Object Detection and Tracking for Football Data Analytics

Conference Paper · January 2024

DOI: 10.4108/eai.23-11-2023.2343216

Guruprakash Jayabalasamy
Amrita Vishwa Vidyapeetham

Object Detection and Tracking for Football Data Analytics

Shankara Narayanan V[1], Syed Ashfaq Ahmed[2], Sneha Varsha M[3], Guruprakash J[4]

{cb.en.u4cse20656@cb.students.amrita.edu[1], cb.en.u4cse20665@cb.students.amrita.edu[2] cb.en.u4cse20659@cb.students.amrita.edu[3], j guruprakash@cb.amrita.edu[4]}

Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore, Amrita Vishwa Vidyapeetham, India[1234]

Abstract. This paper presents a method of quantifying ball possession and its usage in foot- ball sports data analytics by using object detection and object tracking. After comparing the performance of YOLOv5 and YOLOv8 which are two state-of-the-art object detection models, the latter was chosen to be used along with BYTETrack for object detection and tracking. The input will be a video stream of a football game taken from a tactical camera which is passed to the object detection module. The detected objects are individually tracked and ball possession is calculated per player by assigning unique track-id for all players. Finally, aggregating player's individual ball possession into their respective teams provides a way of estimating the team's ball possession.

Keywords: Object Detection, Object Tracking, Ball Possession, Sports Analytics, YOLO Models.

1 Introduction

Since the early 1950s sports analytics has gained more traction amongst sports aficionados and over the course of time, tracking and analytics began playing a crucial part amidst the realm of sports. With the advent of Artificial Intelligence(AI) in our everyday lives, the ever-enthusiastic world of sports did not miss its chance and soon enough sports analysts quickly began using AI to their ad- vantage by utilizing game recordings and existing facts and figures to draw relevant conclusions[1]. Apart from football, studies have been conducted in various other sports such as Cricket where AI has played a major role in producing accurate third umpire decisions [2]. More commonly used statistics today such as pass counts, pass completions and possession came into practice only in the 1990s. These insights provide crucial information about individual player performance as well as team performance, and determine any existing correlation between these data and the final outcome[3].

Ball Possession arose as one of the major factors in soccer analytics wherein the time duration during which the ball remained in any one team's control factored into the end result[4].Recent

IACIDS 2023, November 23-25, Lavasa, India

DOI 10.4108/eai.23-11-2023.2343216

studies have indicated that even though there may not be a direct correlation between ball possession and winning probability, it does factor into establishing the rhythm and pace of the game as well as pressurizing the opposing team to increase their efforts physically and fatigue[5][6]. Furthermore, trends have been noticed about ball possession such as the effect of environmental and contextual factors wherein home teams tend to have a higher ball possession in comparison to their opponents, and teams on the verge of losing tend to have higher ball possession[5][7][8].

Considering the preceding justifications this paper aims to employ a framework for Multi-Object tracking using which the players and the ball can be identified and tracked to provide useful inferences such as ball possession as explained above. Moreover, different YOLO versions were experimented with to find the best one and justified the decision through the precision and recall trade-off.

An individual ball possession model by [9] proposes a framework wherein the Rauch-Tung-Striebel (RTS) [10] method is used to smooth out the trajectories of the ball and players to calculate the distance between the objects. An object detection framework proposed in [11] using YOLOv5 model[12] where the SoccerNet Multi-Object Tracking(MOT) dataset [13] has been used. Two different YOLOv5 frameworks have been used which are the YOLOv5m model and the YOLOv5s model and the results are being compared. A framework to estimate ball possession in a soccer match has been proposed by [14]. The model takes each individual frame and searches for objects in the ball's neighborhood. If an object of the class player is detected, the player is assumed to have possession of the ball[15].In the field of sports analytics, a framework has been proposed by [16] wherein the sound of the referee's whistle is automatically detected and used to extract highlights.

2 Experiments

The experiments performed in this paper include multiple object recognition, object tracking, and ball possession among the players.

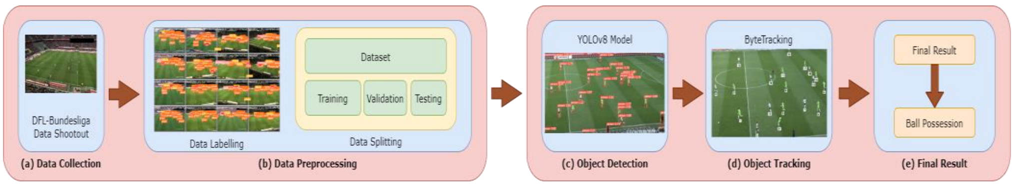


Fig. 1. A Pipeline Diagram depicting the step-wise procedure followed.

As illustrated in Fig. 1, the experiment begins with the collection of data, followed by manual annotation(labeling). The annotated data is then separated into the training, validation, and testing sets. After preprocessing, the YOLOv8s model is used for object detection using pre-defined classes. Furthermore, the BYTETrack model is used to track the identified classes of objects. The rest of the experiment section will be divided into four parts viz Dataset, Object Detection, Object Tracking, and Ball Possession.

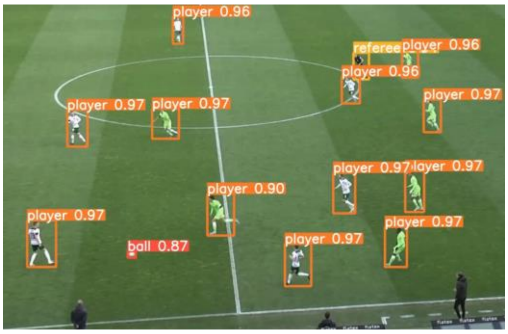


Fig. 2. Object Detection using YOLOv8

Fig. 3. Player Tracking using BYTETrack

2.1 Dataset

The dataset comprises of images taken from 200 football match clips provided by the DFL-Bundesliga Data Shootout Kaggle competition[17]. Upon randomly extracting 3 frames per video, a dataset of 600 images is created for training the YOLO model. For image annotation, the online tool MakeSense.ai[18] was used. The most prominent classes of objects present in a football match will be the players, referee, goalkeeper, and the ball. The annotation classes were labeled as ‘0’ for ‘ball’, ‘1’ for ‘goalkeeper’, ‘2’ for ‘player’, ‘3’ for ‘referee’. The Players are detected in almost all the instances, and rightfully so because they outnumber the other classes of objects by a large margin. Referees are the second highly represented class of objects and this is because of the positioning of the three referees in different parts of the field. Even though the ball is one of the lowest-represented classes, it is present in most of the frames. Finally, the goalkeepers are the lowest represented class.

2.2 Object Detection

YOLO, or ”You Only Look Once” is a prevalent multi-class object detection model. The model works by splitting the image into *m* cells on a matrix and ascertains whether a given cell carries the central coordinates of a classifiable and identifiable object. YOLO models accuracy in detecting objects rapidly helps in situations where the human eye cannot track multiple objects simultaneously[19]. This has many real-world applications such as tracking vehicle number plates [20] and in the field of medicine as well such as the detection of breast cancer [21].

The preference of YOLO models over the existing R-CNN and Fast R-CNN models is due to its efficiency and speed at detecting smaller objects [22] which is especially useful for our problem statement as it involves the detection of a football which is of a relatively smaller size in comparison to the rest of the scene.

The YOLOv5 model was developed in 2020 by Ultralytics. It is a single-state object detector that is comprised of three parts, namely the Model Backbone, Neck, and Head[23]. The Back- bone is used to extract relevant and important features from the provided input image, which is done using Cross Stage Partial (CSP) networks.The Model Neck uses a variant of Spatial Pyramid Pooling(SPP) and a modified version of the Path Aggregation Network by incorporating Bottle- NeckCSP. The Model Head performs the final detection operation by applying anchor boxes on relevant features[23].

The YOLOv8 model provides significant improvement over the existing YOLOv5 model. The YOLOv8 model implements an anchor-free model wherein the center of an object is directly pre- dicted without the use of an offset from an anchor box. Furthermore, the YOLOv8 shows notable improvement in accuracy in comparison to the older models when run on the COCO dataset[24].

Fig. 2 is taken from the output video obtained after object detection using the YOLOv8s model. All the identified objects are tagged with their class and confidence score and the model can be fine- tuned to detect objects only above a threshold confidence measure. For our experiments, a high confidence score of 0.60 was set and inferences were recorded.

2.3 Object Tracking

Multiple Object Tracking is an operation in the field of Computer Vision that encompasses mo- tion tracking of multiple objects over the course of time in a video. Due to its practical applications in our daily routines such as vehicle monitoring[25][26] and tracking of sports players on the field [27][28][29], it has gained a greater appeal over the course of time[30].

BYTETrack is a universal and relatively simple framework used to track objects in a video sequence[31]. The primary notion of this framework is to retain the low-score non-background boxes for an additional secondary association step amid the preceding frame and the succeeding frame[31].

Fig. 3 shows a frame from the output video stream after BYTEtracking the objects identified by the YOLOv8s model. Each of the identified players is tagged with a player tracker-id which will further be used for ball possession estimation.

2.4 Ball Possession

Ball Possession forms one of the most important and major factors in football analytics and has a substantial effect on the final outcome[32]. Following the Object detection and tracking stage, a simple algorithm is used to estimate ball possession by factoring the proximity of players to the ball. The algorithm establishes a threshold of 40 pixels and proceeds to check each bounding box belonging to the class ‘Player’ to investigate whether any player is within the threshold distance from the football’s bounding box. If any player is found to fit this criteria, then that player is assumed to have possession of the ball. A dictionary is created to record individual ball possession. For each subsequent frame that the ball remains in possession of the player, the respective key is incremented and is later converted to real-time (minutes and seconds) using the frames-per-second rate. Aggre- gating the ball possession of each of the team members, we will be able to calculate the team’s ball possession. Similar to the individual ball possession discussed above, the value calculated in the form of a number of frames can be converted into seconds

3 Results and Conclusions

In this section, results drawn from the aforementioned experiments will be discussed to draw useful and relevant conclusions from the same. We will present details about the results of the pre-processing stage as well as the actual experimentation. Finally, we will discuss our overall findings.

3.1 Evaluation Metrics

This section discusses the various evaluation metrics used to measure and compare the performance of the two YOLO models in providing precise class recognition. The two YOLO models used are YOLOv5 and YOLOv8. The Evaluation metrics used are Precision, Recall, Mean Average Precision (mAP), F1-score. Their respective mathematical expressions have been discussed below.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \quad (3)$$

$$F1\text{-}Score = 2 \cdot \frac{Precison \cdot Recall}{Precision + Recall} \quad (4)$$

3.2 Results

Both models were trained for a total of 100 epochs on a dataset of 600 images extracted from the football shootout clips. The performance of each individual model with regard to our dataset and problem statement has been discussed below

Table 1: Performance Metrics

Model	Precision	Recall	mAP	F1-score
YOLOv5	81.02%	**77.71**%	77.54%	79.1%
YOLOv8	**94.96**%	75.95%	**85.61**%	**84.2**%

Confusion Matrix is a measurement of performance used in classification models where correct and incorrect predictions pertaining to each class are displayed in a matrix format. From Fig. 4 and Fig. 5, we can infer that both models perform well in predicting objects belonging to class Player. Both models have a hard time identifying the ball correctly and instead identify it as belonging to the class Background.

This can be ascribed to the fact that when the ball moves at a fast pace it becomes difficult for the models to distinguish it from the background and hence mistakes the object of the class Ball to belong to the class background.

Precision forms an essential evaluation metric for this study as it calculates the positive prediction quality of a model. This means that it evaluates whether an object belonging to a certain class does indeed belong to the class or not. The mathematical expression of precision has been shown in Eq. 1 where TP represents True Positives and FP represents False Positives. Precision plays an important role in sports object tracking as it is absolutely vital that an object, suppose a player, is not categorized as something that it isn't, for example, a ball.

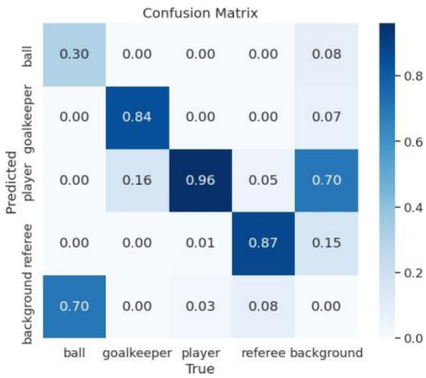


Fig. 4. YOLOv5 Confusion Matrix

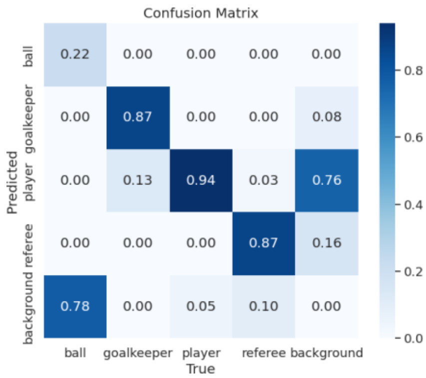
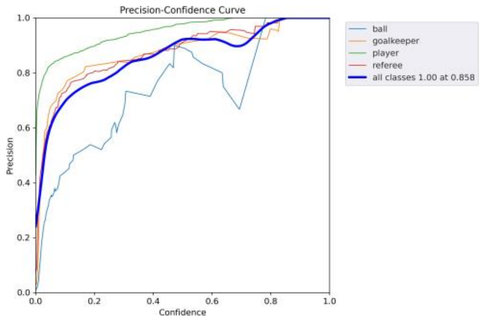


Fig. 5. YOLOv8 Confusion Matrix



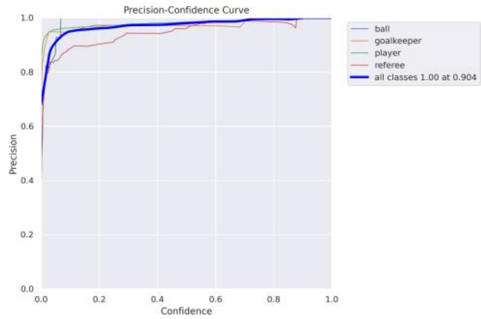


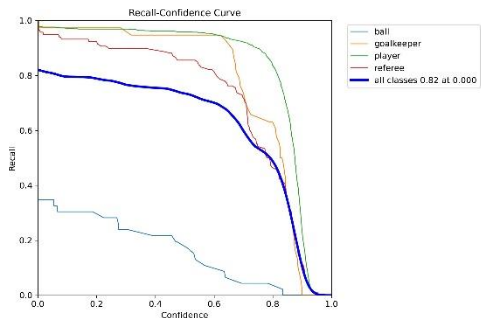
Fig. 6. YOLOv5 Precision Confidence Curve

Fig. 7. YOLOv8 Precision Confidence Curve

As evident from Fig. 6 and Fig. 7 and the results tabulated in Table 1, the precision of the YOLOv8 model is found to be significantly higher than the YOLOv5 model.

Recall calculates the proportion of the true positive count to the total number of objects that actually belong to that class. The mathematical description of Recall has been shown in Eq. 2 where TP represents True Positives and FN represents False Negatives. As shown in Fig. 8 and Fig. 9 the recall of YOLOv5 and YOLOv8 are almost similar. Evaluation metrics tabulated in Table 1 reveal that YOLOv5 has slightly better recall when compared to YOLOv8. This can be attributed to the Precision-Recall trade-off wherein the improvement of one metric usually results in the detriment of the other metric.

The F1-score is a machine learning performance measure that combines both recall and precision metrics as shown in Eq. 4. F1-Score is used as an optimal metric that achieves a balance be- tween precision and recall hence providing a comprehensive measure of the model's performance.



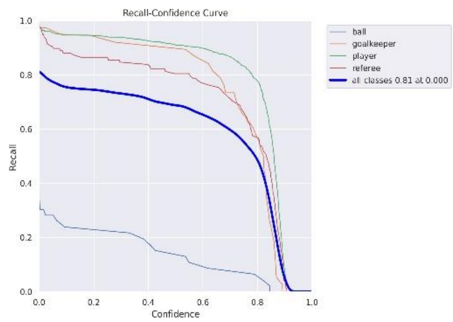


Fig. 8. YOLOv5 Recall Confidence Curve

Fig. 9. YOLOv8 Recall Confidence Curve

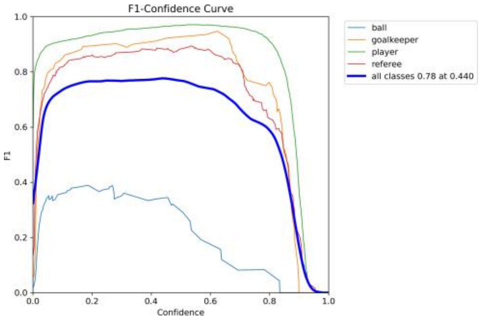


Fig. 10. YOLOv5 F1-Confidence Curve

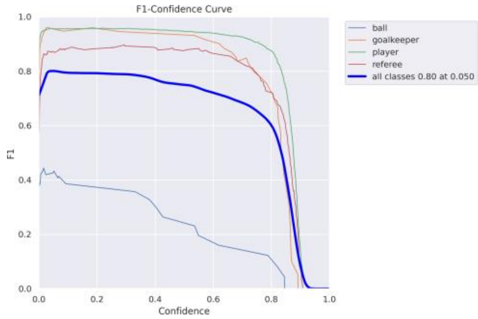


Fig. 11. YOLOv8 F1-Confidence Curve

As displayed in Table 1 and evident from Fig. 10 and Fig. 11, the YOLOv8 model has a better F1-Score than the YOLOv5 model.

Mean Average Precision (mAP) is useful evaluation metric based on Intersection over Union (IoU) which is used to measure the overlapping of the predicted boundary with the ground truth as displayed in Eq. 3. It takes into consideration the trade-off that occurs between recall and precision which makes it an ideal metric for object detection applications. As Table 1 indicates the mAP value for the YOLOv8 model is significantly better in comparison to the YOLOv5 model which indicates the better performance of the former model.

3.3 Object Tracking and Ball Possession

In this section, we discuss the object tracking methodology and the subsequent ball posses- sion data derived from it. The BYTETrack framework has been employed for object tracking. Its superior performance over the state-of-art tracking methods has led to us choosing this as the ideal tracking framework[31]. BYTETrack works by retaining and separating low-score and high-score object detection boxes[31]. The unique algorithm of considering low-score object detection provides BYTETrack an edge over other existing tracking frameworks.

Using the object tracking provided by BYTETrack we propose a framework for ball possession already discussed in the Experiments Section. From the obtained ball possession, players with ball possession that lasted for just a single frame are eliminated as that is caused by occlusion in most cases. Using the discussed algorithm results in the following manner were obtained.

Table 2: Individual Ball Possession

Tracker ID	Team	Number of Frames	Ball Possession
1516	A	24	0.96 seconds
1525	A	17	0.68 seconds
1548	A	133	5.32 seconds
1515	B	13	0.52 seconds
1519	A	18	0.72 seconds
1529	A	4	0.16 seconds

1520	A	4	0.16 seconds
1601	B	5	0.20 seconds
1522	B	19	0.76 seconds
1521	B	27	1.08 seconds
1526	A	4	0.16 seconds
1518	B	13	0.52 seconds
1527	B	4	0.16 seconds

Results obtained in Table 2 can be used to obtain ball possession for each team as shown in 3.

Table 3: Team Ball Possession

Team	Ball Possession
A	8.16 seconds
B	3.24 seconds

In conclusion, this study was able to successfully detect and track relevant objects, and obtain relevant and useful information from the output received. A comparison of the two object detection models illustrated that the YOLOv8 model had better performance and hence was used with the BYTETrack tracking model. The discussed framework would work ideally for game recordings recorded from tactical cameras due to their static nature. This prevents players from exiting the video frame and hence does not affect the model's tracking. The data obtained from this can be used in the future to reveal any possible relationship between ball possession during key moments and their impact in the overall outcome of the game.

References

[1] Morgulev E, Azar OH, Lidor R. Sports analytics and the big-data era. International Journal 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 Á . 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] Lago-Pen˜as C, Dellal A. Ball possession strategies in elite soccer according to the evolu- tion of the match-score: the influence of situational variables. Journal of human kinetics. 2010;25(2010):93-100.
[8] Bradley PS, Lago-Pen˜as C, Rey E, Sampaio J. The influence of situational variables on ball possession in the English Premier League. Journal of Sports Sciences. 2014;32(20):1867-73.
[9] Link D, Hoernig M. Individual ball possession in soccer. PloS one. 2017;12(7):e0179953.
[10] Sarkka S, Solin A, Hartikainen J. Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering. IEEE Signal Processing Magazine. 2013;30(4):51-61.
[11] Dimopoulos GR. Implementing Multi-Class Object Detection in Soccer Matches Through YOLOv5.
[12] Ultralytics. YOLOv5: A state-of-the-art real-time object detection system; 2021. Accessed: 2023-09-03. https://docs.ultralytics.com.
[13] Cioppa A, Giancola S, Delie`ge A, Kang L, Zhou X, Cheng Z, et al. SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2022. p. 3491-502.
[14] 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.
[15] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995;20:273-97.
[16] Kathirvel P, Manikandan MS, Soman K. Automated referee whistle sound detection for ex- traction of highlights from sports video. International Journal of Computer Applications. 2011;12(11):16-21.
[17] Jakub Michalczyk MJMMMRHUD Maggie. DFL - Bundesliga Data Shootout. Kaggle; 2022. Available from: https://kaggle.com/competitions/dfl-bundesliga-data-shootout.
[18] Skalski P. Make Sense; 2019. https://github.com/SkalskiP/make-sense/.