Title

Smart Soccer Insights: Turning Game Footage into Data

Project Category

This project is an application project. It focuses on applying deep learning and computer vision techniques to extract player coordinates and key event information directly from raw soccer footage. The goal is to develop a functional end-to-end system that transforms video into structured, usable analytics data.

Team Details

Daniel Tshiani is the proposed Artificial Intelligence Engineer. Mr. Tshiani will be responsible for all phases of the project. This includes designing, implementing, and evaluating the entire end-to-end system. His responsibilities will include:

- Data Collection & Preprocessing: Gathering raw soccer footage, preparing datasets, annotating data when necessary, and organizing inputs for model training.
- Model Development: Selecting and implementing deep learning models for player detection, tracking, and event recognition
- System Integration: Building the pipeline that converts raw video into structured coordinate and event data, integrating detection, tracking, and analysis components.
- Testing & Evaluation: Measuring model accuracy, validating system outputs, and refining models to ensure reliable performance.
- Deployment & Automation: Developing scripts or tools that allow seamless processing of new footage and that generate analytics outputs automatically.
- Documentation & Reporting: Maintaining clear documentation of methods, experiments, results, and findings, and presenting progress and final outputs.

Introduction/ Motivation

Soccer analytics is rapidly expanding, but many of the most powerful tactical tools rely on detailed player coordinate data and in-game event data. Currently, teams often collect this information using wearable GPS tracking systems, which can be prohibitively expensive for smaller clubs. As a result, advanced analytics are typically limited to professional or well-funded organizations, leaving a large segment of the soccer community underserved.

I connected with a small sports analytics company, DSA-Labs, which has already developed algorithms capable of generating high-level tactical recommendations for coaches. However, these algorithms depend on accurate tracking and event data as inputs. Because many teams cannot afford GPS systems, DSA-Labs is unable to reach a broader market of lower-budget teams—even though these teams often have extensive game footage available.

The problem I aim to tackle is the creation of an automated deep learning pipeline that extracts player coordinates and event data directly from raw video footage. By replacing the need for costly wearable tracking systems, this project will allow DSA-Labs to deliver advanced tactical insights to a much wider range of clients. In essence, my work will help transform ordinary video into actionable analytics, opening an entirely new market segment for the company.

This problem is especially interesting and important within the context of deep learning because recent advances in computer vision have finally made it feasible to analyze free-flowing sports like soccer, which historically have been difficult to model due to their continuous movement and complex interactions. Developing a system that can detect players, track their motion, and identify key events requires integrating modern object detection, tracking algorithms, and temporal modeling—making it an ideal application of real-world deep learning techniques. This project not only demonstrates the practical value of deep learning but also contributes to the broader evolution of data-driven sports analytics.

Methodology

This project will develop a complete computer vision pipeline that converts raw soccer footage into coordinate and event data using modern deep learning techniques. First, I will use a state-of-the-art object detection model such as YOLOv8 or YOLOv9 to identify players and the ball in each frame, since these models offer an ideal balance of speed and accuracy for video analysis and are highly effective at detecting small objects and handling occlusions. To track players consistently across time, I will integrate a multi-object tracking method such as ByteTrack or DeepSORT, which can maintain player identities even during overlaps or fast movements. After detecting and tracking players in pixel space, I will estimate the camera’s homography to the soccer field so that the system can convert pixel positions into real-world field coordinates. This step is crucial because DSA-Labs' tactical algorithms operate on metric field positions rather than image coordinates.

Once trajectories are extracted, I will implement temporal deep learning models—such as LSTMs, GRUs, 1D CNNs, or transformer-based architectures—to recognize events like passes, shots, tackles, and transitions. These models are well-suited for event detection because soccer actions occur over sequences of frames, and temporal architectures can capture the motion patterns and interactions needed to distinguish between different types of events. Finally, I will build an end-to-end post-processing pipeline that smooths trajectories, cleans noisy detections, and formats the outputs for direct use in DSA-Labs' existing tactical recommendation system. Together, these techniques create an integrated deep learning solution capable of turning raw video footage into structured, high-quality analytics data.

Intended Experiments

To evaluate the performance of my approach, I will run a series of experiments that assess each component of the system as well as the full end-to-end pipeline. I will begin by evaluating the object detection and tracking modules independently, testing combinations such as YOLOv8 with ByteTrack or DeepSORT to compare their accuracy, robustness, and consistency in identifying players and the ball. I will then assess the quality of camera calibration by comparing different homography estimation methods—such as using manually annotated field keypoints, automated line detection, or learned keypoint models—and measuring how accurately each method converts pixel coordinates into real-world field positions. For event detection, I will experiment with multiple temporal deep learning architectures, including 1D CNNs, LSTMs/GRUs, and lightweight transformers, to determine which model best captures motion patterns and correctly identifies actions such as passes, shots, and tackles. In addition to these component-level experiments, I will conduct ablation studies to quantify the contribution of individual pipeline elements and perform robustness tests by evaluating the system on matches with varying camera angles, resolutions, and levels of occlusion. Finally, I will run full end-to-end evaluations where the pipeline generates complete coordinate and event outputs, which will then be compared against GPS-based ground truth (when available) and human annotations.

To measure the effectiveness of the system, I will use standard metrics for each subtask. Detection performance will be evaluated with mean Average Precision (mAP), precision, and recall. Tracking quality will be assessed using MOTA, MOTP, IDF1, and identity-switch counts. For coordinate accuracy, I will compute real-world positional errors, such as RMSE and displacement errors, and report the percentage of predictions within key distance thresholds. Event detection models will be evaluated using per-class precision, recall, F1-scores, and temporal overlap measures. For the complete pipeline, I will compare the tactical outputs produced using video-derived data to those produced using GPS data or expert annotations, measuring similarity and agreement. I will also record computational performance metrics such as inference speed and resource usage to ensure the system is practical for real-world use. Together, these experiments and evaluation criteria will provide a thorough assessment of both the accuracy and usefulness of the proposed approach.

Prior Research

One of the most relevant prior works is the paper "Extraction of Positional Player Data from Broadcast Soccer Videos" by Theiner et al. (WACV 2022). In their pipeline, they tackle field registration, shot boundary detection, player detection, and team assignment before converting detections into two-dimensional field coordinates using homography. Their work represents a strong, transparent baseline: they break the system into modular components, evaluate each one, and analyze how errors in earlier stages propagate into the final positional estimates. They also propose novel evaluation metrics to compare the estimated positions to ground truth, because non-visible players and pipeline error accumulation make this a non-trivial problem. This research clearly demonstrates the feasibility of extracting usable tracking data from broadcast video — without relying on wearable devices.

However, while Theiner et al. provide an excellent foundation, their work predates some significant advances in computer vision, particularly in object detection. For instance, YOLOv8 — released in January 2023 by Ultralytics — offers meaningful improvements in both accuracy and efficiency relative to earlier YOLO versions. (Sapkota 2025)

Compared to, say, YOLOv5 or earlier real-time detectors, YOLOv8 achieves higher mAP while maintaining or even improving inference speed. These improvements are especially important for sports analytics: in soccer footage, detecting small, fast-moving objects (like the ball) and coping with occlusion can be very challenging, so using a more modern, high-precision detector could significantly improve overall performance and robustness.

In addition, more recent YOLO models (including YOLOv8) incorporate architectural advancements that can further benefit a tracking and coordinate-extraction pipeline. YOLOv8's design includes anchor-free detection, decoupled heads for objectness, classification, and bounding-box regression, and more efficient feature modules (such as the C2f module), all of which contribute to stronger performance (Sapkota 2025). Given these advances, there is a compelling case to revisit and extend the Theiner et al. pipeline using newer detection models to potentially reduce error in player detection, reduce false positives/negatives, and improve the fidelity of reconstructed coordinate data. By doing so, my project builds directly on solid academic work while leveraging state-of-the-art detection technology to push the boundary of what's possible in video-based soccer analytics.

References

Jonas Theiner, Wolfgang Gritz, Eric Müller-Budack, Robert Rein, Daniel Memmert, Ralph Ewerth; Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2022, pp. 823-833

Roboflow. (2025, August). football-players-detection Dataset. Roboflow Universe. Retrieved from https://universe.roboflow.com/roboflow-jvuqo/football-players-detection-3zvbc

Roboflow. (2024, August). football-field-detection Dataset. Roboflow Universe. Retrieved from https://universe.roboflow.com/roboflow-jvuqo/football-field-detection-f07vi

Roboflow. (2024, July). football-ball-detection Dataset. Roboflow Universe. Retrieved from https://universe.roboflow.com/roboflow-jvuqo/football-ball-detection-rejhg

Sapkota, R., Flores-Calero, M., Qureshi, R. et al. YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series. Artif Intell Rev 58, 274 (2025). https://doi.org/10.1007/s10462-025-11253-3