Simon Wand
Matrikelnr.: **5378012**

Supervisor Jan-Phillipp Tauscher
Institut für Computergraphik
TU Braunschweig

Referee Prof. Dr.-Ing. Marcus Magnor
Institut für Computergraphik
TU Braunschweig

Football Vision: Automated Player and Ball Tracking for Tactical Analysis in Football

Project Thesis

September 9, 2025

Computer Graphics Lab, TU Braunschweig

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Braunschweig, 9. September 2025

Simon Wand

Zusammenfassung

Diese Arbeit stellt eine End-to-End-Pipeline vor, die Einzelkamera-TV-Übertragungen von Fußballspielen in Pfade auf einem metrischen Spielfeld überführt und daraus physische und taktische Erkenntnisse ableitet. Jeder Frame wird von einem domänenangepassten YOLOv8-Detektor verarbeitet. Die Objekte werden klassenweise verfolgt (Spieler, Torhüter, Hauptschiedsrichter, Linienrichter, Ball, Sonstige): BoT-SORT für Personen und BoostTrack für den Ball. Die Teamzuordnung erfolgt über K-Means-Clustering der Trikotfarben. Für stabilere Identitäten liest eine PARSeq-OCR die Rückennummern, gesteuert durch eine Lesbarkeitsprüfung und Mehrheits-Voting. Zur Abbildung der Video-Tracklets auf Feldkoordinaten schätzt PnLCalib eine Homographie pro Frame; ein einfacher Stabilitätscheck unterdrückt plötzliche Sprünge. Die gemappten Daten werden anschließend für Bewegungsbahnen, Geschwindigkeitszonen, Positions-Heatmaps, Ballbesitzschätzung, Ereigniserkennung und Formationserkennung ausgewertet.

Die Pipeline wird an einer vollständigen 45-minütigen Halbzeit eines Spiels der Schweizer Super League getestet. Erkennung und Tracking sind auf Broadcast-Material robust, und die Identitätskonsistenz verbessert sich durch die Trikotnummernerkennung. Die Laufzeit beträgt etwa 341.2 ms/Frame (2.93 FPS) und wird von der Homographieschätzung und der Tikotnummererkennung dominiert. Damit ist das System für Offline-Analysen geeignet und nah an der Einsatzfähigkeit für Live-Overlays im Broadcast-Video ohne Perspektivtransformation mit 21.6 FPS.

Abstract

This thesis introduces an end-to-end pipeline that turns single-camera TV broadcasts of football matches into trajectories on a metric pitch and then into physical and tactical insights. Each frame is processed by a domain-adapted YOLOv8 detector. The individual objects are tracked per class (players, goalkeepers, main referee, side referee, ball, and other): BoT-SORT for people and BoostTrack for the ball. Players are assigned to teams with K-Means clustering on jersey colors. For more stable identities, a PARSeq OCR reads jersey numbers with a legibility gate and temporal voting. To map image tracklets to field coordinates, PnLCalib is used to estimate a per-frame homography and a simple stability check to filter sudden jumps. The mapped data gets analyzed to get movements, speed zones, positional heatmaps, possession estimation, event detection and formation estimation.

The pipeline is tested on a full 45-minute half from a Swiss Super League match. Detection and tracking are robust on broadcast footage and identity consistency is improved through jersey number recognition. The runtime is 341 ms per frame (2.93 FPS), dominated by homography estimation and jerseynumber OCR, which is suitable for offline analytics and close to near-live use for overlays on the broadcast video without perspective transformation with 21.6 FPS.

Contents

1 Introduction **1**

2 Related Work **3**
2.1 Object Detection 3
2.1.1 Traditional Methods 3
2.1.2 Deep Learning 4
2.2 Multiple Object Tracking 4
2.2.1 Tracking-by-Detection 5
2.2.2 Joint Detection & Tracking 5
2.3 Tracking of Football Matches 5
2.3.1 Broadcast Video vs. Multi-Camera Systems 6
2.3.2 Re-Identification of Players 6
2.3.3 Ball-Tracking 6
2.4 Camera Calibration & Homography 7
2.5 Transform Tracking Data into Statistical Insights 8
2.5.1 Physical Performance Metrics 8
2.5.2 Tactical Performance Metrics 8

3 System Architecture & Methods **11**
3.1 Dataset 11
3.2 Pipeline Overview 11
3.3 Getting Tracking Data 13
3.3.1 Object Detection 13
3.3.2 Class-Aware Tracking & ID Stabilization 13
3.3.3 Homography from Broadcast Video to 2D-Coordinates 15
3.4 Data Analysis 16
3.4.1 Data Preprocessing 16
3.4.2 Player Path Visualization & Speed-Zones 16
3.4.3 Positional Heatmaps 16
3.4.4 Event Detection 16
3.4.5 Formation Detection 17

4 Experiments & Results **19**
4.1 Detection Results 20
4.2 Tracking Results 21
4.3 OCR Results 21
4.4 Runtime 21

5 Case Study **23**

6 Discussion & Conclusion **27**
6.1 Limitations 27
6.2 Future Work 27
6.3 Conclusion 28

A Source Code **35**

List of Figures

3.1 Football Vision Flowchart 12
3.2 Live class-aware Tracking output 13
3.3 2D transformed tracking output 15

5.1 Movement path of player 19 of Team 'R' 24
5.2 4x4 positional heatmap for player 19 of team 'R' 25

List of Tables

4.1 Finetuned YOLOv8n results on the SoccerNet Tracking test set. 20
4.2 Finetuned YOLOv8n results on the SoccerNet Tracking challenge set. 20
4.3 Performance of the pipelines individual modules. 21

5.1 Speed-zone distribution for player 19 of Team 'R'. 23
5.2 Event types and counter. 24

Chapter 1

Introduction

Modern football analytics increasingly relies on positional data to quantify physical performance and to analyze tactical behavior. While professional clubs often acquire such data trough manual systems or wearables, access to these solutions is limited by cost and data-sharing constraints. This project adresses that gap by proposing a broadcast-only, end-to-end pipeline that turns TV footage into physical and tactical performance metrics.

Working with broadcast video presents unique challenges compared to controlled manual systems. Camera motion and partial visibility of objects reduce the detection and tracking quality. The ball is a small and fast object, which can lead to many false negatives and identity switches. Player crops change their size when the camera zooms and jersey appearance can change due to lightning or motion blur. In addition to that, even small errors in the pitch homography can lead to drastic inaccuracies in spatial metrics. The core difficulty in this approach is therefore probably not a single component, but the interaction between all components of the pipeline.

Despite these challenges, broadcast-based pipelines offer two advantages. First, they are scalable: historical and live footage is abundant across leagues and levels, enabling analyses without additional hardware. Second, they are cost-effective, which opens the door for amateur and semi-professional contexts.

The work is guided by three questions:

- How reliable are the resulting 2D trajectories of players and the ball?
- How can tracking stability and identity consistency be improved?
- Which physical and tactical metrics can be derived from the generated data?

The thesis makes four contributions:

1. Detector: a YOLOv8n model fine-tuned to football broadcast footage.

2. Tracking: a class-aware tracking stack with team clustering and jersey number recognition.
3. Mapping: a robust, temporally smoothed PnLCalib homography.
4. Analytics: an implementation of physical and tactical metrics.

Chapter 2

Related Work

2.1 Object Detection

2.1.1 Traditional Methods

Before deep learning became popular, object detection methods relied mainly on manually designed features, typically focusing on edges, shapes, and textures. One of the most influential methods was the Scale-Invariant Feature Transform (SIFT), introduced by Lowe (2004). SIFT provided robust features that were stable across changes in scale, rotation, and lighting conditions, which made it particularly useful for matching objects across different views or settings [Low04]. Another widely adopted technique was the Histogram of Oriented Gradients (HOG) by Dalal and Triggs (2005), which effectively captured local shapes using gradient orientation histograms and proved especially successful for detecting humans in images [DT05]. A significant advancement towards real-time detection was made by Viola and Jones (2001), who proposed using simple Haar-like features computed efficiently through an integral image representation, combined with AdaBoost classifiers. Their approach enabled rapid face detection, demonstrating practicality for real-world applications [VJ01]. Later, the Deformable Part Model (DPM) by Felzenszwalb et al. (2008) improved detection capabilities by modeling objects as collections of flexible parts. This allowed the method to handle variations in object poses more effectively, achieving better performance on challenging datasets compared to earlier rigid templates [FMR08]. While these traditional approaches set essential foundations, they often struggled with variations in real-world scenarios and required extensive manual tuning. These limitations eventually motivated the transition toward more powerful deep learning-based methods [Neh+24].

2.1.2 Deep Learning

Deep learning has significantly advanced object detection, primarily through two categories: two-stage and one-stage approaches. Two-stage methods first generate region proposals and then classify these proposals. The seminal work R-CNN introduced a pipeline where convolutional neural networks (CNNs) were applied to region proposals, dramatically improving detection accuracy over traditional methods by leveraging deep, hierarchical features extracted by CNNs [Gir+14]. Building on this, Fast R-CNN optimized this process by sharing computations across region proposals, significantly enhancing training and inference speeds while achieving higher accuracy [Gir15]. Further efficiency was achieved by Faster R-CNN, which introduced Region Proposal Networks (RPNs), integrating proposal generation into the network and achieving near real-time performance while maintaining high accuracy [Ren+15]. Mask R-CNN extended this further, adding a segmentation branch alongside object classification and localization, enabling precise pixel-level instance segmentation and improved detection performance, particularly in cluttered environments [He+17].

In contrast, one-stage methods predict bounding boxes and class probabilities directly from the input image in a single pass, offering a balance of speed and accuracy. YOLO (You Only Look Once) significantly transformed object detection by reframing it as a regression task, directly predicting object locations and classes with remarkable speed suitable for real-time applications [Red+16]. Subsequent improvements in YOLOv4 integrated advanced training techniques and network modifications like CSPDarknet and Mosaic data augmentation, pushing detection accuracy further without sacrificing real-time capabilities [BWL20]. YOLOv8 brought additional refinements through anchor-free approaches and enhanced feature extraction techniques, significantly improving accuracy and efficiency compared to earlier versions [Yas24]. The latest iteration, YOLOv11, introduced novel architectural components such as the C3k2 block and Spatial Pyramid Pooling-Fast (SPPF), further boosting performance, versatility across tasks, and accuracy in object detection [KH24].

2.2 Multiple Object Tracking

Multiple object tracking (MOT) involves identifying and following multiple targets throughout a video sequence, often formulated as a tracking-by-detection task, where detections are first produced independently and then associated over time. Methods can generally be classified into two categories: Tracking-by-Detection and Joint Detection & Tracking.

2.2.1 Tracking-by-Detection

Tracking-by-detection methods first perform detection in each frame and then match these detections across frames to form continuous trajectories. SORT (Simple Online and Realtime Tracking) introduced a framework that employs the Kalman filter for object motion prediction and the Hungarian algorithm for frame-by-frame data association based on intersection-over-union (IoU) of detections, providing a fast and efficient solution for online tracking [Bew+16]. To enhance robustness, Deep SORT integrates visual appearance features with motion cues by using a convolutional neural network to extract appearance embeddings, significantly reducing identity switches compared to the original SORT method [WB18]. ByteTrack introduced a hierarchical data association strategy, first associating high-confidence detections and then low-confidence ones to reduce identity switches and improve tracking robustness, especially in challenging scenarios with occlusion or poor detections [Zha+22]. Similarly, BoostTrack enhances traditional similarity measures and detection confidences, incorporating Mahalanobis distance, shape similarity, and detection-tracklet confidence to boost accuracy in one-stage association setups, outperforming existing online trackers significantly [ST24].

2.2.2 Joint Detection & Tracking

Joint Detection & Tracking methods integrate the detection and tracking components into a unified framework, allowing for simultaneous object localization and tracking. FairMOT advocates for fairness between detection and re-identification (re-ID), highlighting that traditional anchor-based detection approaches can impair the tracking performance by creating ambiguity in re-ID feature learning. Using an anchor-free approach and maintaining the same emphasis on detection and re-ID tasks, FairMOT significantly improves tracking accuracy, especially in crowded environments [Zha+21]. Similarly, BoT-SORT integrates robust appearance information and camera-motion compensation with an advanced Kalman filter state vector to achieve state-of-the-art results on MOT benchmarks. This method successfully balances between detection accuracy (MOTA) and identity preservation (IDF1), demonstrating the importance of carefully fused motion and appearance cues in enhancing MOT performance [AOB22].

2.3 Tracking of Football Matches

Tracking football players and the ball from broadcast or multi-camera footage is crucial for tactical analyses, performance assessment, and strategic decision-making. This task presents unique challenges due to factors such as rapid movements, frequent occlusions, and varying camera angles.

2.3.1 Broadcast Video vs. Multi-Camera Systems

The primary methods used to track football matches include broadcast video tracking and multi-camera systems. Broadcast videos offer a practical and cost-effective approach since no dedicated camera setups are required. However, tracking accuracy can be compromised due to varying camera viewpoints, zoom levels, and missing players outside the broadcasted frame [Seo+97]. Conversely, multi-camera setups, despite their higher complexity and computational cost, provide better coverage and reduce occlusion problems by capturing a more complete view of the pitch. Systems such as Sentioscope successfully employ dual-camera setups to mitigate occlusion effects and improve tracking accuracy under diverse illumination conditions and crowded player scenarios [BD16].

2.3.2 Re-Identification of Players

Re-identification (Re-ID) is a critical component of player tracking, particularly in cases of occlusion or when players move between camera views. Traditional color histogram-based methods struggle with consistent re-identification due to similar team jerseys and varying illumination conditions. Recent methods have adopted machine learning approaches, utilizing robust visual features and deep learning-based re-identification. Approaches such as Deep SORT integrate visual appearance features extracted from convolutional neural networks, significantly reducing identity switches compared to purely motion-based trackers [NFK15] [BD16]. Additionally, methods leveraging appearance embeddings, like FairMOT and ByteTrack, have improved Re-ID capabilities by balancing detection accuracy and feature discrimination, enhancing the robustness of tracking in challenging scenarios [NFK15] [BD16].

2.3.3 Ball-Tracking

Ball-tracking in soccer presents unique challenges due to its small size, rapid motion, and frequent occlusion by players. Traditional methods relying solely on motion or color have limited success. Approaches integrating object detection frameworks such as YOLO (You Only Look Once) combined with tracking algorithms like SORT (Simple Online Real-Time Tracking) have shown promising results. Specifically, a YOLOv3-SORT-based approach effectively classifies and tracks the ball even at high velocities, achieving high tracking accuracy in challenging situations [Seo+97]. Furthermore, enhanced methodologies that incorporate Kalman filters for predicting ball trajectories and deep learning for robust feature extraction demonstrate significant improvement in ball tracking reliability [TB21].

These advancements collectively underline the importance of integrated approaches combining efficient tracking methods, robust player re-identification

techniques, and reliable ball tracking to deliver accurate insights into football match dynamics and tactical performances.

2.4 Camera Calibration & Homography

In broadcast football, mapping image coordinates to a canonical 2D pitch is the bridge between perception and analytics. Three recent directions dominate:

- search-based calibration from synthetic pose dictionaries with learned retrieval and local refinement
- keypoint- and line-driven estimation with non-linear optimization
- keypoint-less, differentiable calibration

Chen and Little propose a highly automatic, single-image sports calibration method that builds a feature-pose database from synthetic edge renderings of the pitch and a physically interpretable PTZ camera model. Key design choices are:

(1) a camera pose engine that reduces the effective degrees of freedom to three significant parameters - focal length, pan, tilt - by exploiting strong broadcast priors on camera placement.

(2) a siamese network that embeds edge images into a compact feature space for nearest-neighbour pose retrieval.

(3) a two-GAN pipeline to segment the playing surface and detect field markings in real images, followed by Lucas–Kanade refinement on truncated distance images to fine-tune the homography.

This yields robust performance across soccer datasets while avoiding heavy manual initialization. Limitations are the reliance on priors (main-camera assumptions) and the need for a sizable synthetic dictionary to cover non-standard views[CL18].

PnLCalib reframes calibration as point-and-line optimization on a 3D field model. The method first predicts keypoints and line extremities (HRNet-based encoder–decoder) to assemble a rich set of geometric primitives: line-line intersections, line-ellipse intersections, ellipse tangents, and additional axial points using SoccerNet's line notation and hierarchical disambiguation. An initial projection provides a coarse estimate, then a nonlinear least-squares refinement jointly fits points and lines, improving both 3D camera calibration and planar homography. Extensive results on SoccerNet-Calibration, WorldCup 2014, and TS-WorldCup show state-of-the-art 3D calibration with competitive homography accuracy, and the pipeline explicitly handles multiview broadcast conditions. Dependencies include accurate line/ellipse extraction and careful disambiguation of key points under occlusion[GA].

TVCalib argues for treating sports field registration as full camera calibration rather than homography-only. After instance segmentation of individual field segments (lines, circle arcs, goal frames), the method defines a differentiable segment reprojection loss and directly optimizes the pinhole camera parameters (FoV, translation, pan/tilt/roll) — and optionally lens distortion — without explicit keypoint correspondences. Practical engineering adds multiple initializations to hedge against local minima and self-verification (loss-based rejection) to improve robustness. On SoccerNet-Calibration and WorldCup 2014, TVCalib reports superior or competitive registration while recovering camera intrinsics and extrinsics in one step[TE22].

2.5 Transform Tracking Data into Statistical Insights

2.5.1 Physical Performance Metrics

The physical demands of football require precise measurement of player performance, often captured through spatiotemporal tracking data. Metrics such as total distance covered, high-intensity running distances, sprint counts, and accelerations provide essential insights into player workloads and their physical output during matches [Fil+17]. Additionally, the physical efficiency index (PEI), capturing the effectiveness of player movements, has been employed to relate player physical output to match outcomes. Recent analyses demonstrate a significant correlation between physical output indicators and team performance, emphasizing the importance of efficient physical exertion in successful match outcomes. Notably, superior physical efficiency, particularly in terms of sprint intensity and distance covered at high speeds, is frequently associated with winning teams[Fil+17].

Moreover, Filetti et al. (2017) reported significant relationships between technical-tactical efficiency and physical efficiency, indicating that although physical performance strongly contributes to successful outcomes, it must be effectively combined with tactical and technical proficiency to maximize success on the pitch [Fil+17].

2.5.2 Tactical Performance Metrics

Tracking data not only captures physical aspects but also reveals deeper insights into tactical behaviors and performance. Tactical performance metrics derived from positional and event data enable a granular analysis of team behaviors, such as formations, role assignments, and tactical flexibility throughout the game. Bialkowski et al. introduced methods that identify and dynamically track players' roles, enabling context-specific analyses of tactical behaviors and formations. Their approach used minimum entropy data partitioning to automatically identify player roles and team formations

directly from positional data, allowing for the detection of strategic patterns and providing valuable context for individual and collective player analysis [Bia+14]. Furthermore, tactical performance analysis extends to evaluating the effectiveness of scoring opportunities through Expected Goals (xG) models, which quantify the probability of goal-scoring chances based on synchronized positional and event data. Anzer and Bauer (2021) presented a high-performing xG model utilizing positional data, demonstrating greater accuracy than traditional methods and providing teams with detailed insights into shot quality and team offensive efficiency [AB21].

Another critical aspect of tactical analysis involves assessing passing effectiveness. Spearman et al. developed physics-based modeling approaches for evaluating pass probabilities, accounting for factors such as defender pressure, spatial positioning, and the technical execution of passes. Their findings reveal that effective passing contributes substantially to team success, with pass probability metrics strongly correlating with successful match outcomes [Spe+17].

Additionally, advanced tactical analytics have focused on quantifying defensive disruption and offensive superiority. Goes et al. (2021) and Meerhoff et al. (2019) developed metrics to analyze team tactics based on disruptions caused by passes and movements of defensive structures, highlighting how effective tactical maneuvers correlate with creating goal-scoring opportunities and overall team success [Goe+21] [Mee+20].

Together, these metrics provide teams and analysts with actionable insights, enhancing their understanding of performance beyond basic statistics. By effectively leveraging physical and tactical performance metrics from tracking data, teams can achieve a competitive edge through optimized training, strategic planning, and in-match adjustments.

Chapter 3

System Architecture & Methods

3.1 Dataset

As the primary benchmark and training corpus, "SoccerNet-Tracking" is used, a multiple-object tracking dataset from the main camera of 12 Swiss Super League matches recorded in 2019 at 1080p/25 fps. The release contains 200 short clips with 30 seconds each centered on challenging actions (e.g., corners, free-kicks, penalties). In total, this amounts to 225,375 frames, 3.65 M bounding boxes and 5,009 unique tracklets across five classes: player, goalkeeper, referee, ball and 'other'. Annotations are frame-level, MOT-style, with persistent identities even after an object leaves and re-enters the frame; for most players and goalkeepers, team side and jersey number are also provided. The dataset is divided into train, test and challenge sets[Cio+22b].

For the jersey number recognition module, "SoccerNet - Jersey Number Recognition" is used. This dataset consists of short video tracklets of football players with their ground truth jersey number. In total the dataset provides 2853 tracklets of players extracted from the "SoccerNet - Tracking" videos and is split into train, test and challenge sets[Cio+22b][Cio+23][Cio+22a][Som+24].

3.2 Pipeline Overview

This thesis implements an end-to-end pipeline that converts raw single-camera TV video into tracking data and pitch-mapped trajectories to gather statistical insights. The system runs frame-by-frame with light temporal memory. The tracking data is then used to extract tactical and physical insights. Figure 3.1 illustrates the data flow.

Each frame is passed to a domain-adapted YOLOv8 detector[Yas24]. The

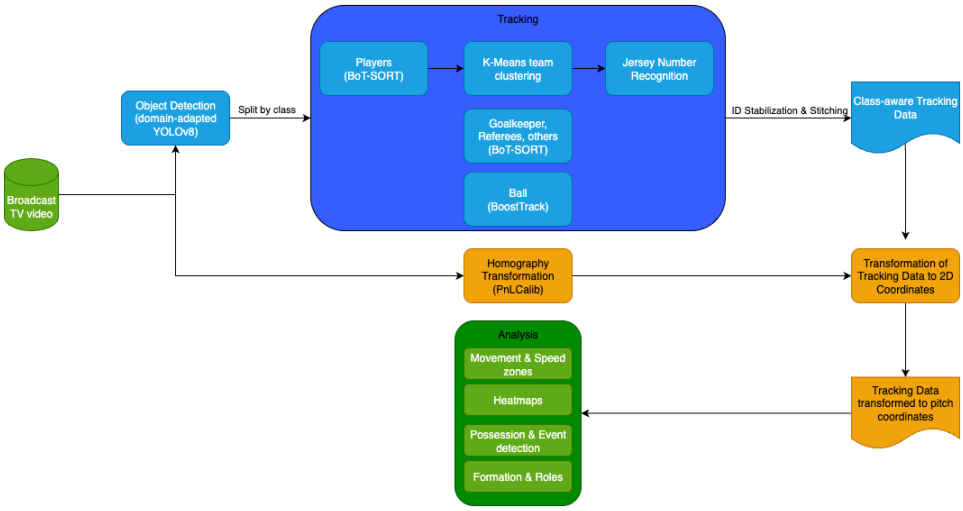

Figure 3.1: Football Vision Flowchart

detector returns class-labeled boxes for players, referees, ball and others. Separate trackers are then used per class to respect different motions and scale patterns. Players and referees use BoT-SORT, while the ball uses a BoostTrack instance. Postprocessing and temporal memory are used to reduce ID switches and create stable tracklets[AOB22][ST24]. During the Tracking process players have to be assigned to a team. This is done by a K-Means clustering. For more stable IDs a PARSeq recognizer is used to predict jersey numbers[KE24]. To map the video trajectories to 2D coordinates PnCalib is used to gather per frame homography matrices[GA]. These are then used to transform all tracklets to pitch coordinates, which are the basis for tactical and physical metrics.

During analysis, first, the data is used to visualize player movement paths and calculate physical metrics like covered distance, top speed, average speed or speed zones[De +18]. Another way to analyze a players positional data is to build their heatmap and visualize it on a pitch to show which areas are primarily covered[Gar+22]. Ball possession is approximated by using a method that assigns the ball to the nearest player, from the ball track and the nearest players events like passes, shots or dribbles are inferred[LH17][KM20]. For analyzing tactical aspects of the data two methods are implemented for formation and role detection[NY19][Bia+16].

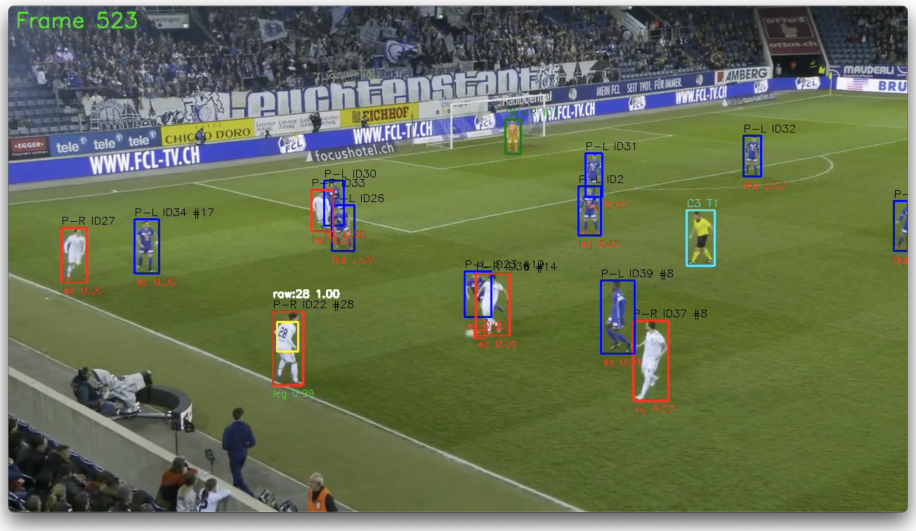

Figure 3.2: Live class-aware Tracking output

3.3 Getting Tracking Data

3.3.1 Object Detection

For Object Detection a YOLOv8n model is finetuned on the training set of the SoccerNet Tracking Challenge. The model is trained on six different classes for players, goalkeepers, ball, main referee, side referees and others. The Training is done in phases. In phase one, the networks early layers are frozen for 20 epochs to stabilise adaption to the football domain. In phase two, all layers are unfreezed to train the entire network for the remaining 60 epochs. This way of training reduces overfitting early on and then refines the features for small, fast moving objects like the ball and similar looking classes like players and referees[Yas24].

At runtime, each frame of the input video is passed through YOLOv8 with the custom weights resulting in an output of tuples containing the bounding boxes, the detection confidence score and the class.

3.3.2 Class-Aware Tracking & ID Stabilization

Broadcast Football contains very different types of class behavior and appearance. Players and referees move like pedestrians, while the ball is often tiny, fast and often occluded. To respect these differences. A tracker-instance is run per class and the results are fused afterwards. Human classes are tracked with BoT-SORT and conservative association settings. BoT-SORTs module for re-identification of objects based on appearance features with a neural network is not used, because tests have shown, that there was no real

benefit due to the similar appearances of players[AOB22]. The ball is tracked using BoostTrack, which is better suited for small, fast targets[ST24]. At inference time, each frames YOLO detections are split by class and passed to the corresponding tracker, which returns local track IDs.

To split the generic player class into two teams an unsupervised, color-based K-Means clustering is used. From the first frame the grass hue value in the HSV color-space is estimated. For every player crop in each frame, the grass color is extracted using a hue window around the estimated color. To get only the most informative kit color information, only the top half of the player crop is kept. The remaining pixels are summarized by their mean color and clustered with K-Means (k = 2). The clusters are refit every 125 frames to adapt to lighting and zoom. To keep the assigned teams stable, the clusters ar labeled as "L" and "R" by their average x-position at the first frame and then, on refits, match the new cluster-centers to the previous centers. Per player tracklet, the assigned team is based on a majority vote done on the per-frame assignments in the latest history. This combination - grass suppression, torso focus, periodic refits and temporal smoothing - makes the team assignment robust to overlaps, occlusion, shadows and camera motion.

The main method for player re-identification after leaving the camera view is jersey number recognition. The jersey number in combination with the assigned team is a unique identifier for a player, therefore it is the best fit for this task. The method used is inspired by the jersey number pipeline introduced by Koshkina et al., which uses a ResNet-34 legibility classifier to identify the player crops that could contain a readable jersey number and a PARSeq model to recognize the actual jersey number[KE24]. The classifier used is the one provided by Koshkina et al. The used PARSeq recognizer is trained on the SoccerNet jersey number challenge dataset by using the legibility classifier to extract the legible player crops from the dataset, before the pretrained PARSeq-tiny weights are finetuned on the data for 32 epochs[BA22]. During inference the legibility classifier is called for every player crop, if the score exceeds a threshold the jersey number pipeline is triggered. First the torso region of interest is extracted by cropping the bounding box to roughly the middle of the upper body. The finetuned PARSeq weights are then used to recognize numbers between 1 and 99, which are only accepted if a minimum confidence is reached.

To stabilize jersey identity over time, a vote table is maintained for each tracklet. Each accepted jersey number contributes a weighted vote, single-digits are slightly penalized to avoid switching between for example "7" and "17". A leaky decay is used to prefer newer recognitions. After all frames have been processed, the final jersey number for each track is computed from the full history and missing frames are back-filled. The combined key of assigned team and jersey number is used to stitch tracklet fragments into a single player identity. Figure 3.2 shows an exemplary frame of the live class-aware tracking output drawn into the broadcast video.

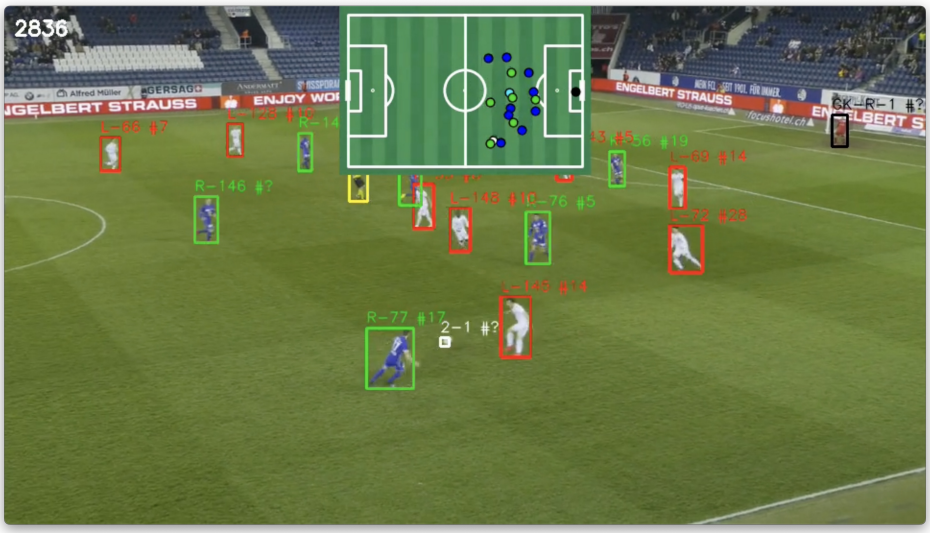

Figure 3.3: 2D transformed tracking output

3.3.3 Homography from Broadcast Video to 2D-Coordinates

To transform the broadcast video tracklets to 2D-coordinates PnLCalib by Gutierrez-Perez et al. with some light refinements is used. PnLCalib predicts soccer field key-points and lines with convolutional neural networks and fits an initial camera position. From the keypoints, lines and camera position the ground-plane homography is read off per frame[GA].

During inference PnLCalib is run per frame resulting in a homography matrix representing the plane extracted from the keypoint and line CNNs. The tracking data and the per-frame homographies are used to map to 2D field coordinates. The matrixs inverse is applied to get x and y coordinates in meters on a 105x68 m pitch. For each bounding box the bottom-center - a players foot point - is transformed. The per-track coordinates are then filtered with a One-Euro-Filter to remove noise and jittering[CRV12]. In addition to that a protection against jumpy homographies is implemented. If the median shift between using the previous vs. current homography across all detections is above 1.5m, the previous homography is used. A new homography is only accepted if the median shift drops below 6.0m. After 5 frozen frames the homography gets unfrozen to avoid dead-locks after fast camera movements.

The pipeline then exports a file containing the IDs, 2D-coordinates, classes and the jersey numbers for each frame to be used for data analysis. Figure 3.3 shows the transformed output visualized on the broadcast video and on a 2D football pitch.

3.4 Data Analysis

This section turns raw tracking data into readable analytics. For this the data is preprocessed and then transformed and visualized by using different methods.

3.4.1 Data Preprocessing

Before physical and tactical analysis is performed, the data needs to be preprocessed. First any detection whose field coordinates are outside the 105x68m pitch are removed. This removes obvious projection errors. In the next step only tracklets that are players, have a valid jersey number and meet a minimum duration threshold are kept. This creates only long and reliable tracklets for stable analysis.

3.4.2 Player Path Visualization & Speed-Zones

Player path visualization can help understand how a player moves on the pitch. Therefore each players path is drawn onto a pitch and colored in 5 different speed zones - walking, jogging, running, high-speed running, and sprinting. From this also key metrics like per-minute high-speed runs and high-speed running distance are extracted to compare players physical performances[De +18].

3.4.3 Positional Heatmaps

Another way to show where a player spends time on the pitch during the game is a heatmap. Heatmaps highlight zones of high presence or activity, they are used to read positional tendencies, roles and tactical patterns. In this pipeline a heatmap extraction methods by Garrido et al. is implemented[Gar+22]. The players positional data is averaged over a M x M grid and then visualized on a colored pitch.

3.4.4 Event Detection

To detect events like possession changes, passes, and shots a simple, rule-based system inspired by Khaustov & Mozgovoy is implemented[KM20].

- A player possesses the ball if it stays within a small radius from the players position. If the ball is far away for a short time but returns to the same player, the full span is still counted as that player's possession.
- A pass starts when the ball leaves the possession of one player and ends when it enters a teammates possession. To stabilize this trajectory and speed changes of the ball are also taken into consideration when the ball enters the vicinity of the receiving player.

- If the ball reaches the goal line and passes close enough to a post, it is interpreted as a shot. If the ball leaves the pitch otherwise it is treated as an unsuccessful pass.

3.4.5 Formation Detection

To get tactical insights about the playing style of a complete team, the real formation can be detected. This is done in two steps following a method introduced by Bialkowski et al.[Bia+14]. Raw player IDs drift, because of swaps and rotations, so players have to be aligned to roles that move with the teams shape.

1. The frames get normalized by centering the team.
2. Role distributions are estimated.
3. Players get assigned to roles per frame with a one-to-one linear assignment.

These role-ordered frames are then used to segment time into tactical states using K-Means clustering. This returns standard football formation like "4-4-2", "5-4-1", "3-4-3" or "4-2-2-2" and the transitions between them. This can also be used to get mean role locations per cluster to get a real average-formation.

Chapter 4

Experiments & Results

This chapter evaluates the individual modules in the pipeline. All experiments have been run in a Google Colab Environment on a NVIDIA L4 GPU and on a M2 MacBook Air. The evaluation of the tracking and jersey number OCR tasks was done by the respective evalAI evaluation server. The metrics used are the following:

- **Precision**: Fraction of predicted positives that are correct
- **Recall**: Fraction of ground-truth positives that are found
- **F1**: Harmonic mean of precision and recall
- **AP@50**: Average precision computed from the precision-recall curve using an IoU threshold of 0.50
- **AP@75**: Average precision computed from the precision-recall curve using an IoU threshold of 0.75
- **mAP@50**: Mean average precision at IoU=0.50, averaged overclasses
- **mAP@50-95**: Mean average precision at IoU=0.50 - IoU=0.95, averaged over classes
- **DetA**: Detection Accuracy - Measures how well objects are found while ignoring identities
- **AssA**: Association Accuracy - Measures how well identities are kept consistent over time
- **HOTA**: Higher Order Tracking Accuracy - Overall tracking score that balances DetA and AssA
- **Accuracy**: Percentage of correct predictions

Class	Images	Instances	Precision	Recall	mAP50	mAP50-95
all	12250	188199	0.757	0.65	0.705	0.419
player	12175	154912	0.85	0.944	0.948	0.616
goalkeeper	6485	6485	0.717	0.761	0.791	0.485
ball	11268	11377	0.621	0.268	0.31	0.108
main referee	9290	9290	0.896	0.639	0.705	0.489
side referee	5677	5698	0.748	0.734	0.814	0.43
other	190	437	0.712	0.554	0.652	0.389

Table 4.1: Finetuned YOLOv8n results on the SoccerNet Tracking test set.

Metric	Score
AP@0.50	0.808
AP@0.75	0.430
Precision@0.50	0.887
Recall@0.50	0.901
F1@0.50	0.894
Precision@0.75	0.640
Recall@0.75	0.650
F1@0.75	0.645

Table 4.2: Finetuned YOLOv8n results on the SoccerNet Tracking challenge set.

4.1 Detection Results

The finetuned YOLOv8n weights have been evaluated on both the test and challenge set of the SoccerNet Tracking challenge. The test set contained detection ground-truth for each class, so here the evaluation was done for each class. Table 4.1 shows the results on the test set.

Overall the results on the test set show that the model is strong on the dominant player class with a high precision and high recall. The performance drops on goalkeeper and referees showing some confusion with players and less consistent localization. The clear weak spot is the ball, a low recall and mAP indicate many missed objects due to the small, fast and motion-blurred objects. The gap between mAP50 and mAP50-95 across classes points to localization precision issues.

The challenge set doesn't contain any ground truth for the individual classes, so the evaluation was done class-agnostic. Table 4.2 shows the results. At loose IoU of 0.50 the model performs really well with good precision and recall, finding most objects while keeping false positives low. At a stricter IoU the accuracy drops, which shows the same behavior as on the test set, that classification is working good while the localization accuracy could be

Step	ms/frame
object detection	13.9
tracking	31.7
jersey number OCR	104.8
homography calculation	188.7
coordinate transformation	2.1
total	341.2

Table 4.3: Performance of the pipelines individual modules.

improved.

4.2 Tracking Results

To evaluate the tracking performance, the complete pipeline including object detection, team clustering, jersey number recognition and postprocessing was applied to the challenge set of the SoccerNet Tracking challenge set. The output was uploaded to the official EvalAI server.

The pipeline achieves HOTA 58.08, with DetA 66.33 and AssA 50.98. That means detection quality is solid and most boxes with reasonable box overlap are found. The bottleneck is clearly the identity association, IDs switch more often than ideal, which is due to players leaving and reentering the screen while their jersey number is not visible for tracklet stitching.

4.3 OCR Results

The jersey number recognition module was evaluated on the SoccerNet Jersey Number challenge. For this, the legibility classifier was used on all player crops of the challenge set. The fine-tuned PARSeq model was then used on all legible crops. The results uploaded to the official EvalAI server scored an accuracy of 82.41%.

4.4 Runtime

To evaluate how fast the system runs five steps that primarily contribute to the runtime must be measured - object detection, tracking, jersey number recognition, homography calculation and coordinate transformation. All following values shown in Table 4.3 are averages achieved on multiple testing scenarios. Because of the slow CPU performance of the Google Colab environment, the tracking and coordinate transformation performances were measured on an M2 MacBook Air.

End-to-end the pipeline runs at 2.93 FPS, which is not real-time for broadcast video at 25 FPS. The two dominant costs are homography estimation and jersey number recognition. Applications that require only object detection and tracking run with 21.6 FPS, which gets closer to real-time but still isn't quite there.

Chapter 5

Case Study

This case study applies the full pipeline to a complete 45-minute half from a Swiss Super League match using only the main broadcast cameras video. The goal is to show what the system delivers end-to-end, from detections to physical and tactical insights, on real footage.

After removing the off-pitch values the tracking pipeline results in an output with 889,796 entries for 1,185 individual player tracklets. For data analysis only the most consistent tracklets, that got assigned a jersey number and got an entry in at least 7,500 frames, are needed. By preprocessing, the data is reduced to 24 unique player tracklets.

From the trajectories, movement paths and speed-zones are derived for each player. Figure 5.1 shows an example of a movement path colored in the different speed-zones for player 19 of team 'R'. Table 5.1 displays the speed-zone distribution of player 19, showing that the player spends most of the time walking or jogging, while the most distance was covered while jogging or running.

For every player a positional heatmap is produced on a 4x4 pitch grid. These maps show where time is spent and thus expose role-typical positions and offer comparisons for players with similar roles. Figure 5.2 shows the heatmap for player 19, with the brighter areas indicating more time spent in those areas of the pitch.

Zone	Time [%]	Distance [%]
walking	51.5	19.1
jog	29.0	34.5
run	13.3	26.7
hsr	2.4	6.3
sprint	3.9	13.4

Table 5.1: Speed-zone distribution for player 19 of Team 'R'.

Type	Count
failed pass	275
out sideline	150
out endline	71
tackle	12
shot wide	7
shot on	4

Table 5.2: Event types and counter.

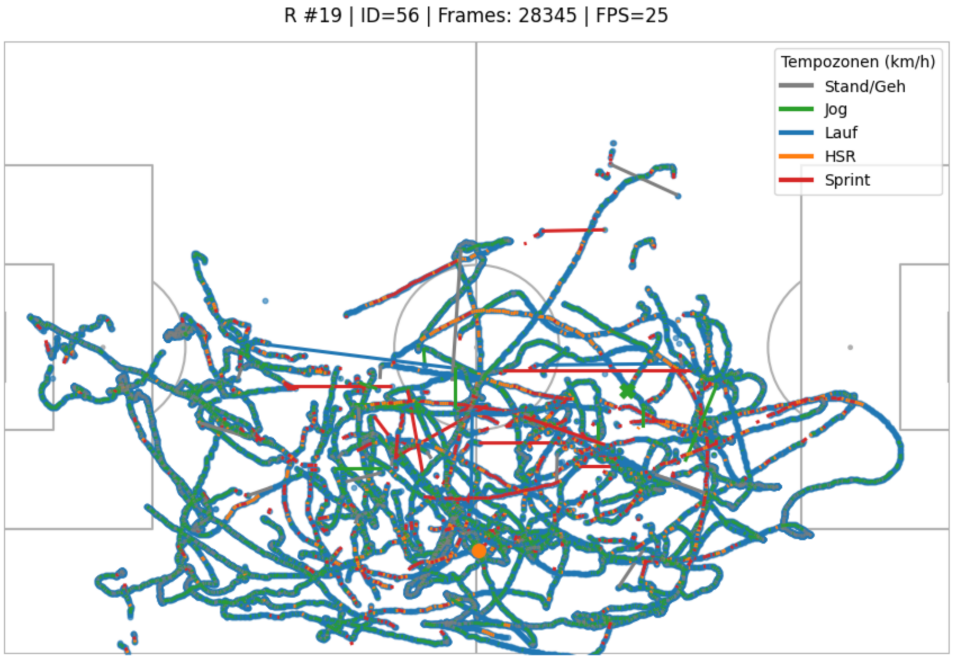

Figure 5.1: Movement path of player 19 of Team 'R'

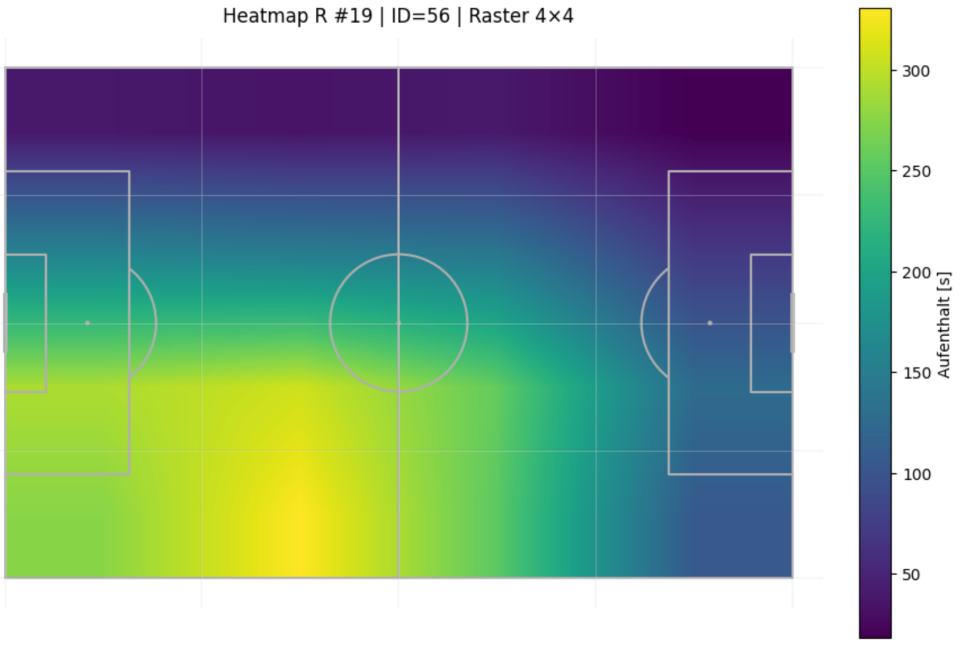

Figure 5.2: 4x4 positional heatmap for player 19 of team 'R'

Ball possession for both teams is calculated using the method described in chapter 3.4.4. This results in 58.6% possession for team 'L' and 41.4% possession for team 'R'. Though, those results must be viewed cautiously, because for an accurate possession calculation a good ball tracking is neccessary. This is not the case when considering the results in chapter 4.1. The same holds for the results of the rule based system for event detections. Table 5.2 shows the detected events and the number of times they occurred, but these results probably have a poor accuracy with the ground truth events due to the poor ball tracking accuracy.

By applying the method described in chapter 3.4.5 clusters the teams in their actual used formations during the match and calculates how often the team switches between those formations. For team 'R' the method returns two used formations - 4-2-2-2 for 52% and 4-2-3-1 for 48% of the time.

In conclusion, this case study shows that for analytics that rely solely on players positional data, the system provides good and usable results for physical and performance metrics. For metrics relying also on ball tracking the results can not be trusted and should be checked by other methods.

Chapter 6

Discussion & Conclusion

6.1 Limitations

This project shows that a broadcast-only pipeline can recover useful tracking and analytics, but several bottlenecks remain. First, identity stability is the main weakness. While detection quality is solid, the overall tracking score indicates that IDs switch more often than ideal, especially when players leave and reenter the frame or the jersey number is not visible. Second, ball detection is fragile. The detector underperforms on the ball and this also shows in event and possession detection.

Third, calibration is sensitive to difficult views. Per-frame PnLCalib with smoothing and temporal safeguards works well on average, but homography jumps can still happen during fast movements of the camera or when keypoints and lines are occluded.

Finally, runtime is not yet real-time end-to-end. The pipeline reaches 2.93 FPS. Homography estimation and jersey number recognition dominate the runtime, limiting live use.

6.2 Future Work

A number of future improvements could raise both accuracy and practicality:

- **Detection & Tracking**: An expanded and rebalanced data set for the ball could improve the detection quality. A domain-adapted appearance-based re-identification module could boost ID stability for player tracklets.

- **Jersey Number Recognition**: Reducing the number of calls and using a lighter backbone for the OCR model, could drastically improve the runtime.

6.3 Conclusion

This thesis delivers a complete pipeline from TV footage to pitch-mapped trajectories and physical and tactical analysis, demonstrating strong results for player-based tasks like movement paths, speed zones, heatmaps or formations. In controlled tests the detector performs well on the dominant classes, the tracking module achieves good accuracy and moderate association. The jersey number recognition performs really well and achieves 82% accuracy. The complete pipeline is promising for offline analysis, while the tracking module is close to practical for live overlays.

The case study on a full 45-minute half confirms that physical and tactical insights based mainly on player positions are already usable, but ball-driven metrics should be treated with caution until detection and tracking improves. Overall, the work establishes a solid,modular baseline and a clear roadmap for future improvements: improve identity consistency and accelerate the heavy modules to move from "works offline" to "works live".

Bibliography

[Seo+97] Yongduek Seo et al. "Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick". In: *Image Analysis and Processing*. Ed. by Alberto Del Bimbo. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 196–203. ISBN: 978-3-540-69586-8 (cit. on p. 6).

[VJ01] Paul Viola and Michael Jones. "Rapid object detection using a boosted cascade of simple features". In: *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*. Vol. 1. Ieee. 2001, pp. I–I (cit. on p. 3).

[Low04] G Lowe. "Sift-the scale invariant feature transform". In: *Int. J* 2.91-110 (2004), p. 2 (cit. on p. 3).

[DT05] Navneet Dalal and Bill Triggs. "Histograms of oriented gradients for human detection". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. Ieee. 2005, pp. 886–893 (cit. on p. 3).

[FMR08] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. "A discriminatively trained, multiscale, deformable part model". In: *2008 IEEE conference on computer vision and pattern recognition*. Ieee. 2008, pp. 1–8 (cit. on p. 3).

[CRV12] Géry Casiez, Nicolas Roussel, and Daniel Vogel. "1 € Filter: A Simple Speed-Based Low-Pass Filter for Noisy Input in Interactive Systems". In: CHI '12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 2527–2530. ISBN: 9781450310154. DOI: 10.1145/2207676.2208639. URL: https://doi.org/10.1145/2207676.2208639 (cit. on p. 15).

[Bia+14] Alina Bialkowski et al. "Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data". In: *2014 IEEE International Conference on Data Mining*. 2014, pp. 725–730. DOI: 10.1109/ICDM.2014.133 (cit. on pp. 9, 17).

[Gir+14] Ross Girshick et al. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587 (cit. on p. 4).

[Gir15] Ross Girshick. "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1440–1448 (cit. on p. 4).

[NFK15] Nima Najafzadeh, Mehran Fotouhi, and Shohreh Kasaei. "Multiple soccer players tracking". In: *2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP)*. 2015, pp. 310–315. DOI: 10.1109/AISP.2015.7123503 (cit. on p. 6).

[Ren+15] Shaoqing Ren et al. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28 (2015) (cit. on p. 4).

[BD16] Sermetcan Baysal and Pınar Duygulu. "Sentioscope: A Soccer Player Tracking System Using Model Field Particles". In: *IEEE Transactions on Circuits and Systems for Video Technology* 26.7 (2016), pp. 1350–1362. DOI: 10.1109/TCSVT.2015.2455713 (cit. on p. 6).

[Bew+16] Alex Bewley et al. "Simple online and realtime tracking". In: *2016 IEEE international conference on image processing (ICIP)*. Ieee. 2016, pp. 3464–3468 (cit. on p. 5).

[Bia+16] Alina Bialkowski et al. "Discovering Team Structures in Soccer from Spatiotemporal Data". In: *IEEE Transactions on Knowledge and Data Engineering* 28.10 (2016), pp. 2596–2605. DOI: 10.1109/TKDE.2016.2581158 (cit. on p. 12).

[Red+16] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788 (cit. on p. 4).

[Fil+17] Cristoforo Filetti et al. "A Study of Relationships among Technical, Tactical, Physical Parameters and Final Outcomes in Elite Soccer Matches as Analyzed by a Semiautomatic Video Tracking System". In: *Perceptual and Motor Skills* 124.3 (2017). PMID: 28514921, pp. 601–620. DOI: 10.1177/0031512517692904 (cit. on p. 8).

[He+17] Kaiming He et al. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969 (cit. on p. 4).

[LH17] Daniel Link and Martin Hoernig. "Individual ball possession in soccer". In: *PLOS ONE* 12 (July 2017), pp. 1–15. DOI: 10.1371journal.pone.0179953. URL: https://doi.org/10.1371/journal.pone.0179953 (cit. on p. 12).

[Spe+17] William Spearman et al. "Physics-based modeling of pass probabilities in soccer". In: *Proceeding of the 11th MIT Sloan Sports Analytics Conference*. Vol. 1. 2017 (cit. on p. 9).

[CL18] Jianhui Chen and James J. Little. *Sports Camera Calibration via Synthetic Data*. 2018. arXiv: 1810.10658 [cs.CV]. URL: https://arxiv.org/abs/1810.10658 (cit. on p. 7).

[De +18] Varuna De Silva et al. "Player Tracking Data Analytics as a Tool for Physical Performance Management in Football: A Case Study from Chelsea Football Club Academy". In: *Sports* 6.4 (2018). ISSN: 2075-4663. DOI: 10.3390/sports6040130. URL: https://www.mdpi.com/2075-4663/6/4/130 (cit. on pp. 12, 16).

[WB18] Nicolai Wojke and Alex Bewley. "Deep cosine metric learning for person re-identification". In: *2018 IEEE winter conference on applications of computer vision (WACV)*. IEEE. 2018, pp. 748–756 (cit. on p. 5).

[NY19] Takuma Narizuka and Yoshihiro Yamazaki. "Clustering algorithm for formations in football games". In: *Scientific Reports* 9.1 (2019), p. 13172. ISSN: 2045-2322. DOI: 10.1038/s41598-019-48623-1. URL: https://doi.org/10.1038/s41598-019-48623-1 (cit. on p. 12).

[BWL20] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection". In: *arXiv preprint arXiv:2004.10934* (2020) (cit. on p. 4).

[KM20] Victor Khaustov and Maxim Mozgovoy. "Recognizing Events in Spatiotemporal Soccer Data". In: *Applied Sciences* 10.22 (2020). ISSN: 2076-3417. DOI: 10.3390/app10228046. URL: https://www.mdpi.com/2076-3417/10/22/8046 (cit. on pp. 12, 16).

[Mee+20] L. A. Meerhoff et al. "Exploring Successful Team Tactics in Soccer Tracking Data". In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Peggy Cellier and Kurt Driessens. Cham: Springer International Publishing, 2020, pp. 235–246. ISBN: 978-3-030-43887-6 (cit. on p. 9).

[AB21] Gabriel Anzer and Pascal Bauer. "A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer)". In: *Frontiers in Sports and Active Living* Volume 3 - 2021 (2021). ISSN: 2624-9367. DOI: 10.3389/fspor.2021.624475. URL: https://www.frontiersin.org/journals/sports-and-active-living/articles/10.3389/fspor.2021.624475 (cit. on p. 9).

[Goe+21] F R Goes et al. "Modelling team performance in soccer using tactical features derived from position tracking data". In: *IMA Journal of Management Mathematics* 32.4 (Apr. 2021), pp. 519–533. ISSN: 1471-6798. DOI: 10.1093/imaman/dpab006. eprint: https://academic.oup.com/imaman/article-pdf/32/4/519/40311948/dpab006.pdf. URL: https://doi.org/10.1093/imaman/dpab006 (cit. on p. 9).

[TB21] Rajkumar Theagarajan and Bir Bhanu. "An Automated System for Generating Tactical Performance Statistics for Individual Soccer Players From Videos". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.2 (2021), pp. 632–646. DOI: 10.1109/TCSVT.2020.2982580 (cit. on p. 6).

[Zha+21] Yifu Zhang et al. "Fairmot: On the fairness of detection and re-identification in multiple object tracking". In: *International journal of computer vision* 129 (2021), pp. 3069–3087 (cit. on p. 5).

[AOB22] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. *BoT-SORT: Robust Associations Multi-Pedestrian Tracking*. 2022. arXiv: 2206.14651 [cs.CV]. URL: https://arxiv.org/abs/2206.14651 (cit. on pp. 5, 12, 14).

[BA22] Darwin Bautista and Rowel Atienza. "Scene Text Recognition with Permuted Autoregressive Sequence Models". In: *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, Oct. 2022, pp. 178–196. DOI: 10.1007/978-3-031-19815-1_11. URL: https://doi.org/10.1007/978-3-031-19815-1_11 (cit. on p. 14).

[Cio+22a] Anthony Cioppa et al. "Scaling up SoccerNet with multi-view spatial localization and re-identification". In: *Scientific Data* 9 (June 2022) (cit. on p. 11).

[Cio+22b] Anthony Cioppa et al. "SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 3491–3502 (cit. on p. 11).

[Gar+22] David Garrido et al. "Heatmaps in soccer: Event vs tracking datasets". In: *Chaos, Solitons &; Fractals* 165 (Dec. 2022), p. 112827. ISSN: 0960-0779. DOI: 10.1016/j.chaos.2022.112827. URL: http://dx.doi.org/10.1016/j.chaos.2022.112827 (cit. on pp. 12, 16).

[TE22] Jonas Theiner and Ralph Ewerth. *TVCalib: Camera Calibration for Sports Field Registration in Soccer*. 2022. arXiv: 2207.11709 [cs.CV]. URL: https://arxiv.org/abs/2207.11709 (cit. on p. 8).

[Zha+22] Yifu Zhang et al. "Bytetrack: Multi-object tracking by associating every detection box". In: *European conference on computer vision*. Springer. 2022, pp. 1–21 (cit. on p. 5).

[Cio+23] Anthony Cioppa et al. "SoccerNet 2023 Challenges Results". In: abs/2309.06006 (2023). DOI: 10.48550/arXiv.2309.06006. arXiv: 2309.06006. URL: https://doi.org/10.48550/arXiv.2309.06006 (cit. on p. 11).

[KH24] Rahima Khanam and Muhammad Hussain. *YOLOv11: An Overview of the Key Architectural Enhancements*. 2024. arXiv: 2410.17725 [cs.CV]. URL: https://arxiv.org/abs/2410.17725 (cit. on p. 4).

[KE24] Maria Koshkina and James H. Elder. "A General Framework for Jersey Number Recognition in Sports Video". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2024, pp. 3235–3244 (cit. on pp. 12, 14).

[Neh+24] Fnu Neha et al. *From classical techniques to convolution-based models: A review of object detection algorithms*. 2024. arXiv: 2412.05252 [cs.CV]. URL: https://arxiv.org/abs/2412.05252 (cit. on p. 3).

[Som+24] Vladimir Somers et al. "SoccerNet Game State Reconstruction: End-to-End Athlete Tracking and Identification on a Minimap". In: June 2024 (cit. on p. 11).

[ST24] Vukasin D Stanojevic and Branimir T Todorovic. "BoostTrack: boosting the similarity measure and detection confidence for improved multiple object tracking". In: *Machine Vision and Applications* 35.3 (2024), p. 53 (cit. on pp. 5, 12, 14).

[Yas24] Muhammad Yaseen. *What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector*. 2024. arXiv: 2408.15857 [cs.CV]. URL: https://arxiv.org/abs/2408.15857 (cit. on pp. 4, 11, 13).

[GA] Marc Gutiérrez-Pérez and Antonio Agudo. "Pnlcalib: Sports Field Registration Via Points and Lines Optimization". In: *Available at SSRN 4998149* () (cit. on pp. 7, 12, 15).

Appendix A

Source Code

The source code implemented for this thesis is uploaded here (https://github.com/Similly/FootballVision).