INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA DE INTELIGÊNCIA ARTIFICIAL



Design and Implementation of a Low-Cost Computer Vision Pipeline for Amateur Football Analysis

RAFAEL NUNO DE SOUSA ALVES
outubro de 2025

Design and Implementation of a Low-Cost Computer Vision Pipeline for Amateur Football Analysis

Rafael Nuno Sousa Alves
Student No.: 1201793

Dissertation for the Master's Degree in Artificial Intelligence Engineering

Supervisor: Paulo Matos
Co-Supervisor: Constantino Martins

Evaluation Committee:
President:
Luís Manuel Silva Conceição, Assistant Professor, Institute of Engineering, Polytechnic of Porto

Members:
José Manuel Ferreira Machado, Full Professor, School of Engineering, University of Minho
Paulo Sérgio dos Santos Matos, Assistant Professor, Institute of Engineering, Polytechnic of Porto

Porto, September 30, 2025

Abstract

The advancement of computer vision and artificial intelligence has opened new possibilities for sports analytics, particularly in football. This dissertation explores the development of an AI-powered multi-platform application designed to track and analyze amateur football matches without the need for wearable sensors. By leveraging computer vision techniques such as object detection, multi-object tracking, and real-time analytics, this research aims to provide an accessible and cost-effective solution for performance analysis in amateur football. The work presents a systematic review of existing methodologies, identifying key challenges such as occlusion, motion blur, and real-time computational constraints. A methodological framework based on the Design Science Research (DSR) approach guides the investigation, ensuring iterative development, validation, and refinement of the proposed system. The findings of this study lay the groundwork for the future implementation of a fully functional AI-based tracking system. Over the next six months, the research will transition into a practical phase, involving model training, system deployment, and real-world testing. By addressing the identified challenges and leveraging recent advancements in AI and computer vision, this project aims to bridge the gap between professional and amateur sports analytics.

Keywords: Computer Vision, Object Tracking, Football Analysis, Multi-Object Tracking, Sports Analytics, AI in Sports

Resumo

O avanço da visão por computador e da inteligência artificial abriu novas possibilidades para a análise desportiva, especialmente no futebol. Esta dissertação explora o desenvolvimento de uma aplicação multiplataforma baseada em IA para rastrear e analisar jogos de futebol amador sem a necessidade de sensores. Através de técnicas avançadas de visão por computador, como detecção de objetos, rastreamento múltiplo de objetos e análise em tempo real, este estudo propõe uma solução acessível e eficaz para a análise de desempenho no futebol amador. A revisão sistemática conduzida identifica os principais desafios existentes, como a obstrução de objetos, o desfoque devido ao movimento e as limitações de computação em tempo real. A investigação é estruturada com base na metodologia de Design Science Research (DSR), permitindo o desenvolvimento iterativo, a validação e a melhoria contínua do sistema proposto. Os resultados deste estudo servirão como ponto de partida para a implementação futura de um sistema completo de recolha de estatisticas baseado em IA. Nos próximos seis meses, o projeto avançará para a fase prática, envolvendo o treino do modelo, a implementação do sistema e testes em cenários reais. Ao enfrentar os desafios identificados e aproveitar os avanços recentes da inteligência artificial e da visão por computador, este projeto visa reduzir a distância entre a análise desportiva profissional e amadora.

Palavras-chave: Computer Vision, Object Tracking, Football Analysis, Multi-Object Tracking, Sports Analytics, AI in Sports

Acknowledgement

I would like to express my deepest gratitude to my supervisor, Professor Paulo Matos, and my co-supervisor, Professor Constantino Martins, for their invaluable guidance, encouragement, and constructive feedback throughout the development of this dissertation. Their expertise and support were fundamental to the successful completion of this work.

I am also profoundly thankful to my family and friends for their patience, understanding, and unwavering encouragement, which gave me the strength to persevere during the most challenging moments. A special word of appreciation goes to my close friends Bruno Pereira and Diogo Machado, whose support and friendship have been an immense source of motivation throughout this journey.

Finally, I would like to thank my twin brother, Nuno Alves, for joining me on this ambitious adventure. Sharing this project with him not only made the journey easier but also transformed it into an experience filled with fun, collaboration, and unforgettable moments.

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence.
AR	Augmented Reality.
CNN	Convolutional Neural Network.
DSR	Design Science Research.
FIFA	Fédération Internationale de Football Association.
fps	frames per second.
GAN	Generative Adversarial Networks.
GDPR	General Data Protection Regulation.
LiDAR	Light Detection And Ranging.
MOT	Multi-Object Tracking.
MPJPE	Mean Per Joint Position Error.
NDA	Non-Disclosure Agreement.
P	Precision.
PR	Precision-Recall.
R	Recall.
R-CNN	Region-based Convolutional Neural Network.
ReID	Re-IDentification.
SLR	Systematic Literature Review.
SORT	Simple Online and Realtime Tracking.
UDP	User Datagram Protocol.
ViT	Vision Transformer.
YOLO	You Only Look Once.

Chapter 1

Introduction

This thesis explores the development and application of Artificial Intelligence (AI) within the domain of amateur sports. It presents a novel multi-platform application designed to capture and analyze player performance data, providing valuable insights and enhancing the overall sport experience. This chapter offers a comprehensive overview of the thesis, encompassing its context, motivations, objectives, contributions, methodology, and structure, ultimately establishing the significance of the research and its proposed solution.

1.1 Context

Sports analytics is used to extract valuable information from games, including insights into gameplay, playing styles, tactics, and individual and team performance. This information is invaluable to professionals, enabling them to identify strengths and weaknesses, optimize strategies, and ultimately increase competitiveness. Existing research in sports analytics spans various sports, including basketball (Pai, ChangLiao, and K.-P. Lin 2017; Tarashima 2021), football (Mavrogiannis and Maglogiannis 2022), padel (Javadiha et al. 2021), and others sports (Pardos, Menychtas, and Maglogiannis 2022).

Within football, there is particular interest in predicting match outcomes (Arntzen and Hvattum 2021; Constantinou 2019), extracting deeper insights and rules from data (Kröckel and Bodendorf 2020), and analyzing specific events like penalty kicks (Habibullah et al. 2020). While much of this research focuses on professional applications, such as coaching, scouting, and entertainment, there is a notable lack of systems that serves the needs of amateur players. Millions of people worldwide engage in recreational sports with friends and family, yet they lack access to affordable and user-friendly tools for performance analysis. This thesis aims to address this gap by developing a computer vision-based solution that democratizes access to performance tracking and analysis for amateur football players.

While professional football is predominantly played in 11-a-side format, the focus of this research lies in enhancing the experience of amateur players participating in the increasingly popular 5-a-side and 7-a-side versions. These formats, often played on indoor or outdoor rented pitches, present a unique opportunity for technological intervention to elevate the playing experience.

The ability to track and analyze performance statistics offers players valuable insights into their gameplay, facilitating skill development and increasing enjoyment. Existing solutions, such as wearable sensors and post-processing video analysis, while more accurate (Kim et al. 2022), present limitations including cost, maintenance requirements, and usability challenges.

This thesis addresses these limitations by proposing a computer vision-based solution. Strategically positioned cameras will capture gameplay footage, enabling automated extraction of performance data without the need for wearable sensors. This approach minimizes equipment-related concerns, reduces manual effort, and offers a cost-effective solution for amateur players.

The proposed multi-platform application is part of a broader system designed to enhance amateur football experiences by providing various features such as performance tracking, game highlights, team management, community-building tools, and future multi-sport support. Within this comprehensive framework, this thesis specifically focuses on the performance tracking component, which aims to deliver detailed statistics on distance covered, top speed, ball velocity, passes completed, and heat maps. The development of this module will be prepare for future expansions that will integrate additional functionalities into the system.

By democratizing access to advanced performance analysis, this application aims to bridge the gap between professional and amateur sports. Furthermore, by incorporating gamification elements like challenges, achievements, and leaderboards, the platform fosters a more engaging and competitive environment for players. This innovative approach not only empowers players but also provides sports facility owners with a valuable tool to enhance their services and attract customers.

1.2 Motivation

Sports hold a vital role in society, contributing significantly to physical fitness, cognitive development, and social interaction. Participation in athletic activities fosters teamwork, encourages healthy competition, and cultivates a sense of community. Among the diverse array of sports, football enjoys unparalleled global popularity and participation. Its accessibility, requiring minimal equipment and infrastructure, makes it a truly inclusive activity enjoyed by individuals from all socioeconomic backgrounds. My motivations for this project stems from a combination of personal, global, and academic motivations. Having played federated football for over 12 years and continuing to play amateur football, I have a deep personal connection to the sport and aspire to enhance the way players at my level experience and improve their game. The global reach of football, which is the sport that is played the most worldwide and accessible across socioeconomic boundaries, emphasizes the importance of creating solutions that are affordable and scalable. Furthermore, as a student of Mestrado de Engenharia em Inteligência Artificial(MEIA) in Instituto Superior de Engenharia do Porto(ISEP), I have built expertise in computer vision through projects like obstacle prediction for warehouse robots and speech-to-text conversion using digital wave analysis. This thesis is a natural extension of my academic journey, combining my knowledge and passion to address a significant challenge in amateur football.

1.3 Objectives

The central aim of this research is to design and implement a computer vision–based system that addresses the limitations of existing football analytics solutions by explicitly targeting the amateur and recreational domain. Unlike the majority of prior work, which assumes broadcast-quality footage from professional matches, this project seeks to provide a low-cost, real-time, and field-agnostic solution that can function effectively under the variable conditions of grassroots football.

The hypothesis guiding this research is that **a dedicated pipeline, trained and evaluated on amateur-level data, can deliver accurate performance tracking and tactical insights in real time using only consumer-grade equipment**. By removing dependencies on green-field assumptions, high-resolution close-ups, or expensive hardware, such a system can extend the benefits of football analytics to millions of recreational players worldwide.

To achieve this goal, the following objectives have been defined:

- `Establish a Cost-Effective and Accessible System:` This objective centers on the system's core design philosophy: to provide a powerful analytical tool for amateur football with minimal cost and infrastructure requirements. The system's ability to operate effectively using consumer-grade equipment, such as smartphones and simple tripods, will be demonstrated.
- `Develop a Robust Computer Vision Pipeline:` This objective focuses on the technical core of the system. The aim is to create a multi-stage pipeline including object detection, player tracking, and field calibration that can accurately and reliably extract performance metrics from low-resolution video footage.
- `Achieve a Real-Time Viability:` This objective addresses the system's performance. The goal is to demonstrate that the system can process video at a rate that approaches or meets real-time requirements, thus providing timely insights. The success of this objective will be measured by the system's ability to maintain a processing rate of at least 20 frames per second (fps) without accumulating a significant delay.
- `Validate the System's Output through External Referencing:` This objective addresses the accuracy and reliability of the final data. The goal is to compare the system's output (e.g. speed and distance traveled) with data from an external, validated source, such as a GPS-enabled sports watch, to quantify and discuss the level of accuracy and potential error sources.
- `Develop a user-friendly application:` The system should be intuitive and easy to navigate, allowing users to initiate match tracking with a maximum of five clicks. This objective prioritizes simplicity and accessibility to ensure a seamless user experience, making performance tracking effortless for players of all skill levels.

These objectives guide the development and evaluation of the proposed application, ensuring alignment with the needs and expectations of amateur football players. By combining **real-time performance tracking with motivational features**, this study aims to demonstrate how technology can enhance player engagement and encourage wider participation in the sport.

1.4 Contributions

This thesis contributes to the advancement of sports technology and the promotion of physical activity by introducing an innovative solution for performance analysis in amateur football. The key contributions of this research are:

- `Increased accessibility to performance analysis:` By leveraging computer vision and a multi-platform application, this research provides a more affordable and accessible solution for performance tracking in amateur football. This contribution democratizes access to advanced analytics, previously limited to professional athletes due to the high cost of existing systems.
- `Promotion of physical activity and sports participation:` The application's engaging features and gamified elements contribute to increased motivation and participation in amateur football. This contribution has positive implications for public health and well-being by promoting physical activity and fostering a sense of community among players.
- `Advancement of computer vision applications in sports:` This research contributes to the growing field of computer vision by demonstrating its potential for automated performance analysis in amateur sports. The developed algorithms and system architecture can serve as a foundation for future research and development in this area.

1.5 Methodology

This research adopts the Design Science Research (DSR) methodology to guide the development and evaluation of the proposed multi-platform application (Vom Brocke, Hevner, and Maedche 2020). DSR focuses on creating and evaluating innovative artifacts to solve real-world problems as ilustrated by figure 1.1. In this context, the artifact is the AI-powered application for performance tracking in amateur football. The DSR process will involve the following iterative stages:

1. `Problem identification and motivation:` This stage involves clearly defining the problem (limited access to affordable performance analysis in amateur football) and establishing the motivation for the research (as outlined in Section 1.2).
2. `Objective definition:` This stage involves specifying the goals and desired outcomes of the research, as detailed in Section 1.3.
3. `Design and development:` This stage focuses on the design and development of the system, which includes both the multi-platform application and the extraction pipeline, guided by the identified needs and objectives. The process involves selecting suitable computer vision algorithms and designing an intuitive user interface. To ensure practical relevance, we emphasize maintaining a cost-effective solution, relying exclusively on personal recordings from outdoor public fields to evaluate and validate the models (Chapter 3).
4. `Demonstration:` This stage involves demonstrating the functionality of the application through testing and simulations(Chapter 3).
5. `Evaluation:` This stage involves rigorously evaluating the artifact's performance against the defined objectives. This will include assessing the accuracy of the performance data capture, measuring user engagement through feedback mechanisms, and evaluating the system's performance in diverse environments(Chapter 4).
6. `Communication:` This stage involves disseminating the research findings and insights through the thesis document and other publications(Chapter 5).

The DSR approach ensures a systematic and rigorous process for developing and evaluating a solution that effectively addresses the research problem as seen in the diagram bellow.

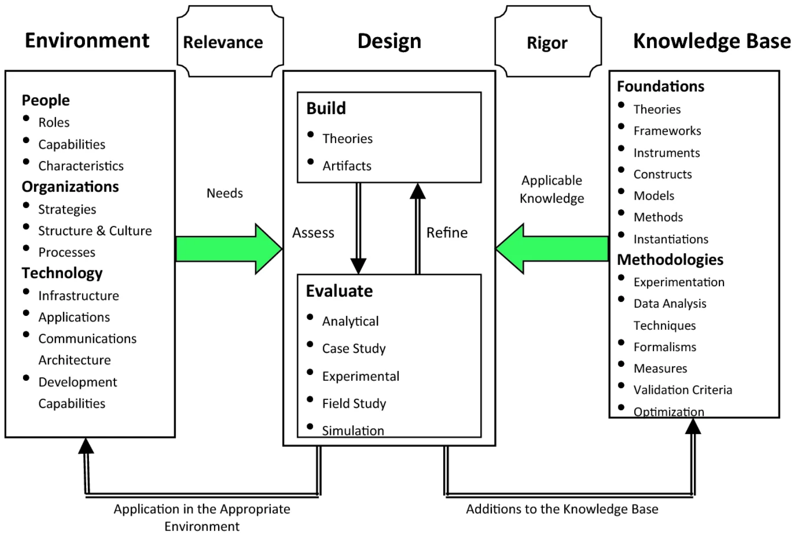

Figure 1.1: DSR flow diagram (Vom Brocke, Hevner, and Maedche 2020)

1.6 Document Organization

This dissertation is structured into four main sections, providing a comprehensive exploration of AI-driven football analytics.

- `Chapter 1:` Introduction – This chapter presents the context, motivation, objectives, and contributions of this research. It also outlines the methodological approach, DSR and provides an overview of the document structure.
- `Chapter 2:` State of the Art – This section reviews relevant literature in the field of AI-based sports analytics, focusing on key concepts such as computer vision, object tracking, and Multi-Object Tracking (MOT). It also presents a systematic review of existing research, identifying major challenges in AI-driven football analysis.
- `Chapter 3:` Implementation - This chapter will be dedicated to a comprehensive, objective exposition of the system's technical implementation, covering the architecture, data acquisition, key components (e.g., object detection, player tracking), and statistical extraction.
- `Chapter 4:` Evaluation - This chapter will present the results of the system evaluation, using the metrics we've established. It will analyze the performance of each component, discuss the significance of our findings, and address the limitations and errors encountered.
- `Chapter 5:` Conclusions – The final chapter summarizes the key findings, discusses the limitations of the current research, and outlines the next steps for future development. It emphasizes the transition from theoretical investigation to practical implementation, setting the stage for the upcoming phases of the project.

This structure ensures a logical flow, guiding the reader through the research process, from conceptualization to practical application, while maintaining clarity and coherence throughout the document.

1.7 Data Protection, Security, and Ethics

Ensuring the privacy, security, and ethical treatment of collected data is a fundamental requirement for this project. Since the system relies on video capture of amateur football matches, particular attention must be given to compliance with data protection regulations such as the General Data Protection Regulation (GDPR) and to broader principles of ethical artificial intelligence. This section outlines the measures taken to safeguard user privacy, maintain data security, and uphold ethical standards throughout the lifecycle of the project.

1.7.1 Data Protection and User Consent

The collection of video data inherently raises issues of personal privacy, as players are directly identifiable within the recordings. To address this, all individuals involved in a recorded match will be fully informed in advance about the purpose of data collection, the way in which the data will be processed, and the duration for which it will be retained. In line with GDPR requirements (Asghar et al. 2019), explicit consent will be obtained from every participant prior to recording. The system will be designed so that if even a single participant refuses consent, the recording process will not commence.

To further limit privacy risks, video data will only be stored temporarily. Recordings will be retained for no longer than twenty-four hours, during which they can be downloaded by authorized users, after which they will be automatically deleted. This approach minimizes the risk of data misuse while still allowing sufficient time for players or coaches to retrieve relevant insights.

1.7.2 Security Measures

In addition to limiting the retention of data, technical safeguards will be applied to prevent unauthorized access. All recorded files will be encrypted and pseudonymized where appropriate, ensuring that even in the unlikely event of a data breach, sensitive information cannot be directly linked to specific individuals (Asghar et al. 2019).

Special care will also be taken in preparing training datasets. To comply with ethical AI practices and to protect player identities, facial regions in training images will be blurred or anonymized before being incorporated into model development (Waelen 2024). Moreover, all data processing will be conducted within a closed environment, with no transfer of data to third-party services or external providers. By keeping data processing entirely in-house, the project minimizes exposure to external risks of leakage or unauthorized distribution.

1.7.3 Ethical Considerations

Beyond legal compliance and technical safeguards, the project is committed to broader principles of ethical AI. The system will be designed in accordance with recognized frameworks of fairness, transparency, and accountability (Waelen 2024). In practice, this means ensuring that algorithms are trained and evaluated in ways that minimize bias in player detection and tracking, particularly avoiding systematic disadvantages to players based on jersey color, body type, or field conditions. Transparency will be pursued by documenting model architectures, training processes, and limitations, enabling informed interpretation of the outputs.

Ethical practice also extends to the software development process itself. Best coding practices will be adopted to prevent vulnerabilities in both mobile and web applications, reducing the risk of exploitation or unauthorized access. By combining robust technical design with ethical awareness, the project seeks to ensure that the benefits of real-time amateur football analytics are delivered without compromising the rights, dignity, or security of participants.

Chapter 2

State of the Art

This chapter provides a concise overview of the research and technology relevant to this thesis, adopting a hybrid approach that combines a review of key concepts with a systematic analysis of existing literature. This hybrid approach offers a balanced review of the state of the art, combining a broad overview of concepts with a focused analysis of existing literature to identify challenges and opportunities in applying computer vision to football.

2.1 Key Concepts

This section introduces the fundamental concepts that form the basis of this thesis, providing a better understanding of the technologies and principles used in the development of the proposed application.

2.1.1 Artificial Intelligence

Defining AI has been a subject of ongoing debate and evolution (Ertel 2024). While a universally accepted definition remains elusive, AI generally refers to the ability of computer systems to perform tasks that typically require human intelligence, such as learning, reasoning, problem-solving, and perception. AI encompasses a broad range of sub-fields, including machine learning, deep learning, natural language processing, and computer vision. Advancements in these areas have led to significant progress in various domains, with AI now playing a crucial role in applications ranging from medical diagnosis (Das et al. 2018) to financial trading (Ferreira, Gandomi, and Cardoso 2021).

2.1.2 Computer Vision

Computer vision, a key subfield of AI, focuses on enabling computers to "see" and interpret visual information from the world, primarily in the form of images and videos. By using sophisticated algorithms and mathematical models, computer vision systems can extract meaningful information from this visual data, allowing them to perform tasks such as object recognition, image classification, and scene understanding (Chai et al. 2021; Szeliski 2022).

Computer vision applications can be broadly categorized into two types:

- `Real-time processing:` This involves analyzing video footage as it is being captured, enabling immediate responses or actions based on the visual input. This is crucial for applications like autonomous driving (L. Liu et al. 2020; J. Mao et al. 2023) and robotics (X. Chen et al. 2020; Zhou, L. Zhang, and Konz 2022), where real-time decision-making is essential.

- `Post processing:` This involves analyzing video footage after it has been recorded, allowing for more in-depth analysis and extraction of insights. This is common in sports analytics, where detailed performance data can be extracted from recorded matches (J. Wang and B. Liu 2023) or fraud detection (X. Mao et al. 2022; Sahni et al. 2020).

This thesis leverages both real-time and post-game processing approaches to achieve its objectives. Real-time processing will be used to capture and analyze performance data as the match unfolds, enabling immediate feedback and potential applications such as score tracking and enhanced broadcasting. Post-game processing will allow for more in-depth analysis and validation of the data, ensuring accuracy and identifying areas for improvement. This dual approach combines the immediacy of real-time feedback with the rigor of detailed post-match analysis, contributing to a comprehensive understanding of player and team performance in amateur football.

2.1.3 You Only Look Once

The You Only Look Once (YOLO) family of models has become the backbone of modern object detection in sports analytics. Its core innovation lies in reframing detection as a single-stage regression task, predicting bounding boxes and class probabilities in one forward pass. Unlike two-stage detectors such as Faster Region-based Convolutional Neural Network (R-CNN), which separate region proposals from classification, YOLO achieves real-time inference while maintaining competitive accuracy. This balance between speed and precision explains its widespread adoption in football analytics pipelines.

Since its introduction in 2015, successive versions of YOLO have refined accuracy, robustness, and computational efficiency. Early generations introduced anchor boxes and multi-scale outputs (YOLOv2–v3), while YOLOv4 and YOLOv5 popularized advanced training strategies and made the models easier to deploy across frameworks. Later iterations expanded scope: YOLOv6 was adopted in robotics, YOLOv7 added pose estimation, and YOLOv8 integrated multiple tasks (detection, segmentation, classification) into a single framework. More recent releases have pushed both efficiency and methodological innovation: YOLOv9 introduced new gradient aggregation strategies (PGI, GELAN), YOLOv10 proposed an end-to-end detection head without Non-Maximum Suppression, and YOLOv11 extended multi-task capabilities across segmentation, pose estimation, and tracking (ultralitics 2024). The most recent YOLOv12 represents a significant architectural shift, integrating attention-centric mechanisms (ultralitics 2025) (e.g., Area Attention and Residual Efficient Layer Aggregation) while preserving real-time speed, thus achieving state-of-the-art performance with fewer parameters(2.1).

Performance

Detection (COCO)

Model	size (pixels)	mAP^val 50-95	Speed CPU ONNX (ms)	Speed T4 TensorRT (ms)	params (M)	FLOPs (B)	Comparison (mAP/Speed)
YOLO12n	640	40.6	-	1.64	2.6	6.5	+2.1%/-9% (vs. YOLOv10n)
YOLO12s	640	48.0	-	2.61	9.3	21.4	+0.1%/+42% (vs. RT-DETRv2)
YOLO12m	640	52.5	-	4.86	20.2	67.5	+1.0%/-3% (vs. YOLO11m)
YOLO12l	640	53.7	-	6.77	26.4	88.9	+0.4%/-8% (vs. YOLO11l)
YOLO12x	640	55.2	-	11.79	59.1	199.0	+0.6%/-4% (vs. YOLO11x)

Figure 2.1: YOLO12 Performance

In addition to version updates, YOLO models are distributed in different sizes, commonly labeled nano (n), small (s), medium (m), large (l), and extra-large (x). These variants correspond to different network depths and widths, offering a trade-off between processing speed and accuracy (Figure 2.2). Smaller models are optimized for edge devices and consumer-grade hardware, often achieving near real-time speeds at the expense of precision, while larger models achieve higher accuracy but require server-level computation.

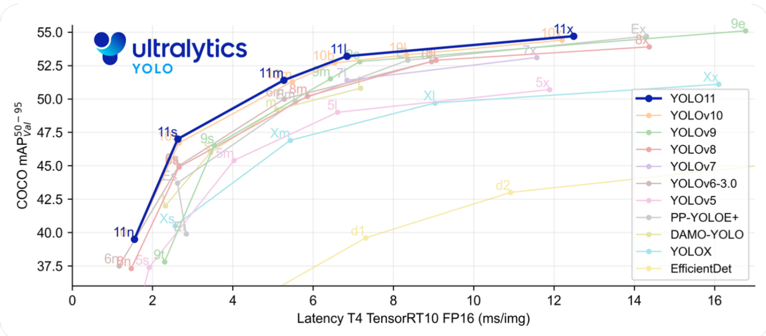

Figure 2.2: YOLO Different Sizes Performance

This scalability distinguishes YOLO from other detection frameworks. It allows researchers and practitioners to select a model that matches the available hardware and application context. In professional settings, large variants may be used for offline tactical analysis with maximum precision, whereas in amateur football, smaller or medium variants are preferable, as they provide a balance between accuracy and real-time performance on modest devices.

Overall, the adaptability of YOLO, across versions, sizes, and deployment platforms, has made it the dominant detection architecture in football analytics research. Its continued evolution, now incorporating attention-based designs, indicates that YOLO will remain central to real-time sports analytics for the foreseeable future.

2.1.4 Key Computer Vision Techniques

In computer vision applications like the one proposed in this thesis, it is common to break down the process into distinct steps to manage complexity and ensure accuracy. A typical approach involves the following five stages:

1. `Court Detection:` The first step is to accurately identify and delineate the boundaries of the playing court within the video frames. This provides a spatial context for subsequent analysis and helps to eliminate irrelevant information from the background. This often involves techniques like edge detection, line detection, and geometric modeling to identify the characteristic lines and shapes of the court (Anand Sharma et al. 2017; J. Chen and Little 2019).

2. `Camera Pose Estimation:` Estimating the position and orientation of the camera relative to the court is crucial for accurately mapping player movements and ball trajectories into a real-world coordinate system. This involves techniques like homography estimation and perspective transformation to account for the camera's perspective and potential distortions in the images (Evangelidis and Psarakis 2008; Homayounfar, Fidler, and Urtasun 2017).

3. `Player and Ball Detection:` This step involves identifying and locating players and the ball within each frame of the video. This typically relies on object detection algorithms like YOLO or Faster R-CNN, which are trained to recognize and classify these objects with high accuracy (Redmon 2018; Ren et al. 2016).

4. `Player and Ball Tracking:`Once players and the ball are detected, their movements must be tracked across consecutive frames. This process enables the system to analyze trajectories, speeds, and interactions between players and the ball. This critical task is handled by a class of algorithms known as multi-object trackers.

   A widely used approach in this domain is DeepSORT(Adžemović et al. 2025; Redmon 2018; Z. Wang, Shan, and Feng 2024), which builds upon the Simple Online and Realtime Tracking (SORT) framework. While SORT leverages a Kalman filter for predicting object locations and the Hungarian algorithm for associating detections with existing tracks based on their proximity, it is prone to identity switches during occlusions. DeepSORT addresses this limitation by integrating a deep association metric into its pipeline. This metric uses a pre-trained convolutional neural network to extract visual features from each object, allowing the tracker to re-identify objects after they have been occluded for an extended period, significantly reducing the number of identity switches Wojke, Bewley, and Paulus 2017.

   More recently, the ByteTrack algorithm has emerged as a robust alternative. ByteTrack's key innovation lies in its two-step association process that retains and analyzes low-score detection boxes, rather than discarding them as most traditional trackers do. It first associates high-score detections with existing tracks, then uses the remaining low-score detections to recover occluded objects by matching them with previously unassigned tracks. This method effectively filters out false positives while ensuring that objects with low confidence scores, often the result of occlusion or blur, are not lost, leading to more complete and continuous trajectories (Y. Zhang et al. 2022).

5. `Analytics, Extraction, and Visualization:` The final step involves extracting meaningful performance data from the tracked trajectories and events. This may include metrics like distance covered, top speed, passes completed, and heatmaps.

The extracted data is then visualized in a user-friendly format to provide insights to players and coaches.

This structured approach ensures a systematic and comprehensive analysis of the video footage, enabling the extraction of accurate and relevant performance data.

2.2 Systematic Review

A Systematic Literature Review (SLR) is a way of bringing together all the available research related to a particular area, topic of study, research question, phenomenon or interest (Budgen and Brereton 2006; Kitchenham 2004). This section presents a systematic review, following () guidelines (Sarkis-Onofre et al. 2021), of the existing literature concerning the application of computer vision in sports, with a particular focus on football. This review aims to identify current trends, challenges, and opportunities in this domain, providing a foundation for the research presented in this thesis.

The PRISMA methodology provides a structured approach to conducting systematic reviews, minimizing bias and ensuring reproducibility. It consists of four main phases: identification, screening, eligibility, and inclusion. The identification phase involves searching electronic databases using predefined search terms and strategies. During the screening phase, researchers review titles and abstracts to remove clearly irrelevant studies. The eligibility phase involves a thorough examination of full-text articles against predetermined inclusion and exclusion criteria. Finally, the inclusion phase incorporates the selected studies into the systematic review. Each step is documented in detail, including the number of studies identified, screened, assessed for eligibility, and finally included in the review, typically presented in a PRISMA flow diagram.

2.2.1 Research Questions

For this SLR, the most recent literature must be explored, filtered, and evaluated based on the four questions outlined in Table 2.1. This review aims to assess the technological advancements in **computer vision applications for football**, including their accuracy, use cases, and integration into both professional and amateur settings.

Additionally, this study examines whether current solutions extend beyond data collection and analysis to incorporate **real-time feedback and gamification features**. Given the growing interest in AI-driven analytics for amateur football, understanding how these technologies impact **engagement, accessibility, and performance tracking** is crucial. The review will assess whether existing systems leverage elements such as **performance challenges, leaderboards, and automated match highlights** to enhance user experience and participation. By analyzing these aspects, this research seeks to identify gaps in current implementations and explore how future developments could make football analytics more interactive and motivating for amateur players.

2.2.2 Search Query

This subsection details the strategy used to identify relevant studies for inclusion in the systematic review. The search process focused on the domains of computer vision, sports analytics, and football (soccer).

Table 2.1: Research Questions

	Research Question
RQ1	What are the current computer vision techniques used in football?
RQ2	How is computer vision being used to analyze individual and team performance?
RQ3	What are the limitations and challenges of implementing computer vision in football?
RQ4	What are the future trends in computer vision applications for football?

Definition of search sources

The first step in SLR research is to identify and define the main research sources to be considered. For this study, five electronic databases were searched, see Table 2.2.

Table 2.2: Eletronic databases used

Identifier	Database	URL
ID1	IEEE Xplore	`ieeexplore.ieee.org`
ID2	Web of Science	`webofscience.com`
ID3	ArXiv	`arxiv.org`
ID4	Science Direct	`sciencedirect.com`
ID5	Springer	`link.springer.com`

Definition of search terms

The second phase involves defining specific search terms and strategies to retrieve relevant articles from the chosen databases. The search terms were carefully selected to reflect the research questions and encompass various aspects of computer vision applications in football. The following search strings were used see Table 2.3.

Table 2.3: Search string

Scope	String
Computer Vision	"Computer Vision" OR "Image Processing" OR "Object detection" OR "Object tracking" OR "Pose estimation"
Sports analytics	"Performance Analysis" OR "Player Tracking" OR "Event Detection" OR "Game Analysis" OR "Tactical Analysis" OR "Ball Tracking" OR "Player Performance" OR "Game Strategy" OR "Metrics" OR "Match Analysis"
Football(soccer)	"Football" OR "Soccer"

PRISMA flow diagram

The results of the search strategy are presented in a PRISMA flow diagram as seen in Figure 2.3. This diagram visually illustrate the different stages of the literature search and selection process. Due to variations in database functionalities and limitations, the specific

Table 2.4: Inclusion criteria

Identifier	Inclusion criteria
IC1	Focused on affordable computer vision applications in sport
IC2	Explored AI applications in amateur football
IC3	Identified benefits of implementing computer vision tools in sport

Table 2.5: Exclusion criteria

Identifier	Exclusion criteria
EC1	Were not written in English
EC2	Were published before 2020
EC3	Were not academic articles, book chapters or conference proceedings
EC4	Lacked direct connection to Computer Vision applications in sport
EC5	Were duplicate entries
EC6	Use body sensors

search procedures and results differed across platforms, requiring some adjustments to ensure a comprehensive yet manageable retrieval of relevant articles.

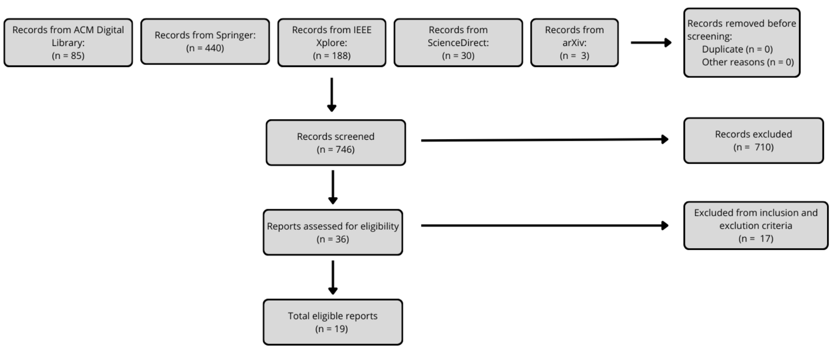

Figure 2.3: PRISMA flow diagram

For instance, IEEE Xplore yielded 188 (25.2%) results with the initial search query, while arXiv produced only 3 (0.4%). Attempts to broaden the search in arXiv resulted in a maximum of 52 results, many of which were not relevant, leading to the retention of the initial query with 3 results. ACM Digital Library returned 85 results (11.4%). ScienceDirect, with its limitation of 8 Boolean operators, necessitated a modified search query that included only the most relevant terms from each string, resulting in 30 results (4%). SpringerLink, on the other hand, initially yielded over 5000 results, exceeding the platform's export limit of 1000 items. To address this, filters such as "Computer Science", "Artificial Intelligence" and a publication date range of "2022-2025" were applied, narrowing down the results to

440 (59%), making a total of 746 results. From this total, no problems relating to format or duplications were found, meaning that all can proceed to the screening phase.

This next process consists of reading the titles and abstracts of the documents obtained to eliminate possible studies outside the area of study, because the first screening return more than 400 documents, what would make the eligibility phase almost impossible, a second screening was conducted with a more restricted approach that result on only 36 documents accepted, this second screening excluded studies that only focused on camera calibration, pitch detection or body and ball trajectory analysis for pass prediction. The third phase involves checking eligibility by meeting the inclusion criteria and then removing by the exclusion criteria. These criteria can be seen in Tables 2.4 and 2.5, respectively. From the 36 studies that passed the screening phase, 17 were abandoned after eligibility checks, the majority because the documents made a hybrid system that combined computer vision with body sensors and the rest because they use drone footage which makes the solution not affordable for amateur football. Leaving 19 documents for the final phase, these selected studies have been included in this thesis for further analysis.

2.2.3 Critical Review of selected papers

Research into football analytics through computer vision has produced a diverse set of approaches, ranging from classical tracking frameworks to deep learning–based detection and recognition systems. A recurrent observation across this literature is the trade-off between methodological sophistication and practical feasibility, especially when applied to real-time scenarios or amateur-level contexts.

Court and Camera Calibration Approaches

Across football analytics research, one of the earliest and most consistent priorities has been the reliable calibration of the pitch and correction of camera distortions. Without a stable geometric model, even the most sophisticated tracking algorithms cannot produce meaningful tactical data. Most contributions fall into two main families: learning-driven line/grass segmentation followed by homography, and direct detection of field markers with perspective transforms.

Pipelines based on generative models illustrate the first strategy, where Generative Adversarial Networks (GAN) are applied to extract line structures and filter grass regions before estimating homographies (Mavrogiannis and Maglogiannis 2022). Such systems achieve a consistent top-down projection of the field but are bottlenecked by line detection, which prevents real-time processing. The assumption of uniformly green pitches further restricts applicability in amateur contexts, where surfaces may vary in color(2.4). By contrast, more recent approaches leverage object detectors directly. YOLO-based frameworks integrate homography by aligning detected field markers or intersections, producing accurate perspective correction with far lower computational overhead (Athanesious and Kiruthika 2024; Karungaru, Tanioka, and Matsuura 2022). These designs reduce reliance on fragile color cues, making them conceptually closer to deployable solutions.

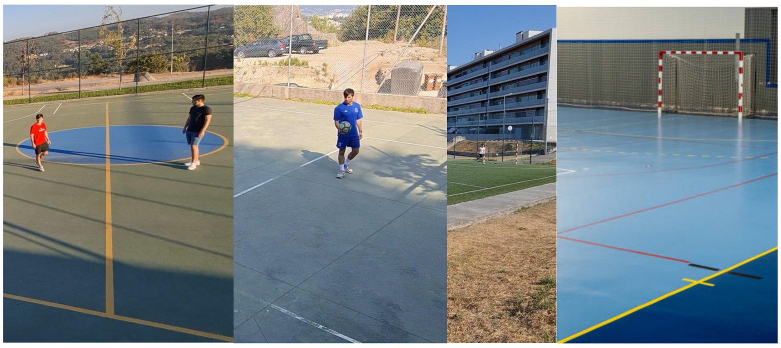

Figure 2.4: Different Amateur Fields

A parallel stream of research seeks to recover elevation information. While most video-only methods remain constrained to two-dimensional corrections, some authors propose augmenting monocular vision with additional sensors such as Light Detection And Ranging (LiDAR) (W. Huang et al. 2024). This resolves the z-axis ambiguity and significantly improves positional fidelity, but it introduces costs incompatible with grassroots football. Taken together, the literature suggests that while homography and keypoint-based calibration are sufficiently mature for two-dimensional applications, extending to full 3D remains an open challenge, with expensive sensing or combination of multiple sensors remain the current option.

Player Detection and Tracking Strategies

If the pitch provides the coordinate system, player detection and tracking define the moving elements within it. Approaches here vary between computationally heavy segmentation networks, modified two-stage detectors, and lightweight trackers aimed at efficiency.

Segmentation methods, such as those adapting U-Net to football (Biliškov et al. 2021), excel under blur and occlusion but are computationally demanding. Attention-augmented detectors like Cascade R-CNN with hybrid modules (Tang et al. 2023) attempt to balance accuracy and robustness, handling dense crowds more effectively than generic detectors. Yet both categories remain difficult to scale to real-time use without substantial hardware. Conversely, lightweight tracking approaches based on Kalman filtering remain attractive for their simplicity (M. Zhang 2023), but they struggle under nonlinear movements characteristic of sport.

This gap has motivated the widespread adoption of tracking-by-detection frameworks. ByteTrack and DeepSORT variants have proven particularly influential, with broadcast-oriented systems combining face recognition for close-ups and feature-based multi-object tracking for wide shots (Qiu et al. 2023). Extensions such as observation-centric distance recovery refine SORT/OC-SORT methods to handle non-linear player trajectories, and introduce post-hoc re-identification to reconnect fragmented tracks (H.-W. Huang et al. 2023). While these refinements yield measurable improvements, they are mostly performed offline, limiting their use in real-time analysis.

A different strand emphasizes portability and low-cost setups. Systems using stitched dual cameras (Kalafatić, Hrkać, and Brkić 2022) demonstrate that structured recording configurations can alleviate occlusion and perspective problems. Such solutions are particularly relevant for smaller clubs, though they still require planned infrastructure. Overall, the detection–tracking literature demonstrates a tension between robustness and efficiency: highly accurate Convolutional Neural Network (CNN)-based methods remain impractical for live use, while efficient trackers often lack resilience in chaotic football scenarios.

Ball Tracking and Possession Estimation

Tracking the ball reliably remains one of the most difficult challenges due to its small size, rapid velocity, and frequent occlusion. A common strategy is to combine detection with motion cues, where optical flow supplements YOLO predictions to ensure continuity (Modi et al. 2024). This approach mitigates frame-to-frame disappearance and achieves strong consistency across datasets. However, most studies report detector accuracy without presenting complete pipeline runtimes, leaving open the feasibility of true real-time performance.

Possession estimation builds on ball tracking but diverges in philosophy. Some frameworks adopt lightweight geometric heuristics, calculating possession from spatial proximity of players to the ball within a calibrated pitch (Athanesious and Kiruthika 2024). This design supports near real-time applications but lacks semantic depth. Others pursue principled definitions, reframing possession in terms of successful passes per Fédération Internationale de Football Association (FIFA) regulations (Sarkar, Mukherjee, and Chakrabarti 2022). Reinforcement learning models offer higher accuracy under this definition but are computationally expensive and more suitable for offline analysis. Meanwhile, hybrid systems that chain detection, team classification, and ball, controller assignment demonstrate how tactical statistics, possession time, successful passes, steals—can be automatically extracted. Yet, even here, the emphasis remains on post-match tactical review rather than in-game insights.

In short, the literature reveals a continuum between lightweight but shallow heuristics and accurate but impractical learning-based solutions. Bridging this divide remains crucial for practical deployment.

Identity Recognition and Player-Level Analytics

Beyond detecting and tracking players, many systems attempt to assign persistent identities or extract higher-level statistics. Identity recognition approaches typically combine multiple cues like faces, jersey numbers, or re-identification embeddings. While combining face and number recognition increases robustness (Diop et al. 2022), both cues rely on high-resolution close-ups unavailable in amateur recordings.

Re-identification frameworks increasingly explore efficiency in training. CNN–Vision Transformer (ViT) hybrids trained with cost-effective schemes have achieved strong results on SoccerNet with significantly reduced GPU demands (Kao, Chou, and C.-H. Lin 2024). These advances lower barriers to model development, but deployment challenges persist, as runtime efficiency and real-time performance are rarely addressed.

Comprehensive analytics platforms such as GreenSea (Sheng et al. 2020) move beyond detection, providing visualization and scoring of player abilities via broad learning systems. This work demonstrates the potential of visual analytics to influence training and tactical decisions, yet it is validated mainly on professional and elite youth settings.

Together, these studies indicate that while individual identification and advanced analytics are technically feasible, their dependence on broadcast-level footage or expensive computation constrains their relevance for grassroots football.

2.2.4 Existing Challenges in AI-Based Football Analysis

The application of AI in football analytics has evolved significantly, leveraging deep learning techniques for player tracking, ball detection, and tactical analysis. However, numerous challenges persist, particularly in achieving accurate real-time tracking, robust player identification, and efficient data processing. This section explores the key challenges identified in recent AI-based football analysis research.

Challenges in Player and Ball Detection

- `Occlusion and Identity Switching:` One of the most significant challenges in football analytics is the issue of occlusion, where multiple players overlap, making it difficult for models to track and re-identify individuals correctly. This often leads to identity switching, especially in scenarios such as corner kicks and dense midfield play (Tang et al. 2023). Despite advancements in hybrid attention mechanisms and multi-frame tracking, identity persistence remains an open problem (H.-W. Huang et al. 2023). LiDAR-camera fusion approaches (W. Huang et al. 2024) have been introduced to improve depth perception and reduce tracking errors, yet occlusion-related misidentifications persist.
- `Motion Blur and Small Object Detection:` Football is a fast-paced sport, making it challenging to detect and track small objects like the ball, especially when motion blur is present. High-speed movements cause tracking failures, leading to frequent object misidentifications (Kao, Chou, and C.-H. Lin 2024). Some studies have proposed using optical flow techniques and YOLO-based object detection (Modi et al. 2024), but accuracy deteriorates when the ball is occluded or moves unpredictably. Additionally, player detection from a single camera setup introduces further inaccuracies (Kalafatić, Hrkać, and Brkić 2022).

Issues in Multi-Object Tracking

- `Identity Reassignment and False Positives:` MOT systems rely on detection models to consistently follow players throughout a match. However, many models fail when players move rapidly across frames or when camera angles change (Cioppa et al. 2022). These tracking errors result in identity reassignment issues and false positives, reducing the effectiveness of automated analysis (Sheng et al. 2020). The fusion of LiDAR and camera data (H.-W. Huang et al. 2023) has improved depth awareness, reducing identity switching but requiring complex calibration.
- `Lack of Robust Datasets for Training:` One of the main barriers to improving AI-based football tracking is the lack of diverse and robust datasets. While the SoccerNet-Tracking dataset (Kao, Chou, and C.-H. Lin 2024) has introduced new benchmarks, many datasets still lack sufficient variations in lighting, camera angles, and occlusion scenarios. This makes it difficult to train models that generalize well across different match environments. The introduction of automated tactical analysis (Theagarajan and Bhanu 2020) has partially alleviated this problem by generating synthetic data through GANs, but further advancements are needed.

Computational Complexity and Real-Time Constraints

- `High Processing Power Requirements:` Many AI-based tracking systems depend on computationally expensive deep learning models, such as YOLO, TrackFormer, and Vision Transformers (Athanesious and Kiruthika 2024). The need for extensive GPU processing makes real-time implementation challenging for lower-end hardware or edge computing applications (Mavrogiannis and Maglogiannis 2022). LiDAR-based tracking solutions (W. Huang et al. 2024) introduce additional computational overhead due to 3D point cloud processing.
- `Latency in Decision-Making:` For AI systems to be useful in live football matches, they must process data in real-time with minimal latency. However, existing methods suffer from frame-by-frame processing delays, making them impractical for applications like automated referee assistance (Athanesious and Kiruthika 2024). Reducing inference time while maintaining accuracy is a major challenge that needs to be addressed. Systems integrating multiple cameras and fusion techniques (Kalafatić, Hrkać, and Brkić 2022) attempt to optimize tracking, yet real-time constraints remain an issue.

Challenges in Tactical and Statistical Analysis

- `Inaccuracy in Tactical Assessments:` Automated tactical analysis remains an unsolved problem due to variations in playing styles, camera perspectives, and dynamic player movements. Many AI systems struggle to accurately detect formations and predict strategies (Benakesh and Rajeev 2024). Spatio-temporal analysis techniques have improved results (Athanesious and Kiruthika 2024), but a deeper contextual understanding is required to match human-level strategic assessments. Tactical evaluation methodologies (Theagarajan and Bhanu 2020) have been developed, yet further validation on real-world match data is necessary.
- `Errors in Possession and Pass Prediction:` Pass detection and ball possession statistics are critical for tactical insights but remain error-prone. Current models, including reinforcement learning-based detection systems (Sarkar, Mukherjee, and Chakrabarti 2022), still suffer from misclassifications due to incorrect team assignments and occlusions. Improved segmentation and tracking methodologies (Theagarajan and Bhanu 2020) are needed to enhance the reliability of these analytics.

While AI-driven football analytics has made notable progress, several unresolved challenges remain, particularly in real-time tracking, identity switching, and dataset limitations. Future research must focus on refining tracking robustness, minimizing identity-switch errors, and developing more diverse datasets. Addressing these issues will be crucial in making AI-based football analytics a viable solution for coaches, analysts, and broadcasters.

2.3 Research Questions

The adoption of computer vision in football has advanced significantly in recent years, with applications spanning player tracking, tactical analysis, and performance evaluation. This response synthesizes recent literature to address the state of implementation of computer vision in football, focusing on the research questions guiding this systematic review.

2.3.1 RQ1: What are the current computer vision techniques used in football?

Modern computer vision applications in football primarily involve player tracking, action recognition, event detection, and performance analysis. Various techniques are employed, including deep learning-based object detection, MOT, and perspective transformation for accurate localization.

- `Object Detection and Tracking:` Player tracking is fundamental for match analysis, with multiple approaches relying on CNNs and deep learning-based detection models. For instance, the SoccerNet-Tracking dataset provides a benchmark for MOT in football, allowing automated tracking of players, referees, and the ball (Cioppa et al. 2022).
- `Real-time Player Re-Identification:` advanced Re-IDentification (ReID) techniques, such as ViT-based models, are used to track individual players across multiple camera views, improving player recognition despite occlusions and lighting changes (Kao, Chou, and C.-H. Lin 2024).
- `LiDAR-Camera Fusion:` Hybrid tracking approaches integrate LiDAR and computer vision to enhance player localization accuracy without requiring wearable sensors, reducing intrusiveness for players (W. Huang et al. 2024).
- `Perspective Transformation for Location Estimation:` Perspective transformation techniques allow real-world positioning of players based on video footage, eliminating the need for GPS-based tracking (Karungaru, Tanioka, and Matsuura 2022).
- `Face Recognition for Player Identification:` Face recognition combined with multi-object tracking enables real-time event analysis and player identification from match footage, increasing automation in game analytics (Qiu et al. 2023).

These advancements illustrate that computer vision has matured to enable precise tracking and recognition, with growing applications in tactical analysis and game event recognition.

2.3.2 RQ2: How is computer vision being used to analyze individual and team performance?

Computer vision plays a crucial role in both individual player assessment and team strategy evaluation, leveraging automated statistics and tactical insights.

- `Automated Tactical Performance Statistics:` Deep learning-based tracking systems automatically generate key performance metrics, such as ball possession duration, successful passes, and steals. These statistics reduce manual video annotation and improve scouting efficiency (Theagarajan and Bhanu 2020).
- `Data-Driven Tactical Analysis:` Systems like GreenSea use broad learning models to assess team formations and player strategies. These systems facilitate tactical evaluations by visualizing key moments and tracking player movements (Sheng et al. 2020).
- `Amateur Football Analytics:` Computer vision provides low-cost analytics for amateur football, extracting game insights from simple video footage without expensive sensors. This democratizes performance tracking for non-professional players and training academies (Mavrogiannis and Maglogiannis 2022).

- `Full-Field Player Tracking and Correction`: Attention-based tracking mechanisms are used to improve accuracy in player localization across full-field videos. These methods allow for better team analysis and strategic decision-making (yang2024soccer).

By integrating these approaches, teams can improve decision-making, enhance training, and optimize game strategies using AI-driven insights.

2.3.3 RQ3: What are the limitations and challenges of implementing computer vision in football?

Despite its progress, several challenges hinder the widespread adoption of computer vision in football:

- `Occlusion and Identity Switching`: Tracking accuracy is affected by frequent occlusions and player identity switches, particularly in crowded areas such as penalty boxes. Multi-object tracking methods still struggle with maintaining consistent player IDs (Kalafatić, Hrkać, and Brkić 2022).
- `Real-time Processing Constraints`: Many tracking and recognition models require significant computational resources, making real-time analysis difficult. Systems using deep learning must balance accuracy with processing speed (Qiu et al. 2023).
- `Limited Adoption in Amateur Football`: While elite teams have access to high-end tracking technology, many amateur clubs lack the financial resources for sophisticated analytics systems. Research on low-cost solutions, such as vision-based tracking without dedicated sensors, is still emerging (Mavrogiannis and Maglogiannis 2022).
- `Inconsistent Video Quality`: Variations in camera angles, lighting, and resolution affect the robustness of detection models, requiring additional dataset augmentation and training methods for generalizability (Cioppa et al. 2022).
- `Need for Ground Truth Data`: Many computer vision applications in football rely on annotated training data, but large-scale labeled datasets remain scarce. Efforts like the SoccerNet dataset are addressing this gap, but further research is needed Cioppa et al. 2022.

These limitations indicate that while computer vision is a powerful tool for football analytics, further improvements are necessary for real-time deployment and accessibility in lower-tier competitions.

2.3.4 RQ4: What are the future trends in computer vision applications for football?

The future of computer vision in football is likely to be shaped by advancements in AI, real-time analytics, and gamification elements that enhance player engagement.

- `AI-Assisted Coaching and Training`: AI-powered systems will provide real-time recommendations for coaches, integrating live match data with historical performance trends to suggest tactical adjustments.
- `Gamification and Player Motivation`: Real-time feedback and gamification elements, such as performance-based challenges and rankings, will enhance player engagement and motivation, particularly in amateur and youth football (Mavrogiannis and Maglogiannis 2022).

- `Integration of Augmented Reality (AR):` AR overlays will allow coaches and analysts to visualize player movement and tactical insights directly on the field using AR glasses or tablet interfaces.
- `Self-Training and Automated Highlights:` AI-driven self-training tools will enable players to receive instant feedback on their performance through automatically generated match highlights and tactical breakdowns.
- `Hybrid Sensor and Vision-Based Systems:` The combination of wearables with computer vision will improve tracking accuracy, especially in scenarios where occlusion remains a challenge.

The implementation of computer vision in football has progressed significantly, with applications ranging from player tracking and performance analysis to tactical evaluation. While challenges such as occlusion, processing constraints, and accessibility remain, emerging trends in AI-driven analytics, gamification, and real-time feedback promise to enhance the sport further. As research continues to refine these technologies, their adoption in both professional and amateur football will expand, transforming the way the game is played and analyzed.

2.4 Work Relevance

The review of existing contributions in football analytics highlights a clear imbalance in the field: while technical sophistication has advanced considerably, the majority of solutions remain tied to the conditions of professional football. Nearly all state-of-the-art methods are trained and validated on broadcast-quality footage, benefiting from high-resolution cameras, controlled lighting, and standardized green pitches. As a result, current pipelines systematically overlook the practical realities of amateur and recreational football, where most players worldwide participate. This project addresses precisely that underexplored domain.

A first limitation in the literature is the overreliance on professional TV broadcasts. Systems designed for such footage are not transferable to grassroots contexts, where recordings are often made with low-cost, low-resolution cameras, most times from mobile devices. The present work deliberately avoids the use of broadcast material and instead focuses on amateur matches captured in such constrained conditions. To further mitigate the absence of a dedicated camera operator, the solution incorporates a multi-angle recording strategy and explores field-stitching methods, ensuring continuous coverage even when the ball moves outside a single frame.

A second persistent issue concerns pitch calibration. Many pipelines depend on color masking or line detection, assuming that the pitch is uniformly green and that field markings are clearly visible. In practice, amateur football is frequently played on multi-purpose surfaces that include overlapping markings for basketball, tennis, or futsal, or on worn-out fields where line visibility is poor. This project proposes to overcome these limitations by using keypoint detection trained on a dataset of amateur fields, removing dependence on simplistic color thresholds and ensuring robustness across heterogeneous environments.

Third, the literature shows a recurring struggle to reconcile accuracy with speed. While detection modules such as U-Net and cascade R-CNN variants achieve high accuracy, they fail to operate in real time; conversely, lightweight filters such as Kalman are fast but unreliable under occlusion and nonlinear motion. In this project, YOLO model with ByteTrack is adopted as the main tracking mechanism, given its demonstrated balance between speed and reliability. To specifically address its weakness under severe occlusions, an additional refinement is developed, ensuring stable player trajectories even in crowded scenes.

Finally, the greatest gap lies in the absence of cost-effective, real-time pipelines explicitly designed for grassroots deployment. Current solutions are either computationally intensive or hardware-dependent, rendering them inaccessible to most of the millions of amateur players worldwide. The contribution of this dissertation is to design and validate a low-cost framework capable of real-time analytics under non-ideal conditions, directly supporting the environments where football is most widely practiced but least technologically served.

By tackling these gaps, broadcast dependency, field calibration assumptions, computational inefficiency, and cost barriers, this project seeks to extend football analytics from the elite minority to the recreational majority, thereby filling an evident and impactful void in the state of the art.

Chapter 3

System Implementation

This chapter presents the technical implementation of the proposed system. The description covers the overall system architecture, the data acquisition and preprocessing strategies, and the methods used for object detection, player tracking, team classification, field calibration, and statistics extraction. The implementation was designed to be scalable, cost-effective, and suitable for real-world amateur football context, where equipment and infrastructure are limited.

3.1 System Architecture

The implemented system is designed as a modular pipeline, with each component performing a specialized function to process video input and generate sports analytics data. The architecture can be conceptualized in three primary stages: **data acquisition, computer vision processing**, and **data analysis**. This modular design ensures a clear separation of concerns, which enhances the system's maintainability and scalability.

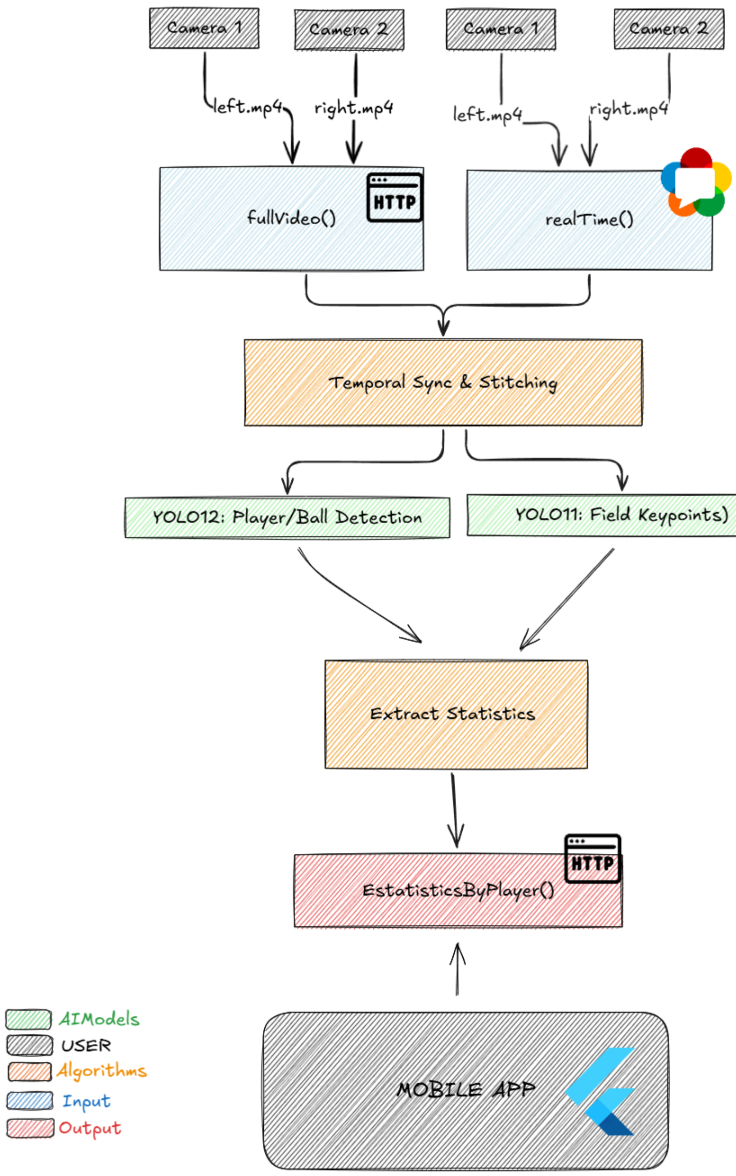

Figure 3.1: System Architecture.

Figure 3.1 illustrates the architecture of the system, which is a fundamental aspect of its design. The process begins with the **data acquisition** module, which ingests raw video footage of a football match. This footage can be from a previously recorded file or a live stream. The system is designed to handle one or two video sources, enabling a user to use two cameras to cover a wider field of view. These clips are then synchronized and stitched into a single, comprehensive panoramic video. The resulting panoramic clip is subsequently processed by two models to extract key features related to the field and players. This data is then passed to subsequent modules for statistical extraction. The final information is made available to the user via a Flutter application, where it can be accessed either during or after the match.

This structure was inspired by the pipeline proposed in Mavrogiannis and Maglogiannis 2022, a clever design illustrated in Figure 3.2, but extended to support real-time features. Our contribution is a practical solution that can already be tested in real-world scenarios, serving as an initial step to demonstrate the system's potential. While further adjustments will be required to make it fully production-ready, the modular design facilitates straightforward improvements and extensions. This aligns with the DSR principle of continuously evaluating and refining systems through iterative development. As a result, this work lays the foundation for a scalable and robust framework for football analytics, with potential applications in other sports as well.

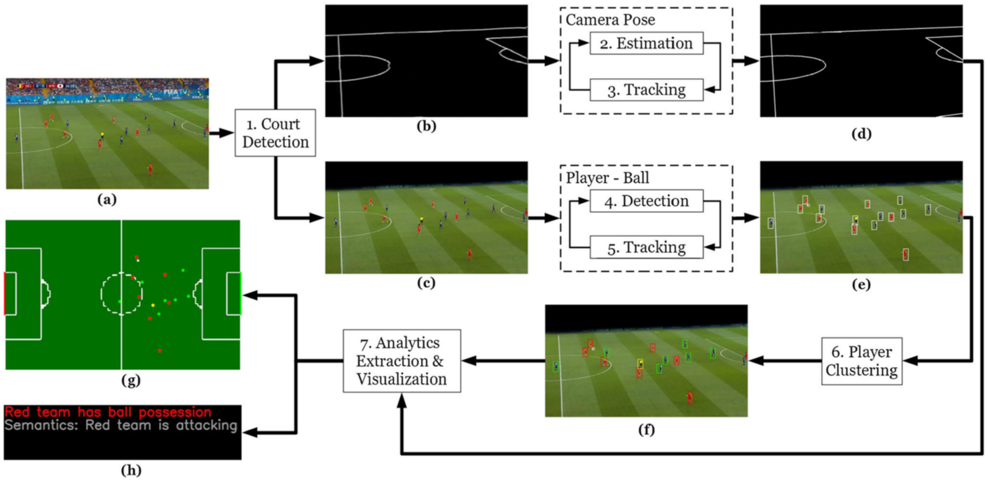

Figure 3.2: Mavrogiannis & Maglogiannis Pipeline

3.2 Data Acquisition

The system's data acquisition strategy was developed specifically for amateur football contexts, where professional equipment is often unavailable. Consequently, two main scenarios were considered to ensure accessibility and cost-effectiveness: fixed-field cameras and smartphone-based recording. In facilities where cameras can be permanently installed, the system leverages higher-quality devices with wired connections. This setup enables stable video transmission, synchronized frame capture, and more reliable internet connectivity. The costs can also be distributed across multiple games, reducing the financial impact on individual players.

For more flexible, open-field settings without existing infrastructure, a cost-effective solution was designed using smartphone mounts. As shown in **Figure** 3.3, this setup consists of two smartphones placed facing opposite sides of the field, positioned close to each other, with a cost approximately 20 euros. This configuration provides full coverage of the field. The tripod feet are malleable, allowing them to be attached to a post or a fence, providing versatility in placement. While we also tested standard tripods, they were found to be more expensive despite being easier to set up.

Figure 3.3: Smartphone holder solution

The system accommodates both live streaming and post-match video upload, each with distinct processing mechanisms. Post-match video upload is the simpler and more reliable version. After a match, the user uploads the recorded clips for both sides of the field via specific endpoints within the application. The system then uses audio features to synchronize the two clips. The stitching process is performed on a frame-by-frame basis, where the system utilizes OpenCV's panorama stitching algorithm to find the best seams between the two frames, ensuring minimal loss of relevant information.

Live streaming functionality is achieved by leveraging the **LiveKit** Python framework and the **WebRTC** protocol, both open-source solutions. While more robust message queueing systems (e.g., Kafka, Apache Flink) could be integrated, this was considered excessive for the proof-of-concept stage. The Flutter application provides a streaming option(Figure 3.4), using WebRTC's User Datagram Protocol (UDP) protocol supported by Flutter libraries on both iOS and Android. This ensures sub-second latency for video transmission. The system's backend, using the LiveKit framework, receives the byte packets and converts them into images. Synchronization is more challenging with UDP due to network instability, so each frame is annotated with a timestamp. The backend uses these timestamps to align frames, which are then passed to the same stitching algorithm. To manage the unreliability of UDP, certain rules were implemented, such as dropping packets with resolutions below 760p and maintaining a 60-frame buffer to handle frames that might arrive out of order.

Figure 3.4: Room solution for multiple camera view

A significant challenge encountered in this process was **parallax distortion**, which is the apparent change in an object's position when viewed from different angles. As shown in **Figure** 3.5, when two cameras are separated by a significant distance, such as two meters, the lines on the field and the position of the player in the middle appear distorted and misaligned. This makes it impossible for the algorithm to find consistent seam points, preventing the creation of a seamless panoramic view.

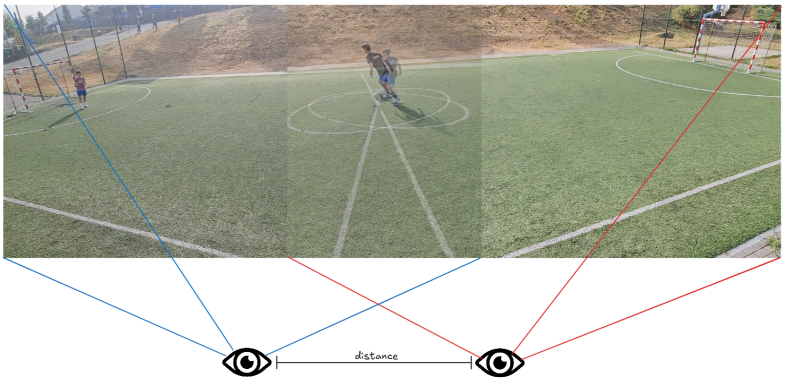

Figure 3.5: Parallax Problem

However, by carefully controlling the distance between the two cameras, positioning them at the same height, and ensuring at least a 35% overlap between both clips, this distortion becomes more manageable. **Figure** 3.6 illustrates a case where the distortion is less notable, and the panorama stitching process is able to produce an almost perfect result.

Figure 3.6: Panorama Stitching

An encouraging confirmation that we are moving in the right direction comes from Veo, a well established company that provides professional teams with video-based statistical analytics. Recently, they launched an amateur-oriented product, Veo GO, which relies on a smartphone setup similar to ours (Fig. 3.7). However, their bundle, consisting of a phone holder and a 2.7 meter tripod, costs around 60 euros, compared to our 20 euro setup. In addition, their solution is restricted to iOS devices, specifically requiring an iPhone 11 or newer. Although this represents a lower entry barrier than their main offerings, with cameras exceeding 1,700 euros, with subscriptions ranging from 500 to 2,000 euros annually, it still does not provide a truly accessible solution for recreational players. Instead, it primarily targets amateur clubs with smaller budgets than professional teams but still with the resources to invest in performance improvement. Nevertheless, Veo's efforts to broaden accessibility reinforce the relevance and value of this type of system, further validating the importance of our approach.

Figure 3.7: Veo GO Phone Holder

3.3 Object Detection

Based on the findings presented in the state-of-the-art chapter, the **YOLO** model was identified as the most promising approach for player and ball detection. Its well-known speed makes it suitable for real-time applications, while its high accuracy ensures reliable object identification.

The initial plan was to leverage the **SoccerNet** dataset, a renowned resource comprising over 550 matches recorded with multiple camera angles. The objective was to evaluate how well models trained on professional broadcast footage could generalize to amateur video recordings. However, despite signing a Non-Disclosure Agreement (NDA), access to the dataset was ultimately denied. As an alternative, we turned to publicly available options such as the Roboflow Universe player detection dataset (Figure 3.8), which contains over 600 annotated frames. Unfortunately, models trained on this dataset performed poorly on our custom video clips, consistently failing to detect players or the ball.

A likely explanation for this issue lies in the differences between datasets: the Roboflow models were trained on high-resolution footage where players appear relatively small due to distant camera placements, whereas our low-resolution recordings capture players much closer to the field, occupying a larger portion of the frame as we can see in Figure 3.8. Attempts to mitigate this mismatch by downscaling our input images proved ineffective, as the resolution loss only further degraded model performance.

Figure 3.8: Professional VS Amateur Player and Ball detection

The project pivoted to using a pre-trained, standard YOLO12 model without custom training. This out-of-the-box model, as seen in **Figure** 3.9, demonstrated impressive results and served as an effective starting point for the system. While its performance could be further optimized with a custom-trained model, the initial results were sufficient to proceed without the extensive manual labor required for labeling a custom dataset.

Figure 3.9: Player and Ball tracking

3.4 Player Tracking and Identity Assignment

Player tracking is one of the most challenging aspects of football video analysis, particularly due to occlusions, rapid movements, and identity re-assignment issues. Many existing solutions rely on high-cost equipment such as high-resolution cameras, LiDAR, jersey number recognition, or wearable sensors. These are impractical for the amateur football context, where accessibility and cost-effectiveness are critical.

To achieve robust tracking under these constraints, the system integrates the ByteTrack algorithm. This state-of-the-art tracker is based on bounding box association and effectively maintains player identities across consecutive frames. A brief comparative test was conducted with the DeepSORT algorithm; however, ByteTrack was selected for its superior speed and accuracy in our low-resolution environment. Like most tracking-by-detection methods, ByteTrack encounters difficulties when players are occluded or temporarily leave the field of view.

To mitigate these challenges, a custom ID assignment policy was implemented. Each new detection receives a unique identifier, and previously used identifiers are not reassigned. This policy prevents potential misidentification when a player reappears after an occlusion. At the conclusion of a match, users are prompted to manually select the identifier(s) corresponding to them. The system then retrospectively infers and matches missing identities by analyzing the overlaps of visible identifiers throughout the game.

For example, if both ID 1 and ID 2 appear in the same frame, and a user selects ID 1, they cannot also select ID 2. If every player is assigned a unique ID, the last remaining identifier is automatically assigned to the last remaining player. This strategy, as illustrated in **Figure** 3.10, reduces statistical errors caused by ID switches and ensures that player-level statistics remain coherent throughout the match, even after temporary losses of a track.

Figure 3.10: Player Id tracking

3.5 Team Classification

After players are assigned unique identifiers, the next step is team classification. To ensure minimal user intervention, this process happens automatically within the first 100 frames of a match. Bounding boxes of all detected players are extracted. The system then filters out both green (common for artificial grass fields) and skin colors to isolate the players' jerseys. The top half of each bounding box, targeting the torso region as shown in **Figure** 3.11, is then used to extract color features.

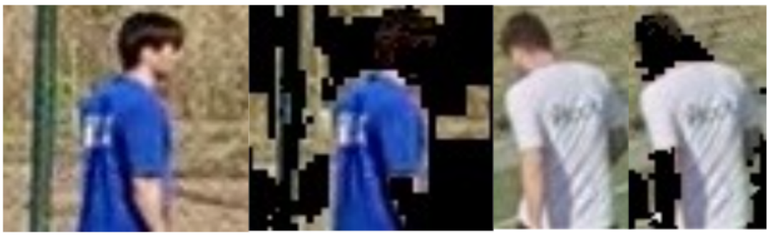

Figure 3.11: Team Assign remove background and skin color

These extracted colors are subsequently clustered using **k-means clustering**, which groups them into two distinct clusters representing the colors of the two teams. These clusters are stored as reference colors for the duration of the match. For any new ID generated during the game, the system attempts to classify the new player's jersey color and predict which team they belong to within a 30 frame window. This approach assumes that teams wear visibly different jersey colors, a practice common in amateur football where colored vests or contrasting shirts are typically used. A minor limitation of this approach is that players cannot wear green, black, or light skin-colored jerseys, as these are filtered out.

3.6 Field Calibration and Position Estimation

Accurate player statistics require mapping detections from 2D image coordinates to real-world 2D field coordinates. This process is essential because cameras introduce perspective distortion. A **perspective transformation using homography** was applied to correct for this distortion.

Homography requires at least four anchor points to delineate the playing court, allowing the system to convert the distorted field view into a standardized 2D coordinate system. An initial attempt using simple Canny edge detection with Hough transform for line detection proved to be ineffective due to the presence of multiple overlapping lines, varying light conditions, and different camera perspectives.

A more successful approach was the use of **YOLO11-Pose estimator** to track key points on the field. The first iteration of this approach used 21 key points, including penalty area boundaries, four corners, and goal posts, along with bounding boxes for the field and each penalty area. This key-point configuration was adapted for the five-a-side amateur fields, which often feature rounded penalty areas instead of rectangular ones.

While this solution was a significant improvement, further optimization was possible. A final, refined attempt still used the YOLO11-Pose model but changed the key point class to only one bounding box for the entire field. It also added more key points in both penalty areas, totaling 27 key points, as seen in **Figure** 3.12.

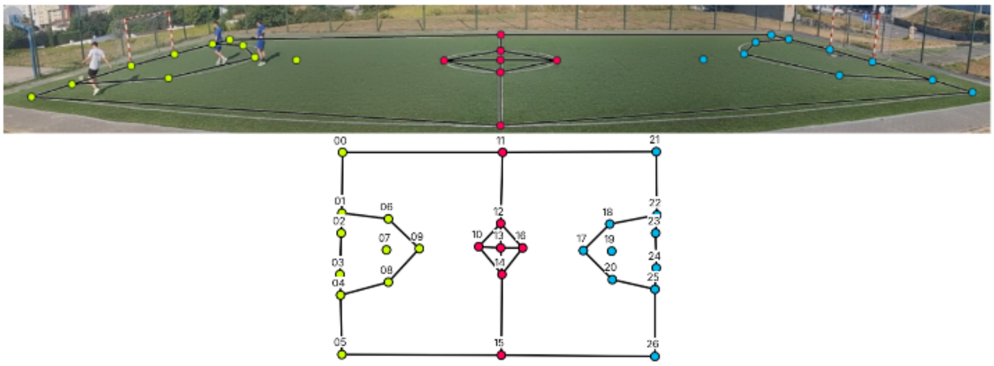

Figure 3.12: Annotated dataset

The custom dataset for this model consisted of over 500 real frames captured from various pitches, angles, and lighting setups(Figure 3.13).

Figure 3.13: Multiple Dataset Frames

These frames were then augmented with transformations such as grayscale, brightness, hue, and horizontal flips, resulting in over 1200 frames for training as seen in Figure 3.14.

Figure 3.14: Dataset Augmentation

Since the cameras remain static during a game, key point detection is performed at intervals rather than continuously. This reduces the computational load without compromising the accuracy of the homography transformation.

3.7 Statistics Extraction

The final stage of the pipeline involves generating individual and team statistics. For this proof-of-concept, the extracted metrics focused on both player and team-level performance. For individual players, the system calculated speed, distance traveled, and heat maps. For teams, it focused on ball possession and heat maps, in **Figure** 3.15 we can see the an illustration of this stats, on the top left corner the team ball possession and each player has a label with distance traveled, top speed and current speed.

Figure 3.15: Statistics extraction

To calculate **player speed and distance traveled**, the system processes the real-world position of each player, obtained via homography, in each frame. The distance traveled between frames is then calculated, and since the video is recorded at 30 frames per second, this distance can be converted into speed. **Heat maps** are generated by using the real-world position of each player and teams over time to indicate the density of their presence across the pitch.

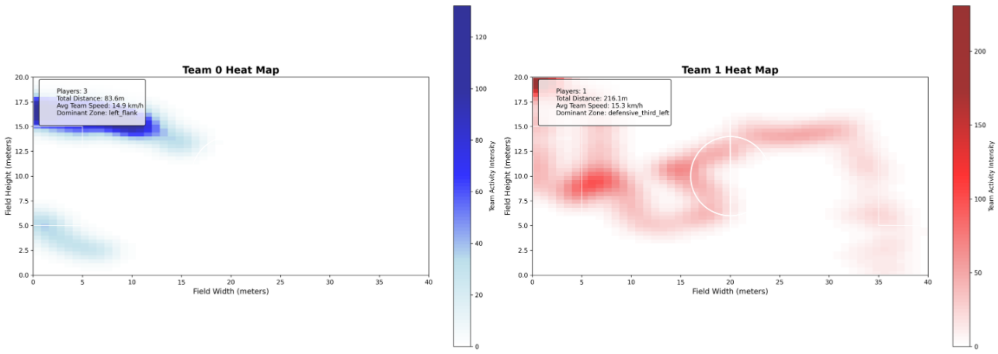

Figure 3.16: Teams Heat Map

Team ball possession is estimated by considering the ball's position relative to the players on the field. The system assumes a player has possession if the ball is within a close proximity to them. Possession is only transferred to the other team when a player from that team is detected to be in possession of the ball. While only a subset of statistics was implemented, the pipeline's modular design allows for future expansion to include more advanced metrics such as pass detection, shot attempts, or event classification.

Chapter 4

Evaluation

This chapter presents a comprehensive evaluation of the proposed system, focusing on the accuracy, robustness, and efficiency of its key components: object detection, player tracking, and statistics extraction. The evaluation was conducted using a dedicated dataset containing both familiar and new scenarios to assess the system's ability to generalize beyond its training data. This discussion not only presents quantitative results but also provides valuable insights into the strengths and weaknesses of the implementation, guiding potential paths for future work.

4.1 Evaluation Setup

Metrics - The evaluation of the models relied on a set of standard computer vision performance indicators, complemented by visual diagnostic tools to better understand their strengths and limitations. Metrics such as **Precision (P)**, which measures the rate of false positives, and **Recall (R)**, which measures the rate of false negatives, were used to quantify the model's accuracy. The **F1 Score**, the harmonic mean of Precision and Recall, provided a balanced measure, particularly important for imbalanced datasets. Additionally, the **Precision-Recall (PR) Curve** was utilized to visualize the trade-off between Precision and Recall at different confidence thresholds. The **Confusion Matrix** was used to compare predicted versus true labels, highlighting systematic misclassifications. Finally, the **Mean Per Joint Position Error (MPJPE)** was calculated to measure the average Euclidean distance in pixels between the ground truth and the predicted locations of key points, offering a more granular measure of positional accuracy. Together, these metrics allowed for a robust evaluation of both quantitative performance and qualitative robustness.

Hardware and Environment - The experiments were conducted on a workstation equipped with an NVIDIA RTX 3090 GPU with 24 GB of VRAM, an Intel Core i7 12th generation processor, and 32 GB of RAM. All data was stored and accessed from a solid-state drive to minimize input/output bottlenecks during training and evaluation. The software environment was based on Python 3.13, running within a Linux container via the Windows Subsystem for Linux (WSL). The training pipeline was implemented with PyTorch 2.8.0, compiled with CUDA 12.9 support, and the Ultralytics YOLO framework version 8.2.0.

4.2 Object Detection - Field Keypoints

The YOLO-based keypoint detection model was a primary focus of our evaluation, as it was custom-trained with two different types of labels. The initial training run, which used a labeling scheme that included separate bounding boxes for each penalty area and a larger one for the entire pitch, was initialized from the pre-trained YOLOv11pose backbone. The model was trained with an image size of 640 pixels, a batch size of 46, and a maximum of 500 epochs. However, the training process converged early at approximately epoch 346, as the "patience" parameter halted training due to a lack of significant improvement over the previous 50 epochs.

This training run presented some predictable challenges, particularly model overfitting, given the relatively small dataset of only 500 frames from four fields and over ten different camera perspectives. While data augmentation was used to expand the dataset, it was clear that achieving high precision without some degree of overfitting would be difficult. The training took roughly two and a half hours, with each epoch requiring approximately 26 seconds. As seen in the **Normalized Confusion Matrix** 4.1, the model was confusing the background with each area, even though the Precision, Recall, F1, and PR values were impressively high, as illustrated in **Figure** 4.2, where the Precision and Recall lines are nearly touching in the upper-right corner. While these metrics suggested a strong balance, a closer look at the predictions revealed that these values were not fully representative of the model's actual performance.

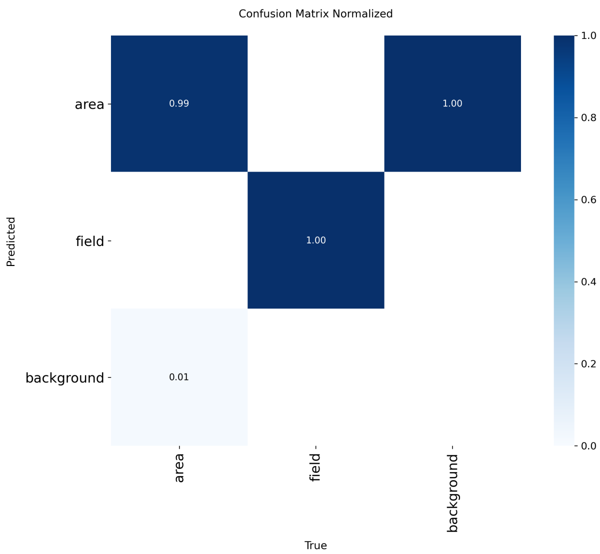

Figure 4.1: Field Keypoint Detection Confusion Matrix

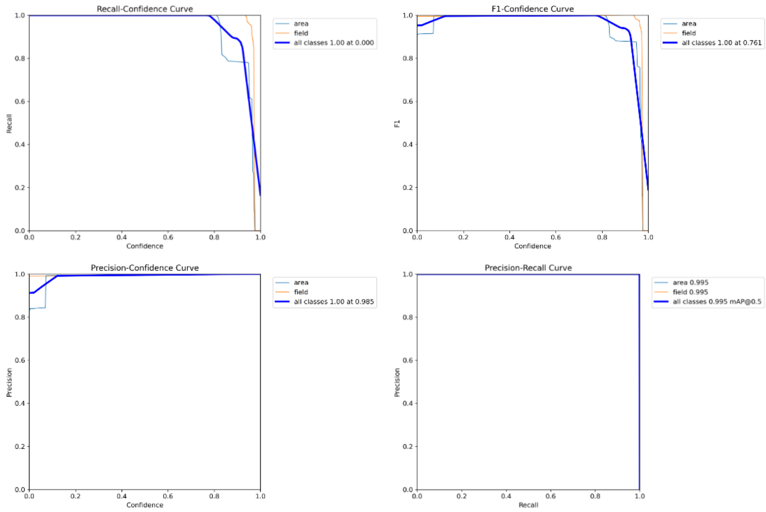

Figure 4.2: Field Keypoint Detection Pose Performance

To address the background confusion issue, a new approach was implemented. The distinction between the pitch and individual areas was removed, and a single, larger bounding box with all points was created. This change completely resolved the background misclassification problem by eliminating the irrelevant correlation between the pitch and area boundaries, as evidenced in **Figure** 4.3. The performance metrics also showed an increase, as seen in **Figure** 4.4.

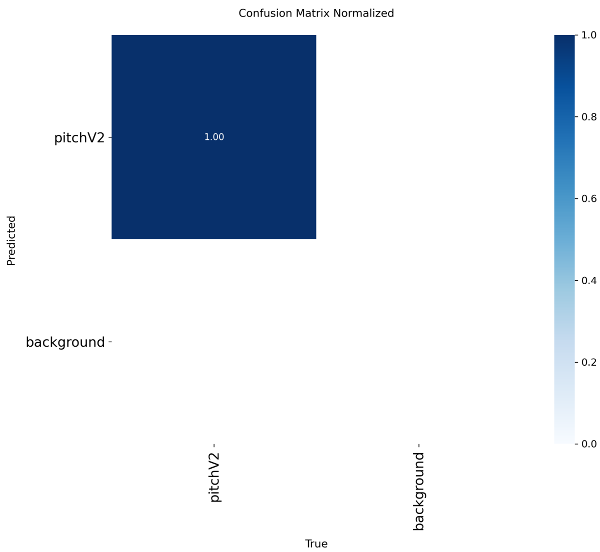

Figure 4.3: Field Keypoint Detection Confusion Matrix V2

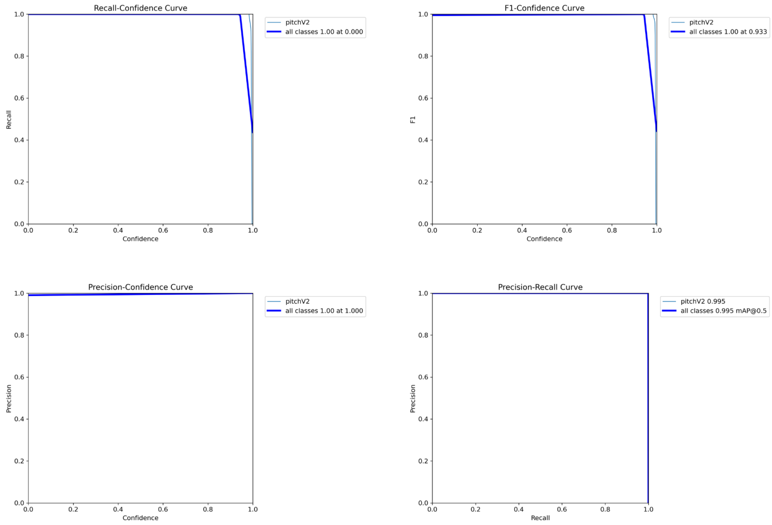

Figure 4.4: Field Keypoint Detection Pose Performance V2

However, these values seemed unrealistically high, leading to the conclusion that the standard metrics were not sensitive enough to the precise coordinates of the key points. **Figure** 4.5 illustrates this with several misplaced points. This issue stemmed from a limitation within the Ultralytics library, which did not allow for fine-tuning the error distance between predictions and ground truth for each point, leading to a premature conclusion of "good" performance.

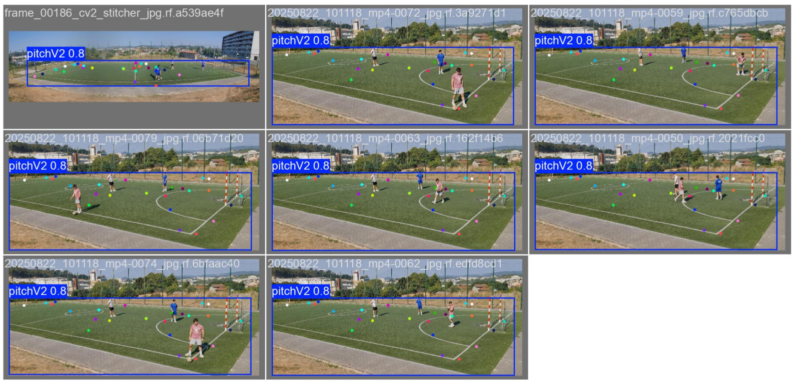

Figure 4.5: Field Keypoint Detection Prediction Batch

To accurately measure the error, a custom Python class was developed to compare the ground truth with the predictions and calculate the mean error in pixels. Running this new metric against the initial model yielded a **MPJPE** of 258.42 pixels. Considering the ground truth bounding box area of 1,189,300 pixels^2, this normalized to 0.1274. Analysis of individual points showed that some, like point 15, had a high MPJPE (687.41 pixels, or 0.3360 normalized), while others, such as point 25, had a much lower error (78.89 pixels, or 0.0385 normalized). This confirmed that on average, the model had a significant error of over 250 pixels.

With this new information, the model was retrained to avoid the premature stop. The patience parameter was set to 0, and the training was extended to 600 epochs. The image resolution was also increased to 960 pixels to provide more detail, which necessitated a reduction in the batch size to 4. The standard metrics remained consistent, as seen in **Figure** 4.6, further proving their insensitivity to minor coordinate discrepancies.

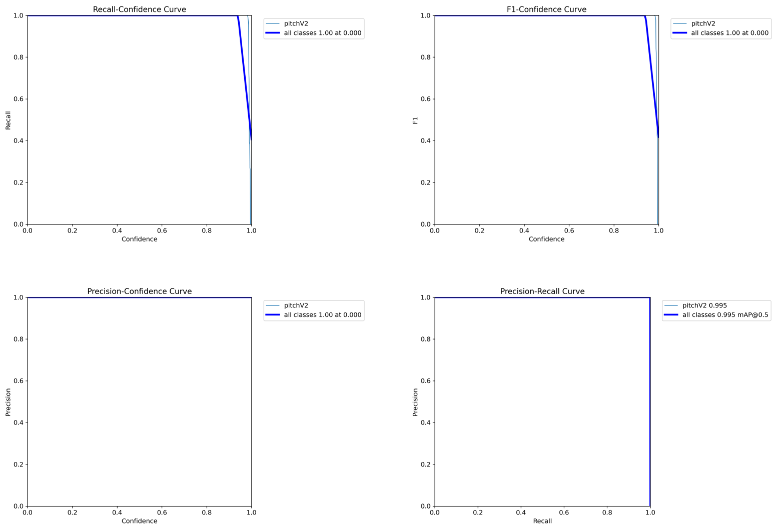

Figure 4.6: Field Keypoint Detection Pose Performance V3

However, the custom MPJPE metric revealed a significant improvement, with the new average error dropping to 86.29 pixels (0.0424 normalized). This represents a 66.6% improvement, visible on the new predictions batch in **Figure** 4.7. Despite this progress, certain outlier points, such as points 21 and 22, still exhibited high errors (431.66 and 423.15 pixels, respectively, or 0.2112 and 0.2069 normalized). These outliers likely indicate that the model still struggles to generalize to the small number of frames with unique camera perspectives that were not well represented in the training data.

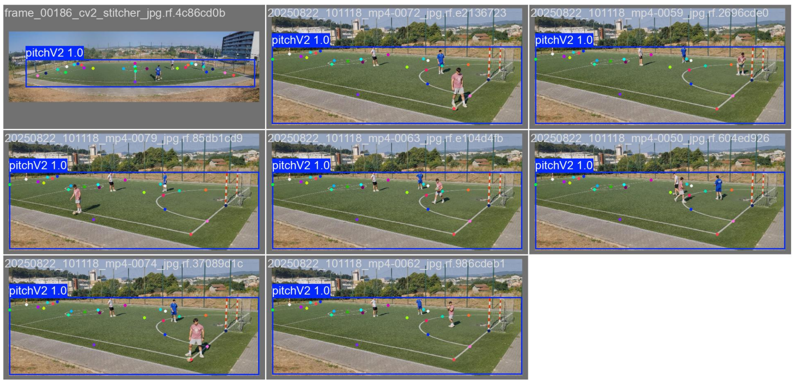

Figure 4.7: Field Keypoint Detection Prediction Batch V3

4.3 Object Detection - Player and Ball

The detection of players and the ball presented several challenges, particularly due to the small size and high speed of the ball. To evaluate the pre-trained model, a custom set of 40 challenging images was labeled, focusing on scenarios where the ball was in the air, on a player's hand, or missing from the frame as shown in **Figure** 4.8, to get an insight how the model would react on the worst scenarios.

Figure 4.8: Player and ball detection validation

The initial results of the base model were not optimal. As shown in **Figure** 4.9, while the precision-confidence curve was good for both the person and sport ball classes, there was a noticeable drop in recall as confidence and precision increased. For instance, at a confidence threshold higher than approximately 0.35, the model lost almost all detections of the sport ball class. This initial test was conducted with an image resize of 640 pixels, which made it difficult to detect smaller objects like the ball.

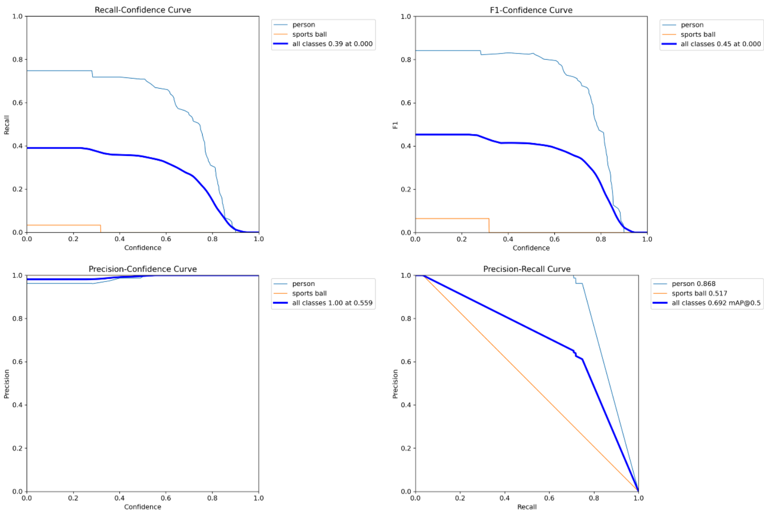

Figure 4.9: Player/Ball Detection Box Metrics

A subsequent test was run with images resized to 960 pixels, and the results showed a significant improvement, as seen in **Figure** 4.10. All metrics improved, particularly the Recall-Confidence and F1-Confidence curves for the sport ball class. This confirmed that higher resolution was crucial for the accurate detection of smaller, fast-moving objects. While this was a positive step, there is still significant room for improvement, which could be achieved in the future with a custom-trained dataset.

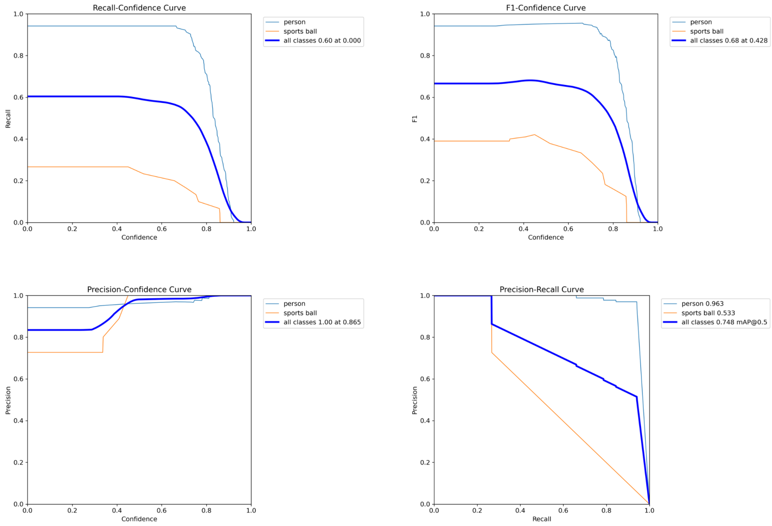

Figure 4.10: Player/Ball Detection Box Metrics V2

4.4 Statistics Extraction

Evaluating the accuracy of the statistics extraction was a challenging part of this project due to the lack of specialized equipment. The accuracy of metrics such as player top speed and distance traveled was assessed by comparing the system's output with data from a **Galaxy Watch7 LTE**, which functioned as a GPS tracker. This watch provided reference values for top speed and distance traveled as seen in **Figure** 4.11.

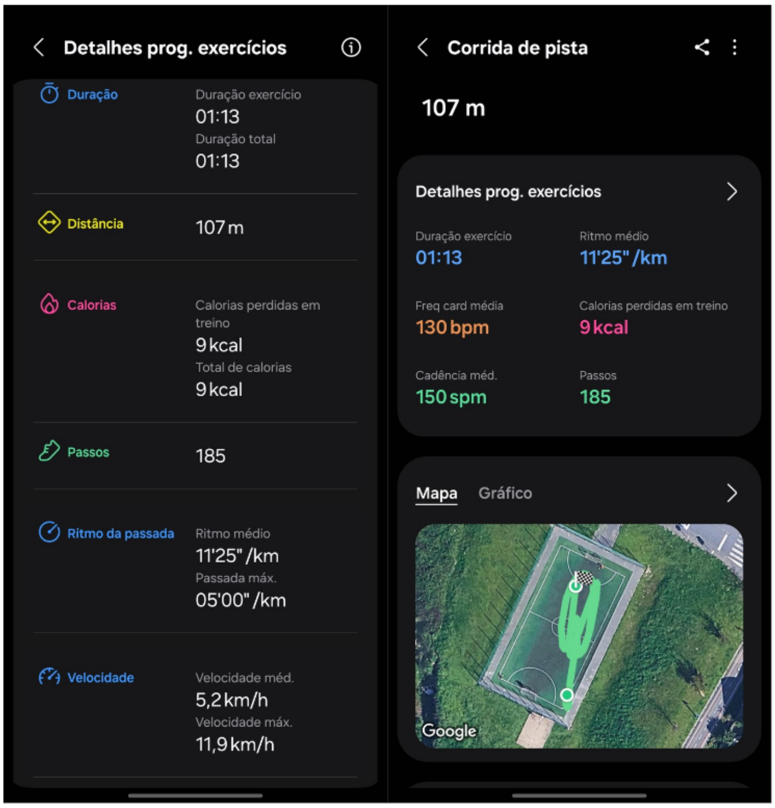

Figure 4.11: Smart Watch Statistics

After a manual comparison, it was found that the system's velocity measurements were consistently higher than the watch's by a margin of approximately 10 km/h with some clear outliers that spike the top speed to 140km/h. This discrepancy in speed also led to a higher calculated distance traveled. Several factors can be attributed to this distortion. First, the perspective distortion of the field itself, particularly when a player is far from the camera, can introduce significant errors. Second, the field dimensions were measured manually with steps, which is an imprecise method. This may have introduced measurement errors that caused players to appear to travel faster than their actual speed. It is also worth noting that the watch itself is not a perfect measurement tool and can introduce its own errors. The watch was used primarily as a consistent reference point rather than an absolute ground truth.

4.5 System-Level Performance

The system was designed with real-time performance in mind, and tests were conducted to evaluate this viability. The average processing time for the core pipeline was approximately **45 ms** per frame. For a video with a rate of 30 frames per second, this translates to a processing time of 1.35 seconds per second of video, creating a delay of 350 ms every second. Over a 10 minute clip, this accumulates to a 3.5-minute delay, and over a one-hour match, the delay would be 21 minutes, making the current setup unviable for real-time live analysis.

To improve performance without a heavy refactoring of the codebase, two key optimizations were implemented. The first and easiest solution was to reduce the frame rate processed by the system from 30 to 20 frames per second. At 20 frames per second, the system can process all frames within 900 ms (20 frames * 45 ms/frame), which is under the one-second window, thus eliminating the delay. The system can even handle up to 22 frames per second without accumulating a delay. The second optimization involves increasing the interval at which the field key points are extracted, as these do not change once a camera is fixed.

It is also important to note that the system has significant room for improvement through parallelization techniques. The Docker container running the inference model was only utilizing approximately 65% of the GPU, 8 GB of RAM, and 10% of the CPU, indicating that there is available headroom to run multiple processes simultaneously. This is a critical consideration for future work, as the current setup processes only one match at a time. The generation of heat maps and video labeling are also performed after the match, as this information is not critical for the user during a live game.

4.6 Summary of Findings

The evaluation confirms that while the system demonstrates the core functionalities needed to serve as a proof-of-concept for amateur football analytics, several key areas require further refinement.

- **Field Keypoint Detection:** The custom-trained YOLO model shows a significant reduction in MPJPE after a refined training strategy. However, outlier errors still exist, indicating that the model struggles to generalize to specific, underrepresented camera perspectives.
- **Player and Ball Detection:** The pre-trained model for object detection shows that a higher resolution (960 pixels) is crucial for accurately detecting small, fast-moving objects like the ball. A custom dataset and retraining could further enhance the model's accuracy.
- **Statistical Accuracy:** The system's statistical outputs for speed and distance traveled show a consistent overestimation when compared to a GPS-enabled sports watch. This is likely due to a combination of perspective distortion from the cameras and imprecise manual measurements of the field.
- **System Performance:** The system is not viable for real-time analysis at 30 frames per second due to a cumulative delay. However, by reducing the processed frame rate to 20 fps, the system can operate with sub-second latency. The low hardware utilization suggests that parallelism could be a key strategy for handling multiple simultaneous matches in the future.

In conclusion, the system successfully validates the feasibility of using a low-cost, AI-powered computer vision pipeline for amateur football analytics. The findings highlight specific areas, data, model training, and performance optimization, which can be the focus of future research to transform this proof-of-concept into a robust and accurate commercial product.

Chapter 5

Conclusions and Future Work

This dissertation proposed and implemented a computer vision–based system for analyzing amateur football matches without the need for wearable sensors. The system integrates object detection, multi-object tracking, team classification, field calibration, and statistics extraction into a coherent pipeline accessible through a mobile application. A modular architecture was designed to accommodate both live streaming and post match video upload, ensuring adaptability to different infrastructures, from smartphone-based recording to fixed-field cameras.

From a methodological perspective, the work applied the DSR framework to ensure iterative development, evaluation, and refinement. Empirical results demonstrated that the proposed models, particularly for field keypoint detection, achieved near-perfect performance, while player detection reached competitive levels in variable scenarios. Ball detection remained the most challenging task due to its small size and frequent occlusions, but even so the results confirmed the feasibility of deploying cost-effective analytics for amateur contexts.

In addition to the technical contributions, this research provided an augmentation pipeline tailored to address overfitting and improve generalization. Techniques such as grayscale conversion, noise injection, and synthetic transformations enabled the models to perform well in test scenarios that intentionally included conditions absent from training. By doing so, the project advances the state of low-cost sports analytics and lays the groundwork for scalable adoption in amateur environments.

5.1 Limitations

Despite the promising outcomes, several limitations must be acknowledged. The reliance on two-camera smartphone setups introduces challenges of parallax distortion and imperfect stitching, particularly in fields with irregular shapes or poor lighting. Player occlusion remains a major obstacle, leading to occasional identity loss even when modern tracking algorithms such as ByteTrack were used. Similarly, lighting variability and motion blur negatively impacted detection accuracy, particularly for the ball.

Another limitation concerns dataset size and diversity. Although approximately 1,200 annotated samples were used, complemented by artificial augmentation, the dataset is still modest compared to large-scale benchmarks. While the augmentation strategies mitigated overfitting, the system remains sensitive to extreme scenarios not represented in training, such as unusual weather conditions or atypical camera perspectives.

Additionally, the absence of reliable ground-truth methods for validating player movement, such as GPS trackers or high-precision motion capture systems, limits the accuracy assessment of the extracted performance metrics. This constraint makes it difficult to fully quantify the system's reliability when compared to established measurement standards.

Finally, the proof-of-concept implementation has yet to be validated extensively in large-scale, real-world deployments, where computational constraints, network instability, and user adoption present additional challenges.

5.2 Future Work

Future research can extend this work in several directions. Enhancing the robustness of tracking remains a priority, particularly regarding identity preservation under occlusion. Promising approaches include transformer-based tracking models and hybrid methods that combine appearance and motion cues. Improving ball detection may benefit from targeted dataset expansion, high-frame-rate recording, or temporal modeling techniques such as optical flow.

Another path concerns event detection and higher-level analytics. Extending the pipeline to automatically recognize passes, goals, or tactical formations would significantly enrich the value of the statistics provided. Pose estimation could also be leveraged to derive biomechanical insights, such as kicking dynamics or player fatigue indicators. In addition, future research could evaluate the applicability of the proposed system to other sports, broadening its relevance beyond football.

At a systems level, scaling the deployment to handle multiple simultaneous games and ensuring low-latency inference remain open challenges. The integration of edge computing or cloud-based GPU resources could support real-time applications at scale. Moreover, from a user perspective, incorporating gamification features, community tools, and comparative leaderboards would enhance engagement and broaden adoption.

Finally, exploring the feasibility of running lightweight versions of YOLO directly on smartphones presents a promising direction. This approach could further reduce infrastructure costs and make the solution even more accessible. Additionally, future improvements will involve testing the system under a wider range of conditions, including matches with more players (5v5 and 7v7 formats) as well as low-light environments during night games, to ensure robustness across diverse real-world scenarios.

5.3 Final Remarks

This research shows that advanced sports analytics do not need to remain exclusive to professional environments. By combining affordable devices with modern computer vision techniques, amateur players can now access insights that were once limited to elite teams. Although technical challenges persist, particularly in small-object detection and robust player tracking, the results demonstrate that a scalable and cost-effective solution is achievable.

Moreover, the methodological contributions of this work, including its modular architecture, data augmentation strategies, and evaluation protocols are transferable to other sports facing similar constraints. In this way, the project contributes not only to the democratization of sports analytics but also to the broader vision of making artificial intelligence accessible in everyday recreational contexts, thereby fostering greater participation, inclusivity, and enjoyment in sport.

References

Adžemović, Momir et al. (2025). “Beyond Kalman filters: deep learning-based filters for improved object tracking”. In: *Machine Vision and Applications* 36.1, pp. 1–22.

Anand Sharma, Rahul et al. (2017). “Automated Top View Registration of Broadcast Football Videos”. In: *arXiv e-prints*, arXiv–1703.

Arntzen, Halvard and Lars Magnus Hvattum (2021). “Predicting match outcomes in association football using team ratings and player ratings”. In: *Statistical Modelling* 21.5, pp. 449–470.

Asghar, Mamoona N et al. (2019). “Visual surveillance within the EU general data protection regulation: A technology perspective”. In: *IEEE Access* 7, pp. 111709–111726.

Athanesious, J Joshan and S Kiruthika (2024). “Perspective Transform based YOLO with Weighted Intersect Fusion for forecasting the Possession Sequence of the Live Football Game”. In: *IEEE Access*.

Benakesh, AH and R Rajeev (2024). “Advancing Football Game Analysis: Integrating Computer Vision, Deep Learning, and Hybrid Techniques for Enhanced Video Analytics”. In: *2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0*. IEEE, pp. 1–6.

Biliškov, Ivan et al. (2021). “Players Detection using U-Net based Fully Convolutional Network”. In: *2021 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, pp. 1–5.

Budgen, David and Pearl Brereton (2006). “Performing systematic literature reviews in software engineering”. In: *Proceedings of the 28th international conference on Software engineering*, pp. 1051–1052.

Chai, Junyi et al. (2021). “Deep learning in computer vision: A critical review of emerging techniques and application scenarios”. In: *Machine Learning with Applications* 6, p. 100134.

Chen, Jianhui and James J Little (2019). “Sports camera calibration via synthetic data”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 0–0.

Chen, Xiaogang et al. (2020). “Combination of augmented reality based brain-computer interface and computer vision for high-level control of a robotic arm”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.12, pp. 3140–3147.

Cioppa, Anthony et al. (2022). “Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3491–3502.

Constantinou, Anthony C (2019). “Dolores: a model that predicts football match outcomes from all over the world”. In: *Machine learning* 108.1, pp. 49–75.

Das, Sumit et al. (2018). “AI Doctor: An intelligent approach for medical diagnosis”. In: *Industry Interactive Innovations in Science, Engineering and Technology: Proceedings of the International Conference, I3SET 2016*. Springer, pp. 173–183.

Diop, Charles-Alexandre et al. (2022). "Soccer Player Recognition using Artificial Intelligence and Computer Vision". In: *2022 IEEE International Conference on Electro Information Technology (eIT)*. IEEE, pp. 477–481.

Ertel, Wolfgang (2024). *Introduction to artificial intelligence*. Springer Nature.

Evangelidis, Georgios D and Emmanouil Z Psarakis (2008). "Parametric image alignment using enhanced correlation coefficient maximization". In: *IEEE transactions on pattern analysis and machine intelligence* 30.10, pp. 1858–1865.

Ferreira, Fernando GDC, Amir H Gandomi, and Rodrigo TN Cardoso (2021). "Artificial intelligence applied to stock market trading: a review". In: *IEEE Access* 9, pp. 30898–30917.

Habibullah, Md et al. (2020). "Player Performance Profiling for Penalty Shootouts in Football Using Video Analysis". In: *Proceedings of the International Conference on Computing Advancements*, pp. 1–8.

Homayounfar, Namdar, Sanja Fidler, and Raquel Urtasun (2017). "Sports field localization via deep structured models". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5212–5220.

Huang, Hsiang-Wei et al. (2023). "Observation centric and central distance recovery for athlete tracking". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 454–460.

Huang, Wenbin et al. (2024). "A Novel LiDAR-Camera Fused Player Tracking System in Soccer Scenarios". In: *IEEE Sensors Journal*.

Javadiha, Mohammadreza et al. (2021). "Estimating player positions from padel high-angle videos: Accuracy comparison of recent computer vision methods". In: *Sensors* 21.10, p. 3368.

Kalafatić, Zoran, Tomislav Hrkać, and Karla Brkić (2022). "Multiple object tracking for football game analysis". In: *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*. IEEE, pp. 936–941.

Kao, Yu-Yung, Po-Yung Chou, and Cheng-Hung Lin (2024). "A Cost-effective Training Scheme for Person Re-Identification in Soccer Matches". In: *2024 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*. IEEE, pp. 455–456.

Karungaru, Stephen, Hiroki Tanioka, and Kenji Matsuura (2022). "Soccer players real location determination using perspective transformation". In: *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*. IEEE, pp. 1–4.

Kim, Hyunsung et al. (2022). "Cost-Efficient and Bias-Robust Sports Player Tracking by Integrating GPS and Video". In: *International Workshop on Machine Learning and Data Mining for Sports Analytics*. Springer, pp. 74–86.

Kitchenham, Barbara (2004). "Procedures for performing systematic reviews". In: *Keele, UK, Keele University* 33.2004, pp. 1–26.

Kröckel, Pavlina and Freimut Bodendorf (2020). "Process mining of football event data: a novel approach for tactical insights into the game". In: *Frontiers in Artificial Intelligence* 3, p. 47.

Liu, Liangkai et al. (2020). "Computing systems for autonomous driving: State of the art and challenges". In: *IEEE Internet of Things Journal* 8.8, pp. 6469–6486.

Mao, Jiageng et al. (2023). "3D object detection for autonomous driving: A comprehensive survey". In: *International Journal of Computer Vision* 131.8, pp. 1909–1963.

Mao, Xuting et al. (2022). "Financial fraud detection using the related-party transaction knowledge graph". In: *Procedia Computer Science* 199, pp. 733–740.

Mavrogiannis, Panagiotis and Ilias Maglogiannis (2022). "Amateur football analytics using computer vision". In: *Neural Computing and Applications* 34.22, pp. 19639–19654.

Modi, Piyush et al. (2024). "Real-time Object Tracking in Videos using Deep Learning and Optical Flow". In: *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, pp. 1114–1119.

Pai, Ping-Feng, Lan-Hung ChangLiao, and Kuo-Ping Lin (2017). "Analyzing basketball games by a support vector machines with decision tree model". In: *Neural Computing and Applications* 28, pp. 4159–4167.

Pardos, Antonis, Andreas Menychtas, and Ilias Maglogiannis (2022). "On unifying deep learning and edge computing for human motion analysis in exergames development". In: *Neural Computing and Applications* 34.2, pp. 951–967.

Qiu, Yuhua et al. (2023). "Real-time analysis scheme of football matches based on face recognition and multiple object tracking". In: *2023 5th International Conference on Robotics and Computer Vision (ICRCV)*. IEEE, pp. 159–163.

Redmon, Joseph (2018). "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767*.

Ren, Shaoqing et al. (2016). "Faster R-CNN: Towards real-time object detection with region proposal networks". In: *IEEE transactions on pattern analysis and machine intelligence* 39.6, pp. 1137–1149.

Sahni, Srishti et al. (2020). "Insurance fraud identification using computer vision and iot: a study of field fires". In: *Procedia Computer Science* 173, pp. 56–63.

Sarkar, Saikat, Dipti Prasad Mukherjee, and Amlan Chakrabarti (2022). "Reinforcement learning for pass detection and generation of possession statistics in soccer". In: *IEEE Transactions on Cognitive and Developmental Systems* 15.2, pp. 914–924.

Sarkis-Onofre, Rafael et al. (2021). "How to properly use the PRISMA Statement". In: *Systematic Reviews* 10, pp. 1–3.

Sheng, Bin et al. (2020). "GreenSea: visual soccer analysis using broad learning system". In: *IEEE Transactions on Cybernetics* 51.3, pp. 1463–1477.

Szeliski, Richard (2022). *Computer vision: algorithms and applications*. Springer Nature.

Tang, Kui et al. (2023). "Cascade RCNN with Hybrid Attention and Dual Pooling for Soccer Player Detection". In: *2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD)*. IEEE, pp. 695–700.

Tarashima, Shuhei (2021). "Sports field recognition using deep multi-task learning". In: *Journal of Information Processing* 29, pp. 328–335.

Theagarajan, Rajkumar and Bir Bhanu (2020). "An automated system for generating tactical performance statistics for individual soccer players from videos". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.2, pp. 632–646.

ultralitics (2024). `https://docs.ultralytics.com/`.

– (2025). `https://docs.ultralytics.com/models/yolo12/`.

Vom Brocke, Jan, Alan Hevner, and Alexander Maedche (2020). "Introduction to design science research". In: *Design science research. Cases*, pp. 1–13.

Waelen, Rosalie A (2024). "The ethics of computer vision: an overview in terms of power". In: *AI and Ethics* 4.2, pp. 353–362.

Wang, Jing and Baiqing Liu (2023). "Analyzing the feature extraction of football player's offense action using machine vision, big data, and internet of things". In: *Soft Computing* 27.15, pp. 10905–10920.

Wang, Zhiyang, Lei Shan, and Lei Feng (2024). "Hyper DeepSORT: Elevating Precision in Multi-Object Tracking through HyperNMS and Adaptive Kalman Filtering Innovations." In: *Engineering Letters* 32.9.

Wojke, Nicolai, Alex Bewley, and Dietrich Paulus (2017). *Simple Online and Realtime Tracking with a Deep Association Metric*. arXiv: 1703.07402 [cs.CV]. url: https://arxiv.org/abs/1703.07402.

Zhang, Mengjie (2023). "Research on target detection and tracking algorithm of football game video based on deep learning". In: *2023 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. IEEE, pp. 116–120.

Zhang, Yifu et al. (2022). *ByteTrack: Multi-Object Tracking by Associating Every Detection Box*. arXiv: 2110.06864 [cs.CV]. url: https://arxiv.org/abs/2110.06864.

Zhou, Longfei, Lin Zhang, and Nicholas Konz (2022). "Computer vision techniques in manufacturing". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53.1, pp. 105–117.