-truth box b_g:

$$\mathrm{IoU}(b_p, b_g) = \frac{|b_p \cap b_g|}{|b_p \cup b_g|}. \qquad (2)$$

A prediction is counted as a true positive if it matches a ground-truth box of the same class with IoU above a chosen threshold; unmatched predictions are false positives

and unmatched ground-truth boxes are false negatives. Using these matches, we report Average Precision (AP) per class and mean Average Precision (mAP) across classes under the selected IoU threshold protocol. For additional interpretability, we also report precision and recall derived from TP/FP/FN counts.

2) Multi-Object Tracking Metrics: **Identity consistency and association quality.** Tracking performance is evaluated based on how well object identities are preserved over time. We report IDF1, which measures identity-based F1 score over the sequence, and identity switches (IDSW), which counts how often the tracker incorrectly changes an object's identity. When applicable, we additionally report trajectory continuity indicators (e.g., fragmentation) to reflect how often tracks are broken during occlusions and camera motion.

3) Team-Level Possession and Passing Metrics: **Possession.** Team possession is computed by aggregating predicted possession states over frames to obtain team possession percentages. If manual possession annotations are available for a subset of sequences, we evaluate frame-level (or window-level) possession predictions using accuracy and/or macro-F1 across possession states (Team A, Team B, no-clear-control). If such labels are not available, we report qualitative examples and compare against a proximity-based baseline to verify consistency with observable match flow.

Passing. Passing is treated as a team-level event derived from ball-control transitions and ball-motion cues. If ground-truth pass annotations are available, pass detection is evaluated using precision, recall, and F1-score. Otherwise, we provide qualitative examples and aggregate summaries (e.g., pass counts and sequences) while avoiding claims of detection accuracy without labeled ground truth.

H. LOCATION AND SAFETY CONSIDERATIONS

Data access and privacy. The system uses publicly broadcast match footage. The project does not collect private personal data beyond what is visible in broadcasts. Output analytics are aggregated at the team/player track level for performance analysis.

Ethical and legal considerations. Footage is used strictly for academic research and evaluation. Any sharing of clips, frames, or derived datasets should follow university guidelines and copyright constraints.

V. INITIAL RUN AND EXPECTED RESULTS/OUTPUTS

1) Initial Run (Technical Summary): **Official Technical Summary: First YOLOv8 Football Detection Results.**

Project Methodology. A one-minute football video (1920×1080, 29.97 FPS) was used, producing 1530 frames. Each frame was manually annotated with YOLO labels for six classes: Team A, Team B, Goalkeeper (A/B), Ball, and Referee. The dataset was split into 80/20% (train/validation). A YOLOv8-nano model pretrained on COCO was fine-tuned on this data. Localization quality was assessed using a ground-truth-driven IoU strategy, where each ground-truth box was matched to the highest-IoU prediction of the same class, while predictions without ground truth were disregarded.

Training Process. Training was performed with 640×640 processing, batch size 2, for 10 epochs on CPU. Training and validation losses decreased uniformly, indicating stable convergence. Precision, recall, and mAP metrics improved steadily across epochs, demonstrating good dataset preparation and effective model adaptation.

Visual Analysis. Training curves (Figure 1) show smooth reduction of box, classification, and DFL losses on validation curves and close similarity to training curves. Precision and recall exceeded 0.9, indicating stable learning.

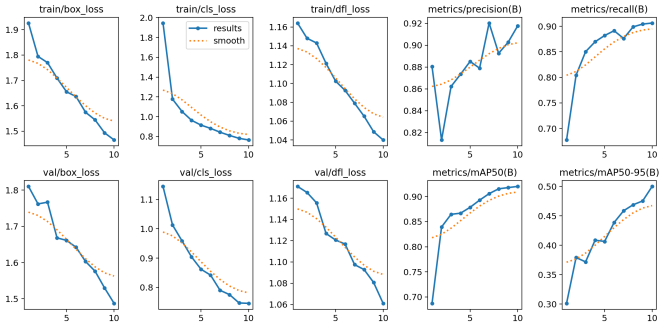

Fig. 4. Training/validation curves and detection metrics across epochs for the initial YOLOv8-nano fine-tuning run.

Label Statistics show high class imbalance: player classes dominate while the ball is small and compact in width–height distribution, explaining the increased recognition difficulty. The normalized confusion matrix (Figure 3) shows high class-specific accuracy for players, goalkeepers, and referees, while the ball class is more frequently confused with background, resulting in lower recall.

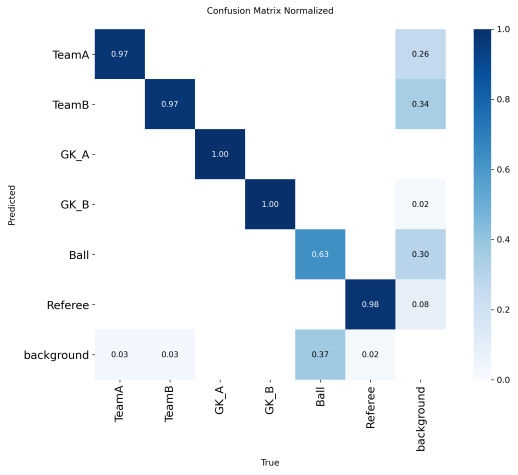

Fig. 5. Normalized confusion matrix for the initial YOLOv8-nano validation results.

Table 1. Table III summarizes overall validation performance.

TABLE III
PERFORMANCE METRICS ON THE VALIDATION SET

Metric	Value
Precision	~ 0.92
Recall	~ 0.91
mAP@0.5	~ 0.92
mAP@0.5:0.95	~ 0.50
Overall Mean IoU	0.7658

Table 2. Table IV provides class-level analysis.

TABLE IV
CLASS-LEVEL DETECTION ANALYSIS ON THE VALIDATION SET

Class	Instances	Norm. Recall	Key Observation
Team A	5461	0.97	Strong detection and localization
Team B	4928	0.97	Strong detection and localization
GK (A/B)	286	1.00	Very high accuracy (low frequency)
Ball	1156	0.63	Small object; most challenging
Referee	1061	0.98	High classification reliability

Conclusion. The preliminary results showed that YOLOv8-nano performs well on football object abstractions in a short CPU-trained dataset. Player and referee detection achieved high accuracy, while ball detection remains limited mainly by scale and motion. The obtained mean IoU of 0.77 provides a solid baseline for future improvements using longer videos, higher resolutions, and more advanced models.

Expected results/outputs. We expect the following outputs:

- A custom annotated dataset for Jordanian Pro League broadcast footage.
- Trained detection models for players and the ball, evaluated under local broadcast conditions.
- Stable multi-object tracks enabling player trajectories and match-level movement statistics.
- Automated possession estimation (player-level and team-level possession percentages).
- A reproducible pipeline and demo results suitable for a graduation project report.

REFERENCES

[1] A. Ferraz *et al.*, "Tracking devices and physical performance analysis in team sports: A scoping review," *Frontiers in Sports and Active Living*, vol. 5, p. 1284086, 2023.

[2] L. Torres-Ronda *et al.*, "Tracking systems in team sports: A narrative review of applications of the data and sport specific analysis," *Sports Medicine - Open*, vol. 8, no. 1, p. 15, 2022.

[3] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, "Soccernet: A scalable dataset for action spotting in soccer videos," in *CVPR Workshops*, 2018.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *CVPR*, 2016.

[5] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3645–3649, 2017.

[6] D. Link and M. Hoernig, "Individual ball possession in soccer," *PLOS ONE*, vol. 12, no. 7, p. e0179953, 2017.

[7] A. Kotthapalli *et al.*, "Yolo-based object detection: A comprehensive survey," *arXiv preprint*, 2025.

[8] M. R. Hartono, C. A. Sari, and R. R. A. Al-Jawry, "Football player tracking, team assignment, and speed estimation using yolov5," *Jurnal Teknik Informatika*, vol. 6, no. 1, pp. 51–62, 2025.

[9] Y.-C. Huang, I.-N. Liao, C.-H. Chen, *et al.*, "Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 11, pp. 3291–3301, 2019.

[10] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3464–3468, 2016.

[11] Y. Zhang, P. Sun, Y. Jiang, *et al.*, "Bytetrack: Multi-object tracking by associating every detection box," in *ECCV*, 2022.

[12] C. A. Casal *et al.*, "Possession zone as a performance indicator in football," *Frontiers in Psychology*, vol. 8, p. 1176, 2017.

[13] A. Peral, C. Redondo-Cabrera, and J. M. Luque, "Ball possessor recognition using spatio-temporal deep learning in soccer videos," 2025.

[14] A. Borghesi *et al.*, "Estimating ball possession in football using deep learning on spatiotemporal data," *Expert Systems with Applications*, vol. 226, p. 119780, 2023.

[15] A. Cioppa *et al.*, "Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos," in *CVPR Workshops*, 2022.

[16] J. Theiner, C. Pio, T. B. Moeslund, *et al.*, "Extraction of positional player data from broadcast soccer videos," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022.

[17] R. A. Sharma and A. Sharma, "Automatic analysis of broadcast football videos using contextual information," *Signal, Image and Video Processing*, 2016.

[18] Z. Cui *et al.*, "SportsMOT: A large multi-object tracking dataset in multiple sports scenes," *arXiv preprint*, 2023.

[19] G. Van Zandycke *et al.*, "DeepSportRadar-v1: A dataset for sports analytics," in *Proceedings of the 13th ACM Multimedia Systems Conference (MMSys)*, 2022.

[20] "APIDIS: Automated production of individualized digital sport (multi-view basketball dataset)." Project/Dataset, 2010. APIDIS project dataset.

[21] Y. Li *et al.*, "MultiSports: A multi-person video dataset of spatio-temporal action detection in sports," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[22] D. Shao, Y. Zhao, B. Dai, and D. Lin, "FineGym: A hierarchical video dataset for fine-grained action understanding," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[23] C. Xu *et al.*, "FineSports: A multi-person sports dataset for fine-grained action understanding," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[24] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[25] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, and A. Zisserman, "The Kinetics human action video dataset," *arXiv preprint*, 2017.

[26] A. Manafifard, H. Ebadi, and H. Ghassemian, "A survey on player tracking in soccer videos," *Computer Vision and Image Understanding*, vol. 159, pp. 19–46, 2017.

[27] B. Sekachev, N. Manovich, A. Zhiltsov, *et al.*, "Computer vision annotation tool (cvat)," 2020. Software project.

[28] Ultralytics, "Ultralytics yolov8 documentation," 2023. Software documentation.

[29] C.-Y. Wang *et al.*, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint*, 2024.

[30] Z. Ge *et al.*, "Yolox: Exceeding yolo series in 2021," *arXiv preprint*, 2021.

[31] A. Wang *et al.*, "Yolov10: Real-time end-to-end object detection," *arXiv preprint*, 2024.

[32] Y. Zhao *et al.*, "Detrs beat yolos on real-time object detection (rt-detr)," *arXiv preprint*, 2023.

[33] J. Komorowski, G. Kurzejamski, and G. Sarwas, "Footandball: Integrated player and ball detector," *arXiv preprint*, 2019.

[34] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," *arXiv preprint*, 2022.

[35] J. Cao *et al.*, "Observation-centric sort (oc-sort): Rethinking sort for robust multi-object tracking," *arXiv preprint*, 2022.
[36] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. on Mathematical Statistics and Probability*, 1967.
[37] A. Cioppa *et al.*, "Scaling up soccernet with multi-view spatial localization and re-identification," *Scientific Data*, vol. 9, p. 388, 2022.
[38] J. Theiner *et al.*, "Tvcalib: Camera calibration for sports field registration in soccer," in *WACV*, 2023.