Detection and Tracking of Football Players in a video clip

Dissertation submitted to

Shri Ramdeobaba College of Engineering & Management, Nagpur

in partial fulfillment of requirement for the award of degree of

Bachelor of Technology (B.Tech)

In

COMPUTER SCIENCE AND ENGINEERING

By

Janhavi Agrawal (05)

Rishi Maheshwari (59)

Shrey Nalode (64)

Yash Prabhat (73)

Of

VI Semester

Guide

Dr. Pravin Sonsare

RCOEM

Shri Ramdeobaba College of Engineering and Management, Nagpur

Department of Computer Science and Engineering

Shri Ramdeobaba College of Engineering & Management, Nagpur 440 013

(An Autonomous Institute affiliated to Rashtrasant Tukdoji Maharaj Nagpur University Nagpur)

April 2024

Detection and Tracking of Football Players in a video clip

Dissertation submitted to

Shri Ramdeobaba College of Engineering & Management, Nagpur

in partial fulfillment of requirement for the award of degree of

Bachelor of Technology (B.Tech)

In

COMPUTER SCIENCE AND ENGINEERING

By

Janhavi Agrawal (05)

Rishi Maheshwari (59)

Shrey Nalode (64)

Yash Prabhat (73)

Of

VI Semester

Guide

Dr. Pravin Sonsare

RCOEM

Shri Ramdeobaba College of Engineering and Management, Nagpur

Department of Computer Science and Engineering

Shri Ramdeobaba College of Engineering & Management, Nagpur 440 013

(An Autonomous Institute affiliated to Rashtrasant Tukdoji Maharaj Nagpur University Nagpur)

April 2024

SHRI RAMDEOBABA COLLEGE OF ENGINEERING MANAGEMENT, NAGPUR

(An Autonomous Institute affiliated to Rashtrasant Tukdoji Maharaj Nagpur University Nagpur)

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the Thesis on **"Detection and Tracking of football players in a video clip"** is a Bonafide work of Janhavi Agrawal, Rishi Maheshwari, Shrey Nalode, Yash Prabhat, submitted to the Rashtrasant Tukdoji Maharaj Nagpur University, Nagpur in partial fulfilment of the award of a Degree of Bachelor of Technology (B.Tech), in Computer Science and Engineering. It has been carried out at the Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur during the academic year 2023-2024.

Date: 06/04/2024

Place: Nagpur

Dr. Pravin Sonsare
Project Guide
Department of Computer Science and Engineering

Dr. R. Hablani
H.O.D
Department of Computer Science and Engineering

Dr. R. S. Pande
Principal

DECLARATION

We hereby declare that the thesis titled "**Detection and Tracking of football players in a video clip**" submitted herein, has been carried out in the Department of Computer Science and Engineering of Shri Ramdeobaba College of Engineering and Management, Nagpur. The work is original and has not been submitted earlier as a whole or part for the award of any degree/diploma at this or any other institution / University.

Date: 06/04/2024

Place: Nagpur

Janhavi Agrawal (05)

Rishi Maheshwari (59)

Shrey Nalode (64)

Yash Prabhat (73)

APPROVAL SHEET

This report entitled "**Detection and Tracking of football players in a video clip**" by **Janhavi Agrawal, Rishi Maheshwari, Shrey Nalode, Yash Prabhat** is approved for the degree of Bachelor of Technology (B.Tech).

Dr. Pravin Sonsare
Project Guide

External Examiner

Dr. R. Hablani
H.O.D, CSE

Date: 06/04/2024

Place: Nagpur

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to all those who have contributed to the completion of this report. We would like to extend a special appreciation to our final year project guide, Dr. Pravin Sonsare, for his invaluable suggestions and encouragement throughout the project. We are thankful to Shri Ramdeobaba College of Engineering and Management, Nagpur for providing us with this great opportunity to be working on this project.

We take this opportunity to acknowledge our profound indebtedness and extend our deep sense of gratitude to Dr. R. Hablani, Head of Department, for his support and guidance for our project. His leadership and expertise played a pivotal role in shaping the direction of the project and ensuring its success. Furthermore, his assistance in resource allocation and networking opportunities would have been instrumental in navigating challenges and securing necessary support.

Lastly, we would also like to express our appreciation for the guidance provided by our other supervisor and the panel members, particularly during our project presentation.

Their constructive comments and advice have greatly enhanced our presentation skills and provided insights into improving our project.

ABSTRACT

In the realm of sports analysis and performance evaluation, accurate detection and tracking of football players in video clips play a pivotal role. Existing systems often face challenges in precisely identifying players amidst dynamic movements, varying lighting conditions, and occlusions. To address these issues, our project, "Detection and Tracking of Football Players in Video Clips," endeavors to develop a robust technology solution aimed at enhancing the analysis and understanding of football matches.

In addition to player detection and tracking, our project incorporates quantitative analysis techniques to provide deeper insights into player performance. By leveraging statistical metrics and machine learning algorithms, we aim to quantify various aspects of player movements, such as speed, acceleration, and positional dynamics. This quantitative analysis will not only enhance the accuracy of player evaluation but also facilitate comprehensive performance comparisons and strategic decision-making for coaches and analysts.

Utilizing state-of-the-art computer vision and machine learning techniques, our system employs advanced algorithms for real-time detection and tracking of football players within video footage. By leveraging object detection and tracking methodologies, our system can accurately identify individual players, track their movements across frames, and provide comprehensive insights into player positioning, interactions, and gameplay dynamics.

This technology holds immense potential for enhancing coaching strategies, tactical analysis, and player performance evaluation in football. By facilitating detailed examination of player movements, formations, and gameplay patterns, our system empowers coaches, analysts, and teams to make data-driven decisions, optimize strategies, and improve overall performance on the field.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 : INTRODUCTION

Sports analytics has gained significant traction since the early 1950s, becoming increasingly essential in the realm of sports. The integration of Artificial Intelligence (AI) into sports analysis has further revolutionized the field, enabling analysts to leverage game recordings and existing data to draw pertinent conclusions. Beyond football, AI has made substantial contributions to various sports, such as Cricket, where it aids in making accurate third umpire decisions.

Modern statistics, including pass counts, pass completions, and possession metrics, emerged in the 1990s, offering vital insights into individual player and team performance, as well as correlations with game outcomes. Ball possession, a crucial aspect of football analytics, directly influences game dynamics and opponent pressure, though its direct correlation with winning probability remains debated. Recent studies have identified trends in ball possession, including the influence of environmental and contextual factors.

In light of these developments, this paper aims to implement a Multi-Player and Ball Tracking framework to identify and track players and the ball, facilitating the derivation of insights such as ball possession. The study explores various versions of the YOLO (You Only Look Once) algorithm to determine the most effective model based on the precision and recall trade-off.

Existing models propose different methodologies for player and ball tracking. For instance, utilizes the Rauch-Tung-Striebel (RTS) method to smooth trajectories and calculate distances between objects. introduces an object detection framework employing YOLOv5 models, with comparisons between YOLOv5m and YOLOv5s variants. Moreover, proposes a method to estimate ball possession, where possession is attributed to players detected in proximity to the ball.

1.1 Background

The experiment performed in the model include multiple object detection, object tracking and ball possesion among the players.

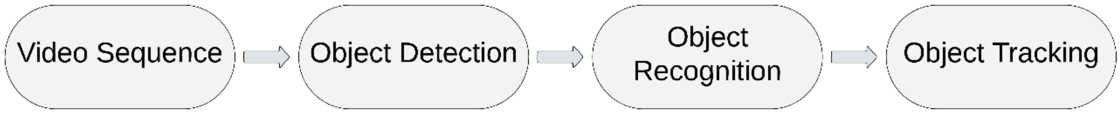

Fig 1.1 A pipeline diagram depicting the step wise procedure followed.

As depicted in Fig 1.1, the experiment commences with data collection, succeeded by manual annotation (labeling). Subsequently, the annotated data undergoes division into training, validation, and testing sets. Following preprocessing, the YOLOv5 model is employed for object detection, utilizing predefined classes. Moreover, object tracking is conducted using the BYTETrack model to track the identified classes of objects. The remainder of the experiment section will be subdivided into four parts: Dataset, Object Detection, Object Tracking, and Ball Possession.

1.1.1 Computer Vision

Computer vision, a branch of AI and computer science, enables computers to interpret visual information from the real world. It involves techniques for processing, analyzing, and extracting insights from images and videos. It basically deals with anything that humans can see and perceive. Computer Vision enables computers and systems to derive complete meaningful information from digital photos and digital videos and other visual inputs.

1.1.2 Dataset

The dataset comprises 255 images extracted from 200 football match clips sourced from the DFL Bundesliga Data Shootout Kaggle competition. Three frames per video were randomly selected for training the YOLO model. Annotation was performed using the Roboflow tool, assigning labels '0' for 'ball', '1' for 'goalkeeper', '2' for 'player', and '3' for 'referee'. Players and referees are prominently represented, reflecting their ubiquitous presence on the field, while the ball, though less prevalent, still appears in the majority of frames. Goalkeepers are the least represented class in the dataset.

1.1.3 Object Detection

Object detection in computer vision involves identifying and locating objects within images or video frames. YOLO, or "You Only Look Once," is a popular deep learning-based model known for its efficiency and speed.
It partitions the input image into a grid of cells, predicting bounding boxes and class probabilities for each cell, enabling rapid and accurate detection of multiple objects simultaneously.

1.1.4 Object Tracking

Object tracking is the process of following the movement of objects across successive frames in a video sequence. It is crucial for applications such as surveillance, vehicle monitoring, and sports analytics. BYTETrack is a versatile framework used for object tracking, which preserves low-score non-background boxes for efficient association between frames. Object tracking facilitates tasks such as player tracking on sports fields, enabling advanced analysis such as ball possession estimation.

1.1.5 Deep Learning

Deep learning, a subset of machine learning, trains neural networks with multiple layers to learn data representations, revolutionizing fields like computer vision. Models like YOLOv5 showcase deep learning's prowess, achieving top-tier performance in object detection tasks, even with smaller objects like footballs in complex scenes.

1.1.6 YOLOv5

YOLOv5, the latest in the YOLO (You Only Look Once) series, strikes a balance between speed and accuracy in object detection. Developed by Ultralytics, it boasts improved performance and efficiency, offering variants from YOLOv5s to YOLOv5x to suit diverse needs. Its single-stage detection architecture enables real-time inference, catering to applications like surveillance, autonomous driving, and medical imaging. YOLOv5's flexibility allows customization to specific tasks and datasets, while its efficient training and inference pipelines, powered by PyTorch, ensure scalability across hardware platforms. Widely embraced across domains, YOLOv5 remains a top choice for cutting-edge object detection tasks.

1.2 Motivation

Object tracking plays a crucial role in sports analytics, offering valuable insights and enhancing the understanding of player dynamics and game strategies. Here are some points highlighting the importance of object tracking in sports,

1. **Team Strategy Evaluation:** By tracking player movements and ball trajectories, object tracking enables the evaluation of team strategies and formations. Coaches can analyze how players coordinate on the field, execute set plays, and adapt to different game situations. This insight helps optimize team tactics and improve overall performance.

2. **Ball Possession and Game Flow:** Object tracking is essential for analyzing ball possession statistics in football, including time in possession and pass completion rates. It enables teams to understand ball movement patterns and possession dynamics, identifying opportunities to maintain control, create scoring chances, and disrupt opponents' strategies. By precisely tracing the trajectory of the ball and players, object tracking provides insights into how teams utilize possession, revealing strengths, weaknesses, and trends in gameplay strategies. This information helps teams adapt tactics, optimize ball control, and capitalize on scoring opportunities effectively, enhancing overall performance and decision-making on the field.

3. **Referee Decision Support**: Object tracking technology can assist referees in making accurate and fair decisions during matches. By providing additional visual information, such as player positioning and ball trajectory, object tracking systems can help referees assess contentious incidents, such as offside calls and foul situations.

4. **Player Performance Analysis:** Object tracking allows for detailed analysis of player movements, positioning, and interactions throughout a match. This information helps coaches and analysts assess individual player performance, identify strengths and weaknesses, and make data-driven decisions for player development and tactical adjustments.

Overall, the utilization of object tracking in sports enhances the depth of analysis, improves decision-making processes, and ultimately contributes to the development and success of teams and individual athletes.

CHAPTER 2 : LITERATURE REVIEW

We studied multiple papers to understand the underlying workings of convolutional neural networks. We also enrolled in some courses to better understand transfer learning models and make our model more fitting for the collected dataset. Some referred papers are mentioned below with a brief discussion of their studies:

2.1 Literature Review Table

Sr. No.	Paper No.	Author	Title	Dataset Used	Conclusion
1	[17]	Simeon Jackman	Football Shot Detection using Convolutional Neural Networks	164 games from the first and second Swedish football league	The study introduces a new pooling method called Deep networks with Temporal Pyramid Pooling (DTPP) which enhances temporal granularity scopes in sports action detection.
2	[11]	Yong-Hwan Lee, Youngseop Kim	Comparison between R-CNN and YOLO models	COCO dataset	YOLOv3 is highlighted as a model providing a good balance between speed and accuracy compared to R-CNN models.
3	[11]	Oluwaseyi Ezekiel Olorunshola, Martins Ekata Irhebhude, Abraham Eseoghene Evwiekpaefe	A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms	Google Open Images Dataset, Roboflow Public Dataset, Local Data	YOLOv5 demonstrates superior performance in terms of precision, mAP@0.5, and mAP@0.5:0.95 compared to YOLOv7, with a 4.0% increase in accuracy. However, YOLOv7 exhibited a higher recall

					value during testing.
4	[12]	Parviz Ahmadov, University of Twente, The Netherlands	Evaluation of Pose Estimation and Object Detection Models for Mini Soccer Ball Project	COCO dataset is large-scale object detection, segmentation, and a captioning dataset containing more than 300,000 images.	YOLOv5 is preferred for object detection due to its advantageous processing speed and sufficiently high mAP@0.75 score, despite a slightly lower mAP compared to other models.
.5	[13]	George R. Dimopoulos, Stanford University	Implementing Multi-Class Object Detection in Soccer Matches through YOLOv5	SoccerNet dataset	Further research is required to determine the superior model for multi-class object detection in soccer videos, as YOLOv5s exhibits higher precision, while YOLOv5m demonstrates better recall
6	[14]	Anthony Cioppa (University of Liege), Silvio Giancola (KAUST)	SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos	SoccerNet-Tracking dataset	The SoccerNet-Tracking dataset serves as a significant resource for multi-object tracking in soccer videos, aiming to advance computer vision methodologies tailored to soccer tracking applications and foster collaborative research efforts.
7	[15]	Narayana Darapaneni, Prashant Kumar, Nikhil Malhotra, Vigneswaran	Detecting key Soccer match events to create highlights using	Dataset of images representing key events in soccer matches	Faster RCNN with ResNet50 as the base model outperforms other models in

		Sundaramurthy, Abhaya Thakur, Shivam Chauhan, Krishna Chaitanya Thangeda, Anwesh Reddy Paduri	Computer Vision		detecting key events in soccer matches, achieving 95.5% class accuracy. The proposed approach successfully reduces a 23-minute match video to a 4:50-minute highlight reel, capturing almost all significant events.
8	[16]	Swetha SASEENDRAN, Sathish Prasad Vetrivel THANALAKSHMI, Swetha PRABAKARAN, Priyadharsini RAVISANKAR	Analysis of Player Tracking Data Extracted from Football Match Feed	World Cup dataset, which contains frames from 20 different football matches broadcast during the 2014 World Cup	The use of data analytics and AI, particularly tracking data in football, empowers clubs to make informed decisions. YOLOv5 and DeepSORT are employed to identify players and their movements. The proposed mathematical model assesses each player's decision-making ability, aiming to enhance player recruitment strategies.

1. **A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms :**

This paper presents a comparative analysis of YOLOv5 and the latest iteration, YOLOv7, with a focus on precision, recall, mAP@0.5, and mAP@0.5:0.95 metrics. Custom models were independently trained using both YOLOv5 and YOLOv7 on a dataset comprising 9,779 images with 21,561 annotations across four classes: Persons, Handguns, Rifles, and Knives. The dataset was compiled from the Google Open Images Dataset, Roboflow Public Dataset, and locally sourced data. Experimental results indicate that YOLOv5 achieved a precision score of 62.6%, a recall value of 53.4%, an mAP@0.5 of 55.3%, and an mAP@0.5:0.95 of 34.2%, while YOLOv7 yielded a precision score of 52.8%, a recall value of 56.4%, an mAP@0.5 of 51.5%, and an mAP@0.5:0.95 of 31.5%. Overall, YOLOv5 demonstrated superior performance in terms of precision, mAP@0.5, and mAP@0.5:0.95, with a 4.0% increase in accuracy compared to YOLOv7. However, YOLOv7 exhibited a higher recall value during testing. The experiment performed showed better performance in favor of YOLOv5.

2. **Implementing Multi-Class Object Detection in Soccer Matches Through YOLOv5 :**

The integration of AI and Deep Learning technologies into sports analysis, particularly in soccer, has garnered significant attention due to its potential to enhance the quality of sports analysis. This paper focuses on implementing multi-class object detection in soccer matches using the YOLOv5 model. Specifically, two variants of the YOLOv5 model—YOLOv5s and YOLOv5m—are trained on the SoccerNet dataset to classify objects on the football pitch and draw boundary boxes around them. While the YOLOv5s model exhibits higher precision, the YOLOv5m model demonstrates better recall. Further research is required to determine the superior model for multi-class object detection in soccer videos. It is worth noting the differences in the success of each model. Although the YOLOv5s model achieved higher precision after ten epochs than the YOLOv5m model achieved after seven epochs (which is to be expected), the YOLOv5m model achieved higher recall after fewer epochs. This indicates that with perhaps a substantially larger number of epochs used to train both models, we may see the YOLOv5m model achieve both higher precision and recall than the YOLOv5s model and therefore out-compete it.

3. **Evaluation of Pose Estimation and Object Detection Models for Mini Soccer Ball Project :**

This research evaluates various pose estimation and object detection models for their suitability in the Minisoccerbal project, focusing on accurately detecting lower extremity joints and sports balls, especially mini soccer balls, in videos captured under specific constraints like stationary cameras. The assessment considers both accuracy and processing speed, aiming for real-time object detection. Among the models tested—YOLOv5, EfficientDet for object detection, and OpenPose, BlazePose for pose estimation—YOLOv5 emerges as the preferred choice due to its favorable processing speed and sufficiently high mAP@0.75 score, despite a slightly lower mAP compared to EfficientDet. The study concludes that YOLOv5 effectively balances accuracy and processing speed, making it the optimal model for object detection in the Minisoccerbal project. Future work will concentrate on integrating YOLOv5 into the project pipeline to enable real-time detection of relevant objects during soccer training exercises.

4. **Soccer Net-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos :**

In response to the crucial need for effective object tracking in soccer videos, this paper presents the SoccerNet-Tracking dataset, a groundbreaking resource designed specifically for multi-object tracking in soccer scenarios. With 200 sequences, each lasting 30 seconds and featuring diverse and challenging soccer scenarios, as well as a comprehensive 45-minute halftime segment for long-term tracking, this dataset addresses the critical gap in representative data for this field. Meticulously annotated with bounding boxes and tracklet IDs, the dataset facilitates the training of multiple object tracking (MOT) models tailored to the complexities of soccer dynamics. Through thorough evaluation and benchmarking on the SoccerNet-Tracking dataset, this paper aims to drive advancements in computer vision methodologies tailored to soccer tracking applications. By pinpointing the shortcomings of current approaches and identifying areas for enhancement, this work endeavors to propel the computer vision community towards the development of robust and effective tracking solutions capable of addressing the intricacies of soccer scenarios. The organization of tracking challenges based on the SoccerNet-Tracking dataset serves as a catalyst for fostering collaborative research efforts and driving progress towards enhanced tracking methodologies, including long-term re-identification in challenging environments.

5. **Detecting key Soccer match events to create highlights using Computer Vision :**

The paper presents a computer vision-based approach for automatically detecting key events in soccer matches to create highlights. The authors employ models based on Faster RCNN and YoloV5 architectures to identify important moments in a match. They train and evaluate these models using a dataset of images representing key events such as fouls, corner kicks, goals, and penalty kicks. Results indicate that Faster RCNN with ResNet50 as the base model outperforms other models, achieving 95.5% class accuracy. The proposed approach successfully reduces a 23-minute match video to a 4:50-minute highlight reel, capturing almost all significant events. Moreover, the study discusses limitations and suggests potential areas for future improvement, including increasing training data diversity, augmenting images, and optimizing model hyperparameters. The research demonstrates the potential of computer vision in enhancing sports broadcasting and viewer engagement through automated highlight creation.

6. **Analysis of Player Tracking Data Extracted from Football Match Feed :**

Data analytics and AI have become extremely relevant in today's football landscape. Data is benefiting clubs in gaining a competitive advantage on and off the field by empowering them to harvest information for improving player performance, decreasing injuries, and increasing commercial efficiency.

Tracking data in football, that is the data of x, y coordinates of all 22 players on the pitch every second adds a lot of value in terms of a team's decision-making and recruitment strategy. Via the use of this data, each player's decision-making ability is also measured. The players and the ball in each frame are identified using YOLOv5, which returns their coordinates. These detections are then passed to

DeepSORT, which assigns IDs to each player and keeps track of the frame by frame, by feeding each player detection to the model. The K-Means model is employed to determine the jersey color of the players to identify the two teams. Finally, the detected coordinates are multiplied with the homogeneous matrix computed using the Sports Camera Calibration through the Synthetic Data paper approach to accomplish the perspective transform. The mathematical model hypothesized and implemented uses pitch control and expected threat to assess each player's decision-making ability, which will be a strive to enhance the recruiting of players.

CHAPTER 3 : PROPOSED METHODOLOGY

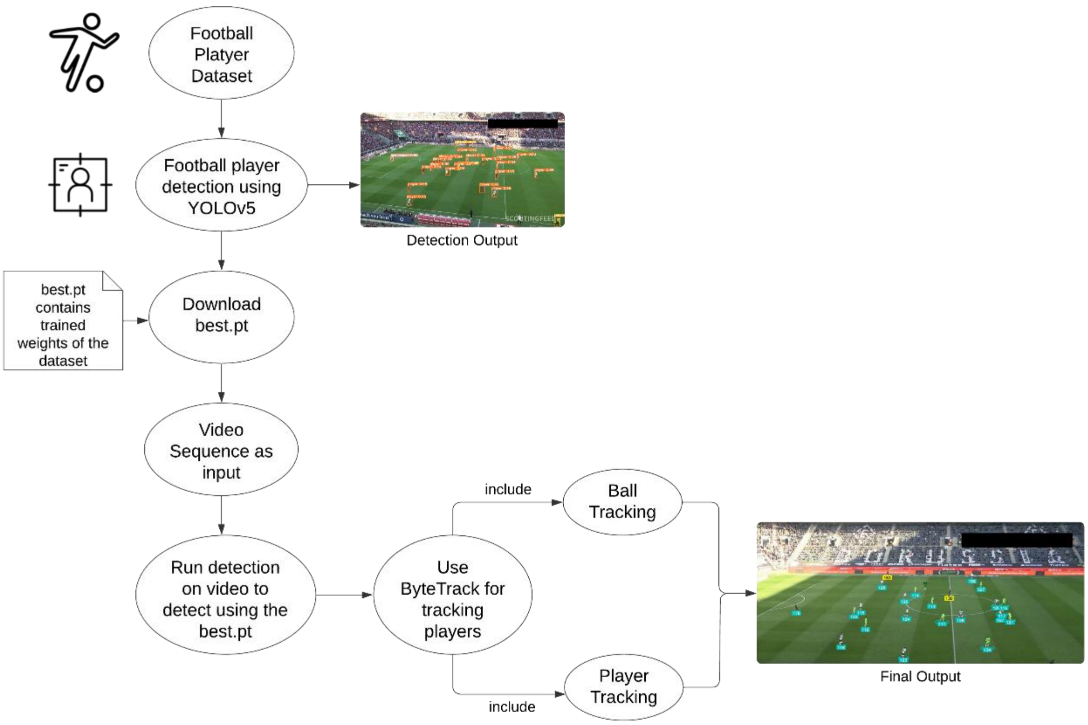

Fig 3.1 Process Flowchart

3.1 Dataset

Dataset Name:

FPD (Football-Players-Detection)

Description:

Developed from 150 clips sourced from Kaggle DFL - Bundesliga Data Shootout competition, the FPD dataset provides comprehensive images of football players derived from multi-source images. The dataset consists of 255 images that have been categorised as train-205, validate-38, and test-12 .

Data Labels:

The label file contains 5 values . The label file will contain normalized values for all 5 parameters so their values will be between 0 to 1. The label for one object looks like as shown below.

<object-class> <x_center> <y_center> <width> <height>

where,

●<object-class> - integer number for an object from 0 to (classes-1)

●<x_center> <y_center> <width> <height> - float values relative to width and height of the image, will be between 0.0 to 1.0for example: <x> = <absolute_x> / <image_width> or <height> = <absolute_height> / <image_height>

●attention: <x_center> <y_center> - are the center of the rectangle (not top-left corner)

Classes and Class-Names:

The FPD dataset consists of 4 distinct scene types, each associated with a class index and a class name, as follows: ball , goalkeeper , player , refree.
These class labels provide a clear and standardized way to categorize and evaluate different scenes in the dataset, facilitating scene classification and land use/cover mapping tasks.

Fig 3.2 Kaggle Dataset

3.2 Object Detection

The next step in our project involves object detection, particularly focusing on identifying football players, the ball, referees, and goalkeepers on the field. Utilizing YOLOv5, a powerful object detection algorithm, we aim to accurately detect and track these crucial elements during gameplay. YOLOv5, known for its efficiency and accuracy, employs a single neural network to directly predict bounding boxes and class probabilities for multiple objects simultaneously, making it well-suited for real-time applications such as sports analysis. With its robust capabilities, we anticipate achieving precise and reliable detection results to enhance our understanding and analysis of football matches.

3.2.1 What is YOLOv5?

YOLOv5 is a model in the You Only Look Once (YOLO) family of computer vision models. YOLOv5 is commonly used for detecting objects. YOLOv5 comes in four main versions: small (s), medium (m), large (l), and extra large (x), each offering progressively higher accuracy rates. Each variant also takes a different amount of time to train.

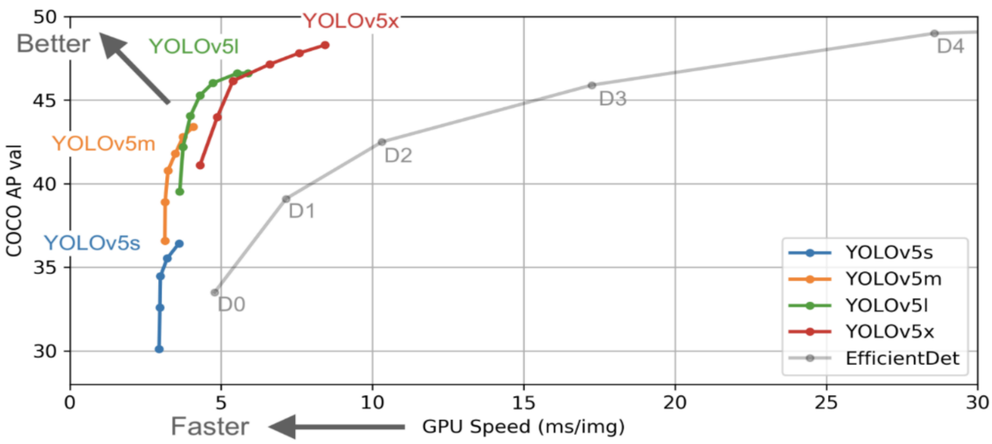

Fig 3.3 Performance of YOLOv5 different sizes models

The chart compares object detection models based on their inference time and accuracy, highlighting YOLOv5's exceptional performance. All YOLOv5 variants train faster than EfficientDet, with YOLOv5x particularly notable for its superior accuracy and faster processing speed compared to EfficientDet D4.This preliminary data underscores YOLOv5's efficiency and effectiveness in object detection tasks, a topic explored in more detail later in the post.

3.2.2 An Overview of the YOLOv5 Architecture

Object detection, a use case for which YOLOv5 is designed, involves creating features from input images. These features are then fed through a prediction system to draw boxes around objects and predict their classes. There are two types of object detection models : two-stage object detectors and single-stage object detectors. Single-stage object detectors (like YOLO) architecture are composed of three components: **Backbone**, **Neck** and a **Head** to make dense predictions as shown in the figure bellow.

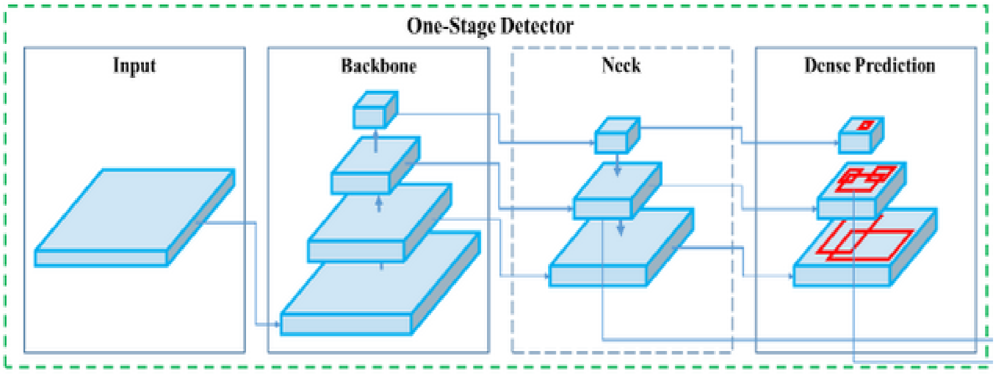

Fig 3.4 YOLOv5 Architecture (Single-Stage Detector)

1. Neck of YOLOv5

A series of layers to mix and combine image features to pass them forward to prediction.

2. Backbone of YOLOv5

A convolutional neural network that aggregates and forms image features at different granularities.

3. Head of YOLOv5

Consumes features from the neck and takes box and class prediction steps.

3.2.3 Learning Bounding Box Anchors

In the YOLOv3 PyTorch repo, Glenn Jocher introduced the idea of learning anchor boxes based on the distribution of bounding boxes in the custom dataset with K-means and genetic learning algorithms. This is very important for custom tasks, because the distribution of bounding box sizes and locations may be dramatically different than the preset bounding box anchors in the COCO dataset.

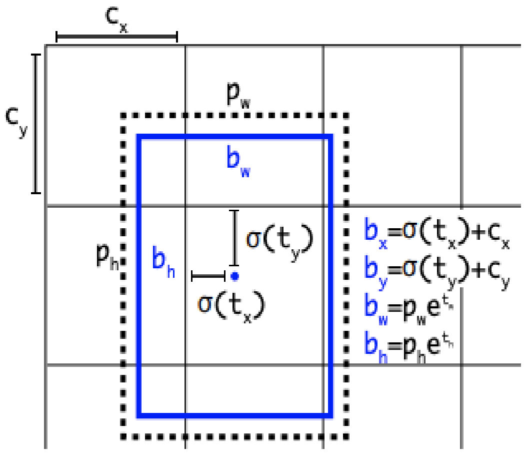

Fig 3.5 Bounding Boxes to determine objects

Equations used to compute the target bounding boxes :

$$b_x = (2 \cdot \sigma\ (t_x) - 0.5) + \ c_x \tag{3.2.3.1}$$

$$b_y = \left(2 \cdot \sigma\ \left(t_y\right) - 0.5\right) + \ c_y \tag{3.2.3.2}$$

$$b_w = p_w\ \cdot\ \left(2 \cdot \sigma\ (t_w)\right)^2 \tag{3.2.3.3}$$

$$b_h = p_h\ \cdot\ \left(2 \cdot \sigma\ (t_h)\right)^2 \tag{3.2.3.4}$$

3.2.4 Activation Function

Choosing an activation function is crucial for any deep learning model, for YOLOv5 the authors went with SiLU and Sigmoid activation function.

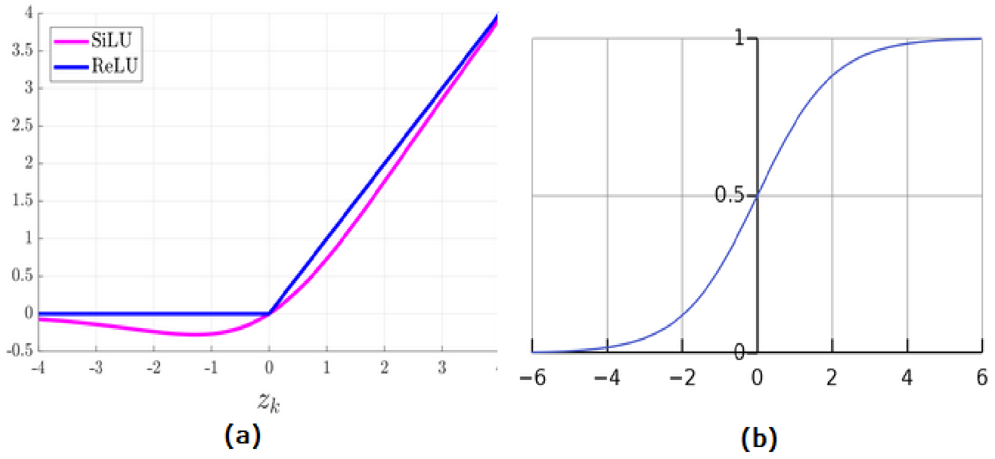

Fig 3.6 (a) SiLU function graph (b) Sigmoid Function Graph

Graph of the activation functions used in YOLOv5. (a) SiLU function graph. (b) Sigmoid function graph. SiLU stands for Sigmoid Linear Unit and it is also called the swish activation function. It has been used with the convolution operations in the hidden layers. While the Sigmoid activation function has been used with the convolution operations in the output layer.

3.2.5 Loss Function

YOLOv5 returns three outputs: the classes of the detected objects, their bounding boxes and the objectness scores. Thus, it uses BCE (Binary Cross Entropy) to compute the classes loss and the objectness loss. While CIoU (Complete Intersection over Union) loss to compute the location loss. The formula for the final loss is given by the following equation.

$$Loss = \Delta_1 L_{cls} + \Delta_2 L_{obj} + \Delta_3 L_{loc}$$

3.3 Object Tracking

Moving forward, our project advances to object tracking, a critical component aimed at monitoring football players, the ball, referees, and goalkeepers during gameplay. By using ByteTrack, an advanced object tracking algorithm, we aim to seamlessly trace these essential elements throughout matches. ByteTrack excels in accurately predicting object trajectories over time, employing cutting-edge techniques in object tracking. Through a blend of deep learning approaches

and motion estimation algorithms, ByteTrack ensures robust tracking performance even in challenging scenarios such as occlusions and rapid movements typical in football matches. Similar to YOLOv5's impact on object detection, ByteTrack promises to revolutionize object tracking by delivering precise and reliable results in real-time, thereby enhancing our sports analysis with comprehensive insights into player dynamics and strategic gameplay.

3.3.1 What is BYTETrack?

The main idea of BYTETrack is simple - keep non-background low score boxes for a secondary association step between previous frame and next frame based on their similarities with tracklets. This helps to improve tracking consistency by keeping relevant bounding boxes which otherwise would have been discarded due to low confidence score (due to occlusion or appearance changes). The generic framework makes BYTETrack highly adaptable to any object detection (YOLO, RCNN) or instance association components (IoU, feature similarity).

The primary innovation of BYTETrack is keeping non-background low confidence detection boxes which are typically discarded after the initial filtering of detections and use these low-score boxes for a secondary association step. Typically, occluded detection boxes have lower confidence scores than the threshold, but still contain some information about the objects which make their confidence score higher than purely background boxes. Hence, these low confidence boxes are still meaningful to keep track of during the association stage.

Fig 3.7 (a) Detection of boxes

Fig 3.7 (b) Tracking by assigning high score detection boxes

Fig 3.7 (c) Tracking by associating all detection boxes

3.3.2 Methodology Breakdown

BYTETrack performs object association in two stages:

1. Initially, high-confidence detection boxes are matched with predicted boxes from previous frames, including active and lost tracklets, using techniques like IoU score or cosine similarity. Linear assignment is confirmed if the matching score exceeds a fixed threshold. Unmatched high-confidence detection boxes are matched with tracklets updated from a single frame or assigned to new tracklets. ellaborate.

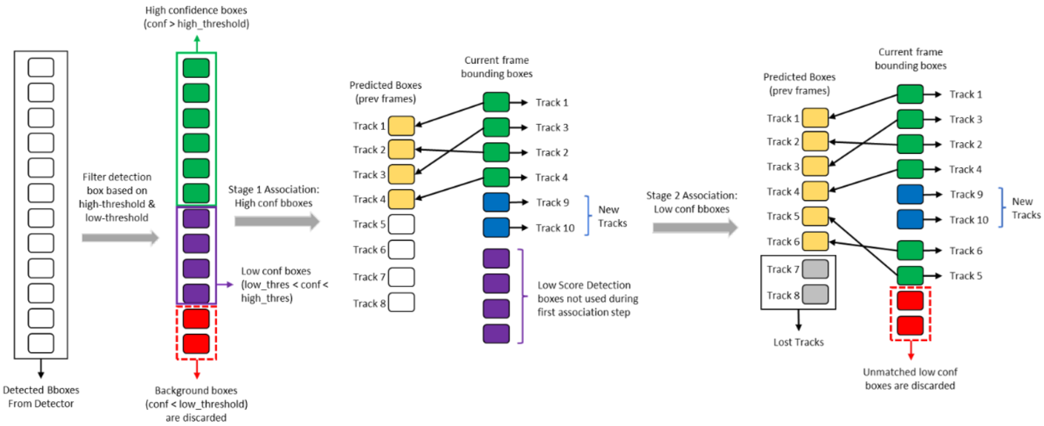

Fig 3.8 BYTETrack Algorithm

2. In the second stage, low-confidence detection boxes are matched against remaining unmatched predicted boxes, with a lower matching threshold to accommodate occlusions. Unmatched predicted boxes become lost tracklets, while unmatched detection boxes are discarded. Lost tracklets are retained for a certain duration and re-added to active tracklets before Kalman filter prediction, aiding in the recovery of temporarily lost objects.

CHAPTER 4 : EXPERIMENTATION

The goal of this project was to develop a football player detection and tracking system. The approach involved the following steps:

1. Dataset Collection: The team gathered a comprehensive dataset of football players and balls from the RobotFlow dataset.

2. Model Training: The YOLOv5 object detection model was trained on the collected dataset to create a robust player and ball detection system. The best model checkpoint, best.pt, was saved for further use.

3. Initial Tracking Attempt: The team first tried to use the YOLOv5 pre-trained COCO model for tracking the players and ball in the video. However, the results were not satisfactory, as the pre-trained model was not optimized for the specific task of football player and ball tracking.

4. Tracking with Trained Model: To improve the tracking performance, the team decided to use the custom-trained best.pt model for the detection stage. This model, which was trained on the RobotFlow dataset, provided more accurate results in detecting the football players and the ball.

5. Detection on Images: The trained YOLOv5 model was tested on individual images, and it was observed that the model successfully detected the football players and the ball in the images.

6. Detection on Videos: The team then applied the trained YOLOv5 model to individual frames of a video input provided by the user. This allowed for the detection of football players and the ball in the video.

7. Tracking: To track the detected players and ball throughout the video, the ByteTrack algorithm was employed. This provided a seamless tracking of the identified objects across the video frames.

The key difference between the COCO pre-trained model and the custom-trained best.pt model is:

1. **Dataset:** The COCO model is trained on the COCO dataset, which is a general-purpose object detection dataset that covers a wide range of object categories. In contrast, the best.pt model was trained on the RobotFlow dataset, which is specifically focused on football players and the ball.
2. **Specialized Performance:** Since the best.pt model was trained on the RobotFlow dataset, it is more specialized and optimized for detecting football players and the ball. The COCO model, being a general-purpose model, may not perform as well on this specific task compared to the custom-trained best.pt model.
3. **Accuracy:** The best.pt model, being trained on the task-specific RobotFlow dataset, is likely to provide more accurate detection and tracking of football players and the ball compared to the COCO pre-trained model.

In summary, the best.pt model is more specialized and tailored for the football player and ball detection task, while the COCO model is a general-purpose model that may not perform as well on this specific use case. The use of the custom-trained best.pt model led to more accurate results for the football player and ball detection and tracking system.

Evaluation Metrics :

As models are classifying applications into thirty image classes, the following metrics of evaluation have been used:

Confusion Matrix: A confusion matrix is a tabular representation of prediction outcomes of any binary classifier, which is used to describe the performance of the classification model on a set of test data when true values are known.

In general, the table is divided into four terminologies, which are as follows:

1. **True Positive(TP):** In this case, the prediction outcome is true, and it is true in reality.

2. **True Negative(TN):** in this case, the prediction outcome is false, and it is false in reality.
3. **False Positive(FP):** In this case, prediction outcomes are true, but they are false in actuality.
4. **False Negative(FN):** In this case, predictions are false, and they are true in actuality.

Accuracy: The proportion of correctly classified instances in the total number of instances. This is a simple and commonly used metric for classification tasks.

$$Accuracy : \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{4.1}$$

Precision: The proportion of true positive instances among all instances predicted as positive. This metric is useful when the goal is to minimize false positives.

$$Precision: \frac{TP}{(TP+FP)} \tag{4.2}$$

Recall: The proportion of true positive instances identified among all actual positive instances. This metric is useful since the goal is to minimize false negatives.

$$Recall = \frac{TP}{(TP + FN)} \tag{4.3}$$

F1 Score: The harmonic mean of precision and recall, which gives equal weight to both metrics. This metric is useful when both false positives and false negatives are important.

$$F1 - score = 2 * \frac{precision * recall}{precision + recall} \tag{4.4}$$

CHAPTER 5 : TECHNOLOGY STACK

The following technologies were used in the development of this football player detection and tracking system:

1. YOLOv5: A state-of-the-art object detection model that was used for detecting the football players and the ball in the images and video frames. Both the pre-trained COCO model and the custom-trained best.pt model were evaluated, with the custom-trained model providing superior results.

2. ByteTrack: A robust tracking algorithm that was utilized to track the detected players and ball throughout the video.

3. Python: The primary programming language used for the development of the entire system.

4. OpenCV: A computer vision library used for video processing and manipulation.

5. NumPy: A numerical computing library used for efficient data manipulation and processing.

6. Matplotlib: A data visualization library used for plotting and analyzing the results.

7. Pandas: A data manipulation and analysis library used for handling and processing the dataset.

8. Torch: A machine learning library used for the implementation and training of the YOLOv5 model.

9. Jupyter Notebook: An interactive computing environment used for prototyping and experimentation.

The integration of these various technologies and libraries, along with the custom training of the YOLOv5 model on the RobotFlow dataset, resulted in a comprehensive football player detection and tracking system that can be used in a variety of applications, such as sports analytics, player scouting, and video production.

CHAPTER 6 : RESULT AND EVALUATION

6.1 Class Description

Below is the table representing class labels and corresponding class names.

Table 6.1 Representing class labels and corresponding class names

Class Number	Class Name
0	Ball
1	Goalkeeper
2	Player
3	Referee

6.2 Confusion Matrix

The following confusion matrix was obtained for the YOLOv5m model.

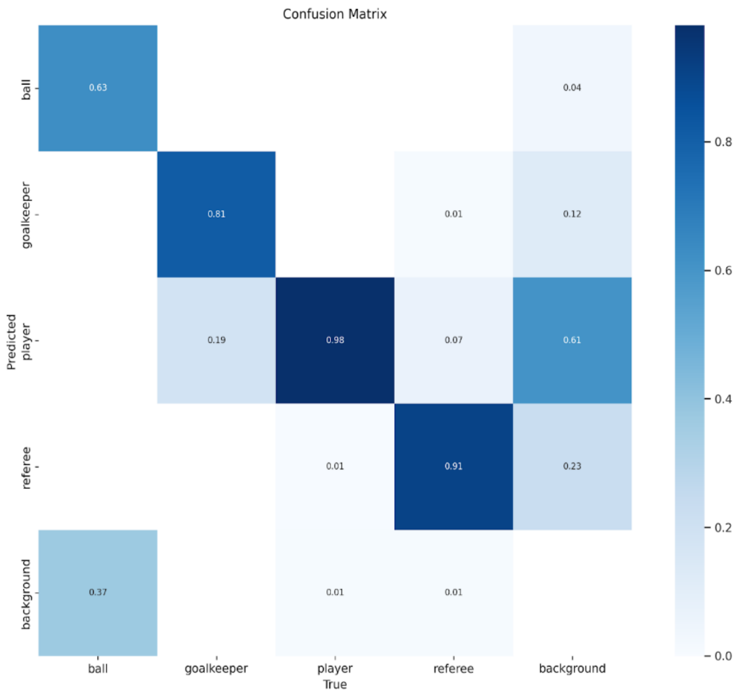

Fig 6.1 Confusion matrix of YOLOv5m model

In the confusion matrix , The matrix displays the predicted classes on the y-axis and the true classes on the x-axis.There are 5 rows and 5 columns , where the rows correspond to the true labels (actual classes), and the columns correspond to the predicted labels (predicted classes).

For instance, in confusion matrix of custom YOLOv5m model(Fig 5.2)

Diagonal elements :

1. The model predicted the "ball" class correctly 0.63 (63%) of the time.
2. The model predicted the "goalkeeper" class correctly 0.81 (81%) of the time.
3. The model predicted the "player" class correctly 0.98 (98%) of the time, which is the highest accuracy among all classes.
4. The model predicted the "referee" class correctly 0.91 (91%) of the time.
5. The model predicted the "background" class correctly 0.8 (80%) of the time.

The off-diagonal elements show the confusions or misclassifications:

1. The model misclassified 0.04 (4%) of the "ball" instances as "background."
2. The model misclassified 0.01 (1%) of the "goalkeeper" instances as "player" and 0.12 (12%) as "referee."
3. The model misclassified 0.19 (19%) of the "predicted player" instances as "goalkeeper."
4. The model misclassified 0.07 (7%) of the "predicted player" instances as "referee" and 0.01 (1%) as "background."
5. The model misclassified 0.01 (1%) of the "referee" instances as "goalkeeper" and 0.23 (23%) as "player."

The confusion matrix helps understand how well the model is performing for each class and which classes might be more prone to misclassification.

By analyzing the values in the confusion matrix, various evaluation metrics can be calculated to assess the performance of the model. Some commonly derived metrics include accuracy, precision, recall, and F 1 score. These metrics provide insights into the model's ability to correctly classify instances and its trade-offs between false positives and false negative.

6.3 Performance

Table 6.2 Model Metrics of YOLOv5m 1280

Class	Images	Instances	P	R	mAP50	mAP50-95
all	38	905	0.904	0.864	0.895	0.669
ball	38	35	0.91	0.576	0.675	0.351
goalkeeper	38	27	0.821	0.926	0.939	0.755
player	38	754	0.969	0.987	0.993	0.86
referee	38	89	0.915	0.966	0.971	0.712

In above table ,

1. Overall Performance ("all" class):
   1. Precision (P) is 0.904, which is reasonably good, indicating a low rate of false positive predictions.
   2. Recall (R) is 0.864, which is also decent but suggests some true positive instances may have been missed.
   3. mAP50 (mean Average Precision at IoU 0.5) is 0.895, which is a good score for object detection tasks.
   4. mAP50-95 is 0.669, which is a bit lower than mAP50, indicating that performance decreases as the IoU threshold becomes stricter.

2. Class-specific Performance:
   1. For the "ball" class, precision (0.91) is good, but recall (0.576) is relatively low, suggesting many ball instances may have been missed.
   2. The "goalkeeper" class has lower precision (0.821) and recall (0.926) compared to other classes.
   3. The "player" class has the highest precision (0.969) and recall (0.987) among all classes, indicating excellent performance in detecting players.
   4. The "referee" class also performs well, with high precision (0.915) and recall (0.966).

4. Relative Performance:
    a. The "player" class achieves the best performance across all metrics, followed by the "referee" class.
    b. The "goalkeeper" class seems to be the most challenging for the model, with relatively lower precision and recall scores.

6.4 Results

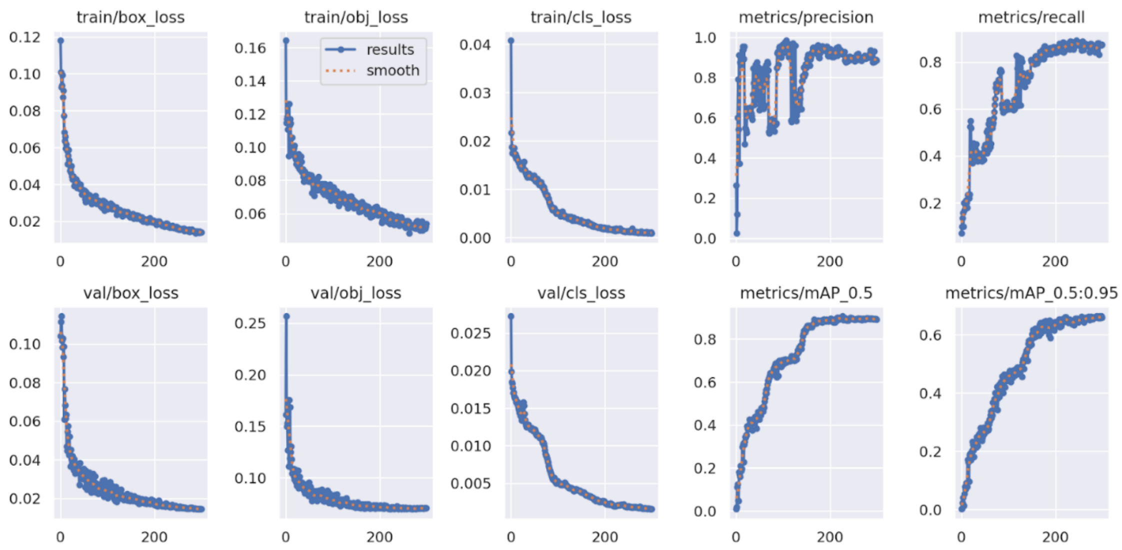

Fig 6.2 Evaluation metrics

Fig 6.2 consists of multiple line plots that show the training progress and evaluation metrics.

The plots are organized into two rows:

1. The top row shows the training losses for box regression ("train/box_loss"), object classification ("train/obj_loss"), and class prediction ("train/cls_loss"). The solid lines represent the actual loss values, while the dotted lines show smoothed versions of the same data. All three losses decrease gradually as training progresses, indicating that the model is learning and improving.
2. The middle row displays the validation losses for box regression ("val/box_loss"), object classification ("val/obj_loss"), and class prediction ("val/cls_loss"). These losses also decrease over time, but they tend to be higher and more noisy than the

training losses, which is expected since the validation set is used to evaluate the model's generalization performance.

3. The bottom row presents various evaluation metrics computed on the validation set, including precision, recall, mean Average Precision (mAP) at an IoU threshold of 0.5, and mAP averaged across IoU thresholds from 0.5 to 0.95. These metrics generally show an increasing trend as training progresses, indicating that the model's performance is improving.

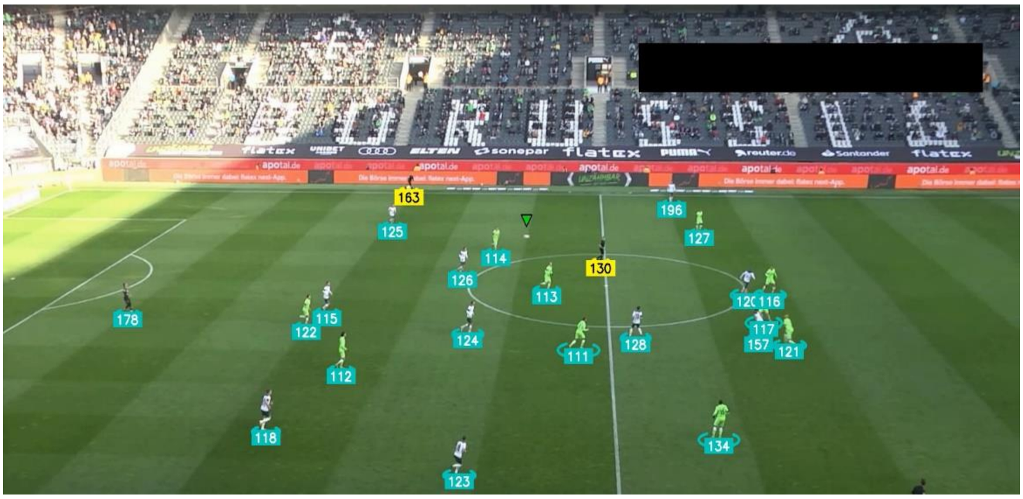

Fig 6.3 Result

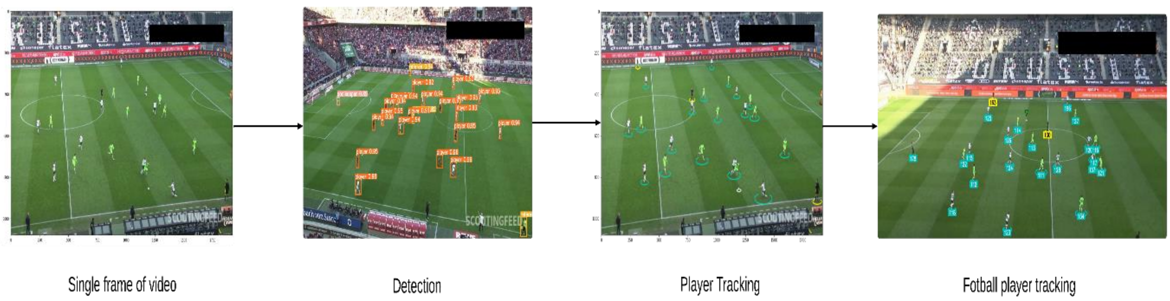

Fig 6.4 The Overall Process

CHAPTER 7 : CONCLUSION AND FUTURE SCOPE

Conclusion

This project introduces a novel approach to football player detection by leveraging the YOLOv5 model for detection and ByteTrack for tracking. Our method involves concatenating the outputs of each layer of the YOLOv5 network and passing them into a fully connected layer. This innovative strategy harnesses the feature extraction capabilities of YOLOv5's deep convolutional architecture to capture comprehensive information about football players.

Future Scope

The proposed approach opens up several promising avenues for future research and application:

1. **Player Performance Analysis:** Use player tracking data to analyze individual and team performance during matches. Metrics such as distance covered, speed, heatmaps of player movements, and positioning can be derived to understand player strategies, strengths, and weaknesses.

2. **Injury Prevention and Rehabilitation:** Implement player tracking to monitor players' physical exertion levels during training and matches. Analyze data to identify fatigue patterns and potential injury risks. This information can be crucial in designing personalized training programs and injury rehabilitation plans.

3. **Tactical Insights and Game Strategy Optimization:** Leverage player tracking data to provide coaches with detailed insights into player positioning and interactions during matches. This data can inform strategic decisions, such as formations, player roles, and game plans.

4. **Fan Engagement and Broadcasting Enhancements:** Utilize real-time player tracking to enhance the viewer experience during broadcasts. Features like

interactive player statistics, virtual replays, and augmented reality overlays can engage fans and provide deeper insights into the game.

5. **Recruitment and Talent Identification:** Scout and analyze player performance metrics using tracking data to identify talented players for recruitment purposes. Objective data on players' speed, agility, and game intelligence can supplement traditional scouting methods.

6. **Gamification and Fantasy Sports:** Integrate player tracking data into fantasy sports platforms to enhance player selection and scoring systems. Gamify the fan experience by allowing users to create virtual teams based on real-time player analytics.

REFERENCES

[1] Morgulev E, Azar OH, Lidor R. Sports analytics and the big-data era. InternationalJournal of Data Science and Analytics. 2018;5:213-22.

[2] Iyer GN, Bala VS, Sohan B, Dharmesh R, Raman V. Automated third umpire decision mak- ing in cricket using machine learning techniques. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE; 2020. p. 1216-21.

[3] Castellano J, Casamichana D, Lago C. The use of match statistics that discriminate between successful and unsuccessful soccer teams. Journal of human kinetics. 2012;31:139.

[4] Mackenzie R, Cushion C. Performance analysis in football: A critical review and implications for future research. Journal of sports sciences. 2013;31(6):639-76.

[5] Lago C. The influence of match location, quality of opposition, and match status on possession strategies in professional association football. Journal of sports sciences. 2009;27(13):1463-9.

[6] Gong B, Cui Y, Zhang S, Zhou C, Yi Q, Go´mez-Ruano M A´ . Impact of technical and physical key performance indicators on ball possession in the Chinese Super League. International Journal of Performance Analysis in Sport. 2021;21(6):909-21.

[7] Link D, Hoernig M. Individual ball possession in soccer. PloS one. 2017;12(7):e0179953.

[8] Dimopoulos GR. Implementing Multi-Class Object Detection in Soccer Matches Through YOLOv5.

[9] Yu X, Xu C, Leong HW, Tian Q, Tang Q, Wan KW. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In: Proceedings of the eleventh ACM international conference on Multimedia; 2003. p. 11-20.

[10] Jakub Michalczyk MJMMMRHUD Maggie. DFL - Bundesliga Data Shootout Kaggle; 2022. Available from: https://kaggle.com/competitions/dfl-bundesliga-data-shootout.