International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

PITCHEYE: YOLOV5 AND BYTETRACK FOOTBALL PLAYER TRACKING SYSTEM IN FOOTBALL VIDEOS

Jiawei Zhou
Khoury College of Computer Sciences Northeastern University
Vancouver, Canada

Abstract—**Football, a universally beloved sport, captivates mil- lions of spectators worldwide. This paper explores the challenges of tracking football action in videos and presents the PitchEye project—a revolutionary initiative integrating YOLOv5 and Byte-Track algorithms for enhanced football video analysis. We detail the methodologies, results, and discussions on the performance of these algorithms, showcasing their strengths and limitations in tracking players, goalkeepers, footballs, and referees. YOLOv5, using the best.pt model, achieves mean detection rates of over 0.8 for goalkeepers, referees, players, and footballs, with notable improvements over default model results. In contrast, ByteTrack excels in accurately predicting trajectories for players and goal- keepers but faces challenges with smaller objects like footballs and swiftly moving referees. Despite these challenges, ByteTrack demonstrates commendable overall performance, showcasing its potential as a valuable tool for comprehensive object tracking in dynamic football video environments.**

Keywords: **football match, YOLOv5, ByteTrack, object detection**

I. INTRODUCTION

Football, universally acknowledged as the most beloved sport worldwide, captures the imaginations of millions of spectators across continents. Its rich heritage, devoted fanbase, and extensive global appeal transcend cultural barriers, uniting individuals in appreciation of athleticism, skill, and team spirit. From the vibrant stadiums of Europe to the sun-soaked fields of South America, the energy and excitement surrounding football matches are palpable, drawing crowds of fervent supporters who eagerly gather to witness the unfolding drama. Whether it's the electric intensity of a local derby or the prestigious spectacle of a World Cup final, football possesses a unique ability to evoke emotions, ignite passionate debates, and create indelible memories that endure through generations. Indeed, the sheer magnitude of football's popularity is staggering. With an estimated 3.5 billion fans worldwide[1], football commands a global audience unparalleled by any other sport. From the bustling streets of Rio de Janeiro to the bustling metropolises of Tokyo, the universal appeal of football transcends borders, languages, and cultures, uniting people from all walks of life under the banner of their favorite teams and players.

Within the realm of widespread passion and enthusiasm for the beautiful game, the act of watching football transcends mere entertainment to become an immersive experience. It evolves into a communal ritual that binds fans together in shared moments of triumph and heartbreak. Whether huddled around a television screen or packed into the stands of a stadium, spectators transform into active participants in the unfolding drama on the pitch. Their emotions ride the highs and lows of their chosen teams' fortunes.

Yet, amidst the allure and excitement, the experience of watching football presents its own set of challenges. The frenetic pace of play, the intricate movements of players, and the ever-shifting dynamics of the match can make tracking the action a daunting task, even for seasoned observers. Accurately monitoring the movements of players and the ball poses a significant challenge—one that has long confounded analysts, coaches, and fans alike.

In response to these challenges, the PitchEye project emerges as an ambitious initiative aimed at revolutionizing the analysis of football videos. Through the integration of cutting-edge computer vision technologies, particularly the YOLOv5 and ByteTrack algo- rithms, PitchEye strives to transform the way viewers interact with football matches captured on video. By harnessing the power of these advanced algorithms, PitchEye aims to provide viewers with an enriched viewing experience, allowing them to follow on-field action with unparalleled precision and depth. Furthermore, PitchEye also serves as a valuable tool for football coaches and aspiring players, enabling them to analyze videos and study player movements in detail.

Before the advent of YOLO and ByteTrack, various studies have explored the complexities of multiple object tracking. Khan et al.
[1] introduce an MCMC-based particle filter for tracking multiple interacting targets. Gelgon et al. [2] present an original and ef- ficient probabilistic multiple hypothesis

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

tracking (PMHT) method tailored for tracking multiple moving objects in image sequences, addressing challenges like occlusions and crossings, and validated through experiments on real-world data. Sullivan et al. [3] propose a Bayesian network-based approach to multi-target tracking, partic- ularly in scenarios with frequent occlusions and interactions, such as football player tracking. Their method constructs a track graph and defines similarity measures between isolated tracks, achieving efficient inference while gracefully reducing complexity in large-scale problems.

Subsequently, artificial intelligence techniques expanded into com- puter vision, with researchers employing modified neural networks such as region-based convolutional neural networks (RCNN), Faster RCNN, You Only Look Once (YOLO), and Single Shot Detector (SSD) for player and ball tracking tasks. Kamble et al. [4] introduce a novel deep learning approach for 2D ball detection and tracking in soccer videos, achieving exceptional accuracy and robustness even in challenging scenarios. Sarwas et al. [5] present DeepBall, a specialized deep neural network for ball detection in long shot videos, achieving state-of-the-art results on the ISSIA-CNR Soccer Dataset. Reno` et al. [6] introduce an innovative deep learning approach for ball detection in tennis games, achieving high accuracy even in challenging conditions. Speck et al. [7] present a real-time ball localization model for RoboCup Soccer scenes, providing publicly available training and test datasets.

In the context of tracking football players in football videos, there has been limited exploration utilizing YOLOv5 and ByteTrack, making it a novel and promising approach. YOLOv5 and ByteTrack offer several advantages that make them suitable for this task.

Firstly, YOLOv5 [8] is known for its speed and accuracy in object detection tasks, making it well-suited for real-time applications such as tracking football players in dynamic video sequences. Addition- ally, YOLOv5's architecture allows for efficient processing of large amounts of data, which is essential for handling the complexities of football videos.

On the other hand, ByteTrack [9] presents a novel approach to multi-object tracking (MOT) by associating nearly every detection box rather than solely relying on high-score detections. Traditional methods often discard low-score detection boxes, leading to missed objects and fragmented trajectories, especially for occluded objects. ByteTrack addresses this issue by leveraging the similarities between low-score detection boxes and existing tracklets to recover true objects and filter out background detections.

At the input stage, YOLOv5 employs several techniques including mosaic data augmentation, adaptive anchor box estimation, and adaptive image scaling. Mosaic data augmentation combines four images into one, enhancing background diversity and improving detection performance for small objects. Adaptive anchor box estima- tion dynamically computes optimal anchor box values by clustering annotated boxes in the training set. Adaptive image scaling reduces computational load and speeds up inference by adding minimal black borders when resizing images to the input size. The backbone network of YOLOv5 features CSP structure and focus modules, while the neck layer utilizes FPN+PAN structure for enhanced feature fusion. In the output stage, GIoU Loss [11] replaces traditional IoU Loss as the bounding box loss function, offering improved detection performance by overcoming IoU's limitations. Its calculation method is:

The paper will include the following sections:

1. Methodology: Details the implementation of YOLOv5 and LGIoU

$$= \text{LIoU} - \frac{|Ac - U|}{|Ac|} \quad (1)$$

ByteTrack for football player tracking, covering preprocessing steps, model configurations, and additional techniques for enhanced accu- racy.

2. Results and Discussion: Presents outcomes of YOLOv5 and ByteTrack application, including tracking accuracy, speed metrics, and qualitative analysis of tracking performance. Discusses potential challenges and limitations.
3. Conclusion: Summarizes findings, underscores the significance of YOLOv5 and ByteTrack for football player tracking, and suggests future research directions.

II. METHODOLOGY

In this section, we will introduce two popular object detection models: YOLOv5 and ByteTrack, and demonstrate how to utilize them for our project.

A. **YOLOv5**

The YOLOv5 [8] model was officially introduced by Ultralytics on June 25th, 2020. It represents an advancement over previous iterations from YOLOv1 to YOLOv4, integrating contemporary techniques in object detection to enhance its performance. The model's implemen- tation is based on the widely-used PyTorch deep learning framework. In contrast to the Darknet framework utilized in earlier versions, YOLOv5 boasts a more mature ecosystem with comprehensive soft- ware and hardware support, facilitating deployment across various devices. YOLOv5 comes in four different sizes—v5s, v5m, v5l, and v5x—each with varying weight file sizes (14, 42, 93, and 170MB, respectively). As the model's size increases, so does its detection accuracy, albeit at the cost of reduced detection speed and increased resource consumption.

The architecture of YOLOv5 can be divided into four main components: input, backbone, neck, and output.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

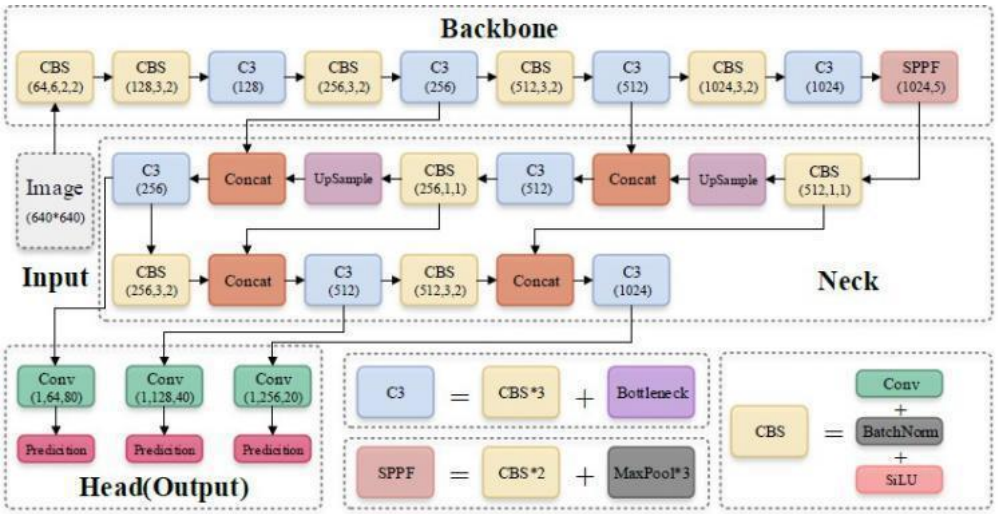

Fig. 1. YOLO v5 Model Architecture [10]

where LIoU stands for the traditional Intersection over Union (IoU), which is the ratio of the intersection area of the predicted box and the ground truth box to their union area. Ac represents the area of the minimum enclosing rectangle of the predicted box and the ground truth box. U denotes the area of the union of the predicted box and the ground truth box.

B. **ByteTrack**

The ByteTrack algorithm represents a cutting-edge advancement in the realm of multi-object tracking, leveraging the capabilities of robust object detection networks like YOLOv8 and YOLOv5 to attain unparalleled precision in object detection tasks. Its foundation rests upon a sophisticated data association strategy, which forms the heart of its functionality. This strategy incorporates an inventive association mechanism designed to seamlessly link objects across successive video frames, ensuring the continual preservation of their identities. Even amidst the complexities of dynamic and cluttered environments, ByteTrack stands resilient, steadfastly maintaining tracking integrity and resisting the common pitfalls of tracking failure.

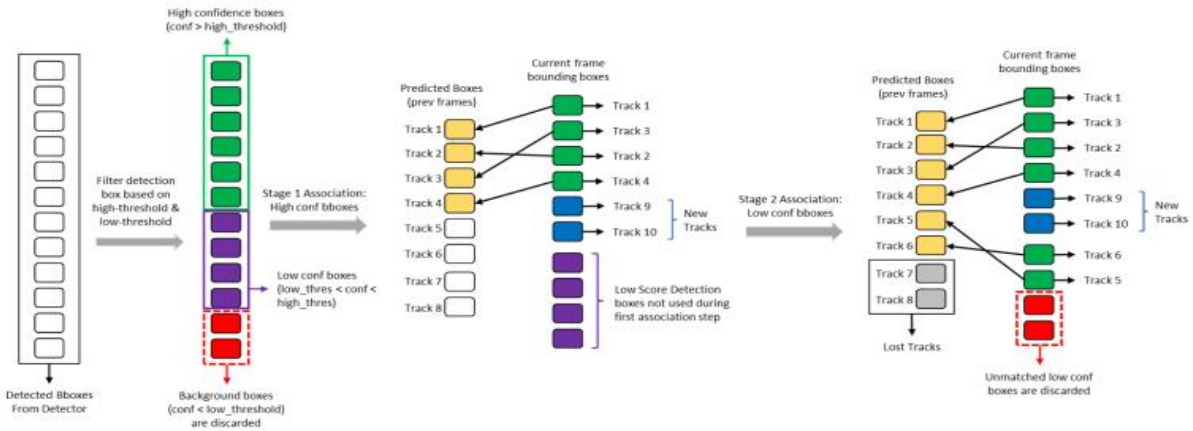

Fig. 2. ByteTrack Model Architecture [12]

In contrast to traditional multi-object tracking methods that typ-ically prioritize high-confidence detection results, ByteTrack stands out by utilizing both high-confidence and low-confidence detection outcomes. This innovative approach allows ByteTrack to exploit motion consistency and appearance information from low-confidence detections, significantly enhancing tracking continuity and robustness, especially in challenging scenarios like

occlusions and dynamic scenes. By retaining low-confidence detection boxes and removing background information, ByteTrack effectively identifies genuine targets, thus reducing missed detections and improving trajectory coherence.

In practical operation, ByteTrack first employs a detection model to identify potential targets in each frame of the video. Subsequently, it utilizes an optimized Hungarian algorithm to match the targets detected in the current frame with existing trajectories. This crucial step meticulously considers the appearance characteristics and motion patterns of the targets, ensuring accurate matching even in scenarios where targets exhibit similarities. For detections that do not match existing trajectories, ByteTrack treats them as new targets and creates fresh trajectories accordingly. ByteTrack incorporates a delay mech- anism during data association, allowing low-confidence detections to be reevaluated in future frames. This capability enables the algorithm to recapture targets shortly after their temporary disappearance, significantly improving occlusion handling capability. Moreover, this mechanism provides a time window for the algorithm to rectify any erroneous matches, further enhancing tracking accuracy.

C. Implementation

In the implementation, we will follow a similar approach to the implementation by Roboflow [13]. You can check the appendix in the end to find a detailed explanation of how to run the code to implement YOLOv5 and ByteTrack on the recorded football video. My project is executed on the Google Colab platform. We utilize the Google Colab Tesla T4 GPU, which is freely available. You can follow this tutorial[2] to setup the GPU. You also need to download version 1.23.5 of NumPy, as my project is incompatible with newer versions. Then, we download the ”DFL - Bundesliga Data Shootout”

[3]dataset from Kaggle. This dataset includes video recordings of nine football games, divided into halves. It comprises three folders and one CSV file. The ”train” folder contains videos for training data, sourced from eight games, while the ”test” folder holds test data. The public leaderboard’s test data includes video recordings from one full game and four half-games, with the rest in the training set. Additionally, there’s a ”clips” folder with short clips from ten games, aiding model generalization. The ”train.csv” file provides event annotations for train folder videos, but we’ll focus solely on using the clips due to their brevity.

Upon configuring YOLOv5, we initiate the process by applying the yolov5x.pt model to our source video. This model file, yolov5x.pt, en- capsulates pre-trained weights and biases, furnishing comprehensive object detection capabilities. By deploying this model, we harness its exceptional accuracy in discerning diverse objects and scenes depicted in the video. This initial step serves as the cornerstone of our detection algorithm, priming the subsequent analysis and processing stages.

Following the application of yolov5x.pt, we proceed to employ best.pt [14] for further analysis of our source video. Best.pt represents a meticulously trained model crafted by Roboflow, meticulously opti- mized for detecting entities such as players, referees, goalkeepers, and balls within images. Leveraging this model facilitates precise object detection within our video footage, providing a robust foundation for subsequent analysis and processing endeavors.

To leverage ByteTrack for football video analysis, our initial step involves defining a pivotal data class known as BYTETrackerArgs. Within this class, we meticulously configure various parameters critical for the ByteTracker algorithm’s functionality. These param- eters encompass essential elements such as the track_thresh, dictating the threshold for object tracking, the track_buffer, specifying the number of frames to retain objects for track- ing purposes, the match_thresh, which determines the thresh- old for matching objects between consecutive frames, and the aspect_ratio_thresh, setting the maximum aspect ratio per- mitted for tracked objects.

Subsequently, we embark on initializing imperative utilities and an- notators indispensable for our video analysis endeavor. Among these utilities are fundamental constructs like ‘Point‘, ‘Rect‘, ‘Detection‘, ‘Color‘, and annotators specifically designed for the visualization of shapes and textual elements on video frames.

Following the setup of utilities, we meticulously configure anno- tators equipped with predefined colors and thickness, meticulously tailored for marking pertinent objects within football video frames. These annotators encompass markers designated for various entities including the ball, players, goalkeepers, and referees, ensuring a comprehensive visual representation of the scene.

With the foundational groundwork laid, we proceed to illustrate the procedural workflow of acquiring video frames, executing ob- ject detection algorithms utilizing a designated model, annotating the frames with detected objects, and subsequently rendering the annotated frames for visualization purposes.

Furthermore, we establish marker colors and dimensions, delin- eating marker contours, and define a proximity distance essential for discerning player possession of the ball, thereby facilitating a nuanced understanding of player dynamics during gameplay.

Moreover, we implement sophisticated functions tailored for cal- culating and drawing possession markers on video frames, alongside mechanisms for discerning the player currently in possession of the ball based on proximity, enhancing the depth of analysis.

Subsequently, we present a detailed demonstration on iterating over frames within a video sequence, conducting object detection operations, tracking objects leveraging the

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

ByteTracker algorithm, annotating frames with markers and textual elements denoting object IDs, and persisting the annotated frames as a new video for further scrutiny.
Finally, we culminate our exploration by showcasing the procedure for writing detection counts for distinct classes such as football players, goalkeepers, footballs, and referees to a CSV file, facilitating comprehensive data analysis and insights extraction.

III. RESULTS

In this section, we present the results of our experiments, starting with the performance evaluation of YOLOv5 followed by ByteTrack.

A. Performance Analysis of YOLOv5

We initiated our analysis by evaluating the performance of YOLOv5, focusing initially on the yolov5x.pt variant. Our obser- vations revealed that the detection rates obtained were comparatively low across the board. Specifically, when detecting individuals in videos, the detection rates for persons typically ranged between 0.3 to 0.7 (Fig. 3).
Our analysis went deeper into the mean detection rates for each label, shedding light on the performance of the yolov5x.pt model across a spectrum of objects. Fig. 4 illustrates the model's capability in identifying various items, encompassing benches, books, chairs, cups, persons, sports balls, tennis rackets, TVs, and umbrellas. However, amidst this breadth of detections, several issues surfaced that are pertinent to our objectives.
Foremost among these concerns is the prevalence of irrelevant detections and misclassifications. Despite the model's capacity to identify a diverse array of objects, many of these detections do not align with the specific objectives of our application. For instance, erroneous identifications of tennis rackets and cups, among others, detract from the model's utility, potentially leading to erroneous conclusions or actions based on flawed data.
Moreover, even within the relevant categories, such as persons and sports balls, the average detection rates remain notably deficient. As previously noted, the average detection rate for persons stands at approximately 0.61, indicative of a substantial number of missed detections within the scene. Similarly, the detection rate for sports balls hovers around 0.39, suggesting a significant portion of these objects remain undetected by the model.
Furthermore, our analysis revealed specific inadequacies in the model's ability to detect certain crucial entities, notably referees and goalkeepers. These shortcomings pose significant challenges, particularly in scenarios where accurate identification and tracking of such individuals are imperative, such as in sports event monitoring or security applications.

Fig. 3. YOLOv5 Video Result by yolov5x.pt

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

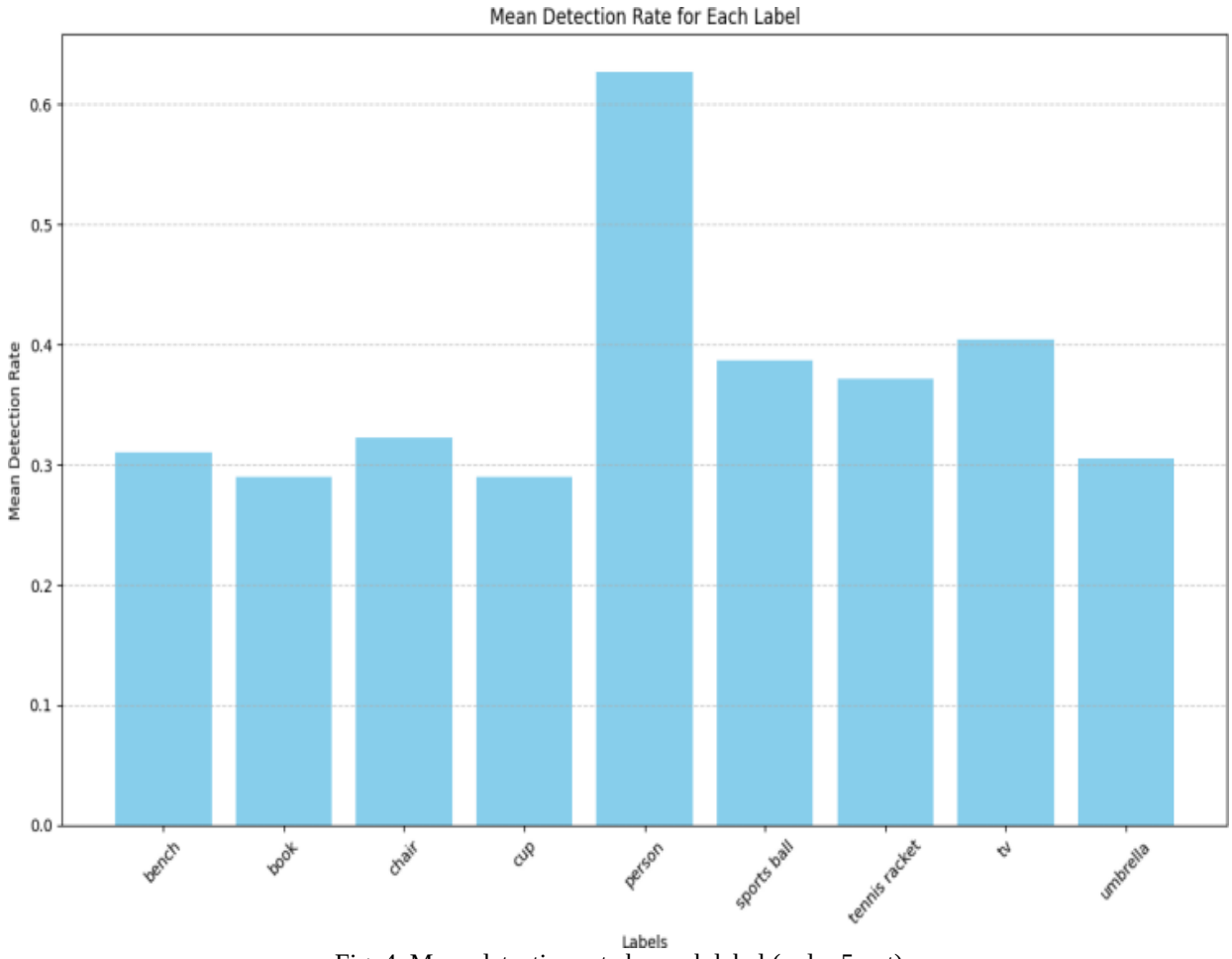

Fig. 4. Mean detection rate by each label (yolov5x.pt)

Upon closer examination of the detection rates for individuals and sports balls separately (Fig. 5), it becomes evident that there are notable shortcomings in the model's performance. This observation is particularly significant given the pivotal role that both persons and sports balls play in the application context.

Firstly, the detection rate for persons exhibits a range from as low as 0.25 to a modest 0.85, with an average hovering around 0.61. This wide range and relatively low average indicate a considerable number of instances where the model fails to detect individuals within the scene accurately. Such inconsistencies suggest that there are instances where persons are present in the scene yet go unnoticed by the model. This could potentially lead to critical oversights, especially in scenarios where the presence and activities of individuals are of utmost importance, such as in surveillance or crowd monitoring applications.

Similarly, the detection rate for sports balls displays a notably low performance, with rates fluctuating between 0.25 and 0.6, averaging around 0.39. This indicates that a significant portion of sports balls within the scene remain undetected by the model. In contexts where tracking sports-related activities or objects is vital, such as in sports analytics or event coverage, this level of inconsistency in detection could lead to incomplete or inaccurate analyses.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

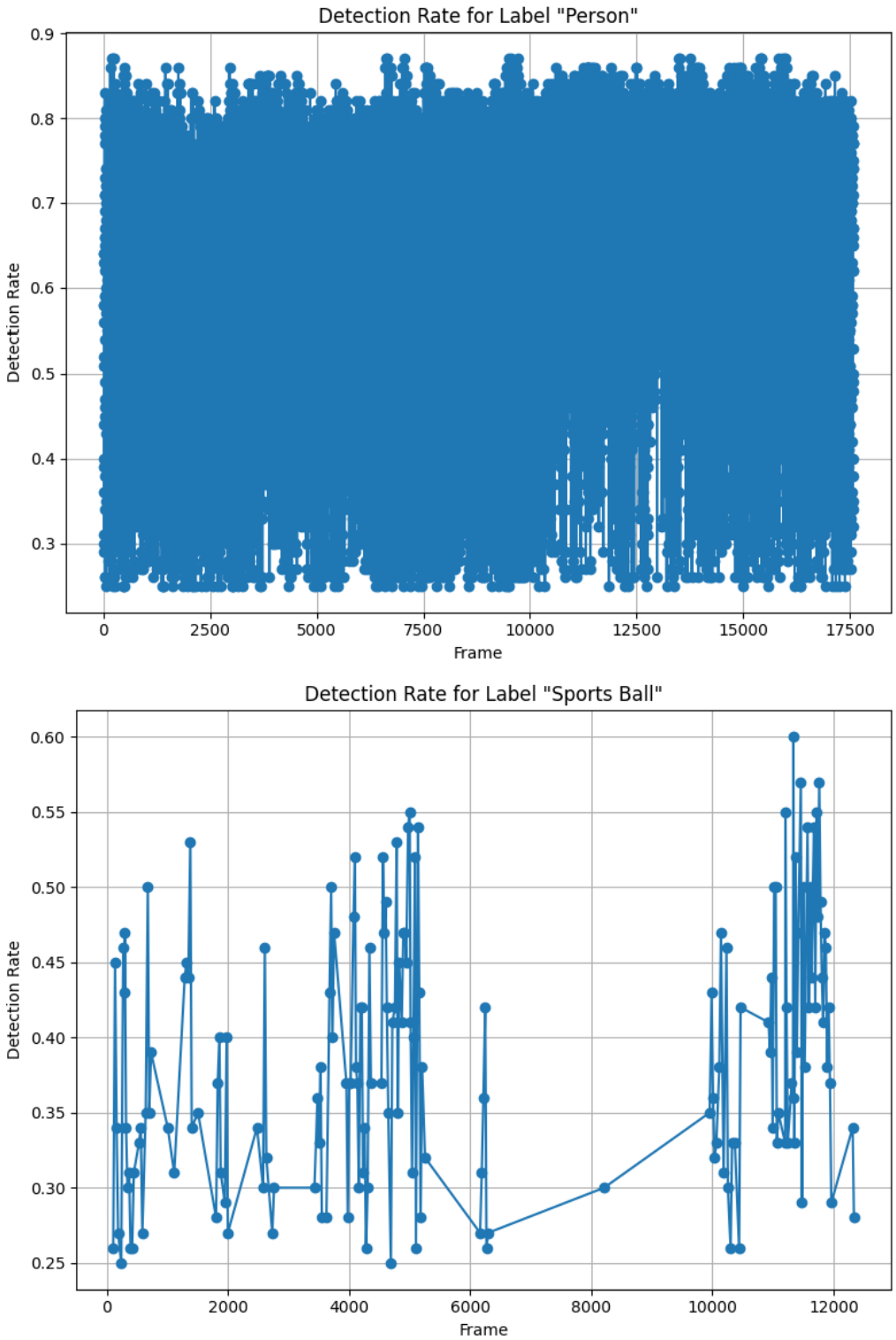

Fig. 5. Detection rate for labels ”Person” and ”Sports Ball”

Based on the aforementioned results, the misidentification of objects by the yolov5x.pt model can be attributed to various factors. Firstly, insufficient training data or the absence of fine-tuning tailored to our specific application context may result in the misclassification of objects that are

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

either underrepresented or share similar visual characteristics with those the model was trained on.
Moreover, occlusions, fluctuations in lighting conditions, or clut-tered backgrounds within the video frames can introduce noise into the input data, posing challenges for the model to accurately differentiate between different objects.
Additionally, the inherent constraints of the model architecture and the complexity of the task at hand may contribute to misidentifica-tions. Despite being a potent tool, YOLOv5 may encounter difficulties with certain object types or scenes, particularly if they significantly differ from the training data.

To enhance object detection accuracy, we utilize the pre-trained best.pt model, which offer improved performance due to its opti-mization and suitability.
We proceeded by utilizing the best.pt model, which was custom- tailored to identify essential elements such as players, balls, referees, and goalkeepers. This precise configuration ensured that only rel-evant objects were identified, effectively eliminating any irrelevant detections. As a result, interference from extraneous objects that could potentially disrupt accurate detection was successfully miti-gated. Consequently, the detection rate for these critical elements experienced a notable enhancement.

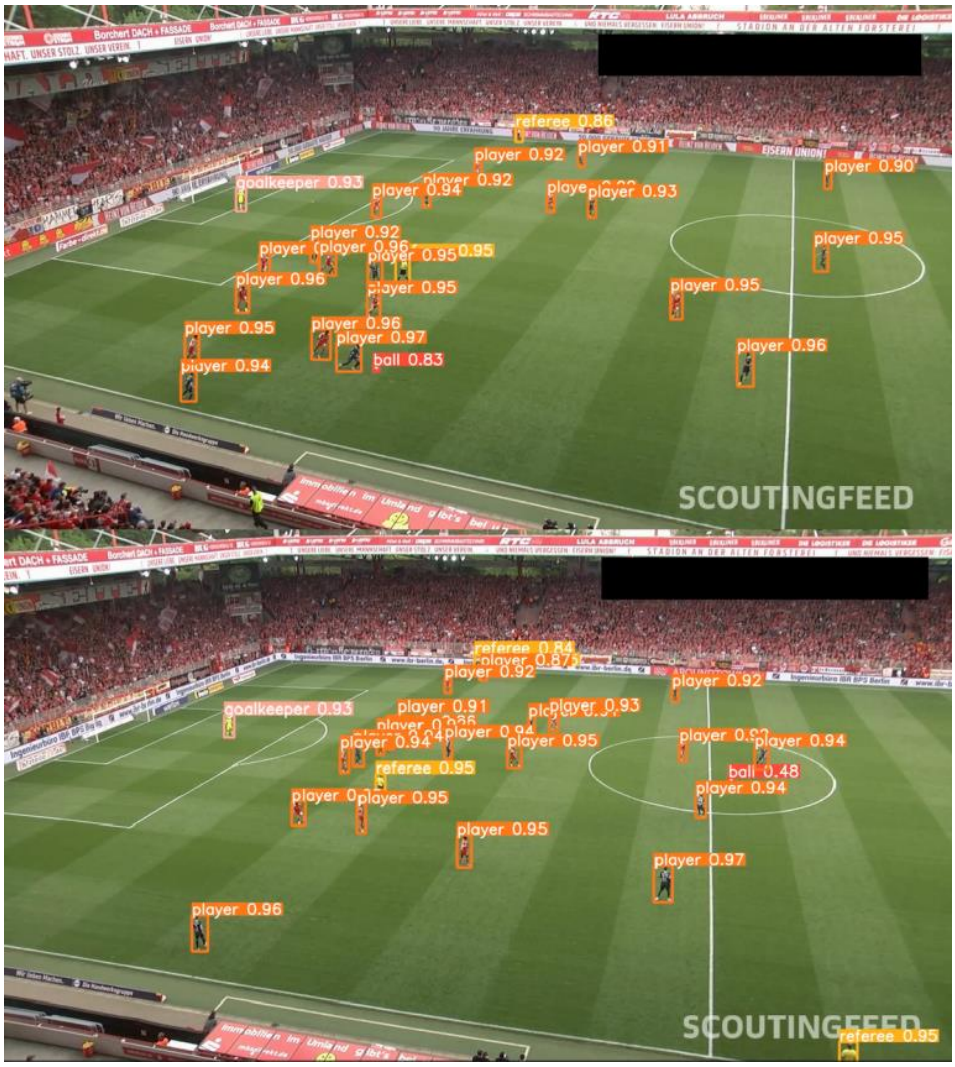

Fig. 6. YOLOv5 Video Result by best.pt

Upon analyzing the mean detection rate for each label, we ob- served that the best.pt model accurately identifies various objects, including balls, players, referees, and goalkeepers (Fig. 7). Notably, the mean detection rates for goalkeepers, referees, and players were all exceptionally high, surpassing 0.8. Furthermore, there was a significant improvement in the detection rate for balls, reaching around 0.7, indicating a substantial enhancement compared to previ- ous results.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

The exceptionally high mean detection rates for goalkeepers, ref- erees, and players, surpassing 0.8, can be attributed to the meticulous training process implemented with the best.pt model. This model underwent a rigorous fine-tuning procedure facilitated by manual labeling, where each instance of the desired labels was meticulously annotated.

The key to the remarkable performance lies in the quality and specificity of the training data provided to the model. Through manual labeling, annotators painstakingly delineated and tagged every instance of goalkeepers, referees, and players within the training dataset. This level of granularity ensured that the model was exposed to a diverse array of visual characteristics and contextual variations associated with these objects across various scenes and scenarios.

By providing the model with precisely labeled data, it gained a deep understanding of the nuanced visual cues and contextual cues

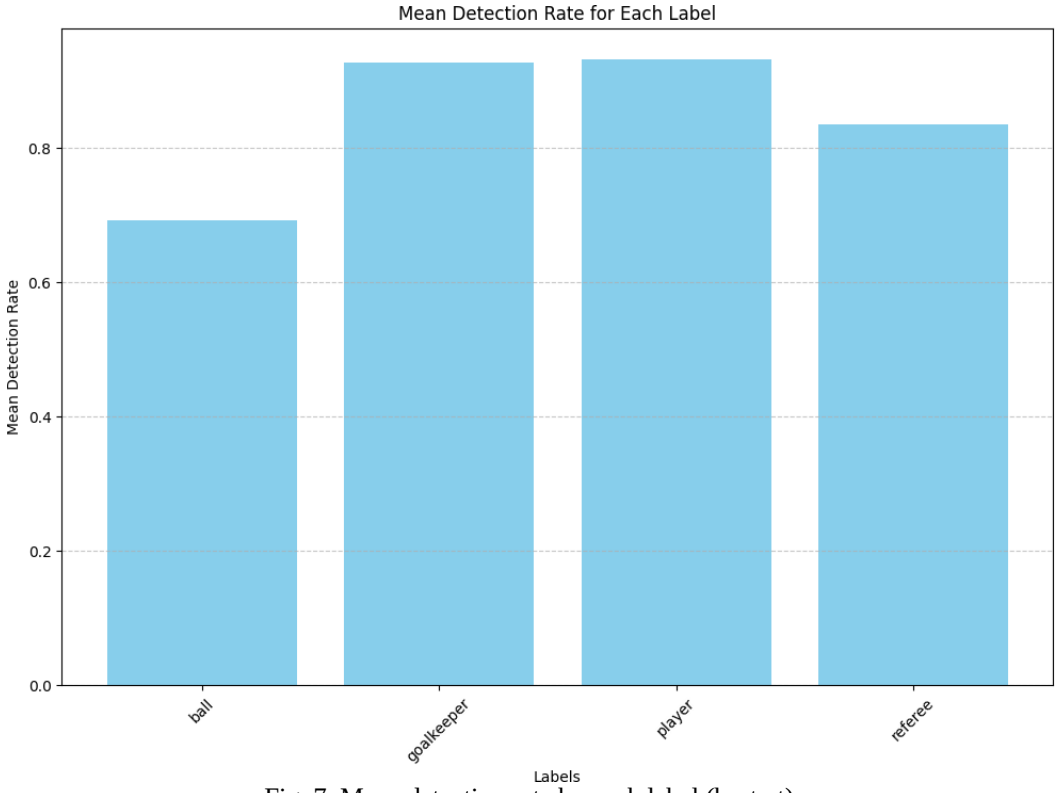

Fig. 7. Mean detection rate by each label (best.pt)

associated with goalkeepers, referees, and players. This enabled the model to discern subtle visual features, such as distinctive attire, unique postures, and characteristic movements, indicative of each object class. As a result, the model became adept at accurately iden- tifying and localizing goalkeepers, referees, and players within video frames, leading to the exceptional mean detection rates observed.

Furthermore, the meticulous fine-tuning process ensured that the model was optimized to prioritize the recognition of these specific object classes. By focusing training efforts on goalkeepers, referees, and players, the model's attention and learning capacity were directed towards refining its ability to detect and classify these objects with a high degree of accuracy.

However, achieving a detection rate above 0.8 for the soccer ball presents additional challenges. Firstly, the size of the ball relative to the frame can make it challenging for the model to detect, particularly in instances where the ball appears small or distant. Moreover, environmental factors such as the condition of the playing field, lighting variations, and video resolution can further impact the model's ability to accurately identify the ball.

Furthermore, the dynamic nature of soccer gameplay introduces complexities that can affect detection performance. For example, during passing or receiving plays, players may occlude the ball with their bodies, shields, or other objects, making it momentarily invisible to the model. Additionally, the rapid movement of the ball and players can introduce motion blur, further complicating detection.

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

These combined factors contribute to the lower detection rate for the soccer ball compared to other labels such as players, goalkeepers, and referees. While efforts can be made to improve detection accu- racy through continued optimization and fine-tuning of the model, addressing these inherent challenges remains essential for achieving consistently high detection rates for all objects of interest.

B. **Performance Analysis of ByteTrack**

ByteTrack is an object tracking algorithm that operates differ- ently from traditional object detection algorithms like YOLO. While YOLO focuses on detecting objects in individual frames, ByteTrack specializes in tracking objects across consecutive frames of a video sequence.

In our setup, we first employ ByteTrack to recognize footballs and players in possession of the ball. Subsequently, we utilize ByteTrack to identify all players on the field. By integrating these two processes, we achieve a comprehensive system capable of recognizing footballs, players in possession, and all players present in the scene. This integrated approach is illustrated in Figure 9

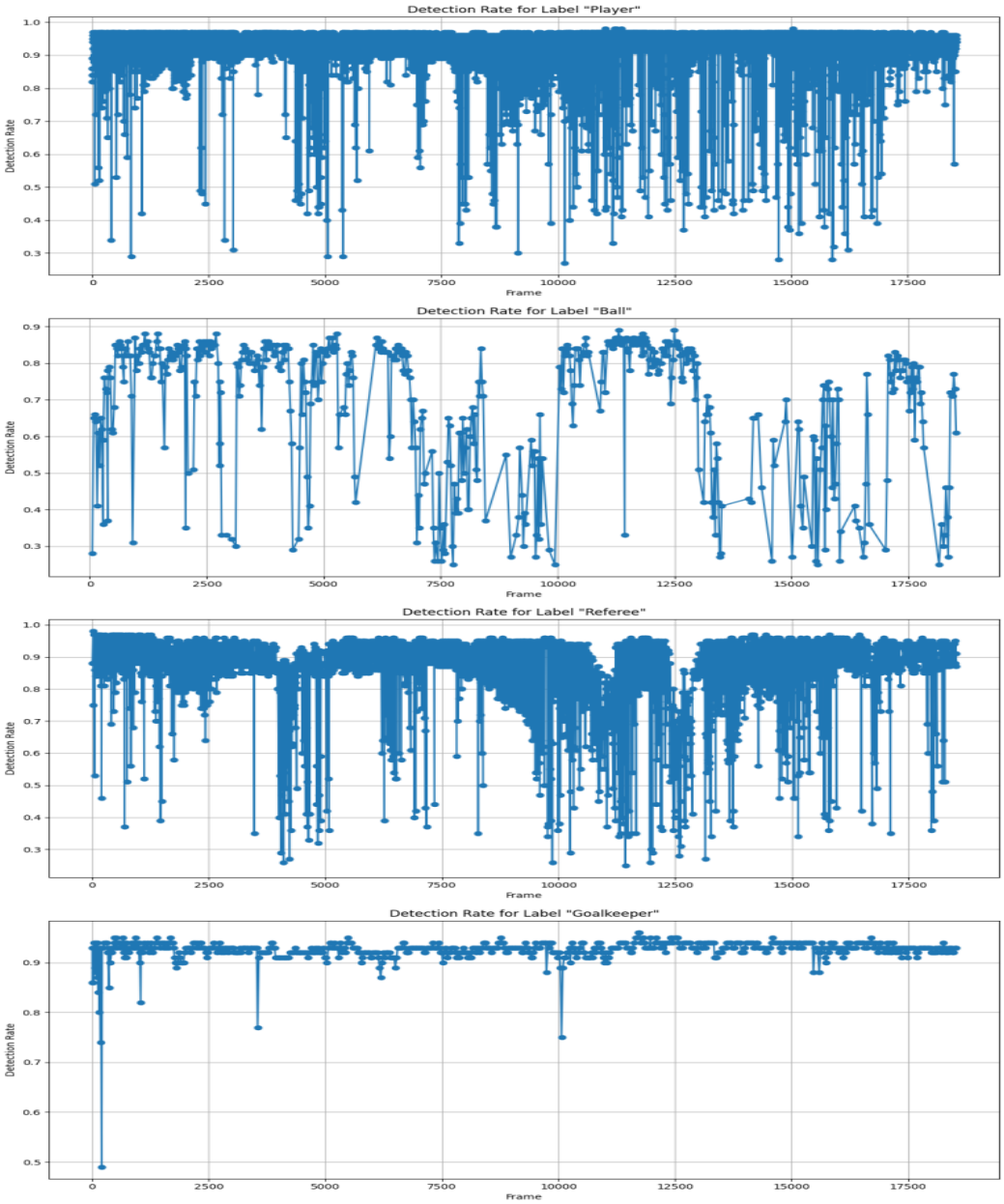

Fig. 8. Detection rate for four labels in best.pt

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

In the analysis of ByteTrack's tracking results, it's evident that while the algorithm excels in accurately predicting the trajectories of football players and goalkeepers, its performance with regards to tracking footballs and referees is somewhat less precise. This discrepancy in performance can be attributed to several factors.

One key factor contributing to the accurate prediction of football players and goalkeepers is the relatively distinct and recognizable appearance of these objects within the video frames. Football players and goalkeepers typically exhibit distinctive attire, such as team jerseys or goalkeeping gloves, which serve as visual cues facilitating their accurate identification and tracking. Additionally, the predictable movement patterns associated with football players and goalkeepers, such as running, dribbling, or positioning within the goal area, further enhance ByteTrack's ability to anticipate their trajectories across consecutive frames.

Fig. 9. ByteTrack Video Result

Furthermore, the strategic positioning of football players and goalkeepers within the field of play often facilitates unobstructed visibility, minimizing occlusions and ambiguities that may impede accurate tracking. This enhanced visibility enables ByteTrack to maintain consistent object identities and trajectories, thereby con- tributing to its high tracking accuracy for these specific object classes. On the other hand, tracking footballs and referees presents dis- tinct challenges due to their smaller size, variable appearance, and susceptibility to occlusions and rapid movements within the scene. Footballs, in particular, may exhibit erratic motion patterns, such as bouncing or spinning, which can complicate their tracking process. Similarly, referees may move swiftly across the field, and their appearance may be less distinct compared to football players, making them more prone to misidentification or confusion with other objects. Despite these challenges, it's essential to acknowledge that Byte- Track's overall performance remains commendable, with efficient tracking capabilities demonstrated across various object classes. By leveraging its strengths in accurately predicting the trajectories of individual frames with remarkable precision and efficiency. Leverag- ing a single-stage architecture, YOLOv5 enables real-time inference, rendering it ideal for applications necessitating swift object detection in static images or video frames. Through its robust training process and sophisticated architectural design, YOLOv5 attains impressive detection accuracy across a diverse spectrum of object classes, encompassing footballs, players, goalkeepers, and referees.

In contrast, ByteTrack sets itself apart by specializing in ob- ject tracking across consecutive frames within a video sequence. By harnessing motion cues and temporal information, ByteTrack enhances its capacity to preserve object identities and trajectories over time, thereby facilitating seamless monitoring and analysis of dynamic

International Journal of Engineering Applied Sciences and Technology, 2024
Vol. 9, Issue 01, ISSN No. 2455-2143, Pages 12-25
Published Online May 2024 in IJEAST (http://www.ijeast.com)

scenes. This unique attribute renders ByteTrack particularly invaluable in scenarios demanding prolonged object tracking, such as sports event analysis or surveillance operations.

Regarding performance comparison, YOLOv5 typically exhibits superior accuracy in object detection within individual frames, espe- cially for smaller or less discernible objects like footballs and refer- ees. Its multi-scale feature fusion and advanced network architecture equip it with the capability to capture intricate details and subtle object nuances with exceptional fidelity.

Conversely, ByteTrack excels in accurately forecasting object trajectories and preserving consistent object identities across frames, particularly for larger and more conspicuous objects such as football players and goalkeepers. Its emphasis on motion-based tracking empowers it to effectively navigate challenges like occlusions, motion blur, and other complexities inherent in video analysis scenarios.

In summary, while YOLOv5 shines in rapid and precise ob- ject detection within static frames, ByteTrack offers complementary prowess in robust object tracking across dynamic video sequences. By synergizing the strengths of both algorithms, it becomes feasible to harness their collective capabilities to achieve comprehensive and precise object detection and tracking across a myriad of real-world applications.

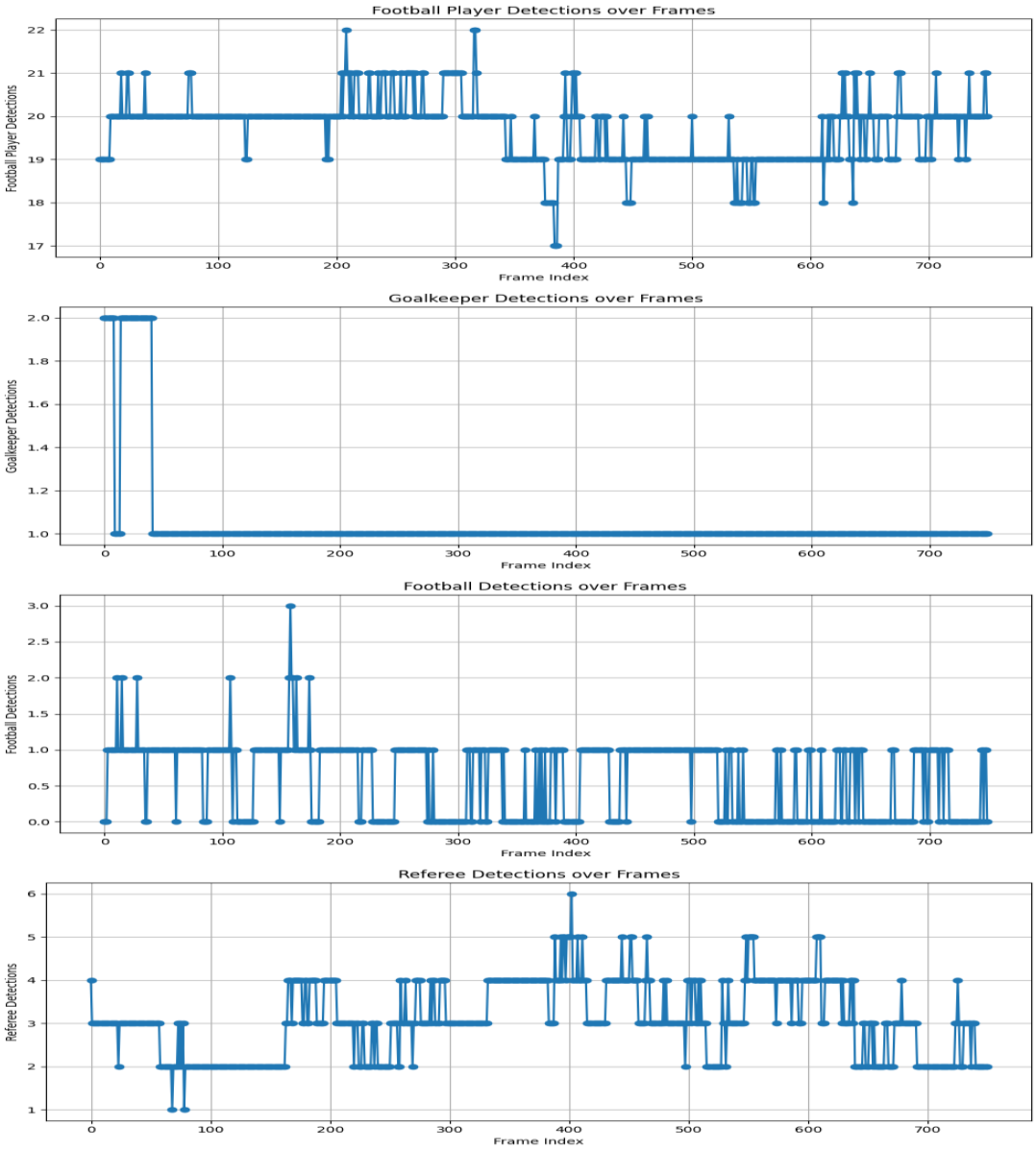

Fig. 10. ByteTrack tracking Result

football players and goalkeepers, while simultaneously addressing limitations in tracking footballs and referees, ByteTrack represents a valuable tool for comprehensive object tracking in dynamic video environments. Continued refinement and optimization efforts can further enhance its effectiveness in diverse real-world applications, ensuring reliable and accurate tracking of objects of interest.

C. Comparison between YOLOv5 and ByteTrack

When comparing ByteTrack with YOLOv5, it's imperative to discern their individual strengths and limitations in the realm of object tracking and detection.

YOLOv5 stands out as a leading object detection algorithm renowned for its adeptness in accurately identifying objects within

IV. CONCLUSION

In conclusion, our comparative analysis between ByteTrack and YOLOv5 underscores the distinct strengths and limitations of each algorithm in the domain of object tracking and detection.

Key findings from our investigation reveal that while YOLOv5 excels in rapid and accurate object detection within individual frames, ByteTrack distinguishes itself with its proficiency in object tracking across consecutive frames within video sequences. YOLOv5 demon- strates superior accuracy in detecting objects, including football players, goalkeepers, and referees, with detection rates exceeding 80%, and footballs with rates surpassing 60%, owing to its advanced network architecture and multi-scale feature fusion. On the other hand, ByteTrack showcases remarkable capabilities in accurately pre- dicting object trajectories and preserving consistent object identities over time, particularly for larger and more conspicuous objects such as football players and goalkeepers.

Important conclusions drawn from our analysis emphasize the complementary nature of ByteTrack and YOLOv5 in addressing different aspects of object tracking and detection tasks. By integrating both algorithms, it becomes possible to harness their collective strengths to achieve comprehensive and precise object detection and tracking across a wide range of real-world applications, ranging from sports event analysis to surveillance operations.

However, it is essential to acknowledge the limitations inherent in both approaches. While YOLOv5 may struggle with tracking objects across consecutive frames, ByteTrack may encounter challenges in accurately detecting smaller or less distinct objects, particularly in complex scenes with occlusions or rapid movements. Specifically, both YOLOv5 and ByteTrack may face difficulties in accurately capturing small footballs, which can sometimes go undetected. Addi- tionally, football players may obstruct the view of the ball, leading to intermittent tracking issues. Moreover, during instances where players contest possession of the ball, the ball itself may become temporarily obscured, further complicating the tracking process.

Moving forward, future research efforts could focus on address- ing these limitations and further refining the performance of both algorithms. For instance, exploring hybrid approaches that combine the strengths of YOLOv5's object detection, particularly in capturing smaller objects like footballs, with ByteTrack's object tracking ca- pabilities, could yield improved results. Additionally, investigating novel techniques for enhancing the robustness and efficiency of object detection and tracking in dynamic video environments remains a promising avenue for future exploration. These techniques may include advanced methods for handling occlusions, refining motion prediction algorithms to better anticipate object trajectories, and optimizing model architectures to improve real-time performance in challenging scenarios. By leveraging innovative strategies and integrating complementary techniques, researchers can pave the way for more accurate and reliable object tracking and detection systems, addressing the evolving demands of various real-world applications. In conclusion, while our study sheds light on the comparative strengths and limitations of ByteTrack and YOLOv5, there remains ample opportunity for continued innovation and advancement in the field of object tracking and detection. By leveraging emerging technologies and methodologies, researchers can strive towards de- veloping more robust and effective solutions to address the evolving demands of real-world applications.

V. REFERENCES

[1]. Khan, Z. Balch, T. Dellaert, F. (2004). An MCMC-Based Particle Filter for Tracking Multiple Interacting Targets. In: Pajdla, T., Matas,J. (eds) Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in Computer Science, vol 3024. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24673-2 23

[2]. Gelgon, M., Bouthemy, P., Cadre, J. (2005). Recovery of the tra- jectories of multiple moving objects in an image sequence with a PMHT approach, Image and Vision Computing, 23, 19-31. 10.1016/j.imavis.2004.07.004.

[3]. Sullivan, J., Nillius, P., Carlsson, S. (2009). Multi-target Tracking on a Large Scale: Experiences from Football Player Tracking.

[4]. Kamble, P., Keskar, A., Bhurchandi, K. (2019). A deep learning ball tracking system in soccer videos, Opto-Electronics Review, 27, 58-69. https://doi.org/10.1016/j.opelre.2019.02.003.

[5]. Sarwas, G., Komorowski, J., Kurzejamski, G. (2019). DeepBall: Deep Neural-

Network Ball Detector. https://doi.org/10.5220/0007348902970304.

[6]. Reno`, V., Mosca, N., Marani, R., Nitti, M., D'Orazio, T., Stella, E. (2018). Convolutional Neural Networks Based Ball Detection in Tennis Games, 1839-18396. https://doi.org/10.1109/CVPRW.2018.00228

[7]. Speck, D., Bestmann, M., Barros, P. (2018). Towards Real-Time Ball Localization using CNNs.

[8]. Jocher, G. (2021). ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations, Zenodo, Apr. 11, 2021. doi: 10.5281/zenodo.4679653.

[9]. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box, arXiv preprint arXiv:2110.06864.

[10]. Tsang, S.-H. (2024). Brief review: Yolov5 for object detection, Medium, https://sh-tsang.medium.com/brief-review-yolov5-for-object-detection- 84cc6c6a0e3a (accessed Apr. 3, 2024).

[11]. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S. (2019). Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression.

[12]. Le, B. (2024). An introduction to bytetrack: Multi-object tracking by associating every detection box, Datature, https://www.datature.io/blog/introduction-to-bytetrack-multi-object- tracking-by-associating-every-detection-box (accessed Apr. 3, 2024).

[13]. Roboflow. (2024). How to Track Football Players, GitHub, https://github.com/roboflow/notebooks/blob/main/notebooks/how-to- track-football-players.ipynb (accessed Apr. 3, 2024).

[14]. Roboflow. (2024). Football-Players-Detection Dataset, https://universe.roboflow.com/roboflow-jvuqo/football-players-detection-3zvbc (accessed Apr. 3, 2024).