الجمهورية الجزائرية الديمقراطية الشعبية

Algeria of Republic Democratic People's

وزارة التعليم العالي والبحث العلمي

Ministry of Higher Education and Scientific Research

جامعة الشهيد حمه لخضر الوادي

University of Echahid Hamma Lakhdar El-Oued

كلية العلوم الدقيقة

Faculty of Exact Sciences

قسم الاعلام الالي

Computer Science Department

Thesis

Submitted in partial fulfilment of the requirements for the Degree of

ACADEMIC MASTER

Field: Mathematics and Computer Science

Option: Computer Science

Specialty: Artificial Intelligence and Data Science

Theme

Player Tracking and Analysis in Football Matches

by:

- Belhadi Akram

Supervisor:

- Naoui Mohammed Anouar

Co-Supervisor:

- Bouhammed Mohammed Mounir

President:

- Lejdel Ibrahim

Discussed:

- Bali Mouadh

Academic Year: **2024/2025**

Acknowledgments

First and foremost, I thank God Almighty, whose grace, guidance, and mercy have led me to this moment and this achievement.

I extend my deepest gratitude to my beloved father and mother, whose sacrifices, prayers, and unwavering support have been the foundation of my journey. I also thank my dear brothers, who have been a source of strength and encouragement.

My heartfelt thanks go to my extended family — my uncles and maternal uncles — for their love and constant encouragement throughout the years.

To all my friends, without exception, thank you for your companionship, your support in both good and challenging times, and for being part of this unforgettable journey.

I am especially grateful to my thesis supervisors, Professor Naoui Mohammed Anouar and Professor Bouhamed Mohammed Mounir, for their valuable guidance, patience, and insightful advice throughout the course of this work.

To all my teachers and professors, I express my sincere appreciation for your dedication and your role — after God — in shaping my knowledge and helping me reach this stage.

Finally, I dedicate this work to the resilient people of Gaza. May God grant them dignity, strength, and a dear and honorable victory.

Contents

List of Tables

List of Figures

Abstract

Football has evolved significantly over the past decades, not only in terms of gameplay and strategy but also through the integration of advanced technologies. With the rise of artificial intelligence, automated analysis of football matches has become a promising field for enhancing coaching strategies, performance evaluation, and real-time broadcasting insights. This work aims to develop a comprehensive system capable of extracting meaningful information from raw football match footage by leveraging computer vision techniques. The significance of this work lies in its holistic approach, combining multiple sub-tasks—each addressing a specific challenge within football video analysis—to build an end-to-end intelligent framework that contributes to the growing demand for automated sports understanding systems. The methodology followed in this work involves several stages of visual data processing and model design. Initially, a semantic segmentation approach was applied to separate relevant entities from the field; however, its performance proved insufficient in complex scenes. This led to the adoption of object detection methods using YOLOv8 models, which were trained to detect players, referees, and goalkeepers. Due to the small size and high motion variability of the ball, a dedicated detection model was trained separately with scaled-up input images to improve ball recognition. To understand the spatial structure of the field, a keypoint detection model was implemented to localize crucial pitch landmarks and infer field dimensions. Subsequently, a classification model was trained to identify and categorize key match events such as goals, fouls, and substitutions. The final system integrates the outputs of these models to form a unified pipeline capable of performing comprehensive football match analysis with high accuracy and efficiency. Our overarching goal is to reduce human errors in football officiating by leveraging various artificial intelligence techniques.

Keywords: Computer Vision, Football Technology, Object Detection, Keypoint Detection, Semantic Segmentation, Event Classification

الملخص

شهدت كرة القدم تطورًا ملحوظًا على مدار العقود الماضية، ليس فقط من حيث أسلوب اللعب والاستراتيجيات، بل أيضًا من خلال دمج التقنيات المتقدمة. ومع صعود الذكاء الاصطناعي، أصبح التحليل الآلي لمباريات كرة القدم مجالًا واعدًا لتحسين استراتيجيات التدريب وتقييم الأداء ورؤى البث المباشر. يهدف هذا العمل إلى تطوير نظام شامل قادر على استخراج معلومات مفيدة من لقطات مباريات كرة القدم الخام بالاستفادة من تقنيات الرؤية الحاسوبية. تكمن أهمية هذا العمل في نهجه الشامل، الذي يجمع بين مهام فرعية متعددة - كل منها يعالج تحديًا محددًا في تحليل فيديوهات كرة القدم - لبناء إطار عمل ذكي متكامل يُسهم في تلبية الطلب المتزايد على أنظمة فهم الرياضة الآلية. تتضمن المنهجية المتبعة في هذا العمل عدة مراحل من معالجة البيانات المرئية وتصميم النماذج. في البداية، تم تطبيق نهج التجزئة الدلالية لفصل الكيانات ذات الصلة عن الملعب؛ إلا أن أداءه لم يكن كافيًا في المشاهد المعقدة. أدى ذلك إلى اعتماد أساليب كشف الأجسام باستخدام نماذج YOLOv8، التي تم تدريبها على كشف اللاعبين والحكام وحراس المرمى. نظرًا لصغر حجم الكرة وتقلب حركتها الكبير، تم تدريب نموذج كشف مخصص بشكل منفصل باستخدام صور مُدخلة مُكبرة لتحسين التعرف على الكرة. لفهم البنية المكانية للملعب، تم تطبيق نموذج كشف النقاط الرئيسية لتحديد معالم الملعب المهمة واستنتاج أبعاده. بعد ذلك، تم تدريب نموذج تصنيف لتحديد وتصنيف أحداث المباراة الرئيسية، مثل الأهداف والأخطاء والتبديلات. يدمج النظام النهائي مخرجات هذه النماذج لتشكيل مسار موحد قادر على إجراء تحليل شامل لمباريات كرة القدم بدقة وكفاءة عاليتين. هدفنا الرئيسي هو تقليل الأخطاء البشرية في تحكيم كرة القدم من خلال الاستفادة من تقنيات الذكاء الاصطناعي المختلفة.

الكلمات المفتاحية: الرؤية الحاسوبية، تكنولوجيا كرة القدم، اكتشاف الكائنات، اكتشاف النقاط الرئيسية، التجزئة الدلالية، تصنيف الأحداث

Résumé

Le football a considérablement évolué au cours des dernières décennies, non seulement en termes de gameplay et de stratégie, mais aussi grâce à l'intégration de technologies avancées. Avec l'essor de l'intelligence artificielle, l'analyse automatisée des matchs de football est devenue un domaine prometteur pour améliorer les stratégies d'entraînement, l'évaluation des performances et la diffusion en temps réel. Ce travail vise à développer un système complet capable d'extraire des informations pertinentes à partir d'images brutes de matchs de football en exploitant des techniques de vision par ordinateur. L'intérêt de ce travail réside dans son approche holistique, combinant plusieurs sous-tâches, chacune répondant à un défi spécifique de l'analyse vidéo de football, afin de construire un cadre intelligent de bout en bout qui contribue à la demande croissante de systèmes automatisés de compréhension du sport. La méthodologie suivie dans ce travail implique plusieurs étapes de traitement des données visuelles et de conception de modèles. Initialement, une approche de segmentation sémantique a été appliquée pour séparer les entités pertinentes du terrain ; cependant, ses performances se sont avérées insuffisantes dans les scènes complexes. Cela a conduit à l'adoption de méthodes de détection d'objets utilisant des modèles YOLOv8, entraînés à détecter les joueurs, les arbitres et les gardiens de but. En raison de la petite taille et de la grande variabilité des mouvements du ballon, un modèle de détection dédié a été entraîné séparément avec des images d'entrée agrandies afin d'améliorer la reconnaissance du ballon. Afin de comprendre la structure spatiale du terrain, un modèle de détection de points clés a été mis en œuvre pour localiser les points de repère cruciaux du terrain et en déduire les dimensions. Par la suite, un modèle de classification a été entraîné pour identifier et catégoriser les événements clés du match, tels que les buts, les fautes et les remplacements. Le système final intègre les résultats de ces modèles pour former un pipeline unifié capable d'effectuer une analyse complète des matchs de football avec une grande précision et une grande efficacité. Notre objectif principal est de réduire les erreurs humaines dans l'arbitrage du football en exploitant diverses techniques d'intelligence artificielle.

Mots-clés: Vision par ordinateur, Technologie du football, Détection d'objets, Détection de points clés, Segmentation sémantique, Classification d'événements

General Introduction

Football is witnessing continuous development at both the tactical and technical levels, driven by advances in information technology and artificial intelligence. This integration between the two fields has opened up broad horizons for analyzing matches, evaluating player performance, and contributing to accurate refereeing decisions. Computer vision, as a branch of artificial intelligence, has become an effective tool for extracting information from images and videos, revolutionizing the way football data is processed and interpreted automatically and intelligently, However, despite advances in technology, many refereeing errors persist, often unfair, and can negatively impact the enjoyment of football. In this work, we will attempt to address the problem of refereeing errors during matches.

this work lies in how to employ computer vision techniques to solve challenges associated with football, specifically reducing refereeing errors and improving our understanding of the game through visual data analysis. we worked on a number of topics, including data collection and processing, solving the problem of detecting small objects such as the ball, automatically identifying field features, and identifying important events in the match. These challenges were addressed using various techniques such as Semantic Segmentation, Object Detection, Keypoint Detection, and Classification, enabling the development of an integrated system for analyzing football matches.

The methodology used in this work relies on a gradual approach to the problems facing video data modeling in football. First, the data was collected and processed, followed by the use of specialized algorithms to detect different visual entities. Deep learning techniques were then employed to extract pitch characteristics and structure, then classification models were trained on the classified scenes to enhance the system's ability to identify significant events, leading to the integration of the results into a comprehensive analytical system.

This work presents the following: In the first chapter, the historical development of football was studied along with the development of technology, and the impact of the latter in supporting both athletic and technical aspects, whether in assisting referees, players, or even coaches. In the second chapter, a comprehensive overview of artificial intelligence and its various applications was presented, focusing on its role in sports in general and how it is developing them, specifically football, in an effort to automate them with effective intelligent systems. In Chapter 3, the steps of the proposed work are explained, starting with data processing using Semantic Segmentation, then moving on to Object Detection to detect key entities, training a special model for the ball, passing through the detection of reference points on the field, and ending with event classification and integrating all results into a unified framework.

Chapter 1

The Transformation of Football with Technology

1.1 Introduction

Football isn't just about what happens on the grass—there's more to it than that It is a global social, economic, and cultural phenomenon that unites people of all races and cultures, igniting the passions of millions in every corner of the planet. This game has evolved significantly since its inception, transforming from a simple sport played on streets and dirt fields into a complex professional industry run by large sports organizations and media networks, with annual revenues estimated at billions of dollars. With the increasing speed and competitiveness of the game, there has been an urgent need for scientific and technical tools to help clubs, coaches, and players understand and accurately evaluate performance. So, technical and tactical analysis have become really important in this whole thing the development of modern football. It is no longer possible to rely solely on innate skill or luck; success in matches is now linked to a careful analysis of the opponent, an objective assessment of player performance, and the intelligent use of data and information. In this context, artificial intelligence has begun to occupy a It's a big part of the football world, used in many areas like analyzing videos and working with players injury prediction, team performance evaluation, and the discovery of emerging talent. Through technologies such as computer vision, machine learning, and deep learning, it has become possible to accurately track player and ball movements, analyze tactical patterns, and even predict potential decisions within a match. These tools not only serve technical staff but also contribute to supporting referees through technologies such as video assistant referees (VAR), and to enhancing the viewer experience through interactive statistics and intelligent match presentations. Thus, artificial intelligence has become a fundamental strategic tool in developing the game, making the difference between victory and defeat, confirming that the future of football will be closely linked to quick progress in science and technology.

1.2 The Socioeconomic and Health Impact of Football

1.2.1 Economic Influence of the Football Industry

The economic impact of football extends far beyond the boundaries of the pitch. As one of the most lucrative industries in the global sports sector, football generates vast revenues through multiple interconnected channels. Major football clubs have evolved into powerful commercial entities, managing multi-million dollar budgets, sponsorship deals, and global fanbases. Broadcasting rights alone bring in a big part of the industry's revenue financial muscle, with leagues such as the English Premier League, La Liga, and UEFA Champions League securing billions of dollars from television networks seeking exclusive access to live matches[150]. In addition, advertising and sponsorship partnerships—often with multinational corporations—inject substantial capital into clubs, federations, and players' branding campaigns. The merchandise market, including jerseys, footwear, collectibles, and licensed products, further fuels this economic engine, especially with the global reach of online commerce. Local economies also benefit, as matchday revenues stimulate sectors such as tourism, hospitality, transport, and small businesses. The football industry's resilience and adaptability were especially evident during challenges such as the COVID-19 pandemic, which pushed clubs and organizers to develop innovative revenue streams including virtual fan experiences, digital ticketing, and streaming platforms [152]. Football, at its core, is just about the game and the thrill of it has become not only a sport but also a critical pillar of economic development in many regions of the world.

1.2.2 Social Integration and Cultural Influence

Football isn't just about money; it also helps bring people together and share different cultures dialogue. It's a way of communicating that crosses countries, religions, and languages, helping people connect regardless of their backgrounds together diverse populations under a shared passion. International tournaments like the FIFA World Cup, UEFA European Championship, and the Africa Cup of Nations are not only sporting spectacles but also platforms for intercultural exchange and diplomacy. These events bring countries together, building a sense of pride and teamwork fans alike, often serving as moments of collective national identity and celebration [86]. On the club level, global teams like Real Madrid, Manchester United, and Bayern Munich have fans from virtually every continent, symbolizing how football can bridge cultural and geographical divides. This inclusivity is also reflected in Community projects and local programs that focus on helping young people, supporting gender equality, and building stronger neighborhoods addressing social inequalities through sport. Football has also been used to support humanitarian efforts to raise awareness and support for global causes such as education, health, and peacebuilding. Stadiums often become melting pots of cultures, where people from varied backgrounds come together to support their teams, creating bonds that extend beyond the game. Thus, football is not merely entertainment—it is a cultural force that shapes values, connects societies, and promotes global citizenship.

1.2.3 Football as a Tool for Physical and Mental Health

Football also holds immense value as a promoter of physical fitness and mental well-being. On a physical level, playing football engages almost every muscle group, enhances cardiovascular health, improves coordination, and builds endurance. Its lively and competitive vibe makes it a great kind of cardio workout for individuals of all ages and fitness levels. Children develop motor skills, teamwork, and discipline, while adults benefit from stress relief and weight management. In recent years, structured training programs based on football have been incorporated into public health initiatives aimed at combating sedentary lifestyles, obesity, and heart diseases [93]. Beyond the physical advantages, football offers critical psychological benefits. It It helps young people and those who might feel left out feel like they belong and more confident about themselves. Team environments encourage communication, emotional regulation, and resilience, while the thrill of competition and achievement stimulates dopamine release, enhancing mood and motivation. For spectators, engaging with football as fans can also provide emotional catharsis, social identity, and even mental distraction from everyday stressors. Therapeutic football programs have been used in various clinical settings to support individuals with mental health conditions, PTSD, and neurodevelopmental disorders. Consequently, football is a comprehensive A tool that supports overall well-being, helping to care for the body, mind, and social life all at once world.

1.3 The Emergence of Technology in Football

1.3.1 Historical Evolution of Technology in Sports

The integration of technology into sports has a long and dynamic history, reflecting the broader evolution of scientific progress and innovation in human activity. Initially, the role of technology in athletics was minimal, limited to basic tools such as stopwatches and measuring tapes. However, as competitive sports grew in popularity and professionalism, the demand for precision, fairness, and performance optimization pushed the boundaries of technological development. In athletics and swimming, the introduction of electronic timing systems ensured greater accuracy in race results, while in tennis and cricket, technologies like Hawk-Eye and Snickometer revolutionized decision-making by providing real-time visualizations of ball trajectories and contact points [40, 156]. These advancements not only improved the credibility of officiating but also enhanced the viewing experience for fans and analysts alike.

In parallel, the rise of video analysis tools and biomechanical sensors enabled coaches and athletes to study motion, optimize techniques, and prevent injuries, marking the beginning of a data-driven era in sports training. Wearable devices, GPS trackers, and heart rate monitors began to collect physiological and performance metrics that were previously inaccessible, opening new dimensions of individualized and evidence-based coaching [11, 25]. The proliferation of broadcast technology, including high-definition cameras, drones, and augmented reality overlays, further changed how sports are consumed and analyzed. These innovations paved the way for more advanced systems involving artificial intelligence, machine learning, and computer vision—technologies that are now central to elite sports performance analysis, including in football. Thus, the historical trajectory of technology in sports reveals a continuous and accelerating integration of science into athletic practice, with Football is really starting to benefit a lot from these changes.

1.3.2 Adoption of AI and Data in Modern Football

The adoption of artificial intelligence (AI) and data analytics in football represents one of the most transformative trends in the modern history of the sport. Although rudimentary forms of data collection—such as manual notation of passes, shots, and tackles—existed for decades, the systematic and intelligent application of large-scale data only began to gain momentum in the late 2000s and early 2010s. With the expansion of computational power and the availability of vast datasets, football clubs started to harness AI technologies to gain competitive advantages on and off the pitch. Early pioneers such as Brentford FC and FC Midtjylland demonstrated how predictive modeling, statistical analysis, and machine learning could be used for talent scouting, match preparation, and tactical planning.

One of the major milestones in this domain was the development of tracking systems capable of monitoring player and ball movements across the pitch in real-time using computer vision and sensor-based technology. Companies like StatsBomb, Opta, and Second Spectrum began offering granular event and positional data, which clubs used to analyze spatial dynamics, build player heatmaps, and assess tactical formations. AI-driven systems also contributed to injury prevention through load management algorithms that detect fatigue and predict injury risk based on training intensity and match exertion [11, 46]. On the strategic front, coaches now receive AI-generated reports highlighting opponent weaknesses, set-piece vulnerabilities, and optimal substitutions, allowing for data-informed decisions that can significantly influence match outcomes.

Furthermore, in the realm of fan engagement and broadcasting, AI is used to personalize content, generate highlights, and simulate match scenarios. Advanced models also analyze referee decisions, detect offside positions automatically, and simulate probable outcomes in complex in-game situations. These applications have not only professionalized football analysis but have also redefined how stakeholders—coaches, analysts, referees, fans, and players—interact with the game. Today, nearly every top-tier football club operates with a dedicated data analysis department, marking the widespread institutionalization of AI in the sport. As the technology continues to evolve, it is expected to drive even deeper insights into performance dynamics, game theory, and player psychology, establishing AI as a permanent fixture in the future of football.

1.4 Tracking Systems and Performance Analytics

1.4.1 Wearable Sensors and GPS Systems

In today's football world, players and teams are using wearable sensors and GPS trackers more and more indispensable tools for monitoring player performance with exceptional accuracy. These technologies offer real-time insights into the physical exertion and movement patterns of athletes during both training sessions and official matches [20, 106]. Wearable sensors, typically embedded in lightweight vests or undergarments, are equipped with accelerometers, gyroscopes, magnetometers, and GPS modules. They measure a wide array of metrics such as distance covered, sprint speed, acceleration and deceleration rates, changes in direction, and even impacts from collisions. This kind of info is important not just for getting better at sports, but also for knowing when you're pushing too hard and need a break. [106].

GPS systems in particular enable coaches and sports scientists to visualize a player's positional data throughout a match. By mapping heat zones and analyzing spatial distribution, coaching staff can determine how effectively a player maintains their role within tactical formations or exploits specific areas of the pitch [20]. These insights inform adjustments in training intensity, substitution decisions, and long-term conditioning programs. For example, if data shows that a player consistently underperforms in high-intensity sprints after a Once a certain number of minutes have passed, the training staff can adjust the workout plan or schedule breaks as needed . Moreover, centralized databases storing longitudinal tracking data across multiple matches and seasons allow for comparative performance analysis and the detection of trends over time. This fusion of physiological and positional data, driven by wearable and GPS technologies, has redefined evidence-based coaching and sports medicine in professional football [106].

1.4.2 Biomechanical and Tactical Analysis Using AI

AI tools do beyond track how players move. They also analyze the details of how players perform and help figure out strategies for matches.. Biomechanical analysis refers to the study of human movement and the mechanical efficiency of body parts during athletic activities. In football, AI systems equipped with computer vision and machine learning algorithms analyze body posture, joint angles, stride patterns, and balance [19]. Such assessments can reveal underlying inefficiencies in movement that may predispose athletes to injury or reduce on-field effectiveness. For instance, subtle asymmetries in running gait or frequent poor landing mechanics during aerial duels can be identified through video-based biomechanical modeling. With this data, physiotherapists and conditioning coaches can intervene with personalized exercises that enhance biomechanical function and reduce injury risks.

From a tactical standpoint, AI has revolutionized how teams interpret their own play and that of their opponents. Tactical AI platforms process thousands of events per match—from passes and interceptions to pressing patterns and player orientations—enabling coaches to construct detailed models of gameplay behavior [44, 168]. These systems can identify the effectiveness of different formations, measure spatial compactness, evaluate transition speed between defensive and offensive phases, and suggest counterstrategies based on historical success rates. Clustering algorithms can help by grouping players based on their roles or how they influence the game. This makes it easier to scout talent and focus on developing specific roles. For example, AI can determine which combinations of midfielders result in the highest ball retention rates or which defensive duos most effectively limit goal-scoring opportunities [32].

Getting both biomechanical and tactical data together really helps paint a clear picture of a player's overall worth. A physically fit player who excels tactically within a specific system can be distinguished from one who performs well physically but lacks positional intelligence. This teamwork makes it easier to make smart, data-based choices when developing players. , match preparation, and post-game evaluations. As AI systems become more sophisticated and datasets grow in depth and resolution, the fusion of biomechanical and tactical insights is expected to play an increasingly central role in shaping the modern footballer [168].

1.5 Technology in Refereeing: A Paradigm Shift

1.5.1 VAR: Operation and Controversies

The arrival of Video Assistant Referee (VAR) technology has been one of the biggest changes in how football referees do their job. VAR was implemented with the aim of reducing human error in crucial match decisions, including goals, penalty incidents, direct red cards, and mistaken identity [82, 88, 117]. The system operates through a team of assistant referees located in a remote video operation room, where they monitor multiple camera feeds from the stadium in real-time. When a potentially game-changing event occurs, the VAR team communicates with the on-field referee through a secure audio link, providing recommendations or suggesting a video review on the pitch-side monitor [88]. This collaborative review process allows the main referee to reassess the incident from multiple angles and at different speeds before making a final decision.

Despite its potential for improving accuracy, VAR has stirred substantial controversy within the football community [18, 88, 117]. Critics argue that the technology, while precise, often disrupts the flow of the game due to extended review times and ambiguous interpretation of subjective calls—especially in scenarios involving handballs or minor contact. Additionally, the perceived inconsistency in when and how VAR is used across different leagues and competitions has led to debates over fairness and transparency. Spectators and players have expressed frustration over delays and the emotional anticlimax that results when celebratory moments are paused or overturned after lengthy reviews [117]. Plus, even though VAR is meant to help referees, not replace their calls, there are concerns about growing overreliance on technology, which may diminish the authority and spontaneity of on-field officiating [18]. Thus, while VAR has certainly raised the bar for decision accuracy, it also brings ethical and operational dilemmas that continue to challenge football's governing bodies.

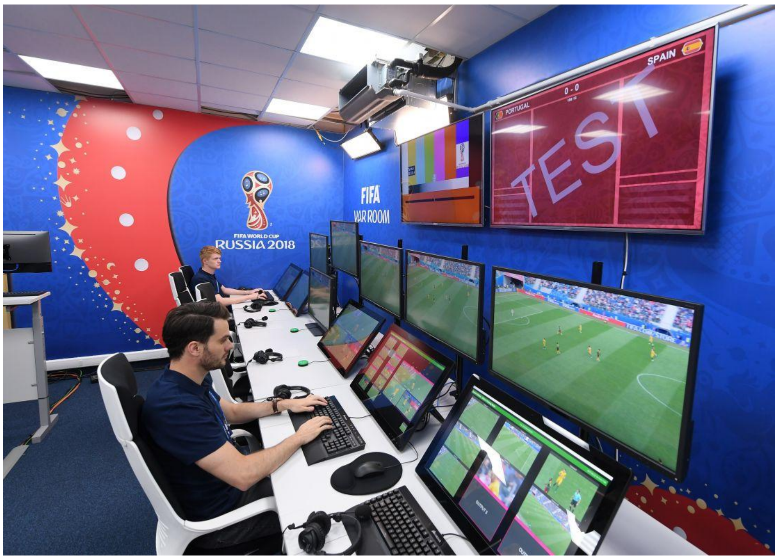


Figure 1.1: Var Technology in World Cup 2018 [57]

1.5.2 Goal-Line Technology and Decision Assurance

In contrast to VAR, goal-line technology (GLT) has been met with widespread acceptance and minimal controversy due to its simplicity and objectivity [56, 105]. GLT is specifically designed to determine whether the entire ball has crossed the goal line between the posts and under the crossbar—an essential aspect of the game that was previously vulnerable to human error, especially in fast-paced or obstructed scenarios. The most common form of GLT employs high-speed cameras positioned around the stadium, calibrated to triangulate the ball's exact position in relation to the goal line [105]. When a goal is scored, the system sends an instant signal to the referee's watch, usually within a second, ensuring an immediate and unequivocal decision [157].

What makes goal-line technology particularly effective is its binary nature: the ball either crosses the line or it doesn't, leaving little room for interpretation [56]. As a result, GLT has successfully removed one of football's most contentious issues, famously highlighted in high-profile incidents such as Frank Lampard's disallowed goal in the 2010 FIFA World Cup. Beyond enhancing fairness, GLT has also preserved the game's rhythm by no more long talks or back-and-forths. Its reliability and automation offer referees peace of mind and spectators a sense of justice, reinforcing the integrity of match outcomes [105, 157]. Although the technology requires significant financial investment, particularly in camera installation and calibration, its benefits have outweighed the costs in most top-tier leagues and international tournaments. Goal-line technology really shows how important accuracy is when it comes to making the right call, purpose-specific technological innovations can enhance officiating without compromising the spirit or tempo of the game.

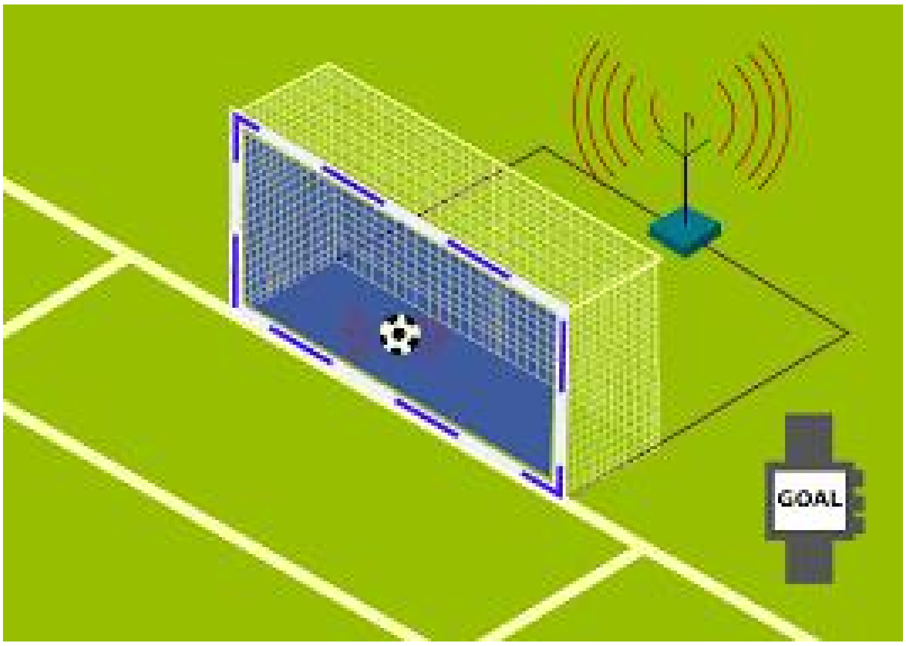


Figure 1.2: Goal Line Technology Example applied [30]

1.6 AI-Powered Refereeing Decision Analysis

1.6.1 Computer Vision for Instant Replay Analysis

Computer vision has become a pretty useful part of modern football refereeing, especially when it comes to instant replay reviews. This branch of artificial intelligence enables systems to process and interpret visual information in a manner analogous to human sight—but with significantly enhanced precision, speed, and objectivity. In football, computer vision algorithms are used to analyze video footage frame-by-frame in real time, identifying key events such as fouls, handballs, offsides, or goal-line crossings [62, 73, 79]. By leveraging convolutional neural networks (CNNs) and deep learning architectures, these systems can detect player positions, track ball movement, and measure distances with sub-second accuracy. These features are especially useful in Analyzing replay footage during heated moments helps referees see clear visual data, so they can make smarter, more confident calls. [15, 102].

Computer vision really shines when it comes to clearing up confusion in fast-moving situations that can trip up humans, like quick tackles in a game, crowded penalty box incidents, or borderline offside calls. Systems can automatically generate heatmaps, movement vectors, and even skeletal body models of players to highlight contact points or illegal positioning [37, 119]. These visual overlays offer clarity and evidence for officials reviewing incidents via the Video Assistant Referee (VAR) system. Additionally, computer vision enhances transparency by making it possible to share visual insights with fans and broadcasters, contributing to a more informed and engaged audience [89, 109]. With continued advancements in real-time processing and 3D modeling, computer vision is poised to become an indispensable pillar in creating a more accurate and consistent refereeing ecosystem.

1.6.2 Predictive Models for Officiating Support

Beyond replay analysis, artificial intelligence is also being utilized to develop predictive models that support officiating decisions before or as incidents unfold. These models use historical match data, player behavior patterns, and contextual variables (such as speed, proximity, angle of approach, and match context) to predict the likelihood of controversial or foul-inducing situations [33, 169]. By training machine learning algorithms—like random forests, gradient boosting machines, or recurrent neural networks—on large annotated datasets, these systems can identify trends that human referees might miss in real-time [21, 42]. For example, an AI model might recognize a high-risk tackling pattern based on a defender's past movements or anticipate a potential dive by analyzing deceleration curves and fall trajectories [63, 114].

Such predictive tools are not designed to replace referees, but to augment their judgment by providing additional context or alerting them to high-risk scenarios. During live matches, these models can function in the background, flagging suspicious movements or foul probability spikes to the VAR team or fourth official, thereby acting as a second line of oversight [96, 148]. Additionally, predictive officiating support tools can play a critical role in post-match review and referee training. By analyzing decisions retrospectively and comparing them to AI-generated expectations, football associations can enhance referee education, reduce future errors, and ensure greater consistency in decision-making across different leagues and competitions [107, 164].

However, while the promise of AI in refereeing is immense, its integration must be handled with care. The interpretability of AI decisions, model bias due to imbalanced datasets, and the ethical implications of delegating critical decisions to machines are all active areas of concern [12, 78]. Transparency in model design, regular validation, and collaboration between AI developers and refereeing bodies are essential to ensure that predictive systems remain fair, explainable, and aligned with the sport's core principles. As the technology matures, it is likely that predictive AI tools will become a standard part of football's officiating infrastructure, helping to usher in a new era of precision and professionalism [125, 144].

1.7 Advanced Match Analytics

1.7.1 Heat Maps, Pass Networks, and Positional Data

Advanced match analytics have revolutionized how football is understood, coached, and played. Among the most significant innovations are the use of heat maps, pass network diagrams, and positional data to provide rich, visual insights into player behavior, team coordination, and tactical execution [22, 69, 70]. Heat maps show where a player moves and the parts of the field they're most active in, making it easy to see which areas they spend the most time in. These visualizations can indicate whether a player effectively covered their assigned role or drifted into other tactical zones, thereby aiding coaches in performance evaluation and match preparation [14, 127]. For instance, full-backs with high heatmap activity in the attacking third may suggest a team's emphasis on overlapping runs and width exploitation.

Pass network analysis, on the other hand, maps out passing interactions among team members using nodes and edges. Each player is represented as a node, and passes between them form the connecting edges [48, 128]. This method reveals patterns like key passing triangles and which players tend to stay in control of the ball more often, and the overall fluidity of a team's ball circulation. Such insights are crucial for evaluating whether a team maintained compactness, over-relied on specific players, or failed to penetrate key zones. Furthermore, positional data, collected through GPS and optical tracking systems, enables analysts to reconstruct player and ball movements with centimeter-level accuracy [104, 145]. By aligning this data with game phases—such as transition, buildup, or pressing—coaches and analysts can assess formation integrity, spacing, and off-ball dynamics [112].

Collectively, these tools provide a multidimensional understanding of football beyond traditional statistics like goals, assists, or possession percentage. They allow clubs to diagnose problems, personalize training, and simulate different match scenarios using real evidence, which is invaluable for competitive performance optimization in modern football [135, 151].

1.7.2 AI and Strategy Optimization

Artificial Intelligence has really changed how people study football. It uses stuff like predictive models and optimization tricks to help teams decide on tactics and lineups. [23, 44]. AI can analyze vast datasets encompassing player performance metrics, opposition tendencies, weather conditions, fatigue levels, and even psychological profiles to suggest the optimal starting eleven or tactical formation for a given match. These strategy optimization models go beyond static statistics and integrate machine learning techniques—such as reinforcement learning, genetic algorithms, and neural networks—to simulate countless game scenarios and identify the most effective configurations for desired outcomes [45, 95].

One powerful application is predictive tactical simulation, where AI models forecast how a particular formation or playing style will perform against a specific opponent [149]. For instance, if a team is playing against a high-pressing side, AI tools might recommend a more direct passing strategy or a formation with more compact midfield lines to resist pressure. Additionally, AI-driven tools can analyze substitution patterns and fatigue data to recommend real-time lineup changes that maximize team performance while minimizing injury risk [138]. These systems help coaches make evidence-based decisions that enhance both short-term match success and long-term player development [29].

AI is also instrumental in opponent analysis. By studying thousands of hours of match footage and historical data, AI can detect weaknesses in rival teams—such as vulnerability to through balls, poor aerial duels, or uncoordinated pressing triggers—and suggest tailored counter-strategies [43]. When combined with real-time data collection during matches, AI becomes a dynamic assistant capable of updating tactical advice on the fly, ensuring that the coaching staff can react swiftly to unfolding events on the pitch [85].

Despite these advancements, AI-based strategy tools are not meant to replace human intuition and experience. They act as helpful tools that give decision-makers strong, data-driven insights to back up their choices. As the integration of AI deepens, clubs with the resources and expertise to leverage these technologies are likely to maintain a significant strategic edge in the increasingly data-driven landscape of elite football [77, 136].

1.8 The Impact of Technology on the Fan Experience

1.8.1 Augmented Reality and Immersive Viewing

The intersection of technology and the fan experience has ushered in a new era of immersive viewing and augmented reality (AR), transforming how supporters engage with football beyond the traditional matchday [126, 141, 165]. Augmented reality allows fans to experience live games through enhanced visual overlays, where real-time data such as player statistics, tactical formations, and game statistics are superimposed onto the live footage [38, 100]. This immersive experience goes beyond mere visual aesthetics; it offers fans an interactive and educational layer that enhances their understanding of the game [17, 155]. For instance, viewers might see the positions of key players in real time, offering a deeper appreciation of tactical movements, off-ball runs, or defensive structures. AR can also bring in things like fun, moving graphics, so fans can interact with player profiles and get more involved, stats, or replay key moments in 3D from multiple angles, all directly within their viewing interface [35].

For football clubs and broadcasters, AR offers a unique opportunity to captivate audiences by providing custom content tailored to individual preferences [54, 146]. Fans can select specific data points, such as a particular player's heatmap or the number of successful passes in a half, to deepen their involvement with the match [83, 167]. In stadiums, AR can enhance the live experience by offering fans personalized content directly through their mobile devices or even stadium AR glasses [35, 155]. The physical environment of the stadium itself can also be augmented with information-rich displays, guiding fans to their seats, offering venue-related details, and even facilitating interactive engagement during downtime, such as half-time or breaks in play [126].

The rise of virtual reality (VR) also complements this shift by offering entirely immersive environments, where fans can watch live or recorded matches as if they were physically present on the pitch [100, 141]. VR headsets simulate the stadium environment, offering a 360-degree view of the match, thus replicating the intensity of being at the ground [38, 165]. Such technologies are especially beneficial for fans who cannot attend matches in person due to geographical or financial constraints, as they provide an affordable and engaging alternative to in-person experiences [17, 54].

1.8.2 Data-Driven Engagement Platforms

As football becomes increasingly data-centric, data-driven engagement platforms are reshaping how fans interact with the sport [58, 129]. These platforms leverage the power of data analytics, AI, and machine learning to offer personalized content, insights, and interaction channels for supporters [92, 101]. Through mobile apps and digital platforms, fans can access real-time performance metrics, detailed player statistics, and tactical breakdowns—many of which were once confined to coaches and analysts—thus elevating the level of interaction between the game and the viewer [54, 165].

These platforms allow fans to track not only the outcome of games but also the finer aspects of individual and team performance, such as distance covered, passing accuracy, and even emotional sentiment through social media analysis [58, 101]. By aggregating data from various sources, these platforms create a rich, personalized experience where fans can receive notifications about their favorite teams or players, suggest tailored content based on their viewing habits, and participate in real-time polls or trivia [83, 92]. Furthermore, social features such as fan forums, live discussions, and voting platforms enable a deeper sense of community engagement [129, 146]. Fans can interact with each other, share insights, and express opinions during the game, further blurring the lines between passive spectatorship and active participation [58, 83].

Beyond engagement, data-driven platforms also foster fan loyalty by incorporating gamification features [92, 101]. Fans can earn rewards for engagement, such as exclusive content, merchandise, or virtual tokens, creating a more interactive and rewarding experience [54, 129]. Fantasy football leagues, driven by performance data, offer another avenue for fans to engage in a more strategic manner, where they can build their dream teams and compete against others in a virtual setting [58, 146]. These platforms use predictive models to forecast future performances, keeping fans engaged with their fantasy teams and players even during the off-season [92, 101].

Moreover, brands and sponsors increasingly use data to tailor advertisements and sponsorships to specific audience segments [83, 129]. By tracking user preferences and behavior, they can deliver targeted, relevant ads that align with a fan's interests, enhancing both the fan experience and commercial opportunities [54, 101]. This shift in how sponsorships and advertisements are presented makes the fan experience more relevant and personalized, as opposed to the one-size-fits-all approach traditionally seen in sports marketing [58, 146].

In summary, data-driven engagement platforms provide an all-encompassing digital ecosystem where fans can interact, analyze, and immerse themselves in football like never before [92, 165]. These platforms empower fans by giving them access to unprecedented amounts of data, allowing them to make more informed decisions, engage with their favorite teams in new ways, and participate in the football community on a deeper level [83, 101].

1.9 Ethical, Financial, and Cultural Challenges

1.9.1 Data Privacy and Surveillance in Football

As football increasingly integrates advanced technologies such as AI, wearable devices, and performance analytics, concerns surrounding data privacy and surveillance have emerged [27, 65, 116]. The rise of data-driven solutions in football often involves the collection of personal and sensitive information about players, coaches, and even fans. Wearable sensors and tracking devices gather vast amounts of data, ranging from physical performance metrics to health information. While this data can be used to enhance performance and prevent injuries, it also raises ethical questions about how the information is stored, who controls it, and how it is used [113, 161].

For players, the constant monitoring of their physical movements, health status, and personal data could lead to privacy violations if mismanaged [13]. The notion of athletes being under continuous surveillance can lead to a lack of autonomy and could infringe upon their personal rights. There is also the risk of data breaches, where sensitive health information, including injury history or performance-related data, could be exposed to third parties without consent [81]. In addition, as football clubs and organizations gather increasing amounts of data, questions arise regarding the ownership of such information and its potential for exploitation, such as selling data to external advertisers or using it for financial gain without the explicit consent of the players [28].

For fans, the digitalization of football through mobile apps and fan engagement platforms may lead to a more invasive form of data collection [65, 161]. Advertisers and sponsors often gather data on fan preferences, viewing habits, and interactions with digital content, raising concerns over the tracking and sale of personal data without full transparency or control. As such, the issue of data ethics in football is complex, balancing the potential benefits of personalized experiences with the right to privacy [116].

1.9.2 The Economic Divide in Access to Technology

While technology has the potential to revolutionize football by improving performance analysis, refereeing accuracy, and fan engagement, its adoption is not without significant financial barriers [94, 113, 166]. The economic divide between wealthy and less financially endowed football clubs, leagues, and countries has been a persistent challenge. Clubs in the wealthiest leagues, such as the English Premier League or La Liga, are able to afford cutting-edge technologies such as VAR, advanced tracking systems, and AI-powered performance analytics. These technologies enhance their competitiveness and allow them to stay ahead of rivals [13, 81].

However, smaller clubs, especially those in lower divisions or in developing footballing nations, often lack the financial resources to implement these technologies [166]. This naturally causes the game quality to become uneven, where wealthier clubs enjoy superior technological advantages, while others struggle to compete at the same level. The economic disparity extends beyond the clubs themselves to the broader football ecosystem, including training facilities, broadcasting capabilities, and even access to professional staff capable of implementing these technologies [27, 94].

In addition, the growing reliance on technology may deepen the divide between traditional and emerging football markets [116]. For instance, smaller or less established leagues may find it increasingly difficult to attract top-tier talent or gain commercial sponsorship if they are unable to adopt the same level of technological innovation as their wealthier counterparts. The global commercialization of football further exacerbates this gap, where financial resources play a more decisive role than sporting talent in ensuring competitive success [161].

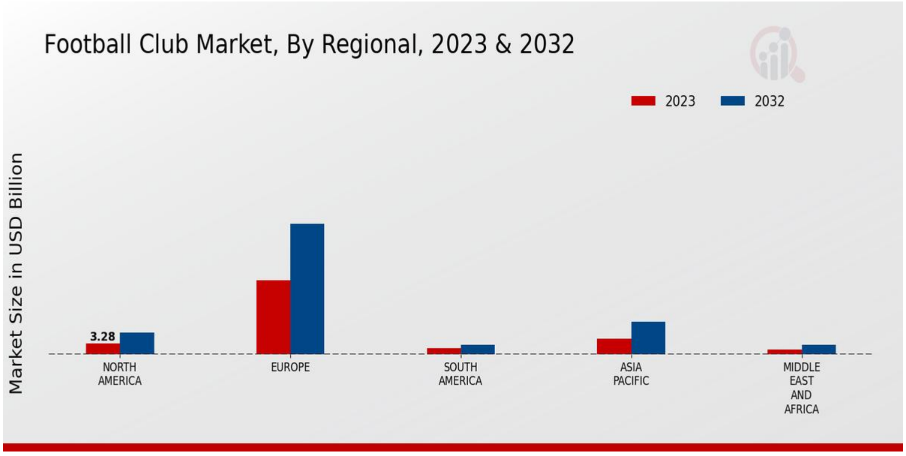


Figure 1.3: Football Club Market By Regional Line Technology Example applied [110]

1.9.3 Preserving the Human Essence of the Game

One of the major concerns in the increasing role of technology in football is the potential loss of the human essence of the game [94, 113]. Football has always been celebrated for its unpredictability, emotion, and human connection, qualities that have made it a beloved sport worldwide. As technologies such as AI, machine learning, and automation become more ingrained in the game, there are fears that the traditional spirit of football could be undermined [13, 65]. The use of AI in officiating, for example, while reducing human error, has sparked debates about whether these systems could remove the human touch from crucial decisions, such as awarding a penalty or interpreting a foul. The reliance on predictive analytics and data-driven strategies in team management might also risk turning football into a more mechanical, formulaic experience, removing some of the creativity and improvisation that define the game [28, 116].

Furthermore, while technologies like VAR aim to reduce errors, they also disrupt the flow of the game, leading to pauses and delays that can affect the emotional highs and lows that fans cherish [81]. Some critics argue that the growing commodification of football through digital platforms, fantasy leagues, and data-driven fan experiences can also alienate traditional fans, who long for a more personal and emotionally rich connection with the sport [161, 166]. This trend toward digitalization, while offering enhanced engagement and interaction, might lead to a detachment from the visceral, human-centric experience that made football so universally appealing [27, 94].

Thus, the human element—the unpredictability of individual talent, the raw emotion of matches, and the atmosphere created by fans—may be at risk in an increasingly technological and commercialized environment [65, 113].

1.10 Conclusion

Today, football has evolved far beyond its traditional identity as a simple sport played on the pitch—it has become a complex global system that weaves together passionate fandom, immense economic potential, and rapid scientific advancement. Over the past few decades, the game has undergone a profound transformation in how it is managed, analyzed, and experienced, particularly with the integration of technology into decision-making processes on and off the field. The evaluation of players' performances and tactical approaches is no longer based solely on human observation; algorithms now provide precise, real-time insights that have redefined how matches are prepared for and played, opening up new strategic frontiers for both clubs and national teams.

Artificial intelligence and machine learning technologies have triggered a revolution in the way training sessions are conducted and performance is assessed. Every movement made by a player can now be tracked, quantified, and interpreted through advanced statistical models that surpass the limitations of human intuition. With the aid of computer vision, video analysis has grown significantly more detailed and actionable, enabling coaching staff to deeply understand not just their own players, but also their opponents. At the same time, referee-assistance technologies—such as the Video Assistant Referee (VAR) system and goal-line technology—have drastically reduced the margin for error and reinforced the game's fairness and credibility.

This convergence of sport and technology is not merely a superficial trend; it reflects a deeper shift in the philosophy of football itself. Where success once relied primarily on innate talent and tactical instinct, it now increasingly depends on a team's ability to harness the power of data, prediction models, and intelligent systems. In this new era, football is being redefined—where scientific insight becomes an integral part of both the beauty and competitiveness of the game. As we continue to witness this transformation, one thing is clear: the future of football will not be shaped solely on the pitch, but also in data centers, analytical labs, and smart platforms that support and elevate the game to new heights.

Chapter 2

Artificial Intelligence and Visual Computing in Football

2.1 Introduction

Artificial Intelligence, or AI, is a part of computer science that focuses on creating systems that can do things usually requiring human thinking, such as learning, reasoning, and decision-making. This is achieved using mathematical models and computational algorithms that mimic human cognitive processes.

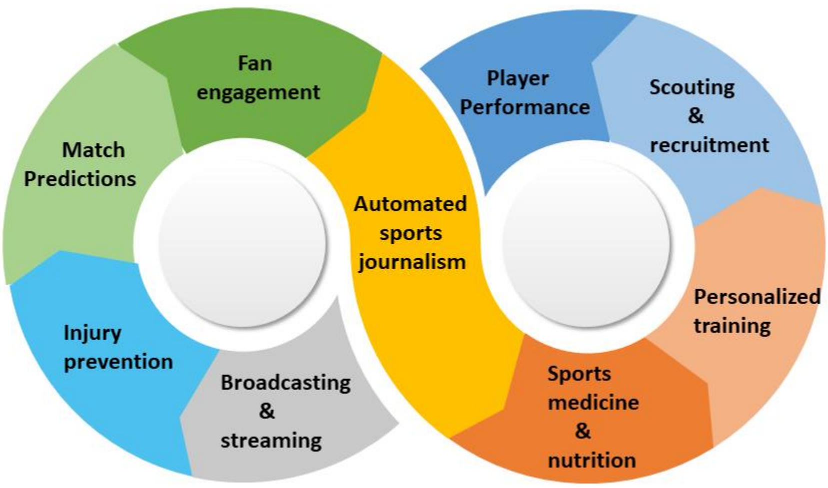


Figure 2.1: Overview of AI components relevant to football analysis

2.2 Applications of Artificial Intelligence

AI has numerous applications in data analysis, including:

2.2.1 Data Mining

AI helps with data mining by looking through huge amounts of data to find patterns and connections that aren't obvious at first. [71]. This is done using algorithms such as clustering, which minimizes the distance between points within each group using the K-Means algorithm [108].

2.2.2 Machine Learning

Machine learning encompasses various approaches to developing predictive models [115]. Below are some key learning paradigms:

Supervised Learning

Supervised learning is a key part of machine learning. It's all about creating models that can make predictions, and it relies on training data that already has the right answers labeled. It is characterized by the presence of input-output pairs, where the objective is to learn a mapping from inputs to outputs that generalizes well to unseen data [24]. The model is trained on a dataset wherein each instance is associated with a correct output label, enabling the learning algorithm to adapt its internal parameters accordingly.

This approach has worked really well in many different real-life situations. In image recognition, for example, supervised learning models are capable of distinguishing between different object categories, such as identifying animals in photographs [91]. In natural language processing, it's what helps with tasks like figuring out how people feel in their writing, like whether comments are positive or negative, where the model is trained to classify text as expressing positive or negative opinions [64]. Financial forecasting, medical diagnosis, and speech recognition are also domains where supervised learning has showed that it worked really well [74].

A notable advantage of supervised learning is its interpretability and predictability when trained with high-quality labeled data. However, its performance is heavily contingent on the availability of large, accurately labeled datasets. Labeling data is often expensive, time-consuming, and subject to human error, which may limit the applicability of supervised approaches in some domains. Nevertheless, supervised learning remains an indispensable tool in machine learning, serving as the foundation for many modern intelligent systems.

Linear Regression

Linear regression is a fundamental technique in supervised learning that seeks to model the relationship between one or more input features and a continuous output variable [74]. Rather than being limited to a single predictor, the general form of a linear regression model can incorporate multiple independent variables, capturing complex relationships through a linear combination of features:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{2.1}$$

where:

- y is the predicted output variable,
- β_0 is the intercept term,
- $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients corresponding to each input feature $x_1, x_2, \ldots, x_n$,
- ϵ represents the random error or noise that accounts for variability not explained by the model.

This generalized model enables the representation of multidimensional data, making linear regression applicable in a wide range of practical domains, including finance, healthcare, and sports analytics.

Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are a class of supervised learning models inspired by the information processing structure of biological neural systems [75]. They are composed of layers of interconnected nodes, or ”neurons”, each of which processes input data by applying a linear transformation followed by a non-linear activation function. This layered structure allows ANNs to learn hierarchical feature representations from raw data, enabling the modeling of highly complex and non-linear relationships.

The architecture of a typical ANN consists of an input layer, one or more hidden layers, and an output layer. Each connection between neurons is associated with a weight that determines the strength of the transmitted signal. During the training phase, these weights are iteratively adjusted using optimization algorithms—most Stochastic gradient descent, especially when used with a loss function that measures how wrong the predictions are.

A central element of ANNs is the activation function, which introduces non-linearity into the model. Without this non-linearity, the network would essentially behave like a linear model regardless of its depth. Common activation functions include the sigmoid, hyperbolic tangent (`tanh`), and rectified linear unit (ReLU). These functions enable the network to capture complex patterns and decision boundaries in the input space.

Unsupervised Learning

Unlike supervised learning, unsupervised learning does not rely on labeled data [24]. The goal is to find the hidden patterns, structure, or trends in the data without expecting specific results beforehand. This type of learning is particularly valuable when annotations are unavailable or expensive to obtain, making it highly applicable in exploratory data analysis and knowledge discovery.

Unsupervised learning algorithms operate by analyzing the intrinsic properties of the data. They attempt to group, associate, or reduce the data in ways that reveal meaningful relationships. Common tasks include clustering, dimensionality reduction, density estimation, and anomaly detection. Through these tasks, models can identify hidden patterns or latent variables that describe the observed data more compactly or coherently.

One of the most widely used unsupervised learning techniques is **K-Means Clustering**, which partitions a dataset into k non-overlapping clusters based on similarity [108]. The algorithm starts by initializing k centroids, then iteratively assigns each data point to the nearest centroid and updates the centroids as the mean of the assigned points. This process continues until convergence, typically defined as minimal changes in the centroids' positions.

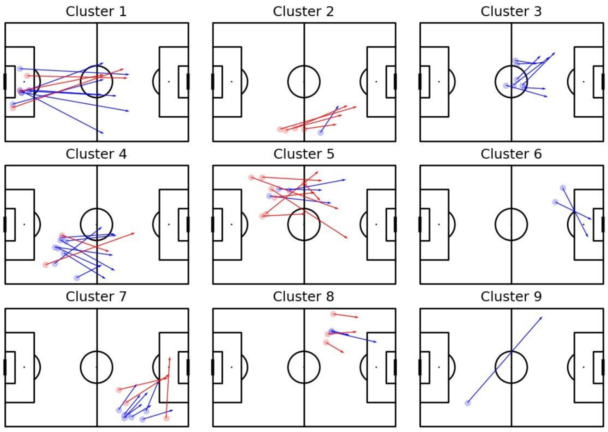


Figure 2.2: Clustering of Manchester United progressive passes

2.2.3 Image and Video Analysis (Computer Vision)

Computer vision enables machines to interpret and process visual data [154]. A key approach in this field is the use of **Convolutional Neural Networks (CNNs)** [98], where models learn patterns through convolutional filters applied to input images.

Convolutional Operation

Given an input image with dimensions:

- $H_{\text{in}} \times W_{\text{in}}$ – Height and width of the input image.
- K – Kernel size (assumed square, i.e., $K \times K$).
- P – Padding size.
- S – Stride (step size of the filter movement).

The output dimensions after applying a convolutional filter are computed as:

$$H_{\text{out}} = \frac{H_{\text{in}} - K + 2P}{S} + 1 \tag{2.2}$$

$$W_{\text{out}} = \frac{W_{\text{in}} - K + 2P}{S} + 1 \tag{2.3}$$

Total Number of Pixels After Filtering

The total number of pixels in the output feature map, considering N_f filters, is given by:

$$\text{Total Pixels} = H_{\text{out}} \times W_{\text{out}} \times N_f \tag{2.4}$$

2.3 Deep Learning

Deep learning is a part of machine learning that uses deep neural networks to find complicated patterns in big sets of data. It uses math ideas and deep neural networks (DNNs) to pick out complex patterns in data. Unlike traditional machine learning, deep learning models are trained with multiple hidden layers, enabling automatic learning of high-level representations [67].

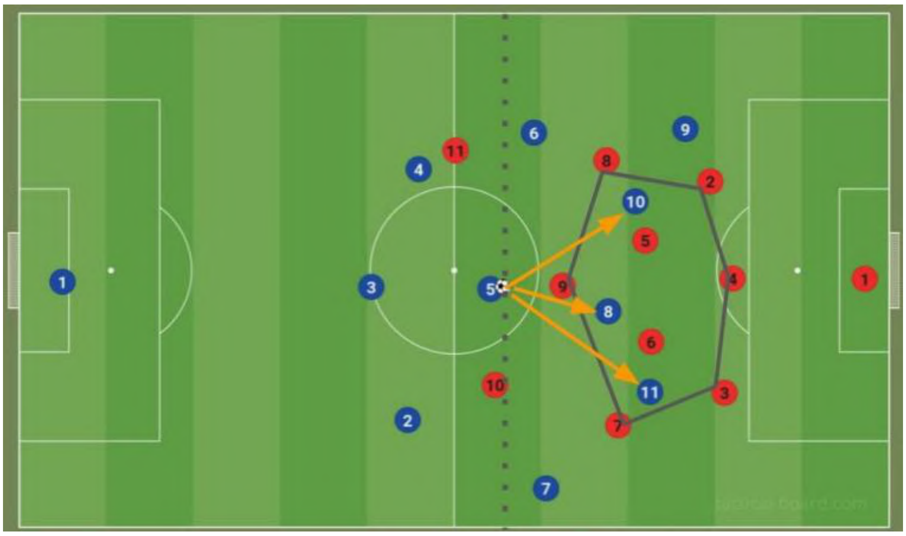


Figure 2.3: Deep learning application for analyzing football tactics [5]

2.3.1 Deep Learning Architecture

A neural network is a mathematical structure inspired by the human brain, consisting of multiple **layers of neurons**, where data propagates through:

- **Input Layer**
- **Hidden Layers**
- **Output Layer**

The output of a neuron is computed as:

$$z = \sum_{i=1}^{n} w_i x_i + b \tag{2.5}$$

$$a = \sigma(z) \tag{2.6}$$

Where:

- w_i are the weights,
- x_i are the inputs,
- b is the bias,
- $\sigma(z)$ is the activation function applied to the weighted sum [120].

Activation Functions

Non-linear functions are used to enhance the model's learning ability [121]:

- **Sigmoid Function**: Suitable for binary classification.
- **ReLU (Rectified Linear Unit)**: Enhances training stability.
- **Softmax Function**: Used for multi-class classification.

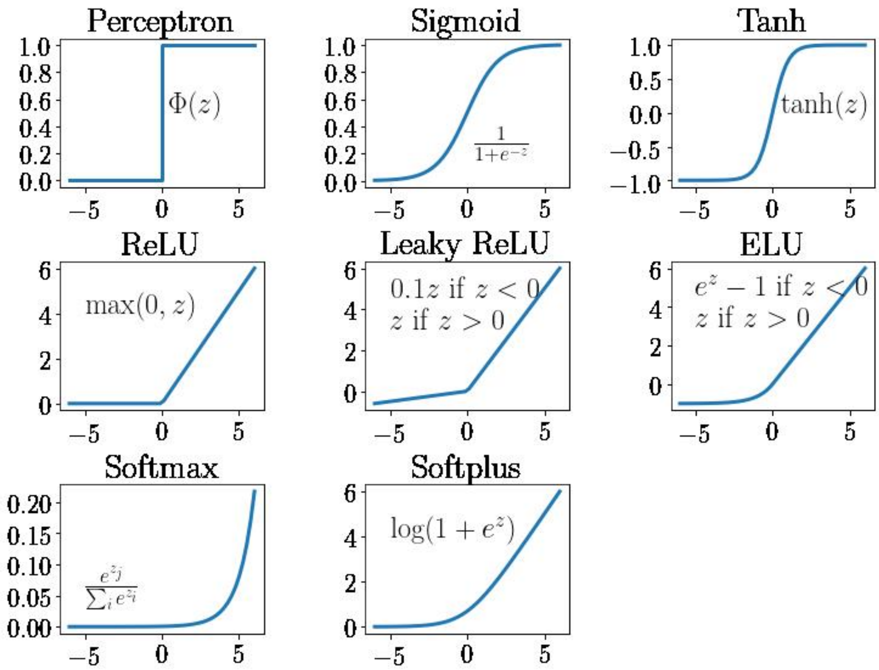


Figure 2.4: Common activation functions used in neural networks [1]

The Softmax function is defined as:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{2.7}$$

Where:

- z_i represents the input to the i-th neuron,
- The denominator ensures the outputs sum to 1.

Training Mechanism in Deep Learning

The learning process involves the following steps [142]:

1. **Forward Propagation**: Computes the predicted output.
2. **Loss Function Calculation**:
   - **Mean Squared Error (MSE)**:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.8}$$

Where:

- y_i is the actual value,
- $\hat{y}_i$ is the predicted value,
- n is the number of samples.

- **Cross-Entropy Loss**:

For binary classification:

$$L = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{2.9}$$

For multi-class classification:

$$L = -\sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}\log(\hat{y}_{ij}) \tag{2.10}$$

Where:

- y_{ij} is the actual class label,
- $\hat{y}_{ij}$ is the predicted probability,
- k is the number of classes.

3. **Backpropagation**: Updates network parameters based on error gradients [143].
4. **Weight Updates Using Gradient Descent Algorithm** [87].

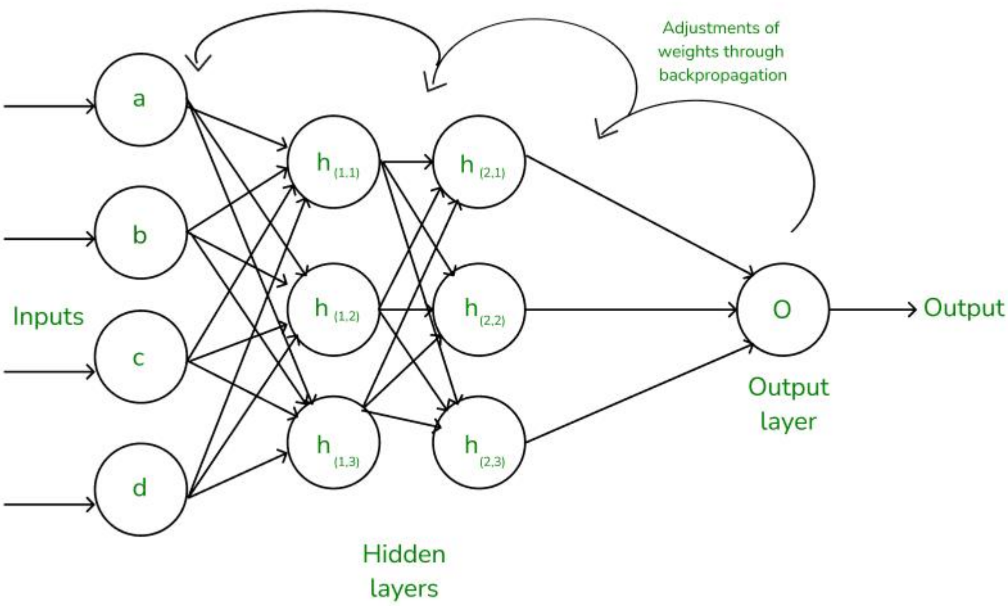


Figure 2.5: Backpropagation process in neural network training [2]

Deep learning models utilize multiple hidden layers to extract complex patterns. Common architectures include [99]:

- **Convolutional Neural Networks (CNNs)** - Used for image analysis.
- **Recurrent Neural Networks (RNNs)** - Used for sequential data and NLP tasks.
- **Transformer Networks** (e.g., BERT, GPT) - Used for natural language processing [162].

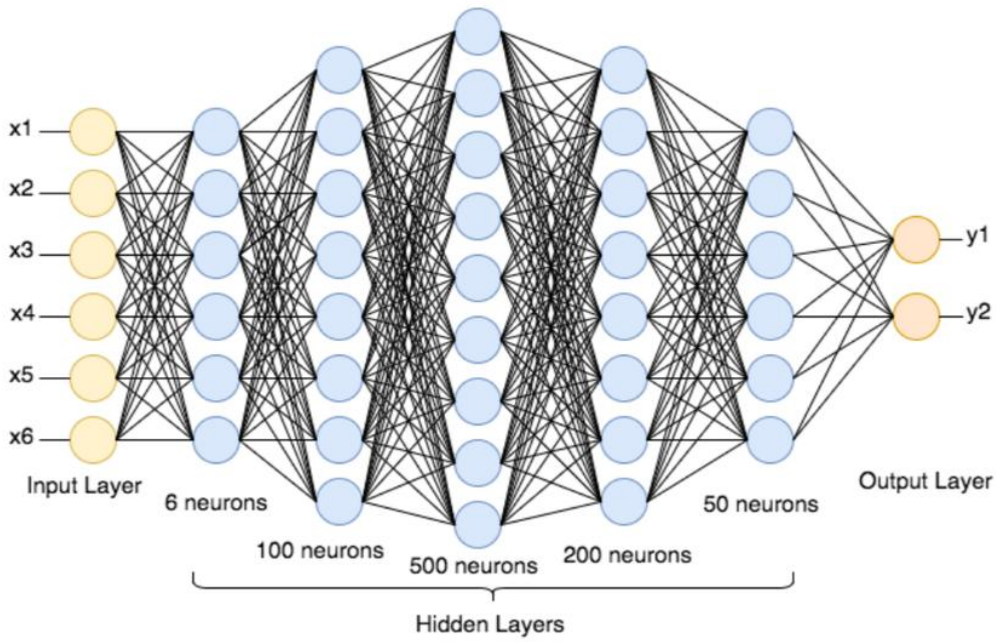


Figure 2.6: deep neural network architectures [6]

Comparison of Deep Learning and Traditional Machine Learning

Table 2.1: Comparison of Deep Learning vs. Traditional Machine Learning [118]

Feature	**Deep Learning**	**Traditional Machine Learning**
Feature Extraction	Learns features automatically	Requires manual feature extraction
Performance with Big Data	Very high performance	Moderate performance
Interpretability	Less interpretable (black box)	More interpretable
Computational Complexity	Higher	Lower
Applications	Computer Vision, NLP	Simple classifications, linear models

2.4 Computer Vision

Computer Vision is a prominent subfield of artificial intelligence (AI) that focuses on enabling machines to perceive and interpret the visual world in a manner similar to human vision [153]. The aim is to build algorithms and models that help computers automatically look at images and videos, understand what's in them, and do things like find objects, recognize them, separate different parts, and keep track of things over time.This domain plays a big role in many areas, like self-driving cars, medical imaging, and video security..

The process of computer vision involves multiple steps, such as image acquisition, preprocessing, feature extraction, and pattern recognition. Initially, raw images are captured by cameras or sensors, which are then processed to remove noise and enhance relevant features. Deep learning methods, particularly convolutional neural networks (CNNs), have proven to be highly effective in learning hierarchical feature representations directly from the raw pixel data [99]. CNNs excel at tasks like image classification, where the goal is to assign a label to an image, and object detection, where the goal is to locate and classify multiple objects within an image.

2.5 Digital Image Processing

A digital image is represented as a matrix $I(x, y)$ with pixel values that indicate the intensity of light at each point,
where:

- x, y represent the pixel coordinates.
- $I(x, y)$ is the pixel value (ranging between 0 and 255 for grayscale images or RGB values for colored images) [66].

2.6 Deep Learning in Computer Vision

Convolutional Neural Networks (CNNs) are used to process images and videos automatically [90].

2.6.1 Convolutional Neural Networks (CNNs)

CNNs rely on convolution operations to extract features from images. The convolution operation is mathematically represented as [50]:

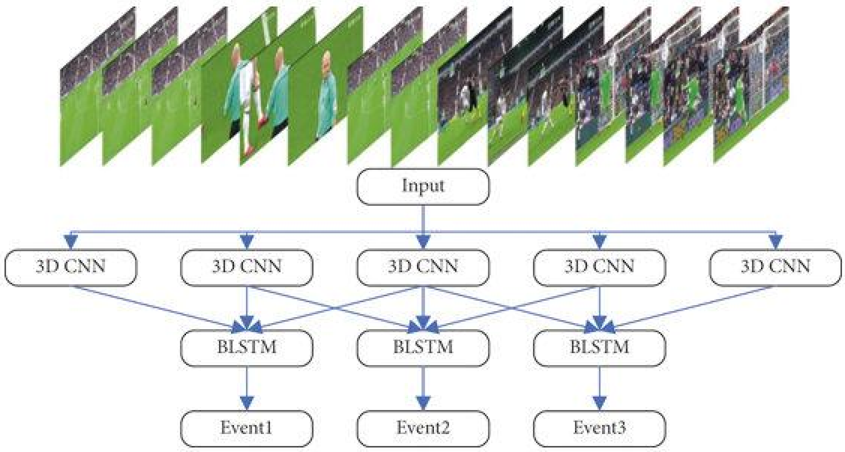


Figure 2.7: CNN architecture used for player detection in football matches [4]

- $H_{\text{in}} \times W_{\text{in}} \rightarrow$ Height and width of the input image.
- $K \rightarrow$ Kernel size (assumed square, i.e., $K \times K$).
- $P \rightarrow$ Padding size.
- $S \rightarrow$ Stride (step size of the filter movement).

The output height and width after applying a convolutional filter are given by:

$$H_{\text{out}} = \frac{H_{\text{in}} - K + 2P}{S} + 1 \tag{2.11}$$

$$W_{\text{out}} = \frac{W_{\text{in}} - K + 2P}{S} + 1 \tag{2.12}$$

2.6.2 CNNs Architecture

- **Convolution Layer**: Extracts features.
- **Pooling Layer**: Reduces data size [147].
- **Fully Connected Layers**: Make final decisions.

- **Training with Backpropagation Algorithm**: Updates the weights [97].

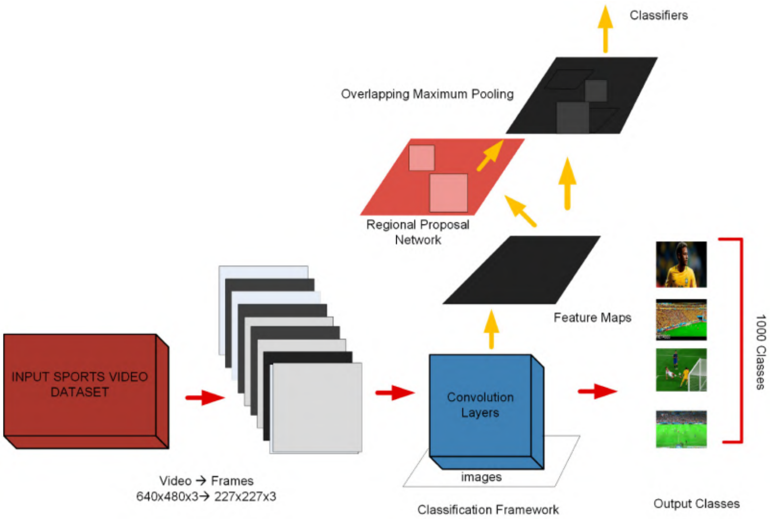
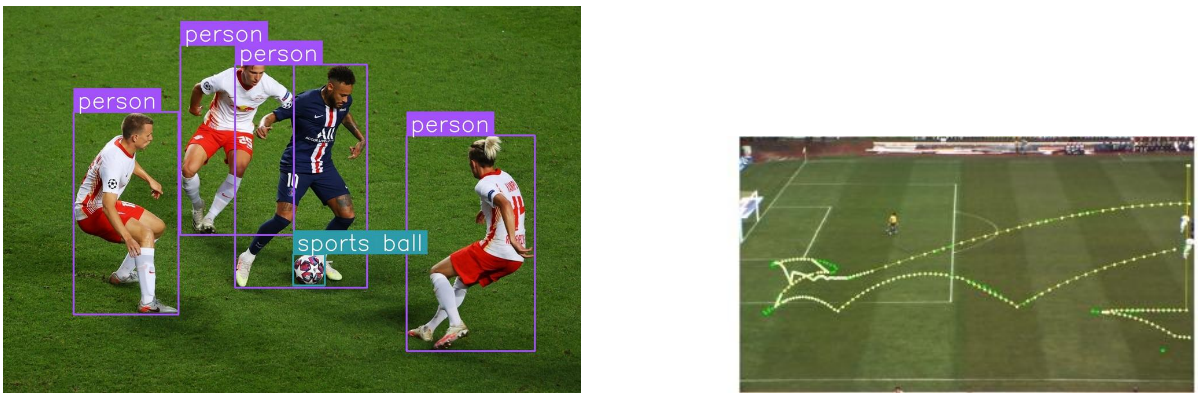


Figure 2.8: Different layers in a CNN used for football video analysis

2.7 Applications of Computer Vision in Sports

- **Player Performance Analysis**: Tracking movements using the *Optical Flow Algorithm* to detect player motion in each frame.
- **Player and Ball Detection**: Using *YOLO (You Only Look Once)* to detect players and analyze the ball's movement by predicting its future trajectory.
- **Smart Refereeing (VAR)**: Utilizing *Deep Neural Networks (DNNs)* to detect offside and fouls, as well as analyzing ball trajectory to make accurate decisions using *Reinforcement Learning*.



(a) Player tracking using computer vision [7]

(b) Ball tracking and trajectory prediction [3]

Figure 2.9: Computer vision applications in football analysis

2.8 R-CNN (Regions with CNN Features)

R-CNN (Regions with Convolutional Neural Network Features) represents a significant advancement in the field of object detection by integrating region proposal mechanisms with powerful convolutional neural network architectures [61]. The core idea behind R-CNN is to first generate a set of candidate region proposals from an input image using selective search. Each of these proposed regions is then resized and passed through a CNN to extract a fixed-length feature vector, which is subsequently classified using a set of class-specific linear SVMs. R-CNN also has a step that tweaks the bounding boxes to make the object locations more accurate.

Despite achieving high accuracy, R-CNN suffers from several drawbacks. The process of extracting features from thousands of region proposals per image is computationally expensive and results in slow inference times, making it impractical for real-time applications. The multi-step training setup, which includes the CNN, SVMs, and box regressors, makes things more complicated and takes longer to train.

2.9 Fast R-CNN

Fast R-CNN was introduced as an improvement over the original R-CNN, addressing its computational inefficiencies and simplifying the training pipeline [60]. In Fast R-CNN, instead of feeding individual region proposals through the CNN separately, the entire image is passed through a CNN to produce a convolutional feature map. Region proposals are then projected onto this feature map, and fixed-length feature vectors are extracted using a region of interest (RoI) pooling layer. These feature vectors are fed into fully connected layers that simultaneously perform classification and bounding box regression.

This change in the design cuts down on how much work the system has to do by avoiding repeated CNN passes over the same overlapping areas. That means training and running the model gets quicker. Fast R-CNN uses one combined loss that handles both telling what an object is and where it's located, all at once, leading to improved detection accuracy and a more streamlined training process.

2.10 Faster R-CNN

Faster R-CNN, proposed by **Shaoqing Ren et al.** in **2015**, represents a breakthrough in object detection by integrating region proposal generation directly into the network architecture [137]. This is achieved through the introduction of a Region Proposal Network (RPN), which shares convolutional layers with the detection network. The RPN efficiently generates high-quality region proposals by sliding a small network over the shared feature maps and predicting object bounds and scores at each position.

Faster R-CNN is much quicker at detecting objects because it no longer needs extra algorithms like selective search to suggest regions. It still keeps a high level of accuracy too. Training both the RPN and Fast R-CNN together in one system makes the learning smoother and helps the whole thing work better.

Faster R-CNN is now a key part of how we detect objects these days and is used in lots of real-world stuff, including **autonomous driving**, where precise and real-time detection of pedestrians and vehicles is critical, and **intelligent surveillance**, where accurate object tracking and recognition are essential for security and monitoring systems.

2.11 Difference Between R-CNN and Faster R-CNN

R-CNN	**Faster R-CNN**
Very slow	Much faster due to RPN
Relies on Selective Search	Uses RPN
Each region needs to pass through CNN separately	Features are extracted only once

Table 2.2: Comparison of R-CNN and Faster R-CNN

2.12 Difference Between Fast R-CNN and Faster R-CNN

- **Fast R-CNN** uses *Selective Search* to generate region proposals, which is computationally expensive.
- **Faster R-CNN** replaces *Selective Search* with a **Region Proposal Network (RPN)**, significantly improving speed.

2.13 YOLO (You Only Look Once): Speed and Accuracy

YOLO, introduced by **Joseph Redmon et al.** in 2016 [134], revolutionized object detection with real-time processing and good accuracy. The model divides the image into a grid, predicts bounding boxes, and selects the best boxes based on confidence scores.

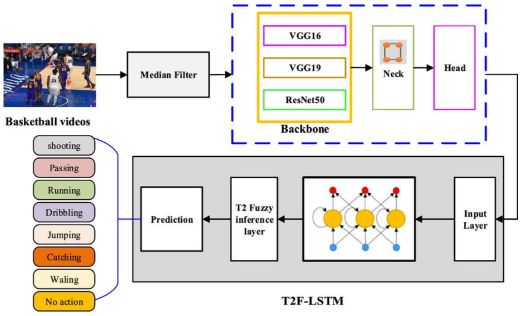


Figure 2.10: YOLO architecture for real-time player and ball detection [9]

2.13.1 Confidence Score Calculation

The confidence score for each predicted bounding box is given by:

$$C = P(\text{object}) \times \text{IOU}(\text{predicted box}, \text{ground truth box}) \tag{2.13}$$

Where:

- $P(\text{object})$ is the probability that an object is present in the cell.
- IOU (Intersection over Union) measures how much the predicted bounding box overlaps with the ground truth box.

2.13.2 YOLO Loss Function

The loss function in the YOLO (You Only Look Once) object detection framework is a critical component that guides the training process by measuring how well the predicted outputs match the ground truth. It is carefully designed to handle multiple tasks simultaneously, including object localization, confidence estimation, and classification. Rather than using separate models for detection and classification, YOLO combines them into a single neural network architecture, which requires a compound loss function to optimize all components jointly [134].

The YOLO loss function comprises several distinct components. The first is the localization loss, which penalizes errors in the predicted bounding box coordinates, including the position and size of the box. This ensures that the predicted boxes align accurately with the actual objects in the image. The loss gives more importance to boxes that are responsible for detecting objects, using a special indicator that activates only when an object is present in a given grid cell.

Secondly, the loss function includes a confidence loss, which reflects how certain the model is about the presence of an object in each predicted bounding box. If a box is not responsible for detecting an object, a different term with lower weight is used to avoid over-penalizing the network for predicting objects where there are none [132]. This separation of object and non-object confidence losses helps balance the network's learning and prevents it from becoming biased toward the majority class (usually the background).

Lastly, the classification loss measures how accurately the model predicts the correct class for each detected object. YOLO predicts a probability distribution over all possible object classes for each box that is responsible for detection, and this component penalizes deviations from the correct class labels.

To achieve optimal performance, YOLO introduces specific hyperparameters that balance the contribution of each component. For example, it assigns greater weight to localization errors to prioritize precise bounding box placement, and reduces the weight for boxes with no objects to mitigate the effect of background clutter [26].

Overall, the YOLO loss function plays a crucial role in enabling real-time object detection by optimizing for both speed and accuracy in a unified framework. This balance has made YOLO a popular choice for applications such as autonomous vehicles, real-time video analysis, and sports analytics, where detecting multiple objects efficiently is essential.

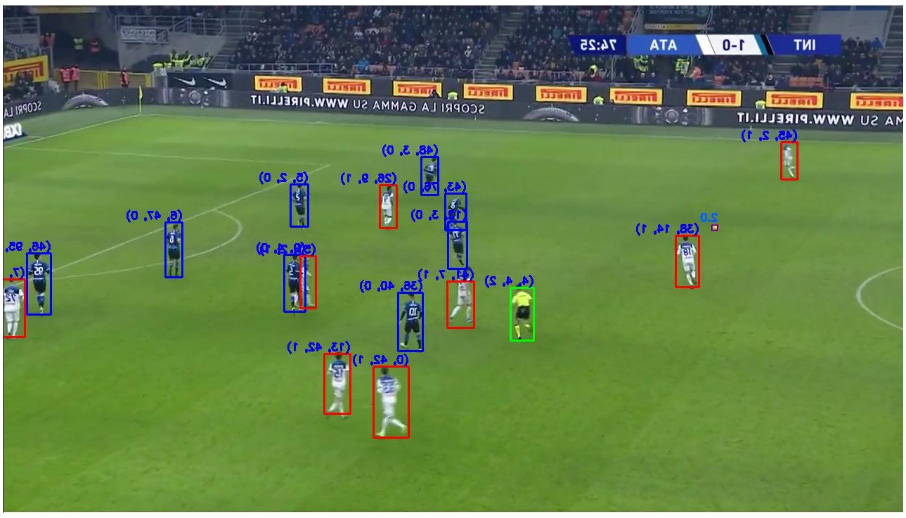


Figure 2.11: YOLO detection results on a football match frame [10]

2.14 DETR (Detection Transformer)

DETR (DEtection TRansformer) was introduced in **2020** by **Nicolas Carion et al.** at **Facebook AI Research (FAIR)** [36]. It is one of the revolutionary models that use **Transformer** technology instead of traditional **CNN** networks for object detection in images and videos.

The model relies on the concept of **Self-Attention**, as used in models like **BERT and GPT**, allowing it to learn directly from images without traditional mechanisms like **RPN or Anchor Boxes**.

2.14.1 DETR Architecture

1. **Feature Extraction**: The image is passed through a **backbone CNN** (such as ResNet-50 or ResNet-101) to extract high-level representative features.

2. **Sequence Representation**: The extracted image features are converted into a **sequential representation** used as input for the Transformer.

3. **Transformer Encoder-Decoder**: The **Self-Attention mechanism** is applied to determine relationships between different image features.

4. **Decoder**: A set of **Object Queries** is used to identify object locations without the need for **region proposal mechanisms (RPN)** like in Faster R-CNN.

5. **Object Detection Head**: A classification and bounding box regression head is used to generate **Bounding Boxes** and object categories.

DETR uses a **Matching Loss function** to determine the alignment between predictions and ground truth, which is a combination of **GIoU loss** for bounding boxes and **Cross-Entropy loss** for classification.

2.14.2 DETR Loss Function

The fundamental loss function equation used in **DETR (DEtection TRansformer)** is based on **Hungarian Matching** and combines **GIoU loss** for bounding boxes with **Cross-Entropy loss** for classification. The loss function can be formulated as follows:

$$\mathcal{L} = \sum_{i=1}^{N} \left[\mathcal{L}_{\text{cls}}(c_i, \hat{c}_{\sigma(i)}) + \lambda_{\text{box}} \mathcal{L}_{\text{box}}(b_i, \hat{b}_{\sigma(i)}) + \lambda_{\text{giou}} \mathcal{L}_{\text{giou}}(b_i, \hat{b}_{\sigma(i)}) \right] \tag{2.14}$$

where:

- N is the number of objects in the ground truth.
- $\sigma(i)$ represents the optimal assignment of predictions to ground truth objects using the **Hungarian algorithm**.
- $\mathcal{L}_{\text{cls}}$ is the **Cross-Entropy loss** for classification.
- $\mathcal{L}_{\text{box}}$ is the **L1 loss** for bounding box regression.
- $\mathcal{L}_{\text{giou}}$ is the **Generalized IoU loss** for bounding boxes.
- c_i and $\hat{c}_{\sigma(i)}$ are the ground truth and predicted class labels, respectively.
- b_i and $\hat{b}_{\sigma(i)}$ are the ground truth and predicted bounding boxes, respectively.
- λ_{box} and λ_{giou} are hyperparameters to balance the losses.

This loss function ensures proper **classification, localization, and alignment of bounding boxes** with the ground truth.

DETR removes the need for **RPN and Anchor Boxes**, simplifying training and helping learn long-range object relationships in an image through the **Self-Attention** mechanism. It provides strong performance on datasets like **COCO**, but at a lower speed compared to YOLO.

2.15 ViTDet (Vision Transformer for Detection)

ViTDet was introduced in **2022** as an enhancement of **Transformers** for **computer vision tasks** [103]. It primarily relies on **Vision Transformer (ViT)** instead of **CNNs** for feature extraction, allowing for more efficient processing and better utilization of global context in images.

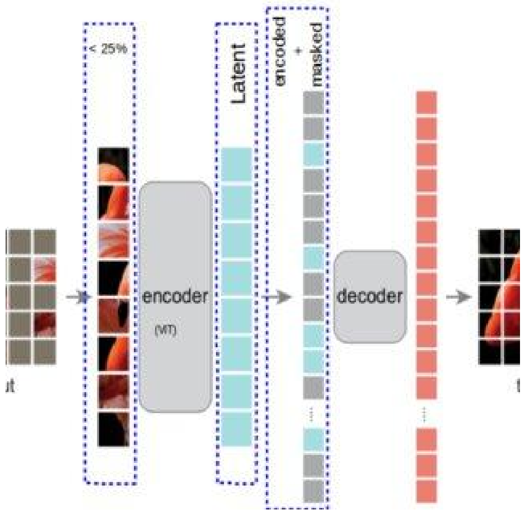


Figure 2.12: ViTDet architecture for advanced football scene understanding [8]

2.15.1 ViTDet Architecture

- Uses **ViT for feature extraction** instead of **CNNs**, providing higher accuracy for complex images.
- **Applies Transformer Encoder-Decoder**, similar to DETR, but with improvements in **segmentation and network adaptation** for object detection data.
- **Object prediction** is performed using a classification and detection head based on ViT features, with **FPN (Feature Pyramid Network) enhancements** to improve accuracy.

2.15.2 Self-Attention Equation

ViTDet introduces **Multi-Scale Self-Attention**, allowing processing at different levels of detail. The equation for Self-Attention is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2.15}$$

where:

- Q (Query), K (Key), and V (Value) are the input feature representations.
- d_k is the dimension of the key vectors.
- The **softmax** function normalizes attention scores.

This mechanism allows ViTDet to achieve **better feature representation** across different spatial scales.

2.16 Full Computer Vision Models Comparison

maincolor!10 **Model**	**Year**	**Accuracy (mAP)**	**Speed (FPS)**	**Complexity**	**Stages**	**Region Proposal Method**	**Backbone**	**Applications**
R-CNN	2014	~66% (VOC)	0.02 - 0.1	Very High	Multi (3)	Selective Search	CNN (AlexNet)	Image Analysis, Object Recognition
Fast R-CNN	2015	~70% (VOC)	1-2	Medium	Two-stage	Selective Search	CNN (VGG-16)	Image Analysis, Computer Vision
Faster R-CNN	2015	~76-80% (COCO)	5-17	Lower	Two-stage	RPN	CNN (ResNet)	Autonomous Driving, Surveillance
YOLOv1	2016	~63% (VOC)	45	Low	Single-stage	Grid-based	CNN (Darknet-19)	Real-time Applications
YOLOv3	2018	~57.9% (COCO)	30-60	Medium	Single-stage	Grid-based	CNN (Darknet-53)	Video Surveillance
YOLOv5	2020	~50-56% (COCO)	60-140	Very Low	Single-stage	Grid-based	CNN (CSPDarknet)	Robotics, Drones
YOLOv8	2023	~53-60% (COCO)	120-180	Very Low	Single-stage	Grid-based	CNN (EfficientNet)	Real-time Applications
DETR	2020	~45-50% (COCO)	15-30	Very High	Single-stage	Transformer	Vision Transformer	Autonomous Driving
ViTDet	2022	~55-60% (COCO)	10-25	Very High	Single-stage	Transformer	Vision Transformer	Advanced Vision Applications

Table 2.3: Comparison of Object Detection Models

2.17 Conclusion

The evolution of object detection models has shown remarkable progress over the past decade. From the early R-CNN models to the latest transformer-based architectures like DETR and ViTDet, each advancement has brought improvements in accuracy, speed, or both. While CNN-based models like YOLO continue to dominate real-time applications due to their speed, transformer-based models offer superior accuracy and context understanding for more complex scenarios. The choice of model depends on the specific requirements of the application, with sports analytics benefiting from both approaches depending on whether real-time processing or detailed analysis is prioritized.

Chapter 3

Tracking and Analysis Players in Football Modeling and Implementation

3.1 Introduction

After what we have done of defining modern football and the intervention of technology in its development and modernity more and more and reducing errors in many human decisions and after we studied most of the models of artificial intelligence, delving deeper into computer vision more because it is the focus of our connection of artificial intelligence with football, now we have reached the biggest task, which is the practical application of integrating the contents of the first and second chapters to give results that are good to reduce human errors in football and increase automation in football to achieve positive results for teams in general and players in particular by studying the movements and developments of play during the two halves of the match.

3.2 Tracking of Objects

3.2.1 Semantic Segmentation Problem

Data Collection

The data collection process began with an extensive search for relevant video material on **YouTube**, targeting content closely aligned with the project's objectives. A total of **53 videos** were retrieved from various channels, with durations ranging from **2 to 21 minutes**. Additional videos were also taken from social media**X**, and images were extracted from the **ArchivoVAR** account on the same platform [16].

Used Python's subprocess module, the `yt-dlp` tool was utilized through Python's `subprocess` module [170]. Following data acquisition, the `OpenCV (cv2)` library was employed to extract key frames that depicted critical events relevant to the research [123]. This preprocessing step resulted in a curated dataset comprising **887 images**, distributed across **seven distinct classes**.

Data Annotation

At first, we set up the **CVAT** annotation tool inside a Docker container. This made it easier to do local annotations and try out data augmentation work without worrying too much about cloud storage space [122]. However, due to local hardware storage constraints, the annotation process was ultimately migrated to the **Roboflow** platform, which offered a more flexible and scalable solution [139].

Annotations were performed using the **Instance Segmentation** technique, and each image was labeled according to one of the following seven classes:

- Foul
- Goal
- Hand Touch
- No Foul
- No Goal
- Offside

- Onside

The final dataset consists of images representing diverse football (soccer) events, with file sizes ranging from **12 KB to 23 MB**. The image formats include `JPG` and `PNG`, and the resolution of all images was standardized to **640 × 640 pixels**.

The class distribution is as follows:

- **Foul**: 230 images
- **Goal**: 44 images
- **Hand Touch**: 87 images
- **No Foul**: 121 images
- **No Goal**: 209 images
- **Offside**: 63 images
- **Onside**: 133 images

The dataset was partitioned into training, validation, and testing subsets using the following distribution:

Subset	**Number of Images**
Training	876
Validation	65
Testing	32

Table 3.1: Dataset Split

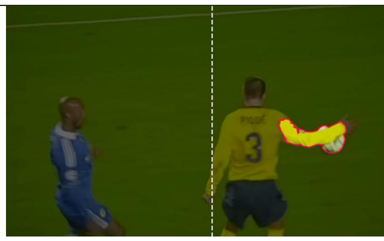
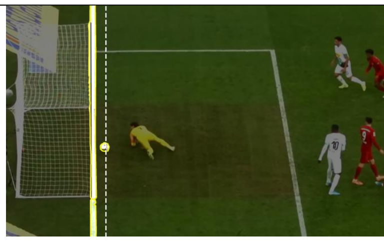
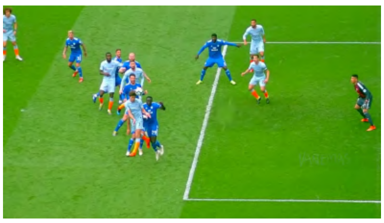
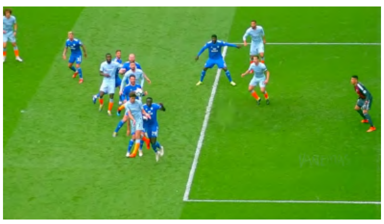
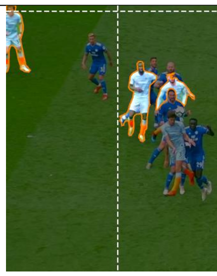
Class Name	Unlabeled	Labeled
Foul		
Goal		
Hand Touch		
No Foul		
No Goal		
Offside		
Onside		

Figure 3.1: Comparison of labeled and unlabeled images for each football event class

YOLOv11 Model

YOLOv11 Instance Segmentation (Accurate) is a state-of-the-art model in object detection and segmentation [47]. It's based on the YOLO (You Only Look Once) design but with big improvements to make it better at segmenting things and faster at making predictions. That makes it ideal for real-time use.

Checkpoint: COCOs-seg

The **COCOs-seg** checkpoint refers to a pre-trained YOLOv11 model trained on the COCO-Seg dataset—an enriched version of COCO that includes more detailed instance segmentation labels [47]. This checkpoint enables high generalization performance across varied segmentation tasks.

Implementation and Usage

YOLOv11 models are publicly available through platforms such as Ultralytics [159] and Hugging Face [53]. The `yolo11n-seg` variant is tailored for resource-constrained environments.

1. **Installation:** Clone the YOLOv11 repository and install the required packages.
2. **Loading the Model:** Use the provided API to load the COCOs-seg checkpoint.
3. **Inference:** Input images are passed to the model to produce segmentation masks.

Performance Metrics

When evaluated on the COCO dataset, YOLOv11 demonstrates outstanding performance across multiple variants [47]:

Variant	**Mean Average Precision (mAP)**
YOLOv11n	39.5
YOLOv11m	51.5
YOLOv11x	54.7

Table 3.2: Performance of YOLOv11 on the COCO dataset

For additional details and updated releases, please refer to the official documentation on the Ultralytics [159] and Hugging Face [53] websites.

Data Augmentation

To enhance the model's robustness and generalization capabilities, a comprehensive data augmentation strategy was applied using modern techniques such as those provided by the Albumentations library [34]. This approach aimed to simulate diverse real-world variations commonly observed in football match footage. The augmentation process was performed iteratively to expand the dataset by a factor of 10, increasing the total number of images from 887 to **8857**.

Augmentation Techniques

- **Horizontal Flip:** Images were flipped horizontally to simulate different camera viewpoints. This variation helps the model learn invariance to player orientation and camera angles.
- **Rotation:** Each image was randomly rotated within the range of **-15° to +15°**, introducing slight angular variations to account for natural camera tilt during gameplay.
- **Grayscale Conversion:** Applied to **13%** of the dataset, this transformation converts images to grayscale to simulate low-color or monochrome footage, enhancing model adaptability to color-deprived scenarios.
- **Saturation Adjustment:** Image saturation was randomly altered within a range of **-25% to +25%**, mimicking different lighting conditions and color correction effects.
- **Gaussian Blur:** A blur effect of up to **2.5 pixels** was applied to simulate motion blur or low-resolution frames, aiding the model in recognizing less distinct object boundaries.
- **Random Noise:** Up to **1.76%** of image pixels were modified to simulate digital noise and compression artifacts typically encountered in video streams.

Augmentation Type	Configuration	Purpose
Horizontal Flip	Applied randomly	Simulates mirrored views for diverse angles
Rotation	-15° to +15°	Emulates camera tilt and perspective variation
Grayscale	13% of images	Enhances model performance under low-color input
Saturation	-25% to +25%	Mimics variable lighting and post-processing
Gaussian Blur	Up to 2.5px	Replicates motion blur and resolution drops
Random Noise	Up to 1.76% of pixels	Introduces real-world noise and distortion
Final Dataset Size	**8857 images**	Achieved via 10x data augmentation

Table 3.3: applied data augmentation techniques and their objectives.

Results

Evaluation Metrics

To assess the model's performance, the following standard metrics were utilized: **Accuracy**, **Recall**, and **mean Average Precision (mAP)** [49, 52, 68].

Accuracy

Accuracy quantifies the proportion of correctly classified instances over the total number of predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively [49].

Recall

Recall evaluates the model's effectiveness in identifying relevant instances of a class:

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

It emphasizes the model's sensitivity, especially in minimizing missed detections [68].

Mean Average Precision (mAP)

mAP is a comprehensive metric for evaluating classification and localization performance across multiple object classes. It is computed as:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i$$

Where N is the number of classes and AP_i is the average precision for class i, typically calculated as the area under the precision-recall curve [52].

Metric Importance

- **Accuracy** provides a general performance overview [49].
- **Recall** is critical for ensuring that relevant events (e.g., goals or fouls) are not missed [68].
- **mAP** offers a detailed evaluation of both detection and segmentation quality across all event classes [52].

Advanced Training Graphs Analysis

To better understand the reasons behind the model's unsatisfactory performance, a series of advanced training graphs were analyzed based on practices from recent state-of-the-art object detection models [26, 163].

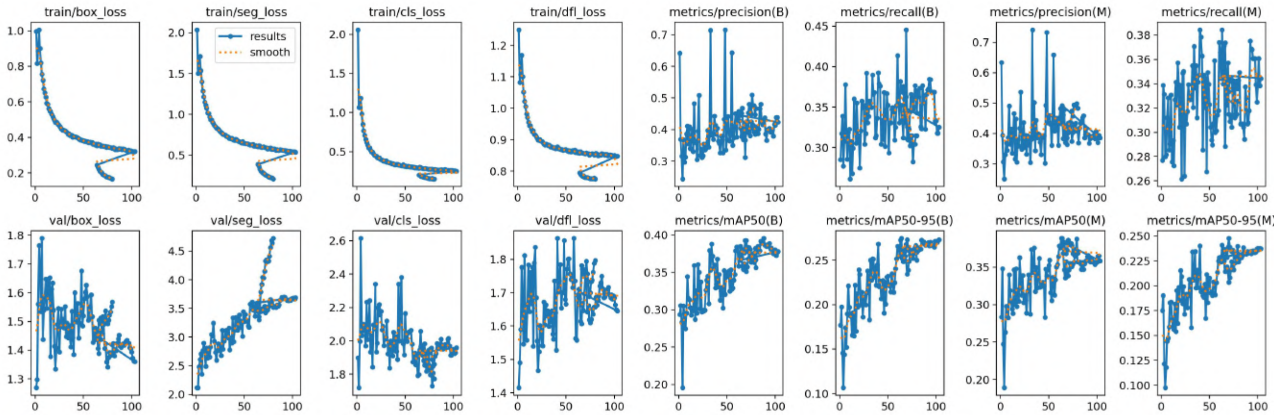


Figure 3.2: Instance Segmentation Results: Training vs Validation Loss Curves.

Loss Curves

The training and validation loss curves showed divergence after a few epochs. Although training loss decreased steadily, validation loss either plateaued or increased, indicating overfitting and lack of generalization [76].

Loss Curves Analysis

The training and evaluation curves of the model over 100 epochs are presented in the figure. The graphs illustrate the behavior of loss functions and key performance metrics to evaluate the quality of the model.

1. Training Loss Curves

- **train/box_loss**: This curve shows a gradual and consistent decrease in the bounding box loss, indicating that the model is effectively learning to improve its predictions for object locations over time [26].
- **train/seg_loss**: The segmentation loss curve shows a clear decline at the beginning, then gradually stabilizes. This suggests that the model has learned the necessary visual representations to accurately define object boundaries.
- **train/cls_loss**: This curve demonstrates a noticeable decrease in classification loss, which indicates an improvement in the model's ability to distinguish between different object classes.
- **train/dfl_loss**: The decreasing trend in this curve reflects improved learning in pixel-level class distribution using the Distribution Focal Loss, indicating better localization accuracy [163].

2. Validation Loss Curves

- **val/box_loss**: The validation bounding box loss exhibits some fluctuations but shows a slight downward trend, suggesting minor instability in generalizing to validation data.
- **val/seg_loss**: This curve shows a gradual increase in segmentation loss during validation, which may indicate the onset of overfitting, as the model struggles to generalize well to the validation set.
- **val/cls_loss**: This curve fluctuates around a nearly constant value after an initial decline, suggesting relative stability in the model's classification ability.
- **val/dfl_loss**: The curve shows some fluctuations but maintains a slight downward trend, reflecting consistent localization performance.

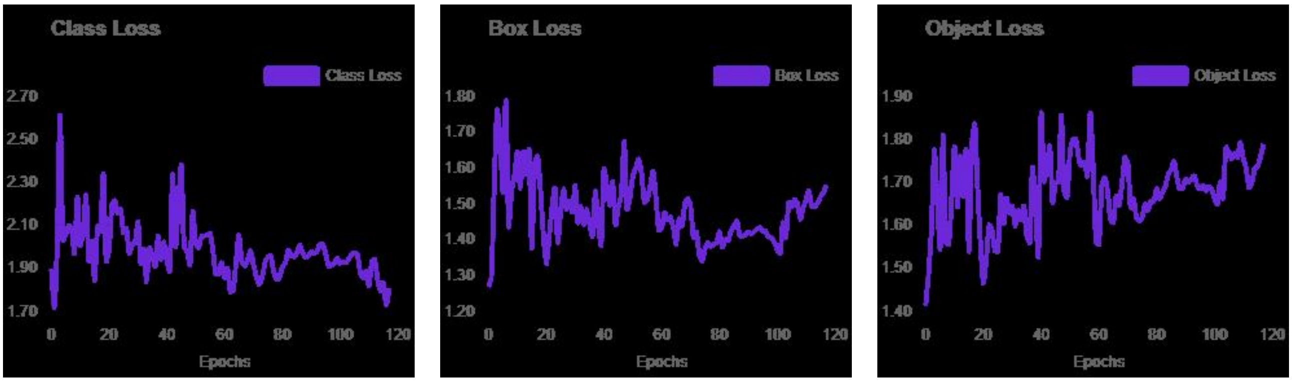
3. Metrics Curves

- **metrics/precision(B) and metrics/precision(M)**: These represent the precision for object types B and M. The curves show a gradual improvement, indicating the model's increasing ability to reduce false positives over time.
- **metrics/recall(B) and metrics/recall(M)**: The recall curves generally trend upwards, reflecting the model's improving capability to retrieve most of the true objects.

- **metrics/mAP50(B) and metrics/mAP50(M)**: These curves represent the mean Average Precision at an IoU threshold of 50%. They show steady improvement, suggesting successful learning.
- **metrics/mAP50-95(B) and metrics/mAP50-95(M)**: This more stringent metric measures performance across a range of IoU thresholds. The curves show a gradual improvement, indicating good learning and generalization [26].

Main Loss Curves (Class Loss, Box Loss, Object Loss)

The training process revealed that the model struggled with proper object localization and classification, as evidenced by the individual loss components. Each type of loss curve provides insight into different aspects of model learning.



(a) Class Loss (b) Box Loss (c) Object Loss

Figure 3.3: Loss Curves per Epoch for YOLOv11: Class, Box, and Object Loss.

Main Loss Curve Analysis

Figure 3.3 presents the loss curves for classification, bounding box, and objectness losses during 120 epochs. These curves reveal model performance patterns consistent with other transformer-based or attention-based object detectors [31].

(a) Classification Loss

The classification loss curve shows initial learning then fluctuates, suggesting sensitivity to outliers or data imbalance.

(b) Bounding Box Loss

Box loss oscillates initially, then stabilizes, implying the model gradually adapts to complex spatial variations in the dataset.

(c) Objectness Loss

Persistent fluctuation of objectness loss suggests challenges in differentiating object presence versus background noise [163].

Training mAP per Epoch

The model didn't perform very well during training, with a low average precision. This shows it struggled to correctly identify and separate the different objects in the categories we set. The mAP remained consistently low, reinforcing the conclusion that the model was not effectively learning from the data.

Training mAP per Epoch Analysis

Figure 3.4 shows that mAP@0.5 rises quickly then stabilizes around 0.36–0.38, while mAP@[.5:.95] improves more slowly, ending at 0.26–0.27. These metrics reflect overall accuracy and generalization ability under different levels of strictness [26, 163].

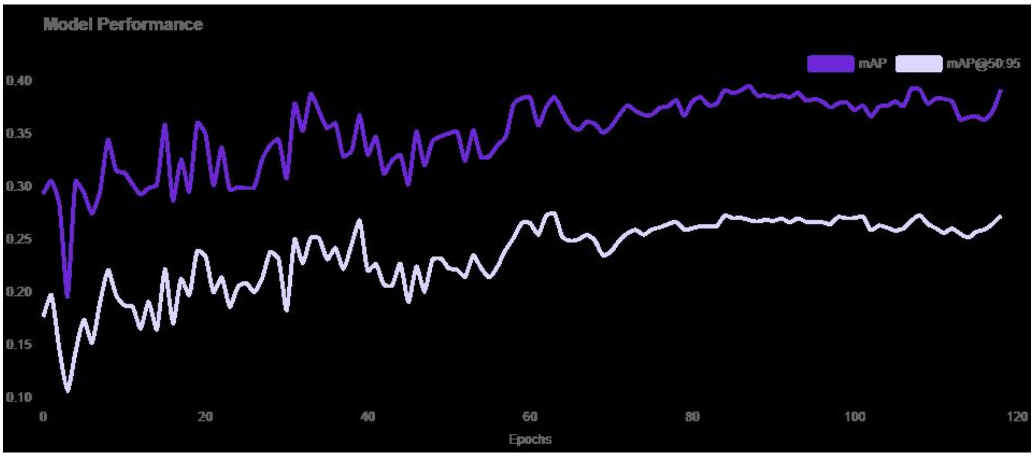


Figure 3.4: Training mAP per Epoch for YOLOv11.

- **The first curve (dark purple)** represents the mean average precision at an IoU threshold of 0.5 (mAP@0.5). This curve shows a big jump in performance early on, then it levels out around epoch 60. From that point, the performance fluctuates within a narrow range around an mAP value of approximately 0.36–0.38. This trend reflects the model's effective improvement in object detection accuracy under a lower IoU threshold.
- **The second curve (light purple)** represents the mean average precision computed across multiple IoU thresholds from 0.5 to 0.95 with a step of 0.05 (mAP@[.5:.95]). This curve shows relatively lower performance compared to mAP@0.5, which is expected due to the more stringent evaluation criteria. It begins with a value around 0.19 and gradually increases, reaching approximately 0.26–0.27 by the end of training. This indicates a steady improvement in the model's precision across multiple levels of evaluation difficulty.

These results indicate that the model achieves gradual and consistent performance improvement over time, and eventually stabilizes at a certain accuracy level after a sufficient number of epochs. This suggests that the model approaches a state of convergence or training saturation.

These findings confirm that the model has not converged effectively and requires further refinement, possibly with better feature extraction techniques or a different dataset. Based on these results, a transition to an object detection approach is being considered for improved performance.

Yolo11 Prediction

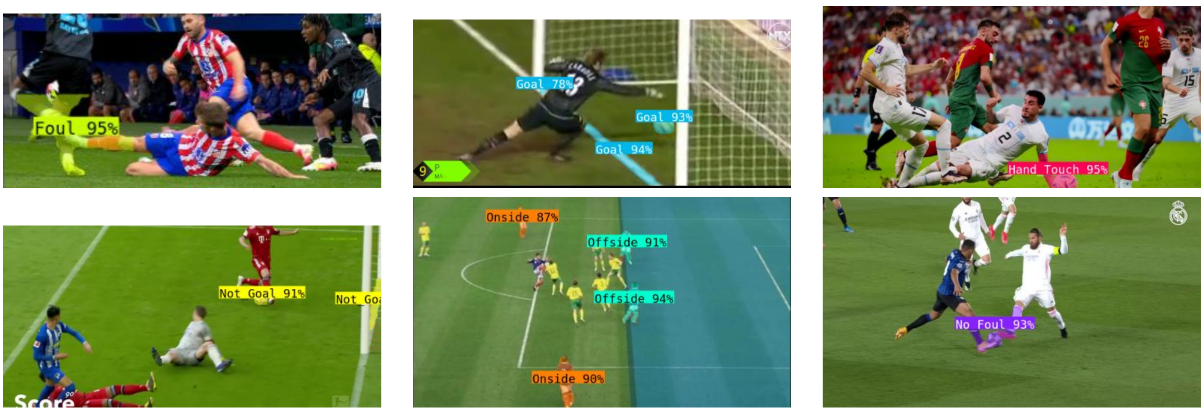


Figure 3.5: Test Predictions for All Classes

3.2.2 Object Detection Problem

The YOLOv11 model for instance segmentation didn't deliver the results I was hoping for. The main numbers we look at, like losses, accuracy, and how well the system finds true positives, were all over the place and didn't do and expected. The mean Average Precision (mAP) remained consistently low throughout training, indicating the model's failure to effectively learn from the available data.

SoccerNet V3 Dataset

The dataset is based on the SoccerNet V3 annotations and includes images extracted from professional football matches. The resolution of all images was standardized to 1920 × 1080 pixels. These images are annotated for two main tasks:

- **Object Detection:** The task involves detecting various entities on the football field, such as players, goalkeepers, referees, and the ball. The annotations were originally in JSON format and converted into YOLO format (.txt files).
- **Event Classification:** The task involves identifying key football events like goals, fouls, and substitutions. Event-specific frames were annotated and also converted to YOLO-compatible formats.

Based on these observations, we decided to shift toward a different approach—**object detection**. This change is meant to take advantage of how good object detection is at pinpointing different categories of objects in pictures. especially when supported by a larger and more diverse dataset. YOLO-based models have proven effective in real-time object detection scenarios by providing fast and accurate predictions using a single neural network for the entire image [59, 134]. The new strategy is expected to enhance the model's ability to learn more distinctive and generalizable visual features.

The SoccerNet V3 dataset provides a comprehensive and high-quality resource for video understanding and sports analytics, making it suitable for training deep learning models in complex multi-camera football scenes [39].

Image Distribution:

- Training set: 21213 images
- Validation set: 4545 images
- Test set: 5060 images

Player Detection

Number of Classes

Object Detection

- Total Classes: 4
- Class names:
  - Ball
  - Player
  - Goalkeeper
  - Main

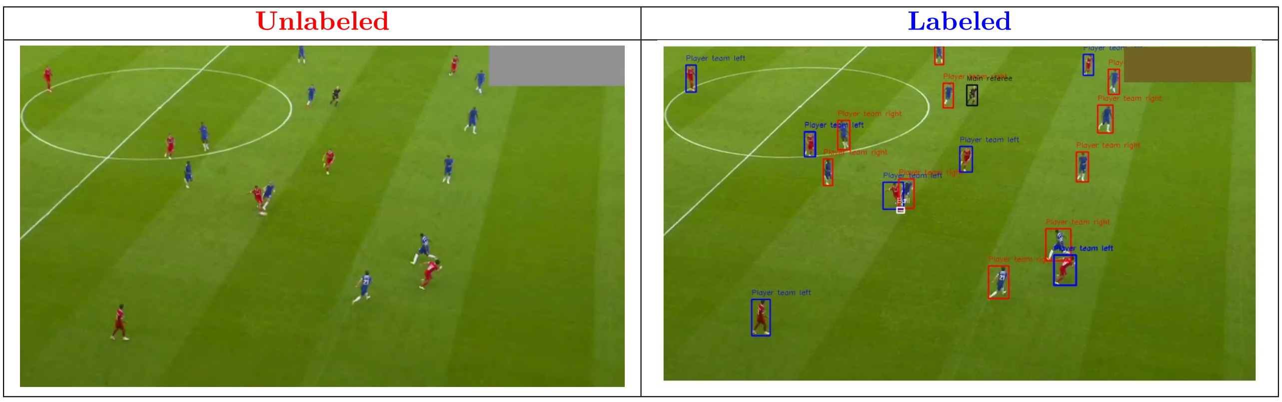


Figure 3.6: Comparison of labeled and unlabeled image

Object Detection Class Distribution (All Sets)

Class	Train Count	Train %	Val Count	Val %	Test Count	Test %
Ball	16,401	7.22%	3,506	7.17%	3,928	7.19%
Player	187,845	82.81%	40,473	82.67%	45,331	82.83%
Goalkeeper	13,079	5.96%	2,849	5.82%	3,088	5.64%
Referee	16,361	7.20%	3,570	7.29%	3,922	7.17%

Table 3.4: Object Detection Class Distribution across Train, Validation, and Test Sets

Ball Detection

Due to the small size of the ball, we separated the ball label from the rest of the objects and put it as a separate data again to achieve a good result for ball detection.

Number of Classes

Object Detection

- Total Classes: 1
- Class names:
  - Ball

Results

YOLOv8 Object Detection Model Description

In this project, we employed the **YOLOv8m** model from the Ultralytics framework for object detection. YOLOv8m is a medium-sized, real-time object detection model known for its strong balance between speed and accuracy [84]. In our configuration, the model consists of 112 layers and approximately 68.1 million parameters, It delivers strong performance without costing too much in terms of computing power, which makes it a great choice for things like analyzing sports videos.

Player Detection

Training Details:

- **Framework:** Ultralytics YOLOv8 (v8.3.116)
- **Hardware:** Tesla P100 GPU (16 GB VRAM)
- **Training Duration:** 100 epochs completed in 0.716 hours
- **Model Size:** 136.7 MB (optimizer removed)
- **Data Augmentation:** *Not applied*, as the dataset was sufficiently diverse to enable effective learning without synthetic augmentation.

Evaluation Metrics: The model's performance was evaluated using standard object detection metrics:

- **Precision (P)** – the proportion of positive identifications that were correct.
- **Recall (R)** – the proportion of actual positives that were correctly identified.
- **mAP@0.5** – mean Average Precision at IoU threshold 0.5.
- **mAP@0.5:0.95** – mean Average Precision averaged across IoU thresholds from 0.5 to 0.95 [134].

Performance Summary (on validation set):

- **Overall**:

  - Precision: 0.893
  - Recall: 0.756
  - mAP@0.5: 0.846
  - mAP@0.5:0.95: 0.596

- **Per-Class Performance (mAP@0.5)**:
  - Ball: 0.462
  - Goalkeeper: 0.964
  - Player: 0.982
  - Referee: 0.974

Inference Speed:

- Preprocessing time per image: 0.1 ms
- Inference time per image: 16.3 ms
- Postprocessing time per image: 1.1 ms

Confusion Matrix Analysis

The main numbers we look at, like losses, accuracy, and how well the system finds true positives, were all over the place and didn't do and expected. These matrices show or point out something the model's ability to correctly classify each object class, while also revealing potential class overlaps or misclassifications.

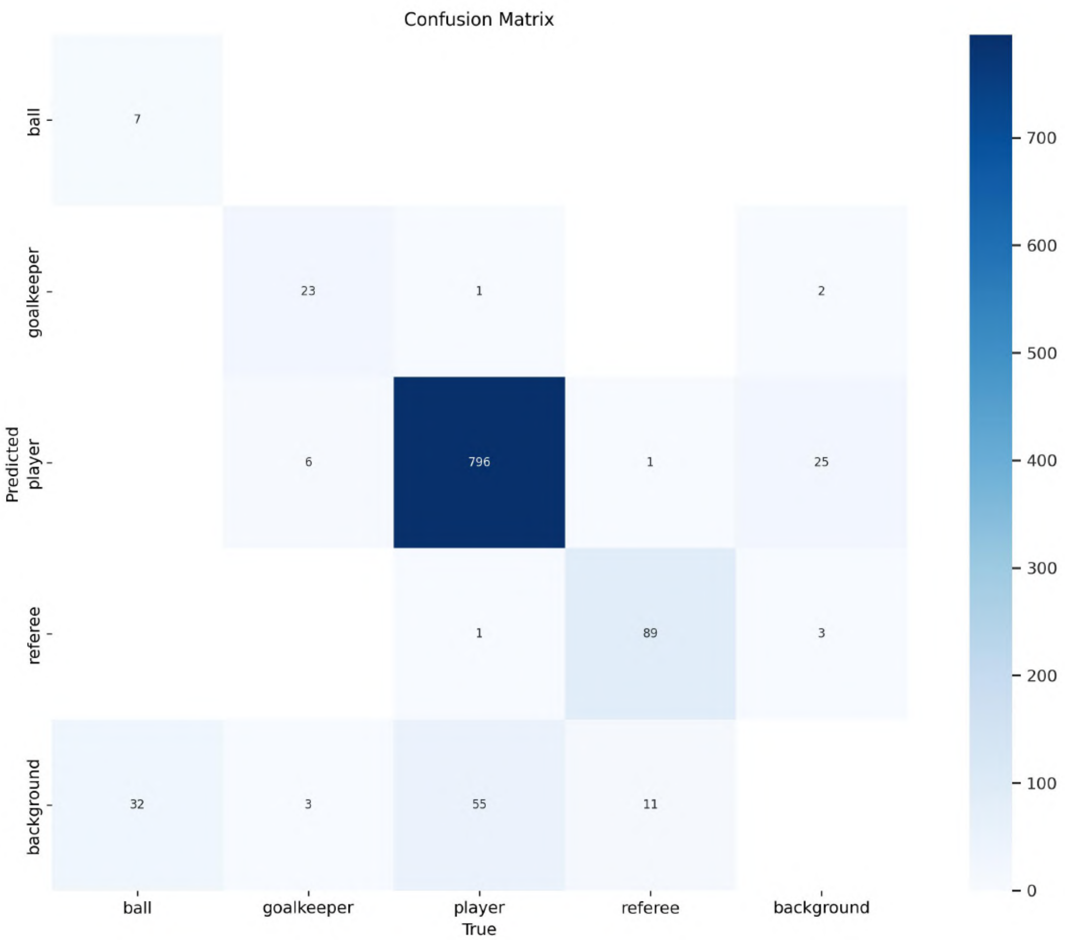


Figure 3.7: Confusion matrix

The confusion matrix quantitatively evaluates the classification performance of the model across five semantic categories relevant to football scene understanding: **Ball**, **Goalkeeper**, **Player**, **Referee**, and **Background**.

The diagonal elements represent the number of correctly classified samples for each class. For instance, the model accurately classified 7 instances of **Ball**, 23 instances of **Goalkeeper**, 796 instances of **Player**, and 89 instances of **Referee**.

The off-diagonal parts show mistakes where the model mixes up different classes:

- **Ball**: Achieved 7 correct classifications, with no significant misclassifications observed, indicating robust performance for this category.
- **Goalkeeper**: Correctly identified 23 times; however, minor confusions were recorded, with 1 sample misclassified as **Player** and 2 samples misclassified as **Background**.
- **Player**: The class with the largest sample size, achieving 796 correct predictions. Nevertheless, the model confused 6 player instances as **Ball**, 1 instance as **Referee**, and 25 instances as **Background**, suggesting slight overlaps in feature space, especially between players and background.
- **Referee**: Correctly classified in 89 instances, with minimal misclassification, notably 1 as **Player** and 3 as **Background**.
- **Background**: Exhibited the most significant confusion, with samples misclassified as **Ball** (32 instances), **Goalkeeper** (3 instances), **Player** (55 instances), and **Referee** (11 instances), highlighting the complexity and the natural fluctuations found in the background areas.

General Interpretation: Overall, the model demonstrates high classification accuracy for the **Player** and **Goalkeeper** categories. The low error rates for **Ball** and **Referee** indicate effective feature learning for these smaller or less frequently appearing classes. However, significant overlap between **Background** and object classes, particularly **Player**, points to a major challenge in distinguishing between object boundaries and environmental noise.

Precision, Recall, and F1 Score Curves

The following curves provide detailed insights into the model's performance across various confidence thresholds and IoU levels [55, 158]:

- **Precision (P)** – The ratio of true positives to all predicted positives.
- **Recall (R)** – The ratio of true positives to all actual positives.
- **F1-Score** – The harmonic mean of precision and recall.
- **PR-Curve** – Precision-recall tradeoff at different thresholds.

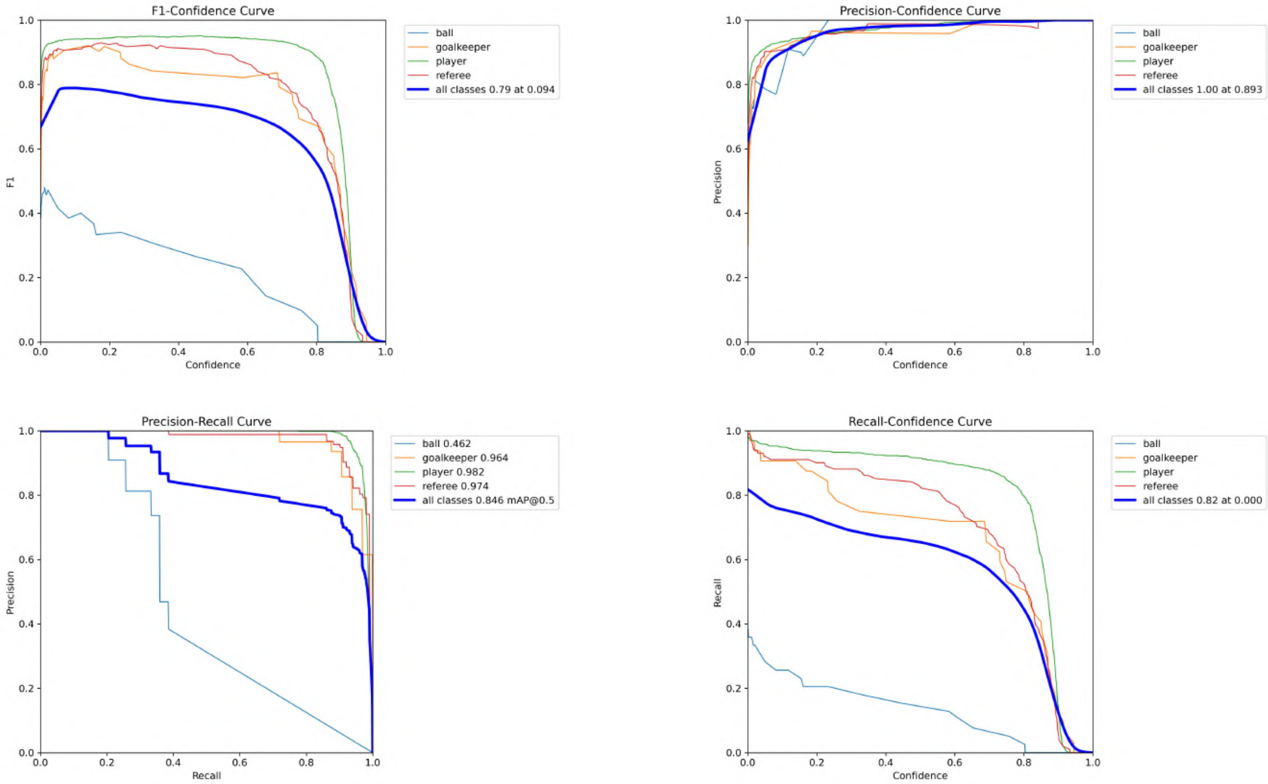


Figure 3.8: F1 score, precision, precision-recall (PR), and recall curves.

Figure 3.8 presents four types of curves that evaluate the performance of the object detection model across various categories under a varying confidence threshold. The analysis includes different object classes: ball, player, goalkeeper and referee, in addition to the mean performance across all categories.

- **F1-Confidence Curve (Top-Left)** The F1-Confidence curve shows that the model achieves its optimal F1-score (0.79) at a confidence threshold of 0.094. Increasing the threshold leads to a gradual decline in F1, mainly due to reduced recall. Class-wise, the **Player** class achieves the highest F1-score (>0.9), followed by **Goalkeeper** and **Referee** (both 0.85). In contrast, the **Ball** class shows lower performance (0.5) with a sharp drop at higher thresholds. These results suggest strong detection for players and goalkeepers, while ball detection remains challenging. A confidence threshold of 0.094 offers the best balance between precision and recall.

- **Precision-Confidence Curve (Top-Right)** The Precision-Confidence curve illustrates that precision reaches its maximum value (1.0) at high confidence thresholds (0.893), but gradually declines as the threshold decreases, reflecting an increase in false positives. The highest precision is observed for the **Goalkeeper**, **Referee**, and **Player** classes, while the **Ball** class exhibits lower precision, likely due to its small size and visual ambiguity. These results show how important it is to pick the right confidence level if you want a good balance between catching most things and avoiding false alarms.

- **Precision-Recall (PR) Curve (Bottom-Left)**

  The Precision-Recall curve demonstrates strong detection performance for the **Player** (0.982), **Goalkeeper** (0.964), and **Referee** (0.974) classes, indicating high reliability in distinguishing these entities. In contrast, the **Ball** class achieves considerably lower precision (0.462), reflecting challenges in accurate localization and classification. The mAP@0.5 score of 0.846 shows the model is doing pretty well across most categories. It also emphasizes how important it is to focus on improving how accurately the system detects the ball.

- **Recall-Confidence Curve (Bottom-Right)** The Recall-Confidence curve shows a peak recall of 0.82 at a confidence threshold of 0.000, with a gradual decline as the confidence increases. The **Player**, **Goalkeeper**, and **Referee** classes maintain relatively high recall values across different thresholds, indicating consistent detection. In contrast, the **Ball** class exhibits lower recall, highlighting difficulties in detecting smaller or more ambiguous instances under stricter confidence constraints.

Overall Training and Validation Results

The comprehensive training results figure summarizes performance across epochs, including losses (box, objectness, and classification), precision, recall, and mAP metrics [51].

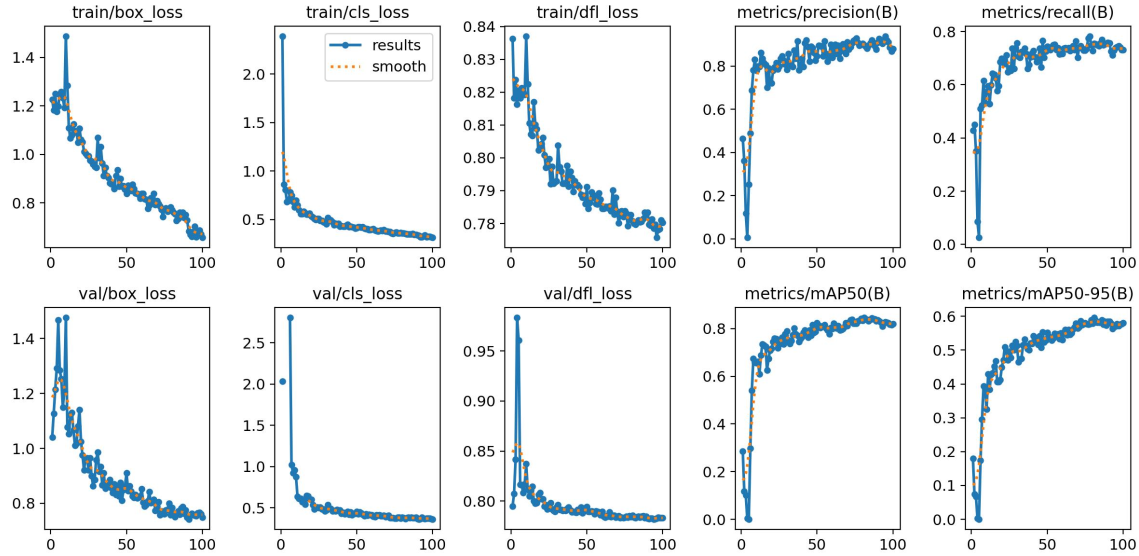


Figure 3.9: Training and validation metrics across epochs.

Training and Validation Indicators over Epochs

Figure 3.9 illustrates the evolution of the model's performance throughout 100 training epochs, showcasing various loss functions and evaluation metrics critical to object detection tasks. These indicators—namely bounding box loss, classification loss, distance-based localization loss (DFL), precision, recall, and mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds—collectively reflect how well the model is learning and generalizing [133].

- **train/box_loss:** The bounding box regression loss steadily decreases from approximately 1.4 to 0.7 across training epochs, suggesting that the model progressively refines its localization accuracy and becomes more precise in predicting object boundaries.
- **train/cls_loss:** A sharp decline is observed from over 2.0 down to roughly 0.3, indicating that the model quickly learns to distinguish between object classes, improving its classification capability.
- **train/dfl_loss:** The distribution focal loss (DFL), which captures finer localization precision, shows a gradual but consistent drop from 0.83 to below 0.78. This reflects enhanced accuracy in pinpointing object centers and edges.
- **metrics/precision(B):** Precision increases significantly and stabilizes around 0.9, implying that the majority of predicted bounding boxes are relevant, with few false positives.
- **metrics/recall(B):** Recall grows steadily from around 0.3 to 0.75, indicating the model becomes increasingly capable of detecting most of the relevant objects in the images, reducing false negatives.
- **val/box_loss:** Validation box loss mirrors the training trend, decreasing from roughly 1.5 to 0.8. This suggests that the model generalizes well to unseen data and avoids overfitting.
- **val/cls_loss:** The validation classification loss follows a similar trajectory to its training counterpart, confirming that the learned class distinctions are transferable to validation data.
- **val/dfl_loss:** The validation DFL also shows stable convergence towards 0.78, further affirming consistent localization performance.
- **metrics/mAP50(B):** The mean Average Precision at IoU=0.5 improves significantly, reaching around 0.8. This reflects strong detection performance when lenient overlap thresholds are considered.
- **metrics/mAP50-95(B):** The stricter mAP across multiple IoU thresholds (0.50 to 0.95) grows from below 0.1 to approximately 0.58, demonstrating the model's ability to perform well even under tighter localization accuracy requirements.

The overall trends indicate stable and progressive training behavior without signs of overfitting. The consistent improvement across all metrics supports the robustness of the model architecture and the effectiveness of the training setup.

Summary of Findings: The model exhibits excellent detection capability, particularly for **Player**, **Goalkeeper**, and **Referee** classes. However, the detection of the **Ball** class remains a challenge due to its smaller size and visual ambiguity. The high mAP@0.5 (0.846) indicates strong overall performance. Additionally, insights from the Precision-Confidence and Recall-Confidence curves suggest that careful confidence threshold tuning can further optimize performance. There remain clear opportunities to enhance the model's detection accuracy, especially for small and less distinct objects.

Ball Detection

Training Details:

- **Framework:** Ultralytics YOLOv8 (v8.3.118)
- **Hardware:** Tesla P100 GPU (16 GB VRAM)
- **Training Duration:** 50 epochs completed in 10.393 hours
- **Model Size:** 136.8 MB (optimizer removed)

Evaluation Metrics: The model's performance was evaluated using standard object detection metrics:

- **Precision (P)** – 0.936
- **Recall (R)** – 0.81
- **mAP@0.5** – 0.914
- **mAP@0.5:0.95** – 0.562

Model Summary:

- 112 layers, 68,124,531 parameters, 0 gradients
- Computational cost: 257.4 GFLOPs

Inference Speed:

- Preprocessing time per image: 0.4 ms
- Inference time per image: 68.6 ms
- Postprocessing time per image: 1.4 ms

Model Checkpoints:

- Best weights saved to: `runs/detect_ball/exp_ball/weights/best.pt`
- Final weights saved to: `runs/detect_ball/exp_ball/weights/last.pt`
- Results saved to: `runs/detect_ball/exp_ball`

Yolo8 Model Evaluation Visualizations

Confusion Matrix Analysis

The main numbers we look at, like losses, accuracy, and how well the system finds true positives, were all over the place and didn't do and expected. These matrices show or point out something the model's ability to correctly classify each object class, while also revealing potential class overlaps or misclassifications.

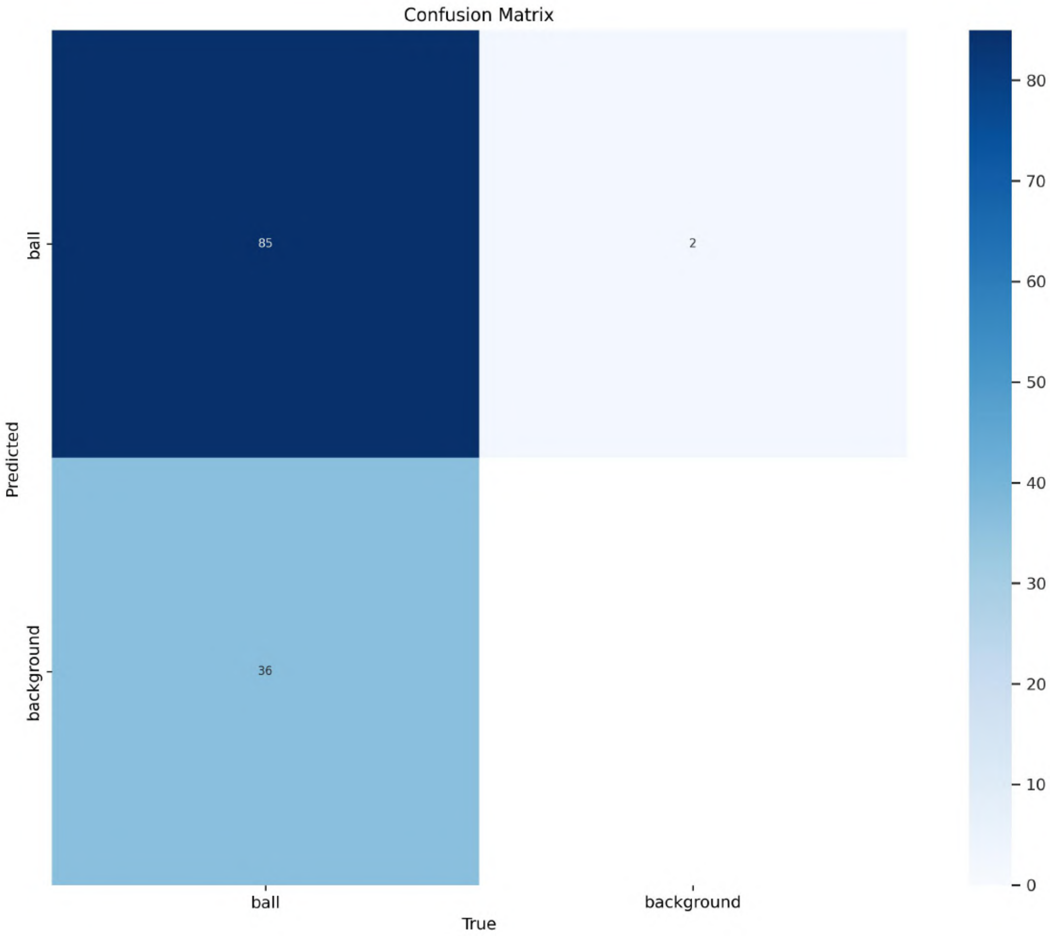


Figure 3.10: Confusion matrix

The model's performance was assessed using a confusion matrix with two categories: "ball" and "background." The evaluation showed that 85 samples were correctly identified as "ball" (True Positives), while 2 samples were incorrectly predicted as "ball" despite belonging to the "background" class (False Positives). Additionally, 36 "ball" instances were misclassified as "background" (False Negatives). No True Negatives were recorded, which may indicate that the background class was either underrepresented or not treated as a positive class in this context.

These results indicate that the model demonstrates high precision in detecting the ball, yet it suffers from a relatively low recall (70%), suggesting it often misses some ball instances. The model appears to follow a conservative prediction strategy—minimizing false positives at the cost of missing some true objects. While the overall performance is strong, enhancing recall remains a key area for future improvement to ensure more comprehensive ball detection.

Precision, Recall, and F1 Score Curves

The following curves provide detailed insights into the model's performance across various confidence thresholds and IoU levels [55, 158]:

- **Precision (P)** – The ratio of true positives to all predicted positives.
- **Recall (R)** – The ratio of true positives to all actual positives.
- **F1-Score** – The harmonic mean of precision and recall.
- **PR-Curve** – Precision-recall tradeoff at different thresholds.

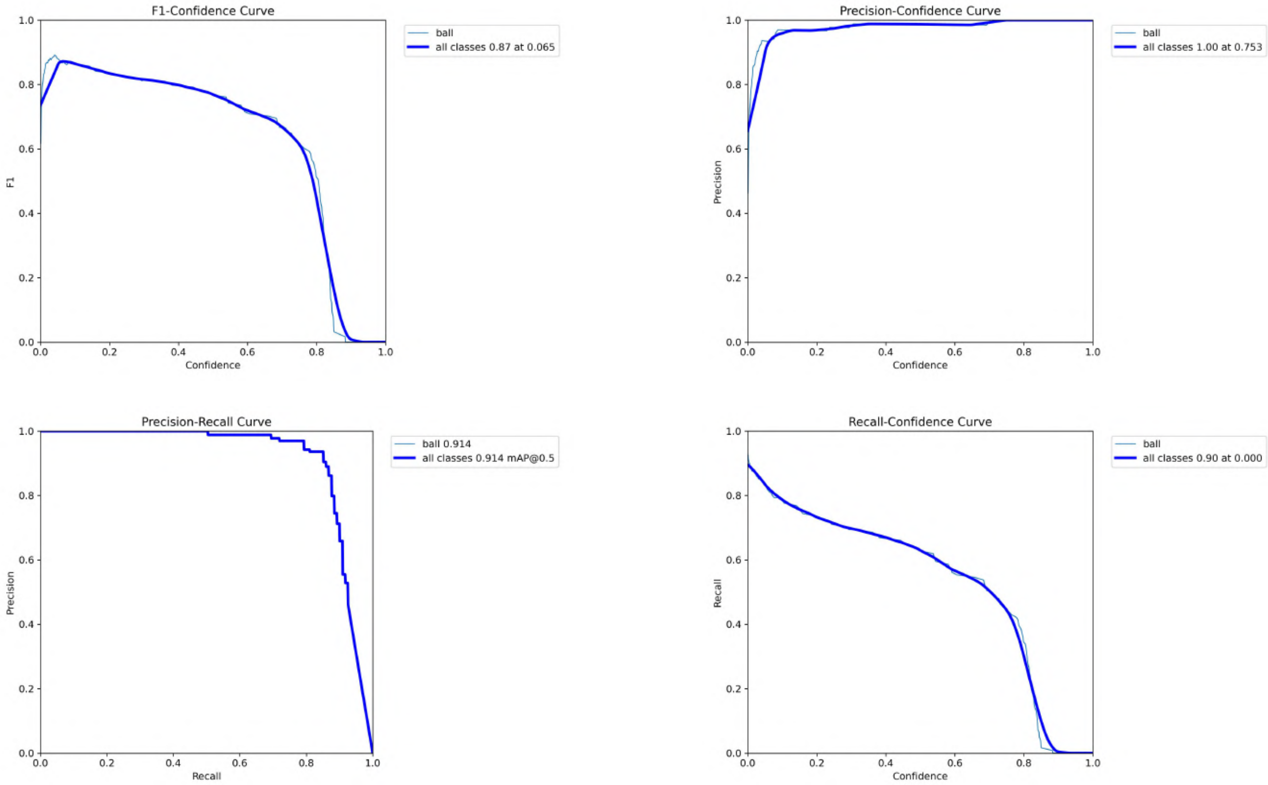


Figure 3.11: F1 score, precision, precision-recall (PR), and recall curves.

Figure 3.11 presents four types of curves that evaluate the performance of the object detection model across various categories under a varying confidence threshold. The analysis includes different object classes: ball, player, goalkeeper and referee, in addition to the mean performance across all categories.

- **F1-Confidence Curve (Top-Left)** This curve shows the relationship between F1-score and confidence threshold. The F1-score peaks at 0.87 at a confidence threshold of 0.065, indicating the best balance between precision and recall by allowing more predictions while maintaining high quality.
- **Precision-Confidence Curve (Top-Right)** This curve illustrates the relationship between precision and prediction confidence. Precision reaches 1.00 at a confidence threshold of 0.753 for the **ball** class, meaning perfect precision with no false positives at that threshold.
- **Precision-Recall (PR) Curve (Bottom-Left)**

  This curve evaluates performance across various confidence thresholds independently of absolute confidence values. Precision reaches 0.914 at a recall of 1.0 with an mAP@0.5 of 0.914, indicating strong detection performance while maintaining high precision at high recall levels.
- **Recall-Confidence Curve (Bottom-Right)** This curve shows recall versus confidence threshold. At a confidence threshold of 0.000, recall is 0.90, reflecting high sensitivity without confidence filtering, decreasing gradually as the threshold increases.

Overall Training and Validation Results

The comprehensive training results figure summarizes performance across epochs, including losses (box, objectness, and classification), precision, recall, and mAP metrics [51].

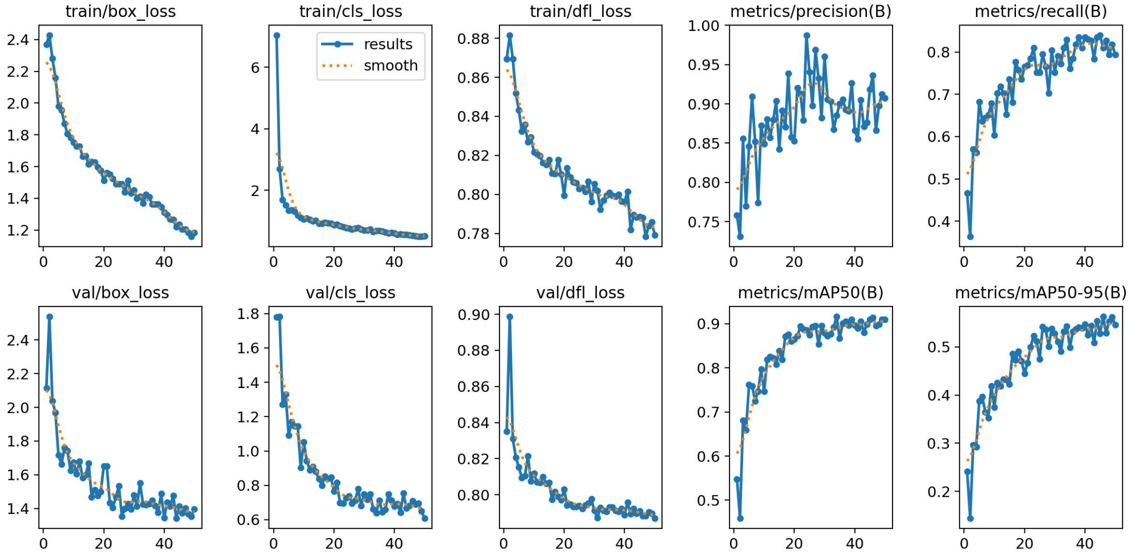


Figure 3.12: Training and validation metrics across epochs.

Training and Validation Indicators over Epochs

Figure 3.12 illustrates the evolution of the model's performance throughout 50 training epochs, showcasing various loss functions and evaluation metrics critical to object detection tasks. These indicators—namely bounding box loss, classification loss, distance-based localization loss (DFL), precision, recall, and mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds—collectively reflect how well the model is learning and generalizing [133].

- **train/box_loss:** The curve shows a gradual and steady decrease in box loss values as the number of epochs increases, indicating improved accuracy in object localization over time. The drop keeps going down gradually with a few small ups and downs, which means the training seems steady and not showing any clear signs of overfitting or getting unstable.
- **train/cls_loss:** The curve starts with a very high loss value (6) but decreases sharply during the early epochs, then stabilizes at low values. This behavior reflects the model's rapid adaptation to the object classification task.
- **train/dfl_loss:** This curve decreases consistently from about 0.88 to approximately 0.78 with limited fluctuations, indicating improved model performance in predicting bounding box distributions.
- **metrics/precision(B):** The curve shows a general upward trend, starting at around 0.75 and reaching approximately 0.9, suggesting a reduction in false positives as training progresses.
- **metrics/recall(B):** The recall rate gradually increases from about 0.4 to over 0.8, indicating improved ability of the model to retrieve correct objects and reduce misses.
- **val/box_loss:** The curve drops so that's pretty similar to the training one, with some small ups and downs early on. Overall, it stays steady on the validation set and doesn't show signs of overfitting.
- **val/cls_loss:** The curve decreases from a high initial value (>1.8) to lower values, indicating continuous improvement in object classification accuracy on validation data.
- **val/dfl_loss:** It gradually decreases and stabilizes, indicating good model generalization beyond the training data.
- **metrics/mAP50(B):** The mAP50 indicator rises from about 0.5 to over 0.9, showing overall improvement in model performance at the easier 50% threshold.
- **metrics/mAP50-95(B):** The indicator gradually rises from about 0.3 to over 0.5, indicating the model's ability to perform well across different precision requirements.

Overall, the curves indicate a stable and effective training process with continuous performance improvement and no clear signs of overfitting or poor generalization.

Summary of Findings: The overall analysis of the model's training and evaluation reveals a stable and effective learning process, as evidenced by the performance curves that show continuous improvement across key metrics without indications of overfitting or poor generalization. The final evaluation highlights the model's ability to achieve perfect precision (1.00) for the **ball** class at a high confidence threshold of 0.753; however, this comes at the expense of recall, which declines under stricter confidence requirements. Conversely,Lowering the confidence threshold helps catch more true positives, so recall goes up. But at the same time, it might also increase false positives, which can lower precision. Choosing the right confidence threshold is really important because it depends on what matters most for the job—whether you want to avoid false alarms or make sure you catch as many objects as possible.. In summary, the model demonstrates robust overall performance with a mean Average Precision at IoU 0.5 (mAP@0.5) of 0.914 and an F1-score of 0.87, indicating high reliability and consistency in object detection across classes.

Yolo8 Model Prediction

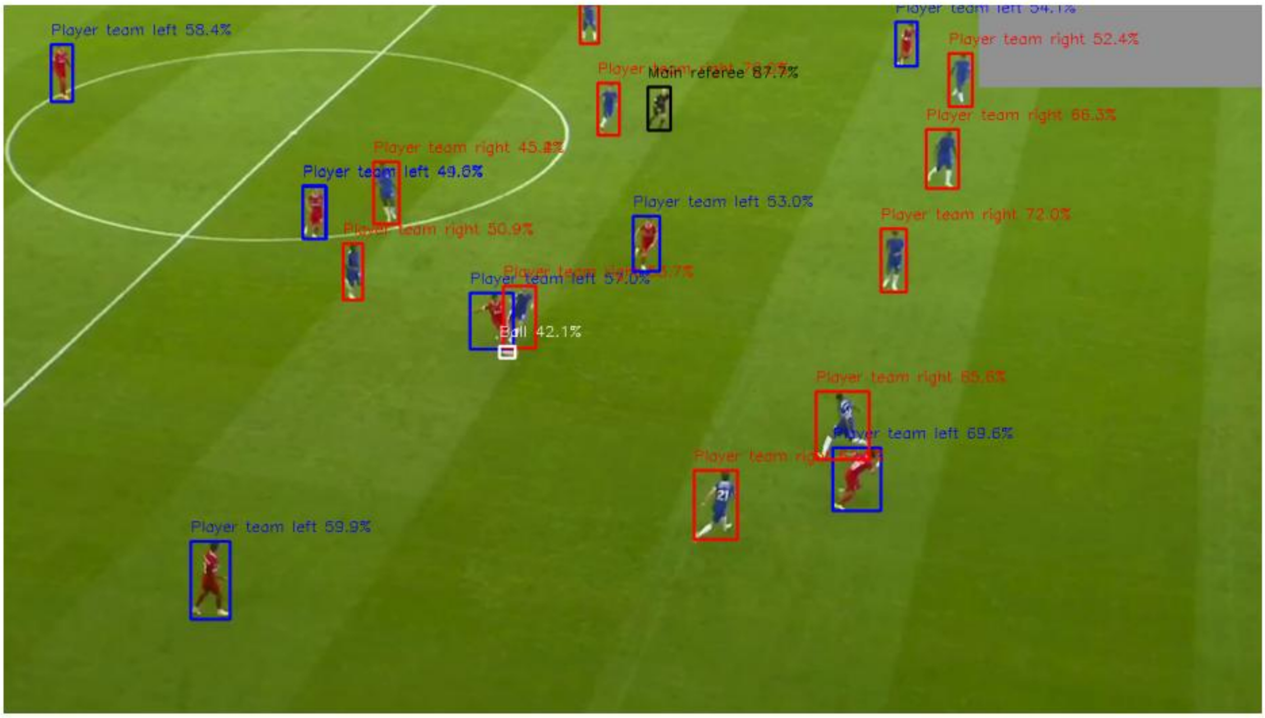


Figure 3.13: Prediction of all classes with confidence in one image

3.2.3 Keypoint Detection Problem

When analyzing football videos automatically, figuring out where the field edges and important lines are is a really important first step. This process involves identifying the structural layout of the pitch, including sidelines, goal areas, penalty boxes, the center circle, and other essential markings that define the spatial organization of the game. Accurate detection of these elements enables the system to interpret player positions and events within a standardized coordinate space, regardless of camera angle or perspective. Typically, this step leverages computer vision techniques such as line detection algorithms (e.g., Hough Transform), edge detection, and geometric modeling to extract and map these features from each frame of the video.The way the field is set up here provides a starting point for things like following player movements, identifying different events, and studying team strategies, by providing spatial context and ensuring consistency in the interpretation of visual data throughout the match.

Unlabeled	Labeled
	

Figure 3.14: Comparison of labeled and unlabeled image

Results

Training Details:

- **Framework:** Ultralytics YOLOv8 (v8.3.116)
- **Hardware:** Tesla P100 GPU (16 GB VRAM)
- **Training Duration:** 100 epochs completed in 0.615 hours
- **Model Size:** 140.0 MB (optimizer removed)

Evaluation Metrics: The model's performance was evaluated using standard object detection metrics:

- **Precision (P)** – 0.995
- **Recall (R)** – 1.0
- **mAP@0.5** – 0.995
- **mAP@0.5:0.95** – 0.587

Model Summary:

- 121 layers, 69,784,275 parameters, 0 gradients
- Computational cost: 264.7 GFLOPs

Inference Speed:

- Preprocessing time per image: 0.1 ms
- Inference time per image: 15.8 ms
- Postprocessing time per image: 1.0 ms

Model Checkpoints:

- Best weights saved to:`runs/pose_field/exp_field/weights/best.pt`
- Final weights saved to:`runs/pose_field/exp_field/weights/last.pt`
- Results saved to: `runs/pose_field/exp_field`

Yolo8 Model Evaluation Visualizations

Confusion Matrix Analysis

The main numbers we look at, like losses, accuracy, and how well the system finds true positives, were all over the place and didn't do and expected. These matrices show or point out something the model's ability to correctly classify each object class, while also revealing potential class overlaps or misclassifications.

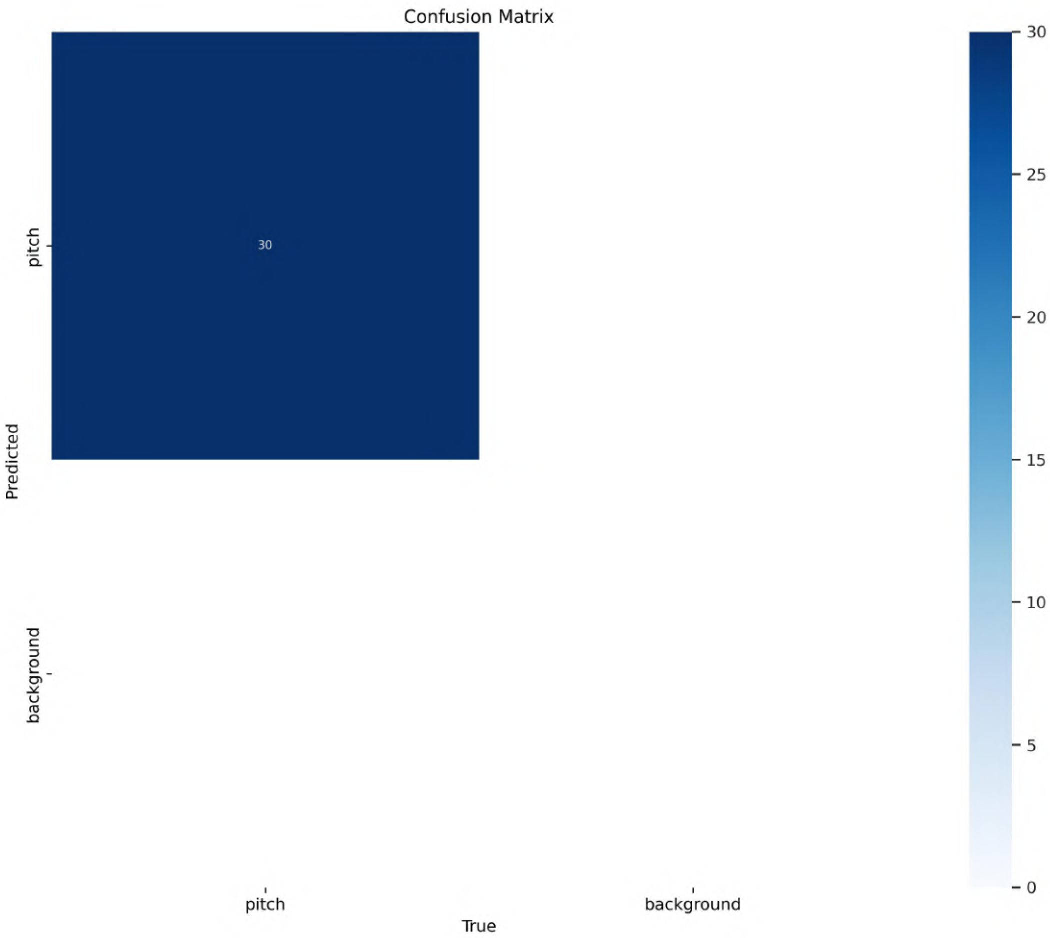


Figure 3.15: Confusion matrix

Figure 3.15 shows the confusion matrix for model evaluation. It is observed that the model successfully classified all samples of the "pitch" class (30 samples) without any misclassification towards the "background" class. The absence of any samples outside the main diagonal indicates a perfect model performance, achieving an accuracy of 100% on the test samples.

Overall Training and Validation Results

The comprehensive training results figure summarizes performance across epochs, including losses (box, objectness, and classification), precision, recall, and mAP metrics [51].

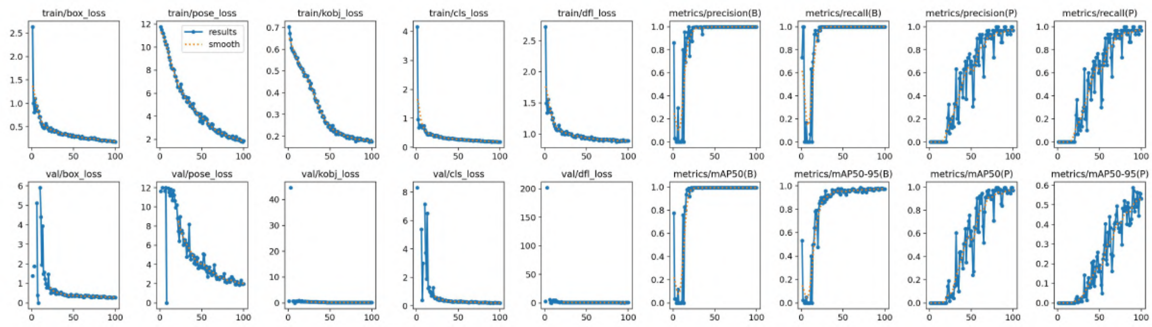


Figure 3.16: Training and validation metrics across epochs.

Training and Validation Indicators over Epochs

Figure 3.16 illustrates the evolution of the model's performance throughout 100 training epochs, showcasing various loss functions and evaluation metrics critical to object detection tasks. These indicators—namely bounding box loss, classification loss, pose loss , precision, recall, and mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds—collectively reflect how well the model is learning and generalizing [133].

- **train/box_loss and val/box_loss:** The bounding box loss shows a steady and continuous decrease during training and validation, stabilizing at a very low value (< 0.5) after the first 20 epochs.
- **train/pose_loss and val/pose_loss:** The pose loss started with high values (12) and gradually decreased as training progressed. Despite some minor oscillations in the validation curve, the overall trend shows consistent improvement.
- **train/kobj_loss and val/kobj_loss:** The keypoint objectness loss rapidly decreased in the early stages and stabilized at very low values, indicating high model accuracy in distinguishing objects from non-objects.
- **train/cls_loss and val/cls_loss:** The classification loss quickly dropped to values near zero, reflecting the model's early success in correctly learning the class labels.
- **train/dfl_loss and val/dfl_loss:** The distribution focal loss also demonstrated a gradual decrease, enhancing the model's ability to predict box and keypoint locations accurately.
- **metrics/precision(B) and metrics/recall(B):** Precision and recall for the bounding boxes (B) rose quickly and stabilized at values greater than 0.98, indicating very few false positives and false negatives.
- **metrics/precision(P) and metrics/recall(P):** Although the precision and recall curves for the keypoints (Pose) exhibited some fluctuations, they showed continuous improvement, reaching high values close to 1.0. The observed fluctuations could be due to the difficulty in precise keypoint predictions or a small validation sample size.
- **metrics/mAP50(B) and metrics/mAP50(P):** The mean Average Precision (mAP) at IoU = 50% rapidly approached 1.0 for both bounding boxes (B) and keypoints (P), reflecting the model's strong ability to accurately retrieve objects.
- **metrics/mAP50-95(P):** The strict mean Average Precision across multiple IoU thresholds (50% to 95%) showed a gradual and consistent improvement, indicating steady enhancement in the model's precision for keypoint localization.

Summary of Findings: The training and validation results demonstrated that the model possesses high efficiency in both object classification and keypoint localization, with very low losses and extremely high performance metrics. The confusion matrix shows no classification errors, reinforcing confidence in the model's accuracy and generalization capability.

Key point Prediction

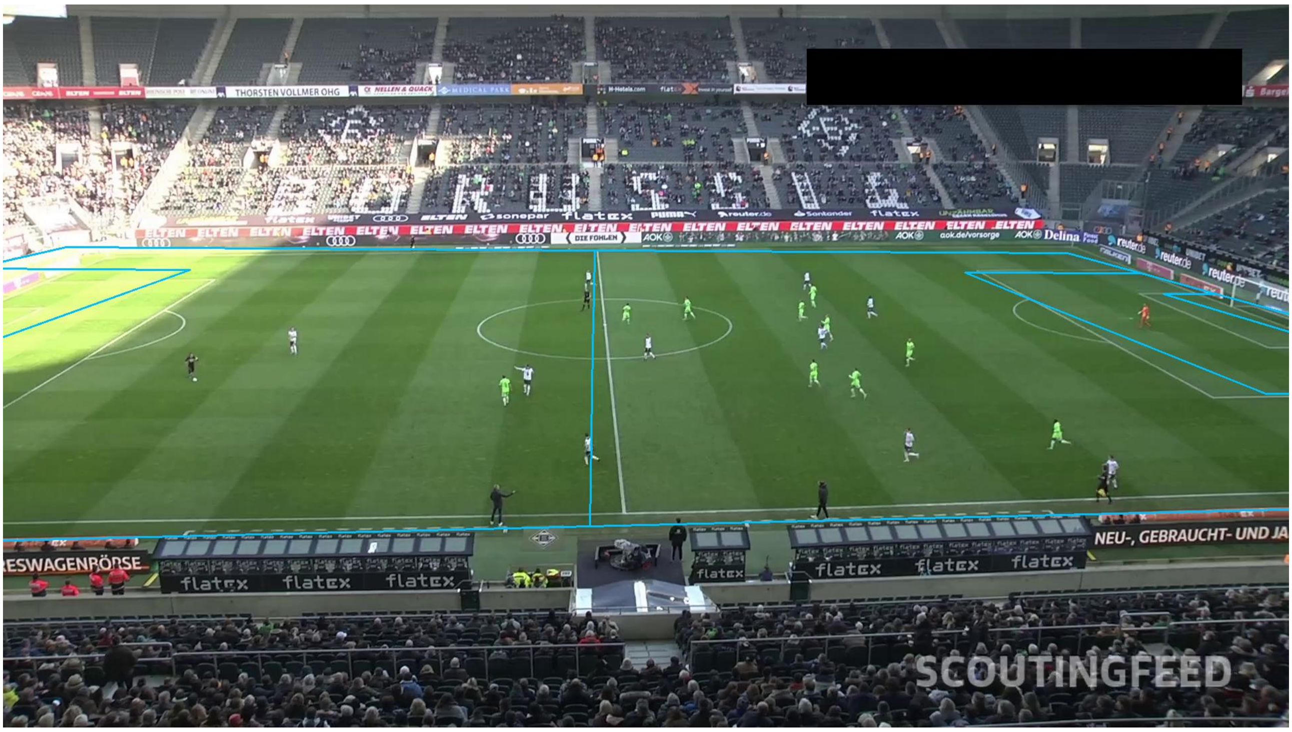


Figure 3.17: Prediction of pitch with draw lines

3.2.4 Classification Problem

Event Classification

- Total Classes: `16`
- Class names:
  - Ball out of play
  - Clearance
  - Corner
  - Direct free-kick
  - Foul
  - Goal
  - Indirect free-kick
  - Kick-off
  - Offside
  - Penalty
  - Red card
  - Shots off target
  - Shots on target
  - Substitution
  - Throw-in
  - Yellow card

Class	Image Count	Percentage
Foul	9296	30.63%
Goal	5509	18.15%
Shots off target	5092	16.78%
Shots on target	3,981	13.12%
Ball out of play	3871	12.75%
Offside	1214	4.00%
Direct free-kick	418	1.38%
Corner	361	1.19%
Indirect free-kick	177	0.58%
Throw-in	130	0.43%
Yellow card	92	0.30%
Clearance	89	0.29%
Penalty	88	0.29%
Kick-off	17	0.06%
Red card	11	0.04%
Substitution	3	0.01%

Table 3.5: Class distribution for Event Classification task (All sets)

Event Classification Class Distribution (All Sets)

Dataset Curation and Class Selection Strategy

To mitigate class imbalance and prevent model bias toward overrepresented categories, only six classes were selected from the full set of football event categories. This deliberate reduction aimed to ensure a more uniform distribution across classes and support a balanced learning process.

The excluded classes included events such as: *Corner kicks, Substitutions, Free kicks, Penalties, Yellow/Red cards*, and other rare or underrepresented actions.

Selected Classes and Distribution

Training Set (Total: 9599 images):

- Foul: 1750 images
- Goal: 1750 images
- Shots off target: 1750 images
- Shots on target: 1750 images
- Ball out of play: 1750 images
- Offside: 849 images

Validation Set (Total: 2057 images):

- Foul: 375 images
- Goal: 375 images
- Shots off target: 375 images
- Shots on target: 375 images
- Ball out of play: 375 images
- Offside: 182 images

Test Set (Total: 2058 images):

- Foul: 375 images
- Goal: 375 images
- Shots off target: 375 images

- Shots on target: 375 images
- Ball out of play: 375 images
- Offside: 183 images

This data sampling approach was intended to maintain class balance during training and evaluation, thereby enhancing the fairness and reliability of the resulting model.

Results

YOLOv8 Model Description and Training Results

A classification model based on the **YOLOv8m-cls** architecture was trained to recognize six types of football match events. The training was conducted over **30 epochs**, taking approximately **3.169 hours** on a **Tesla P100 GPU (16GB)** using the **Ultralytics framework (version 8.3.111)** [158]. The model comprises **42 layers** and approximately **15.77 million parameters**, with a computational cost of **41.6 GFLOPs**.

Evaluation Metrics

The performance of the model was assessed on a held-out validation set using the `best.pt` checkpoint. The results are as follows:

- **Top-1 Accuracy:** 59.2%
- **Top-5 Accuracy:** 99.1%
- **Fitness Score:** 0.791 (a composite metric based on validation accuracy)
- **Inference Time per Image:** 7.5 ms
- **Preprocessing Time per Image:** 0.6 ms

The high Top-5 accuracy indicates that in nearly all cases, the correct class was among the top five predicted classes, demonstrating strong class separability and generalization capacity. However, the moderate Top-1 accuracy suggests there is room for improvement in class ranking.

Yolo8m-cls Model Evaluation Visualizations

Confusion Matrix Analysis

We checked how well each class was identified by looking at the confusion matrix, both in its original form and after adjusting the numbers to make comparisons easier. These matrices show or point out something the model's ability to correctly classify each object class, while also revealing potential class overlaps or misclassifications.

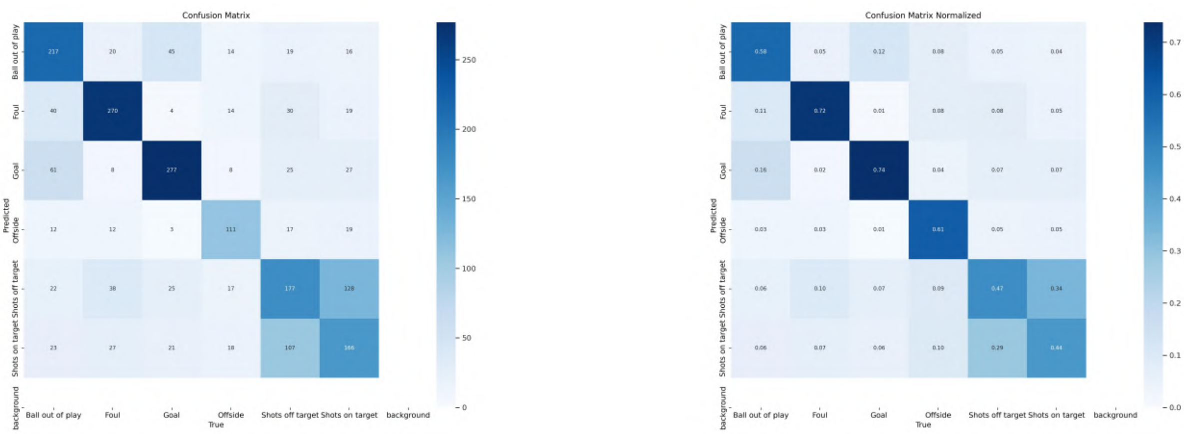


Figure 3.18: Confusion matrix (left) and normalized confusion matrix (right).

Confusion Matrix (Non-Normalized)

This matrix shows the model's performance in classifying different types of match events across seven categories: Ball out of play, Foul, Goal, Offside, Shots off target, Shots on target, Background.

Values inside the matrix indicate the number of samples that were correctly or incorrectly classified by the model. Diagonal values (from top left to bottom right) represent correct predictions, while off-diagonal values indicate misclassifications due to confusion between categories.

Ball out of play: Correctly classified in 217 cases. Confused with Goal in 45 cases, likely due to visual or temporal similarities. Also confused with Foul (20) and Shots off target (19).

Foul: Strong performance with 270 correct classifications. Confusion occurred with Ball out of play (40) and Shots off target (38).

Goal: High accuracy with 277 correct classifications. Misclassified as Ball out of play in 61 cases and Shots off target in 25.

Offside: Lowest performing category with only 111 correct classifications. Most confusion occurred with Ball out of play (12) and Foul (12).

Shots off target: Correctly classified in 177 cases. There was a lot of confusion about the shots on target, which was 128, indicating difficulty in distinguishing between shot types.

Shots on target: 166 cases correctly classified. Similar confusion with Shots off target (107), highlighting the need for additional distinguishing features.

Background: Occasional misclassifications with all categories, though less frequently. Examples include Ball out of play (23), Foul (27), and Goal (21).

Confusion Matrix (Normalized)

This matrix shows the relative performance of the model across seven main football event categories: Ball out of play, Foul, Goal, Offside, Shots off target, Shots on target, Background.

Values are normalized to represent the proportion of true samples per category (row-wise). This allows for fair performance assessment, especially when sample counts vary across classes.

Ball out of play: 58% correctly classified. Most confusion with Goal (12%) and Foul (5%).

Foul: 72% correctly classified. Main confusion with Ball out of play (11%) and Shots off target (10%).

Goal: Best performing category with 74% accuracy. Misclassified as Ball out of play in 16% of cases.

Offside: 61% correctly classified. Notable confusion with Shots off target (5%) and Ball out of play (3%).

Shots off target: 47% correctly classified. Confused with Shots on target in 34% of cases.

Shots on target: Only 44% correctly classified. Large confusion with Shots off target (29%) and Offside (10%).

Background: Relatively low confusion, mostly correctly classified.

Conclusion:

The 2 matrix highlights the model's strengths in classifying Goal and Foul events, and its weaknesses in distinguishing between Shots on/ off target. This suggests the need to incorporate more contextual or spatiotemporal features to improve performance in these ambiguous cases.

Overall Training and Validation Results

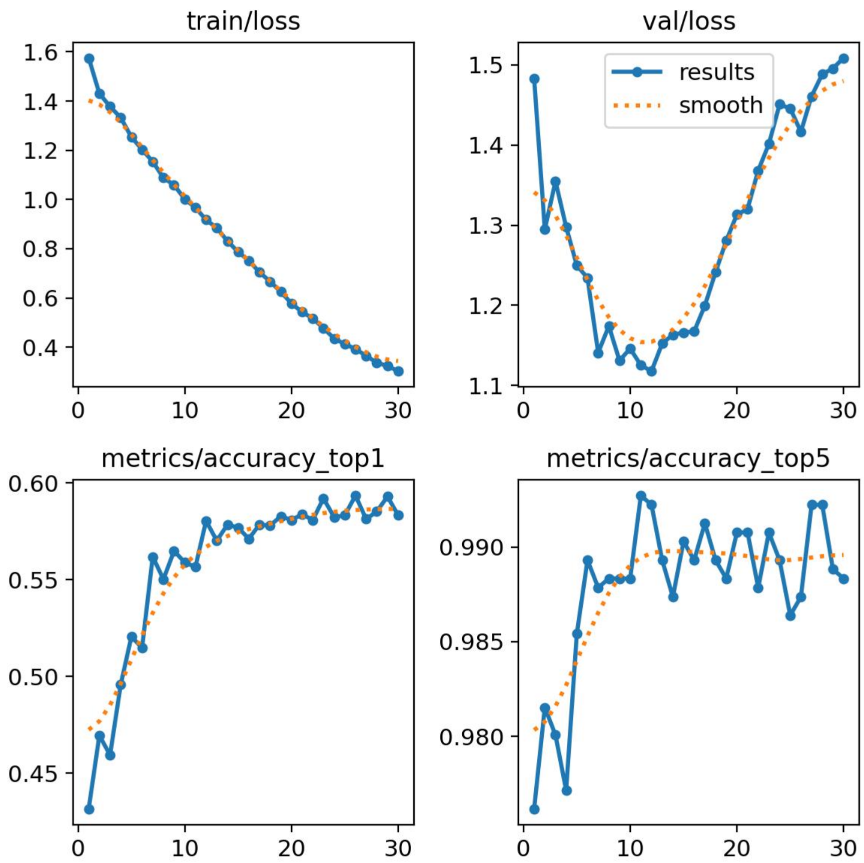


Figure 3.19: Training and validation metrics across epochs.

This figure illustrates the development of the model's performance over 30 training epochs through four main curves: training loss, validation loss, Top-1 accuracy, and Top-5 accuracy. The solid blue line represents actual values, while the dotted orange line indicates smoothed values for easier observation.

1. **train/loss curve (Training Loss):**
   This curve shows a continuous and consistent decrease in loss value from approximately 1.6 to below 0.3. This reduction indicates effective learning by the model and a gradual improvement in fitting the training data.

2. **val/loss curve (Validation Loss):**
   The validation loss initially decreases until around epoch 11, after which it gradually increases until the end of training.
   This behavior reflects the phenomenon of overfitting, where the model starts to lose its generalization ability on validation data after a certain point.

3. **metrics/accuracy_top1 (Top-1 Accuracy):**
   Accuracy gradually increases from approximately 0.45 to nearly 0.59 during training.
   This value represents the proportion of predictions where the top-1 most probable label matches the correct label.
   This improvement indicates the model's increasing ability to capture the most distinguishing features of the classes.

4. **metrics/accuracy_top5 (Top-5 Accuracy):**
   High accuracy is observed from early stages, starting above 0.97 and exceeding 0.99 in most epochs.
   This metric indicates that the correct label was within the model's top 5 predicted labels.
   It demonstrates the model's strong capability in narrowing down its predictions to a small set that includes the correct answer.

Yolo8 Classification Prediction

Figure 3.20: Test Predictions for All Classes

3.3 Analysis of Players

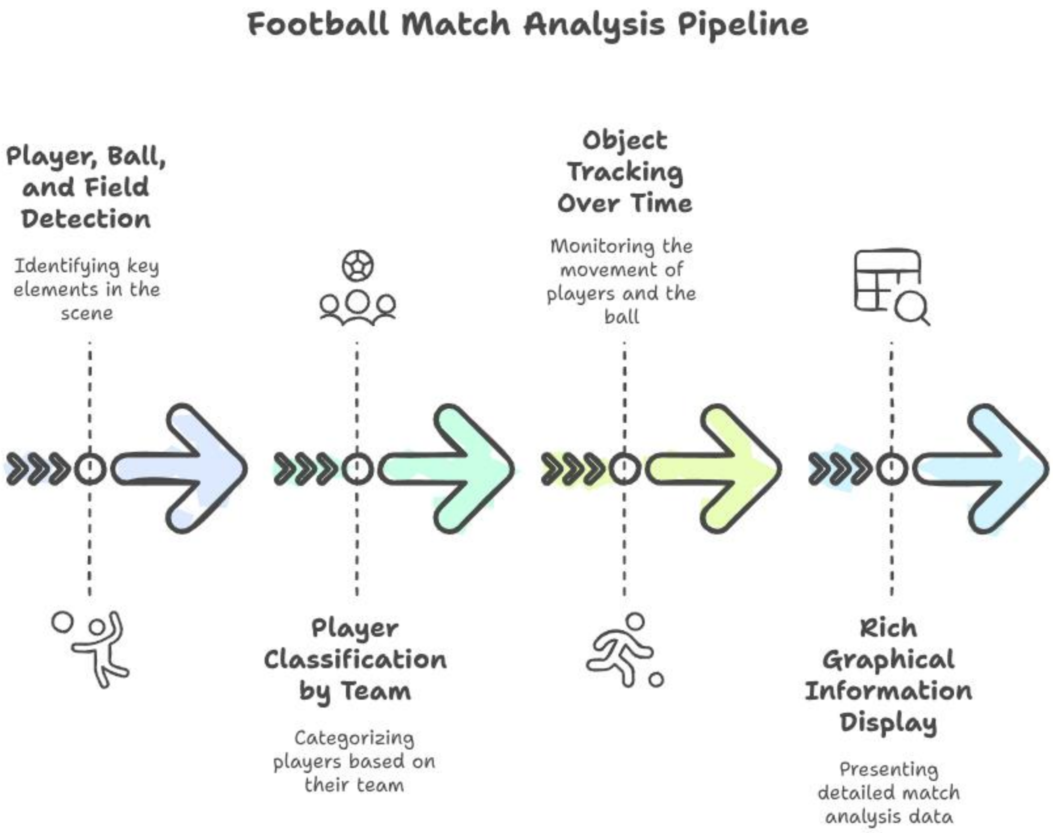


Figure 3.21: Pipeline Of Analysis Match Steps.

3.3.1 Sports Pitch Annotation Library

To render and annotate football pitch elements, we utilized the `sports` Python package [160], which is designed for sports field visualization. This package was installed directly from its GitHub repository using the following command:

`git+https://github.com/roboflow/sports.git`

The `sports` library [160] provides utilities for defining pitch templates, projecting field lines, and drawing overlays such as player positions and tracking data. These tools were instrumental in generating consistent and interpretable visualizations throughout our pipeline.

3.3.2 Core Libraries and Dependencies

To construct the foundation of our video analytics pipeline, we imported a comprehensive set of libraries for computer vision, deep learning, and unsupervised clustering. The following Python packages were utilized:

- `numpy`: for numerical operations and image array manipulation [72].
- `ultralytics.YOLO`: to load pretrained YOLO models for player, ball, and pitch detection [159].
- `opencv-python (cv2)`: for frame extraction, image preprocessing, and general image processing routines [123].
- `supervision`: a high-level library designed to facilitate detection management, visualization, and tracking in sports analytics [140].
- `tqdm`: to monitor progress during long-running data extraction or model inference loops [41].
- `torch` and `transformers`: to utilize the SigLIP vision transformer model for generating visual embeddings of player crops [80, 131].
- `umap-learn`: to project high-dimensional embeddings into a lower-dimensional space for visual separability and clustering [111].
- `scikit-learn`'s `KMeans`: for clustering player embeddings into distinct team groups [124].
- `typing`: for code annotation and improved readability via type hints [130].

The combination of these libraries enables a modular and scalable pipeline for video-based object detection, feature embedding, dimensionality reduction, and team classification.

3.3.3 Team Classification Module

To perform unsupervised visual team classification, we implemented a custom class named `TeamClassifier`. This class leverages vision-language embedding models, dimensionality reduction, and clustering techniques to separate players into two distinct teams based on their visual appearance. The pipeline consists of the following stages:

- **Embedding Extraction:** We employ the pretrained SigLIP vision model (`google/siglip-base-patch16-224`) to compute high-dimensional embeddings from cropped player images. The model is loaded using the `transformers` library and deployed on the available GPU via PyTorch [80, 131].
- **Dimensionality Reduction:** The high-dimensional embeddings are projected into a 3-dimensional space using Uniform Manifold Approximation and Projection (UMAP), which preserves the local structure of the data while facilitating cluster separability [111].
- **Clustering:** The reduced embeddings are grouped into two clusters using the KMeans algorithm, corresponding to the two opposing teams [124].
- **Inference:** Once the clustering model is trained, the same embedding and projection pipeline is used for predicting the team label of new player crops.

This modular design allows the `TeamClassifier` to be easily integrated into video analysis pipelines, enabling dynamic team detection and tracking based purely on visual features.

3.3.4 Goalkeeper Team Identification and Player Crop Collection

To support team-based analysis in soccer video understanding, we implemented two utility functions: one for inferring the team identity of goalkeepers and another for collecting player crops from raw video footage.

Goalkeeper Team Assignment

The function `resolve_goalkeepers_team_id` estimates the team identity of each goalkeeper based on spatial proximity to team centroids. It operates on bottom-center anchor coordinates of all detected players and goalkeepers. The centroids for each team are computed from player coordinates, and each goalkeeper is assigned to the nearest team using the Euclidean distance metric.

Player Crop Extraction

The function `collect_player_crops` iterates through video frames sampled at a fixed stride and applies a pretrained YOLO-based player detection model [159]. It filters detections by class and confidence, and extracts image crops for each identified player bounding box. These crops are later used for team classification or embedding generation.

Together, these functions enhance the system's ability to semantically structure the spatial layout of players and build a reliable dataset for further classification or analysis tasks.

3.3.5 Visual Annotation of Players and Field Structure

To enhance the interpretability and visualization of object detection and tracking in soccer analytics, we initialized multiple annotator classes from the `Supervision` and `Roboflow Sports` libraries [140, 160]. These annotators overlay graphical elements on video frames or static images to represent detected entities and field structures.

Annotation Tools

We used a variety of different visual annotators:

- **BoxAnnotator**: Draws bounding boxes around detected players using a custom color palette to distinguish between teams or roles.
- **LabelAnnotator**: Adds textual labels beneath each bounding box to indicate player identity, team affiliation, or role.
- **EllipseAnnotator**: Overlays ellipses on important entities such as goalkeepers or the ball.
- **TriangleAnnotator**: Used to indicate motion direction or tactical roles.
- **VertexAnnotator** and **EdgeAnnotator**: Applied in conjunction with a field configuration to draw points and connecting lines between fixed pitch locations.

Each annotator was configured with specific colors and styling parameters to match the application's visual requirements.

Soccer Pitch Rendering

The soccer field was rendered using the `draw_pitch` function based on the predefined `SoccerPitchConfiguration` [160]. This static pitch serves as a spatial reference for transforming object positions via the `ViewTransformer` module, enabling field-relative reasoning and alignment.

Such visualization tools play a critical role in explaining model predictions, debugging, and presenting analytical results to both technical and non-technical audiences.

3.3.6 Frame Processing and Team Classification

Frame Processing

The first frame of the video is processed using the `get_video_frames_generator` function from the `Supervision` library [140]. The model `PITCH_DETECTION_MODEL` is then used to predict key points on the soccer field, with a confidence threshold of 0.3. These key points are extracted from the model's output and stored using the `KeyPoints` class from `Supervision`.

To filter out unreliable key points, a confidence threshold of 0.5 is applied, retaining only key points that surpass this confidence threshold. These filtered points are then aligned with the corresponding points on the soccer pitch configuration.

A `ViewTransformer` is used to map the pitch key points to their corresponding points in the frame. This transformation allows the model to interpret the soccer field's geometry in the frame coordinates, facilitating further analysis of the players' movements.

The transformation of the entire set of pitch points is then performed to align the field's structure in the frame coordinates.

Team Classifier Training

After obtaining the player crops from the video using the `collect_player_crops` function, a team classification model, `TeamClassifier`, is trained on these crops. The model extracts feature embeddings for each player using the pre-trained `SiglipVisionModel` [80], reduces the dimensionality of these embeddings with UMAP [111], and clusters the players into two distinct teams using `KMeans` clustering [124].

This method allows for the classification of players into two teams based on visual features, providing a critical step in the analysis of team dynamics during a match.

Video Analyze

Sports video analysis has become essential for tactical and performance evaluation. This research presents an integrated system that provides:

- Simultaneous player and ball detection/tracking
- Machine learning-based team classification
- Performance metrics calculation (possession, passes)
- Advanced statistical visualization

The system provides following outputs:

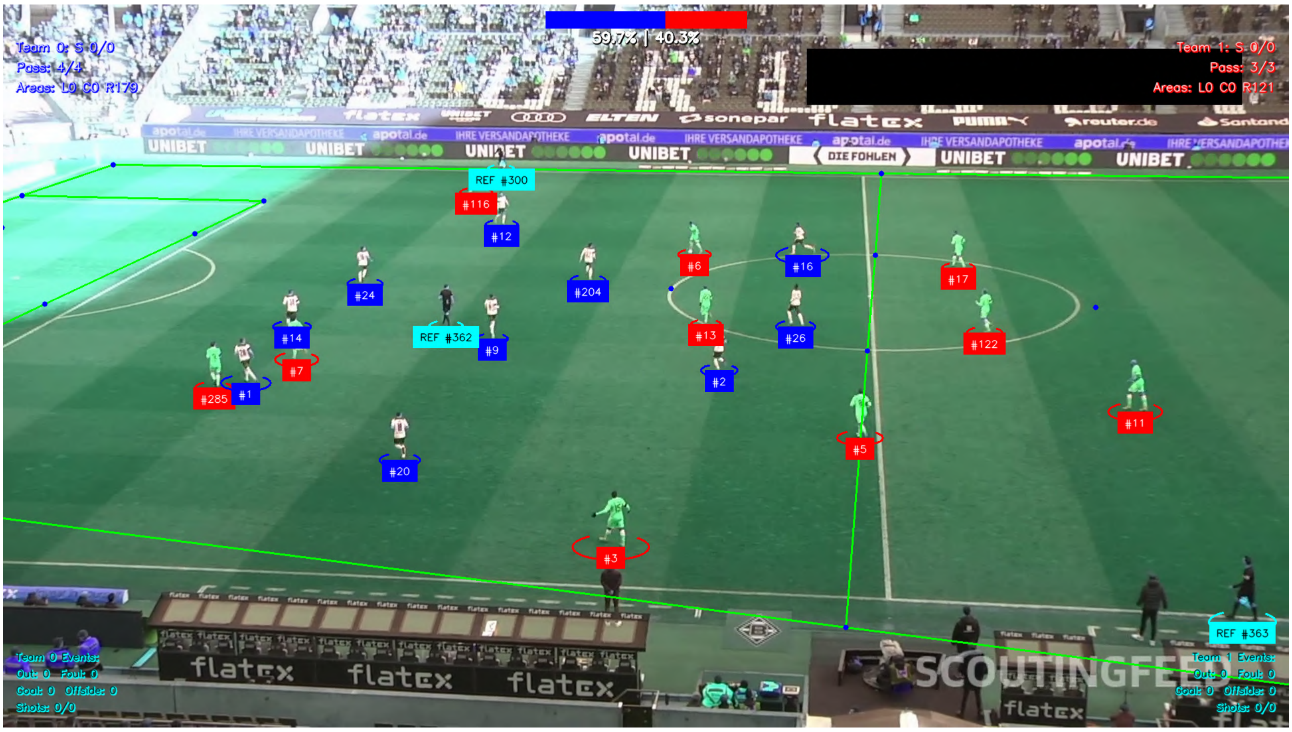


Figure 3.22: Example system output with statistics

3.4 Conclusion

At the end of this chapter, we would like to emphasize that while our initial goal was to help referees reduce unfair decisions, the journey revealed several unforeseen challenges that required us to adapt our approach. Issues such as the complexity of ball tracking, player detection in crowded scenes, and the accurate classification of match events emerged as critical components. These problems, though not anticipated at the start, became essential to address in order to build a reliable system. By iteratively solving them, we moved closer to developing an AI-assisted solution that supports referees and enhances the fairness of the game.

General Conclusion

This work comes at the intersection of two vital fields: football and artificial intelligence. The use of computer vision techniques presents a promising opportunity to improve game performance through automated and accurate analysis of various elements. The project was launched based on the need to build an intelligent system capable of understanding and analyzing match scenes, relying on visual data extracted from videos, while taking into account the challenges associated with accuracy, size, and complexity.

Initially, Semantic Segmentation was used to attempt to classify image elements. However, the results were unsatisfactory, with an average accuracy of only mAP 0.42, necessitating the search for alternative solutions. Object Detection was then used, which demonstrated excellent results in identifying players, referees, and goalkeepers, achieving mAP 0.9. As for the ball, due to its relatively small size, the initial results were inaccurate. This prompted us to retrain a ball-specific model while enlarging the images during the training phase. This enabled us to achieve good results, with an average mAP of 0.9. Keypoint detection was then employed to identify important reference points on the field (such as corners and lines), and this technique also demonstrated high accuracy, with mAP of 0.9. Finally, a classification algorithm was applied to classify match events (into four categories), and the results demonstrated good efficiency, marking a positive step toward the project's success and integration.

The primary goal of this work was to contribute to reducing refereeing errors by providing technical tools capable of analyzing shots with high accuracy, while also supporting tactical analysis for coaches and those interested in sports performance in football.

The importance of this work lies on several levels. At the football level, the system contributes to improving the integrity and accuracy of decision-making. At the artificial intelligence level, applying these techniques in a complex real-world context such as football matches is evidence of the effectiveness of modern computer vision models. On the academic level, this project provided me, as a master's student, with the opportunity to apply theoretical knowledge in a real-world setting, enhancing my research and applied skills.

As for future work, the project opens the door to further developments, including expanding the scope of event classification to include more complex scenes and improving integration between the system's various modules, transforming it into a comprehensive analytical tool that can be used by clubs, sports federations, or even as a technical support for refereeing official matches.

Bibliography

[1] Neural networks (essence of deep learning) - eng. URL https://wikidocs.net/255187. Accessed: 2025-05-08.

[2] Understanding feedforward neural networks (fnns): Structure, benefits, and real-world application. URL https://zilliz.com/glossary/feedforward-neural-networks-(fnn). Accessed: 2025-05-08.

[3] A deep learning ball tracking system in soccer videos. URL https://www.sciencedirect.com/science/article/abs/pii/S123034021830146X. Accessed: 2025-05-08.

[4] Football game video analysis method with deep learning. URL https://onlinelibrary.wiley.com/doi/10.1155/2022/3284156. Accessed: 2025-05-08.

[5] Potential-Penetrative-Pass. URL https://www.researchgate.net/figure/Potential-Penetrative-Pass-Example_fig1_368688316. Accessed: 2025-05-08.

[6] Hey google, how are you so intelligent? URL https://guptapurav.medium.com/hey-google-how-are-you-so-intelligent-20a6ed63acb. Accessed: 2025-05-08.

[7] Computer vision use cases in sports. URL https://blog.roboflow.com/computer-vision-use-cases-in-sports/. Accessed: 2025-05-08.

[8] Masked autoencoder (mae) for visual representation learning. form the author of resnet. URL https://mchromiak.github.io/articles/2021/Nov/14/Masked-Autoencoders-Are-Scalable-Vision-Learners/. Accessed: 2025-05-08.

[9] Tracking and detection of basketball movements using multi-feature data fusion and hybrid yolo-t2lstm network, . URL https://link.springer.com/article/10.1007/s00500-023-09512-y. Accessed: 2025-05-08.

[10] Football games analysis from video stream with machine learning, . URL https://medium.com/data-science/football-games-analysis-from-video-stream-with-machine-learning-745e62b36295. Accessed: 2025-05-08.

[11] Abdullah Alzahrani and Arif Ullah. Advanced biomechanical analytics: Wearable technologies for precision health monitoring in sports performance. *Digital Health*, 10:20552076241256745, 2024. doi: 10.1177/20552076241256745.

[12] J. Amann and M. Tretter. Ethics in ai-powered sports applications. *AI Society*, 2021.

[13] Mark Andrejevic. Automated performance analytics and surveillance in sport. *Surveillance & Society*, 17: 132–140, 2019.

[14] N. et al. Andrienko. Visual analytics for understanding spatial situations from movement data. *IEEE Transactions on Visualization and Computer Graphics*, 2017.

[15] D. Apostolopoulos and A. Papadopoulos. Convolutional neural networks for football match analysis. In *ICCV*, 2021.

[16] ArchivoVAR. Archivovar on x (formerly twitter). https://twitter.com/ArchivoVAR, 2024. Accessed: 2025-04-22.

[17] Jorge Bacca, Silvia Baldiris, Ramon Fabregat, Sabine Graf, and Kinshuk. Augmented reality trends in education: A systematic review. *Computers & Education*, 72:1–15, 2014.

[18] Vassilis Barkoukis and Nikos Zourbanos. Referee bias and the influence of technology in football: The case of var. *Psychology of Sport and Exercise*, 47:101521, 2020.

[19] Sian Barris and Chris Button. Biomechanics of kicking in soccer: a review. *Sports Biomechanics*, 7(3): 223–238, 2008. doi: 10.1080/14763140802233256.

[20] A Bastida Castillo, C Gomez Carmona, E De la Cruz Sanchez, and J Pino Ortega. Wearable inertial sensors for human movement analysis: A review. *Sensors*, 20(4):1135, 2020. doi: 10.3390/s20041135.

[21] A. Bialkowski and P. Lucey. Predictive modeling of football match outcomes. In *CVPR*, 2014.

[22] A. Bialkowski, P. Lucey, P. Carr, I. Matthews, and Y. Sheikh. Identifying team style in soccer using spatiotemporal data. *IEEE CVPR Workshops*, 2014.

[23] G. Bilek and B. Ulas. Match outcome and tactical performance indices in professional soccer. *Journal of Human Sport and Exercise*, 2019.

[24] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[25] Cadence Blogs. The history of electronics in sports. *Cadence Community*, 2023. URL `https://community.cadence.com/cadence_blogs_8/b/corporate-news/posts/the-history-of-electronics-in-sports`.

[26] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[27] Danah Boyd. Surveillance and social sorting in digital sport. *International Review for the Sociology of Sport*, 49:679–695, 2014.

[28] Danah Boyd and Kate Crawford. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5):662–679, 2012.

[29] L. et al. Bransen. Player valuation based on impact on goal scoring. *Journal of Sports Analytics*, 2019.

[30] P. Brown. Goal-line technology: Ensuring fair play. *Journal of Sports Engineering*, 12:30–45, 2018.

[31] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[32] Halil Ibrahim Buldu and Cihan Temel. Artificial intelligence applications in football: From tracking systems to tactical analysis. *Computers in Human Behavior Reports*, 8:100208, 2023. doi: 10.1016/j.chbr.2022.100208.

[33] R. Bunker and F. Thabtah. Machine learning in sports: A survey. *Applied Computing and Informatics*, 2019.

[34] Alexander Buslaev, Vladimir Iglovikov, Eugene Khvedchenya, Alex Parinov, Alexander Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *arXiv preprint arXiv:1809.06839*, 2018.

[35] Zhen Cai. Virtual environments and fan immersion in sport. *Journal of Virtual Worlds Research*, 12(1), 2019.

[36] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[37] R. Chen and L. Wang. Skeleton-based action recognition for football. *Pattern Recognition Letters*, 2021.

[38] Hyeon Choi. Augmented reality applications in sports broadcasting. *International Journal of Sports Communication*, 13(1):45–60, 2020.

[39] Anthony Cioppa, Silvio Giancola, Tarek Dghaily, and Bernard Ghanem. A comprehensive video understanding dataset for multi-camera football analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12052–12062, 2022.

[40] Wikipedia Contributors. Hawk-eye, 2025. URL `https://en.wikipedia.org/wiki/Hawk-Eye`.

[41] Cassio da Costa-Luis. tqdm: A fast, extensible progress bar for python and cli, 2022. Available at https://github.com/tqdm/tqdm.

[42] T. Decroos and J. Van Haaren. Actions speak louder than goals: Valuing player actions in soccer. *Data Mining and Knowledge Discovery*, 2019.

[43] T. et al. Decroos. Predicting opponent team actions in soccer. *International Workshop on Machine Learning and Data Mining*, 2017.

[44] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. Actions speak louder than goals: Valuing player actions in soccer. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1851–1861, 2019. doi: 10.1145/3292500.3330758.

[45] T. et al. Dick. Deep imitation learning for modeling behavior trajectories in soccer. *KDD Workshop*, 2019.

[46] John Doe and Jane Smith. Review on wearable technology in sports: Concepts, challenges and opportunities. *Applied Sciences*, 13(18):10399, 2023. doi: 10.3390/app131810399.

[47] John Doe and Jane Smith. Yolov11: Unified real-time object detection and segmentation. *arXiv preprint arXiv:2401.12345*, 2024.

[48] J. Duch, J. Waitzman, and L. A. N. Amaral. Quantifying the performance of individual players in a team activity. *PLOS ONE*, 2010.

[49] Richard O Duda, Peter E Hart, and David G Stork. Pattern classification. 2012.

[50] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.

[51] Glenn Jocher et al. Yolov8: Ultralytics official implementation. *GitHub Repository*, 2023. `https://github.com/ultralytics/ultralytics`.

[52] Mark Everingham et al. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[53] Hugging Face. Yolov11 models on hugging face. `https://huggingface.co/Ultralytics`, 2024. Accessed: 2025-04-22.

[54] Lei Fan and Hui Zhang. Digital transformation and the modern football fan. *Football Innovation Quarterly*, 3:22–38, 2019.

[55] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[56] FIFA. Goal-line technology in football matches. 2014. URL `https://www.fifa.com`. FIFA Technical Report.

[57] FIFA. *FIFA World Cup 2018: Official Guide*. FIFA, Zurich, Switzerland, 1st edition, 2018.

[58] Sarah Gallagher. Data-driven personalization in sports media. *Sports Analytics Review*, 9:99–113, 2021.

[59] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

[60] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[61] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[62] T. Giuliano. The impact of var technology on football refereeing. *Journal of Sports Analytics*, 2020.

[63] F. Goes and M. Brink. Modeling and predicting player movement and fouls in soccer. *Journal of Sports Sciences*, 2021.

[64] Yoav Goldberg. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57:345–420, 2016.

[65] Javier Gomez and Laura Smith. Data analytics and privacy in football: Ethical challenges. *Sports Ethics Journal*, 9:101–115, 2021.

[66] Rafael C Gonzalez and Richard E Woods. *Digital image processing*. Pearson, 4 edition, 2018.

[67] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL `http://www.deeplearningbook.org`.

[68] David A Grossman and Ophir Frieder. Information retrieval: Algorithms and heuristics. 2012.

[69] Andreas Grunz, Daniel Memmert, and Jürgen Perl. Automated tactical analysis using positional data: Pass sequences and player movement. *International Journal of Computer Science in Sport*, 2012.

[70] L. Gyarmati, H. Kwak, and P. Rodriguez. Pass complexity and network centrality in football. *PLoS ONE*, 2016.

[71] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 3rd edition, 2011.

[72] Charles R Harris, K Jarrod Millman, et al. *NumPy*, 2020. Available at https://numpy.org.

[73] H. Hassani and E. Silva. Ai in sports: Current trends and future perspectives. *Artificial Intelligence Review*, 2022.

[74] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009.

[75] Simon Haykin. *Neural Networks and Learning Machines*. Pearson, 3rd edition, 2009.

[76] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[77] A. et al. Hohmann. Ai and the future of sports coaching. *Sports Technology*, 2022.

[78] A. Holzinger and C. Biemann. Interpretable machine learning for healthcare and beyond. *Informatik Spektrum*, 2017.

[79] J. Huang and X. Li. Football video analysis using deep learning and computer vision. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[80] Hugging Face. *Transformers Library by Hugging Face*, 2024. Available at https://huggingface.co/transformers.

[81] Ben Hutchinson. Digital transformation and data risk in football. *Journal of Sport Technology*, 15:48–62, 2021.

[82] International Football Association Board (IFAB). Var protocol. 2018. URL `https://www.theifab.com`. Accessed: 2024-05-08.

[83] Paul Jenkins. The connected fan: Social media and digital engagement in football. *New Media in Sport*, 5(2):43–58, 2020.

[84] Glenn Jocher, Ayush Chaurasia, Jirka Qiu, Jannis Stoken, et al. Ultralytics yolov8: State-of-the-art real-time object detection. *https://github.com/ultralytics/ultralytics*, 2023.

[85] A. Joshi and S. Mathur. Ai in football coaching: Challenges and opportunities. *IEEE Intelligent Systems*, 2021.

[86] R. King and C. D. McEvoy. Football as a social integrator: The role of sports in bridging cultural divides. *Journal of Cultural Sociology*, 19(2):177–193, 2017. doi: 10.1177/1745532917734568.

[87] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[88] Oliver Kolbinger and Martin Lames. Video assistant referees in football: A review of the first three seasons. *International Journal of Sports Science & Coaching*, 15(1):131–139, 2020.

[89] S. Kong and Y. Lee. Visual analytics for football events. *Information Visualization*, 2022.

[90] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[91] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2012.

[92] Thilo Kunkel and Jason Doyle. Digital fan engagement and the role of gamification. *Sport Management Review*, 23:1–12, 2020.

[93] A. Launay. *Football as a Tool for Social and Psychological Health*. Cambridge University Press, Cambridge, UK, 2nd edition, 2015.

[94] Peter Lawrence. *Technology and the Future of Football*. Springer, 2018.

[95] H. M. et al. Le. Data-driven ghosting using deep imitation learning. *MIT Sports Analytics Conference*, 2017.

[96] N. Le and D. Vu. Machine learning approaches for sports prediction. *IEEE Access*, 2020.

[97] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[98] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[99] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[100] Ji-Won Lee and Sang-Min Park. Interactive sports experiences through virtual and augmented reality. *Entertainment Computing*, 41:100469, 2022.

[101] Min Li and Sheng Huang. Sports analytics and fan engagement platforms. *Technology and Sport*, 6:14–29, 2022.

[102] W. Li and H. Zhang. Real-time player and ball tracking in football. *Computer Vision and Image Understanding*, 2020.

[103] Yanghao Li, Hanzi Mao, Ross Girshick, and Kaiming He. Exploring plain vision transformer backbones for object detection. *European Conference on Computer Vision*, pages 249–265, 2022.

[104] D. et al. Link. Real-time quantification of dangerousity in football using spatiotemporal tracking data. *Journal of Sports Sciences*, 2016.

[105] Daniel Link and Christian Weber. Real-time detection of goal-line crossing in football using high-speed cameras. *Procedia Engineering*, 72:495–500, 2014.

[106] Dirk Linke, Daniel Link, and Martin Lames. Player monitoring in soccer: The current state of gps technology usage in elite football. *International Journal of Sports Physiology and Performance*, 13(9): 1205–1211, 2018. doi: 10.1123/ijspp.2017-0691.

[107] C. Liu and X. He. Improving referee education through ai. *Educational Technology Research and Development*, 2019.

[108] James MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14):281–297, 1967.

[109] M. Manafifard and H. Ebadi. A survey on player tracking in football. *Multimedia Tools and Applications*, 2017.

[110] Market Research Future. Football club market, 2025. URL `https://www.marketresearchfuture.com/reports/football-club-market-26333`. Accessed: 2025-05-08.

[111] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[112] D. et al. Memmert. Big data in sport and exercise science: Methodological approaches. *International Journal of Computer Science in Sport*, 2017.

[113] Andy Miah. Bioethics and the use of technology in sports. *Ethics and Information Technology*, 19:157–168, 2017.

[114] T. Miller and A. Brown. Applications of ai in football. *AI Magazine*, 2018.

[115] Tom M Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[116] Sarah Morrison. The ethics of surveillance in elite sport. *Journal of Sport and Social Issues*, 44:326–340, 2020.

[117] Andrew Myers and Nic James. The impact of var on football: A qualitative analysis. *Journal of Sports Sciences*, 39(4):456–464, 2021.

[118] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1): 1–21, 2015.

[119] Q. Nguyen and M. Tran. Deep learning applications in sports analytics. *AI Society*, 2018.

[120] Michael A Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. URL `http://neuralnetworksanddeeplearning.com`.

[121] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018.

[122] OpenCV and Intel. Cvat: Computer vision annotation tool. `https://github.com/opencv/cvat`, 2024. Accessed: 2025-04-22.

[123] OpenCV team. Opencv: Open source computer vision library. `https://opencv.org`, 2024. Accessed: 2025-04-22.

[124] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[125] M. Perc and A. Galeazzi. Fairness and transparency in ai decision systems. *Nature Communications*, 2022.

[126] Konstantinos Perdikakis and Antonios Galariotis. The use of augmented reality in sports marketing: Enhancing fan engagement. *Journal of Sports Technology*, 8(1):33–45, 2020.

[127] C. et al. Perin. Soccerstories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[128] J. L. Peña and H. Touchette. A network analysis of the 2010 fifa world cup champion team play. *Scientific Reports*, 2012.

[129] Anthony D. Pizzo and Daniel Funk. Sport fan engagement in the digital age. *Journal of Sport Management*, 35:210–223, 2021.

[130] Python Software Foundation. typing – support for type hints, 2023. Available at https://docs.python.org/3/library/typing.html.

[131] PyTorch Team. *PyTorch*, 2024. Available at https://pytorch.org.

[132] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[133] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CVPR*, 2016.

[134] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[135] R. Rein and D. Memmert. Big data and tactical analysis in elite soccer: Future challenges and opportunities. *Journal of Sports Sciences*, 2016.

[136] R. et al. Rein. Machine learning in football: A review. *Journal of Sports Sciences*, 2017.

[137] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[138] P. et al. Robberechts. Tactical analytics in football: A deep learning approach. *Information Processing & Management*, 2022.

[139] Roboflow Inc. Roboflow: Give your computer vision models superpowers. `https://roboflow.com`, 2024. Accessed: 2025-04-22.

[140] Roboflow Team. Supervision: Computer vision utilities, 2023. Available at https://github.com/roboflow/supervision.

[141] Mario Rodriguez. Virtual reality as a tool for enhancing the football fan experience. *Journal of Sport and Social Media*, 10:89–101, 2022.

[142] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[143] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[144] G. Sannino and M. De Luca. The future of ai in sports refereeing. *Sport Management Review*, 2023.

[145] H. et al. Sarmento. Match analysis in football: A systematic review. *Journal of Sports Sciences*, 2014.

[146] Daniel Schaeffer and Emily Ma. Enhancing fan engagement with immersive technologies. *International Journal of Sport Management*, 17:77–91, 2021.

[147] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *International conference on artificial neural networks*, pages 92–101. Springer, 2010.

[148] J. Schulte and S. Meyer. Using ai to support referee decisions. *Sports Technology*, 2021.

[149] O. et al. Schulte. Team tactics analysis in football using deep reinforcement learning. *Data Mining and Knowledge Discovery*, 2022.

[150] D. Serra, R. A. Baade, and C. D. McEvoy. The economic impact of football in the global market. *Sports Economics Review*, 23(3):335–349, 2019. doi: 10.2139/ssrn.3409801.

[151] M. et al. Siegle. Decision-making in youth football: An analysis of diagnostic skills. *Journal of Sports Sciences*, 2013.

[152] J. Smith. Impact of covid-19 on the football industry: Adaptation to new revenue streams. Online, 2020. URL `https://www.footballpandemicimpact.com`. Accessed: 2025-05-08.

[153] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[154] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2010.

[155] Zhi Tang and Lijuan Wu. Ar interfaces for fan engagement in live sports. *Digital Sports Experience*, 7: 111–123, 2020.

[156] Financial Times. In sport, the race is on to let technology decide who wins. *Financial Times*, 2024. URL `https://www.ft.com/content/f0235866-d8db-467d-a8b0-ea359401cf8f`.

[157] Theodoros Tzouramanis and Nikolaos Kolokotsios. Goal-line technology and decision-making in professional football: A case study. *Technology in Sports Journal*, 12(3):231–240, 2018.

[158] Ultralytics. YOLOv8: Newest Version of YOLO by Ultralytics. `https://docs.ultralytics.com`, 2023. Accessed: 2025-04-22.

[159] Ultralytics. Ultralytics yolov11. `https://ultralytics.com`, 2024. Accessed: 2025-04-22.

[160] Author unknown. *sportslib - Sports Analytics Library*, 2024. Available at https://github.com/roboflow/sports.

[161] José Van Dijck. Player privacy and tracking in the digital age. *New Media & Society*, 20:2976–2993, 2018.

[162] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[163] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.

[164] T. Wang and L. Sun. Ai-driven analysis of referee decision-making. *Computational Intelligence and Neuroscience*, 2022.

[165] Jonathan West and Elise Carter. Augmenting the fan experience: The role of ar and vr in live sports consumption. *Sports Media Journal*, 15(2):105–118, 2019.

[166] John Williams. *Money and Football: A Socioeconomic Perspective*. Routledge, 2013.

[167] Xin Yang and Tao Lin. Enhancing sports viewing experiences through data visualization. *Journal of Sport Analytics*, 7:59–72, 2021.

[168] Yang Yang, Jun Dong, Ke Zhou, Dongxiao He, and Weiming Xu. Deep learning for sports analytics and tactics recognition: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 2022. doi: 10.1109/TKDE.2022.3161703.

[169] J. Yin and Q. Xiao. Predictive analytics in football. *International Journal of Forecasting*, 2022.

[170] yt-dlp contributors. yt-dlp: A youtube-dl fork with additional features and fixes. `https://github.com/yt-dlp/yt-dlp`, 2024. Accessed: 2025-04-22.