Automated Detection and Classification of Soccer Field Objects using YOLOv7 and Computer Vision Techniques

Jafar AbuKhait, Murad Alaqtash, Ahmad Aljaafreh, Waleed Othman
Dept. of Computer and Communications Engineering
Tafila Technical University
Tafila, Jordan

Abstract—**In the last two decades, many technologies have been deployed and utilized in Soccer games (Football) as a result to the huge investment of Federation of International Football Association (FIFA). These technologies aim to monitor and track all soccer match objects including players and the ball itself in order to measure the player performance, and tracking the players' positions and movements at the field. Latest emerging artificial intelligence and computer vision techniques are being used recently in many systems and deployed in different scenarios. Identifying all field objects automatically has to be the first step in the monitoring process of soccer games. In this paper, we are proposing an automated system that has the ability to detect and track the ball and to detect and classify players and referees on the soccer field. The proposed system implements a detection model using a real-time object detection model YOLOv7 to detect the ball and all humans on the field after building a labeled dataset of 1300 different soccer game frames. It also deploys Improved Color Coherence Vector (ICCV) features to classify all humans on the field to five classes (Team1, Team2, Goalkeeper1, Goalkeeper2, and Referee) using K-Nearest Neighbor algorithm. The proposed system has achieved high accuracy in both the detection and classification modules.**

Keywords—Soccer game; football; YOLOv7; human detection and classification; ball detection; improved color coherence vector

I. Introduction

In the realm of sports, where every moment carries significance, the automated computer vision detection and recognition of soccer match field objects has emerged as a pivotal technological advancement [1]. Soccer, often referred to as football in many parts of the world, stands as one of the most globally celebrated and passionately played sports. Over the years, it has not only evolved in terms of gameplay but has also embraced technology to analyze and elevate the sport to new heights. The integration of automated computer vision has ushered in a transformative era, fundamentally reshaping how we perceive and comprehend soccer matches.

The soccer field itself serves as a dynamic canvas, where players, the ball, and various other elements such as goalposts, corner flags, and boundary lines converge to create a complex and fast-paced spectacle. Traditionally, the task of monitoring and dissecting these elements fell to human operators, a process fraught with potential errors and subjectivity. However, with the advent of automated computer vision, we have witnessed a profound shift in our ability to capture, process, and leverage data from soccer matches [2]. This technology empowers us to identify and track field objects in real-time, providing invaluable insights to coaches, players, analysts, and fervent fans.

In this context, this paper delves into the profound significance of automated computer vision in the detection and recognition of soccer match field objects. We explore the tangible applications of this technology, its impact on game analysis, player performance evaluation, automatic offside detection, and fan engagement. Furthermore, we delve into how it is poised to redefine the future of soccer as we know it. Through this exploration, it becomes increasingly evident that automated computer vision is not just a tool but a transformative force that is redefining the very essence of soccer analysis and appreciation.

In the past two decades, the world of soccer (or football) has witnessed a profound transformation, fueled by substantial investments from organizations like the Federation of International Football Association (FIFA) [3]. These investments have ushered in a new era of technology-driven enhancements within soccer games, aimed at monitoring and tracking various aspects of the game, including player performance and positional data [4]. Recent advancements in artificial intelligence and computer vision techniques have played a pivotal role in this transformation.

The initial step towards automating the monitoring process of soccer matches involves the automatic detection and classification of all relevant objects on the field. In this context, several Convolutional Neural Network (CNN) architectures were suggested and deployed to detect the ball and the players on the soccer field [5, 6]. In addition, several research works have addressed the detection of soccer events, ball events, actions on the soccer game, and team tactics estimations [7, 8].

In general, several researchers have addressed the detection process of soccer field objects for various applications but, it is also important to classify these objects to ease the monitoring process of each soccer team player. In this context, the class of each human on the soccer field should be determined to enable tracking of individual players and team movements.

In this paper, we present an automated system designed to detect and track the soccer ball and classify players and referees on the field using state-of-the-art techniques. In this work, we aim to: 1) construct an annotated detection dataset that consists of 1300 soccer game matches' images; 2) detect soccer field objects, ball and humans using YOLOv7; 3) classify every human on the soccer field to Team1, Team2, Goalkeeper1, Goalkeeper2, and Referee using Color Coherence Vector and k-NN classifier; and 4) improve the detection precision and the classification accuracy by implementing a cascaded detection and classification systems. Overall, this system aims to provide an accurate and efficient method for detecting and classifying soccer game objects, which can be beneficial for coaches, broadcasters, and analysts in different computer vision applications such as team performance monitoring, automated offside detection and tactics estimation.

This paper is organized as follows. Section II provides a theoretical background of the techniques being used. Section III demonstrates the architecture of the proposed system. Section IV presents the experimental results and discussion. Finally, conclusions are drawn in Section V.

II. Theoretical Background

A. YOLO

You Only Look Once (YOLO) is convolutional neural network architecture for object detection in one shot [9, 10]. YOLO partitions the input image into N grids, each with equal dimensions. Each grid is responsible for the detection and localization of the object that it contains. In general, YOLO networks extract features through a backbone. The extracted features are combined and mixed in the neck, and then they are passed along to the head of the network to predict the locations and classes of objects around which bounding boxes should be drawn [11, 12].

The YOLOv7 algorithm, an upgraded version of YOLO object detectors, surpasses all known object detectors in both speed and accuracy [13]. YOLOv7 was trained only on MS COCO dataset from scratch without using any other datasets or pre-trained weights. It improved real time object detection accuracy without increasing the inference cost.

YOLOv7 has extended efficient layer aggregation networks (E-ELAN). It also has model scaling for concatenation-based models. The YOLOv7 algorithm also uses a technique called anchor boxes to improve the accuracy of object detection. Anchor boxes are pre-defined shapes that the algorithm uses to predict the location of objects in an image. By using anchor boxes, the algorithm is able to detect objects of different sizes and shapes with greater accuracy.

B. Improved Color Coherence Vector

Color Coherence Vector (CCV) is a color feature extractor that encodes information about color spatial distribution. It classifies each pixel in the image as either coherent or incoherent. Coherent pixels belong to a big connected component (CC) while incoherent pixels belong to a small connected component. CCV aims to build a low dimensional representation of the image through the following steps [14]:

1) Blur the image by averaging.

2) Quantize the image colors into n distinct colors.

3) Classify each pixel either as coherent or incoherent by:

a) Finding the connected components for each quantized color.

b) Determining the Tau's value which is typically about 1% of image size. Pixels are considered coherent if they belong to any connected component with number of pixels are more than or equal to tau.

4) For each color compute two values: α which is the number of coherent pixels, and β which is the number of incoherent pixels.

Improved Color Coherence Vector (ICCV) has more spatial information with respect to CCV and thus; it is more efficient in comparing image contents [15]. In addition to (α, β) pairs that were computed by CCV, the mean of position coordinates (rows and columns) for the maximum connected region in the coherent pixels (γ) is computed using ICCV. Each quantized color would be described accordingly by four values in the form of (α, β, γ_x, γ_y). The size of feature vector would be the number of quantized colors multiplied by 4.

C. K-Nearest Neighbor Algorithm

The k-nearest neighbor algorithm (k-NN) is one of the simplest machine learning algorithms. Any test object is assigned to its nearest neighbors' class by adopting majority voting. Nearest neighbors are determined by measuring a distance metric which could be the Euclidean distance, the hamming distance, or the correlation. Number of nearest neighbors is defined by K which is a problem dependent parameter [16, 17]. k-NN has two stages; determining the nearest k neighbors and determining the class by majority voting.

III. The Proposed System

The proposed system shown in Fig. 1 is composed of three modules:

- Dataset Preparation: in which soccer game field images of on-going games are collected, pre-processed, annotated, and augmented to build the dataset for training, evaluating, and testing the YOLO detection model.
- Human and Ball Detection: in which a YOLOv7 deep learning model is selected and trained on both the training and validation datasets and then, evaluated on the test dataset. The detection model has the capability of detecting the ball and all humans on the soccer field
- Humans' Classification: in which k-NN classifier is being used to classify detected humans to team1, team2, goal keeper, and referee using Color features of detected human. Human templates of team1, team2, goal keeper, and referee are used in the classification based on Color Coherence Vector (CCV).

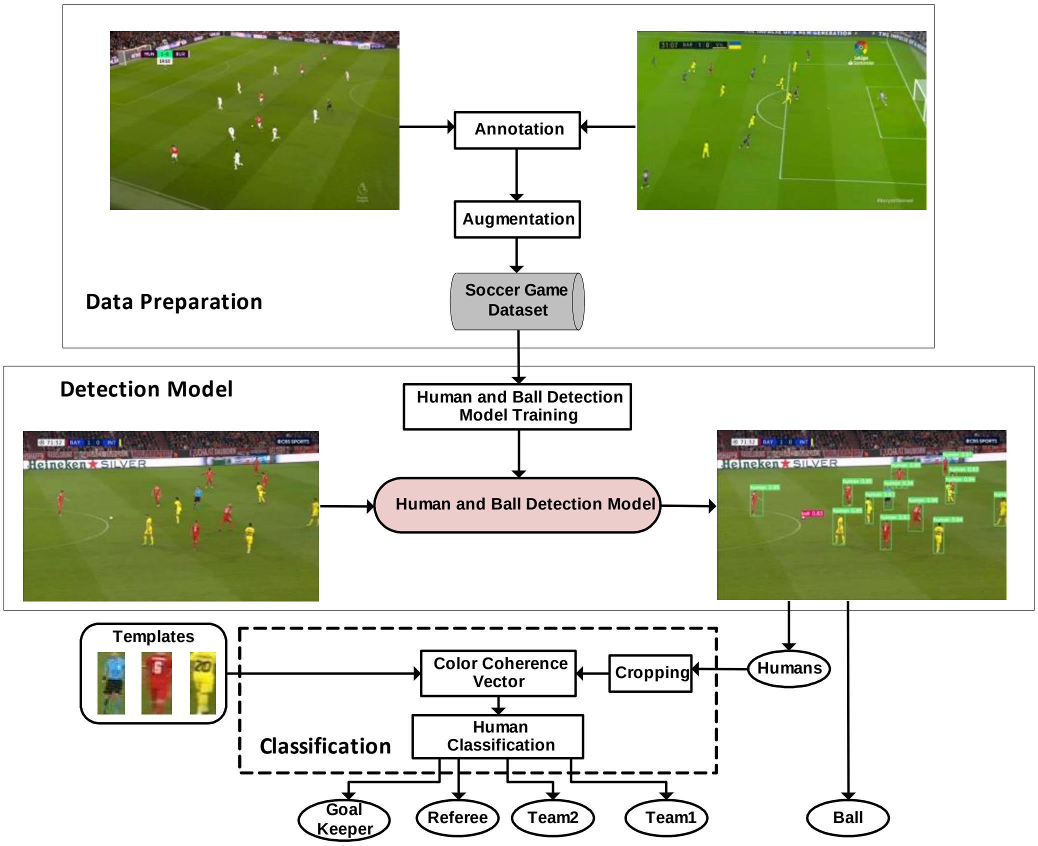


Fig. 1. The proposed system.

A. *Dataset Preparation*

In this module, a labeled dataset of 1300 different soccer match images was constructed. This dataset will be used for training, validation, and testing the Human and Ball detection model.

1) Image collection: A set of 1300 different images of different soccer matches was constructed. 810 images were obtained from multiple videos of soccer game matches and 490 images were obtained from the online dataset at [18]. All images were captured from a single camera position covering one half of the field. Several images were selected for each soccer game. Fig. 2 shows samples of soccer matches' images.

2) Annotation, preprocessing and augmentation: In this step, the collected images were uploaded to Roboflow platform, where it was labeled into two classes: human and ball using Roboflow's annotation tool. Fig. 3 shows some annotations of ball and humans on the original images.

The annotation task of the original 1300 images has resulted into 25636 different labels of both ball and human labels as shown in Table I.

TABLE I. SUMMARY OF ANNOTATED IMAGES

Number of Images	Annotations	Ball Annotation	Human Annotation
1300	25636	1174	24462

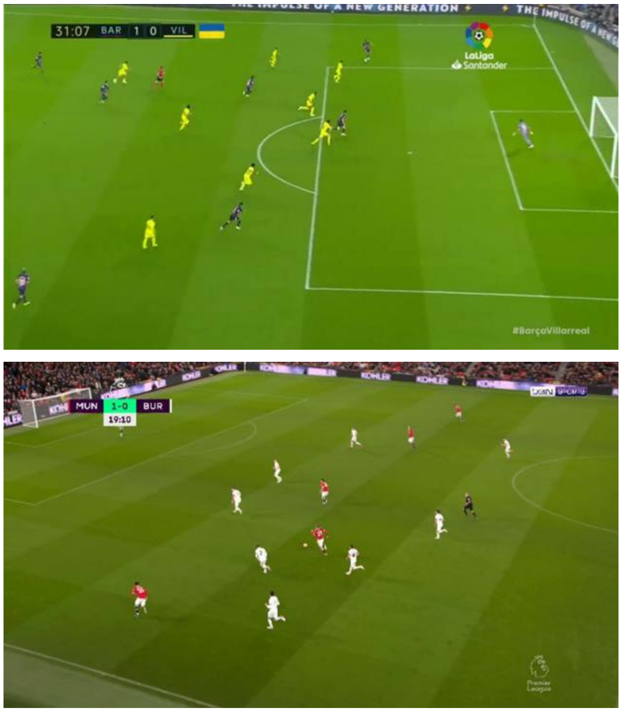


Fig. 2. Sample images of soccer game matches.

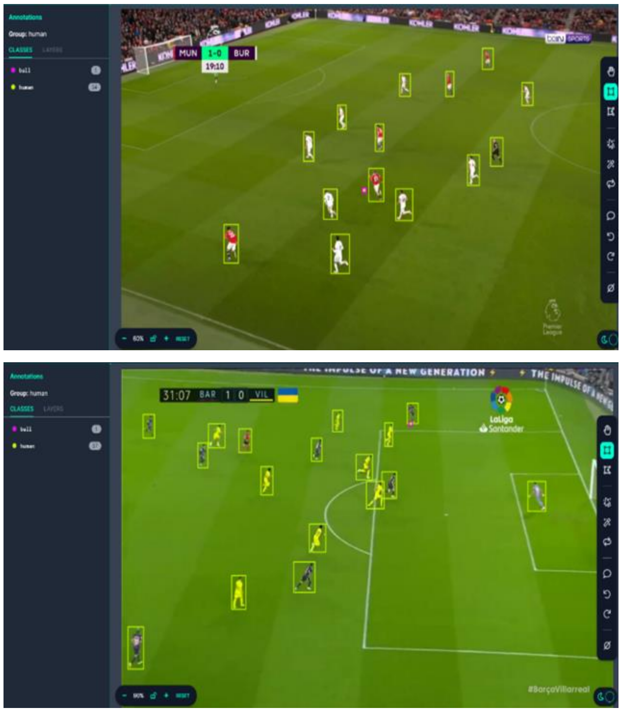


Fig. 3. Sample images of soccer game matches

After labelling, some preprocessing and augmentation tasks of the images were applied as follows:

- Preprocessing:
  - Auto Orient: Applied
  - Resize: Fit within 640x640
- Augmentations:
  - Grayscale: Apply to 25% of images
  - Saturation: Between -25% and +25%
  - Brightness: Between -30% and +30%

After Augmentations, a set of 2203 images were obtained. These images were splitted into 83% (1826), 12% (273), and 5% (104) for training, validation, and testing, respectively.

B. Human and Ball Detection Model

In this module, YOLOv7 algorithm is used for human and object detection. Model training was achieved on Google Colab notebook. Once the model was trained, it was tested on a separate set of validation images to evaluate its performance.

1) Model training and evaluation: The model was trained using YOLOv7. This model is evaluated for detecting two classes: human and ball. Two epoch choices were used to train the model; 100 and 180.

The model detection performance was evaluated using mean average precision (mAP), recall and precision. The evaluation metrics that were used to evaluate the model are explained as follows:

Precision is a measure of a network's ability to accurately identify targets at a single threshold, calculated by:

$$\text{Precision} = \frac{\text{Tp}}{\text{Tp+Fp}} \qquad (1)$$

Recall is a measure of the network's ability to detect its target, calculated by:

$$\text{Recall} = \frac{\text{Tp}}{\text{Tp+Fn}} \qquad (2)$$

Where:

- Tp: are the Bounding Boxes (BB) that the intersection over union (IoU) with the ground truth (GT) is above 0.5.
- Fp: two cases (a) BB that the IoU with GT is below 0.5 (b) the BB that have IoU with a GT that has already been detected.
- Tn: there are not true negative, the image is expected to contain at least one object.

Fn: images containing an object were the method failed to produce a BB.

Intersection over Union (IoU) is a method used to compare two arbitrary shapes, i.e., object widths, heights, and location of two boxes into the original region. This will evaluate the precision of the object detector on particular dataset [19] as in (3). Fig. 4 shows how IoU is calculated diagrammatically.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad (3)$$

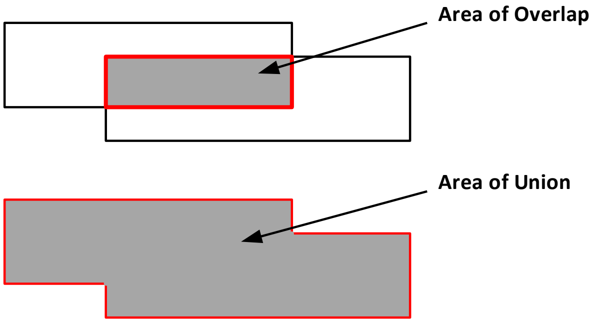


Fig. 4. Diagrammatic example intersection over union (IoU) calculation.

Average precision is a method combining recall and precision for the entire ranking. It is the average of precision in a single ranking [20].

$$AP = \frac{1}{|Class|} \sum_{c \in class} \frac{TP(c)}{TP(c)+FP(c)} \qquad (4)$$

Mean average precision (mAP) is the average of precision values at the rank where there is a relevant document [21]. It is calculated from precision, recall and interception over union IOU.

$$mAP = \frac{AP}{Total\ number\ of\ class} \qquad (5)$$

C. Humans' Classification

This module classifies the detected humans in the previous module to team1, team2, goal keeper, and referee using shirt colors of detected human. Color Coherence Vector (CCV) is used to describe the color features and k-NN is deployed for classification. The classification has been achieved by comparing each detected human with stored templates of humans from the same soccer match.

1) Template preparation and preprocessing: Template Preparation and Preprocessing have been done prior to features extraction and classification. At first, manual cropping of humans on the soccer match images was achieved for image group of each distinct soccer match. Five suitable templates of team1, team2, goalkeeper1, goalkeeper2 and referee were selected. Each five templates were obtained from different image frames of the same soccer match to choose the best body orientation of each class. Each single template represents the body area of the human excluding the head and lower part of leg to concentrate on distinct color features for classification. All templates were resized to 25x50. Fig. 5 shows the five templates of humans selected from different images of the same soccer match.

Next, each detected human in the detection module is cropped automatically to exclude most of the soccer field background and body parts that haven't distinct color features. This process has been done by cropping 15% from each side of the detected humans' area as shown in Fig. 6.

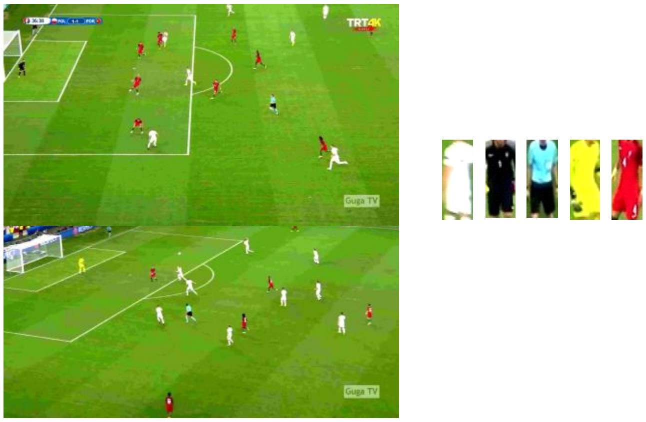


Fig. 5. Example of template preparation.

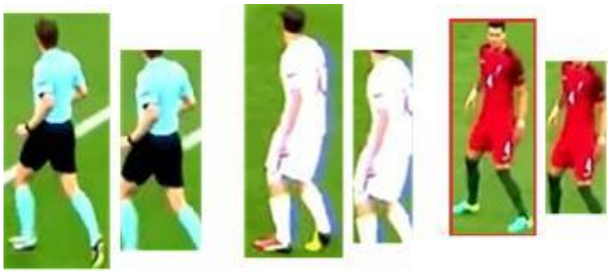


Fig. 6. Examples of cropping detected humans' images.

2) Color feature extraction using ICCV: Color features of human templates and all cropped images of all detected humans are extracted using Improved Color Coherence Vector (ICCV). The output of this stage is a feature vector of size 64x1 based on the following steps:

- Quantize the color-space into 16 distinct colors.
- Classify each pixel either as coherent or incoherent by finding the connected components for each quantized color and determining the tau's value to be 5% of image's size. Any connected component with number of pixels more than or equal to tau then its pixels are considered coherent otherwise they are incoherent.
- For each color, compute the number of coherent pixels (α), the number of incoherent pixels (β), and the mean of position coordinates (rows and columns) for the maximum connected region in the coherent pixels (γ).

Each quantized color would be described accordingly by four values in the form of (α, β, γx, γy). For the 16 colors, we would gain a feature vector of size 64x1.

3) Classification using k-NN: Each detected human would be classified to team1, team2, goalkeeper1, goalkeeper2 or referee based on 3-NN classifier. Nearest neighbors are determined by measuring the Euclidean distance and determining the class by majority voting.

Classification accuracy can be calculated according to the following formula:

$$\text{Accuracy} = \frac{\text{Tp+Tn}}{\text{All Elements}} \tag{6}$$

Where Tp and Tn are the elements correctly classified by the model.

IV. Results and Discussion

In this section, we demonstrate the experimental results of both the detection and classification modules. We have tested the detection model on 104 images that were picked randomly from various soccer matches. In the classification module, we have achieved the testing on five different matches because we use color features which are different on each single match. A total number of 51 different images have been selected to validate the classification results.

A. Detection Results

In this section, we evaluate the performance of the YOLOv7 detection model, which plays a crucial role in automatically identifying and tracking soccer field objects such as the ball and players. The results provide insights into the model's accuracy and effectiveness in object detection.

Fig. 7 serves as a comprehensive illustration of the YOLOv7 model's performance in identifying and localizing objects of interest, including both players and the ball. This figure utilizes three distinct metrics—Accuracy, Precision, and mAP@0.5—to evaluate the model's precision in object localization and its proficiency in correctly classifying objects.

Fig. 8 presents a valuable snapshot of the quantitative metrics used to evaluate the performance of the detection model throughout the training process. These metrics include precision, recall, and mean average precision (mAP@0.5), which provide insights into the model's effectiveness in detecting objects of interest in soccer matches.

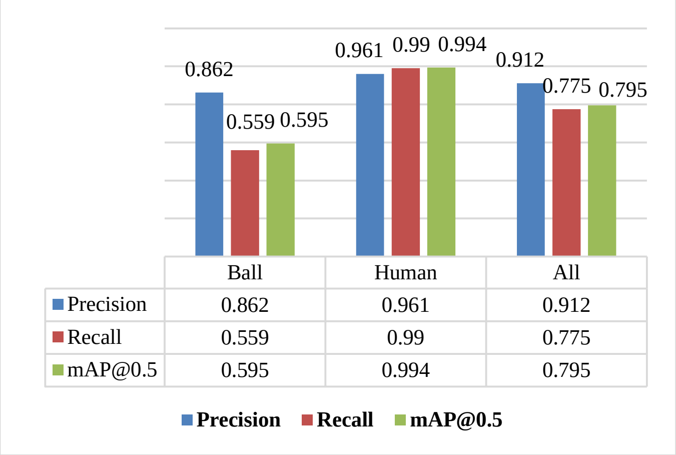
Precision measures the accuracy of positive predictions made by the model. It is a crucial metric for object detection, as it assesses the model's ability to correctly identify objects without generating too many false positives. Recall evaluates the model's ability to detect all relevant objects in the dataset, minimizing false negatives. It is particularly important in ensuring that no objects of interest are missed. mAP@0.5 is a comprehensive metric that combines precision and recall across different object classes. It considers precision-recall trade-offs and provides an aggregate assessment of the model's performance.

The confusion matrix, as shown in Fig. 9, provides a detailed breakdown of the model's performance, including true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for each object class (e.g., ball, humans and background).

Fig. 10 offers an insightful comparison between different models' training using varying numbers of training epochs (specifically, 100 and 180). The key takeaway from this figure is that it underscores the importance of training the model beyond 100 epochs for achieving optimal performance.

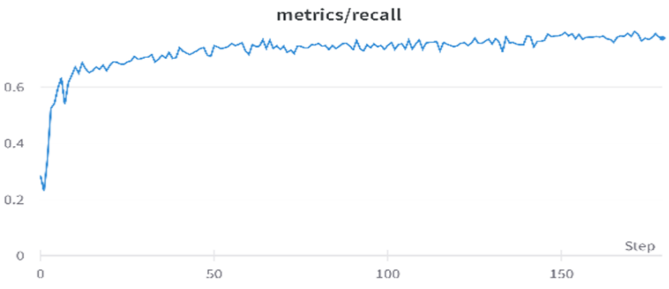
To perform model testing, we loaded the saved weights of the trained model and passed the test dataset through it. Fig. 11 shows some examples of detection model Inference.

The results of the model testing showed very good accuracy and performance in detecting objects of interest in the images. Overall, the model testing was successful in demonstrating the accuracy and effectiveness of the trained model in detecting objects of interest in the images.

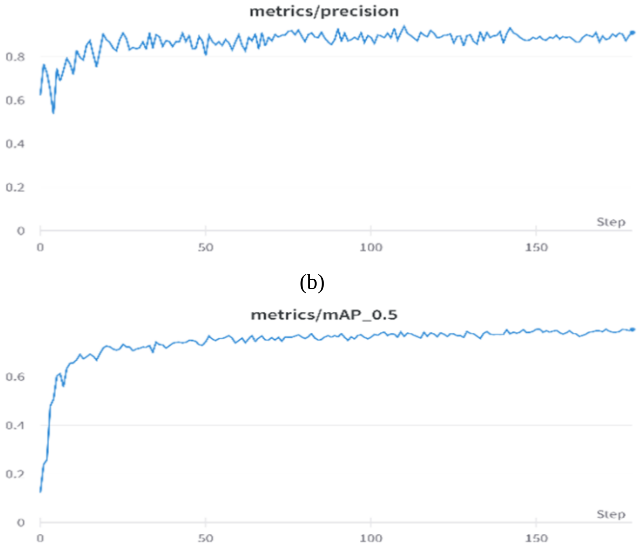


	Ball	Human	All
■ Precision	0.862	0.961	0.912
■ Recall	0.559	0.99	0.775
■ mAP@0.5	0.595	0.994	0.795

Fig. 7. YOLOv7 model's performance.



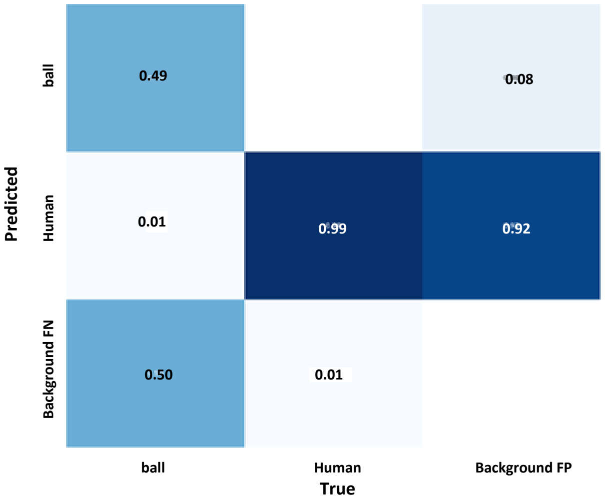
(a)



(b)

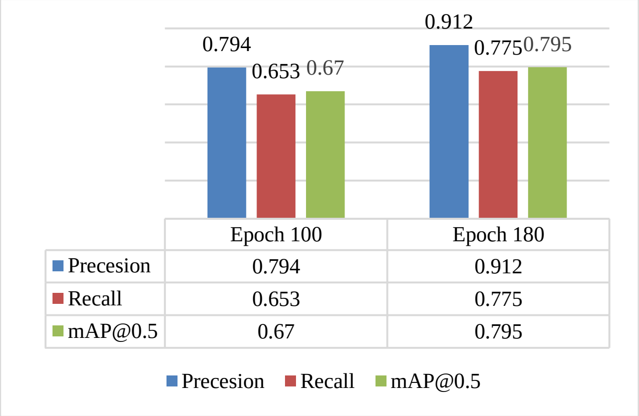
(c)

Fig. 8. Performance of the detection model throughout the training process: a) recall; b) precision, c) mAP@0.5.



	ball	Human	Background FP
ball	0.49		0.08
Human	0.01	0.99	0.92
Background FN	0.50	0.01	

Fig. 9. Confusion Matrix of the detection model.



	Epoch 100	Epoch 180
■ Precesion	0.794	0.912
■ Recall	0.653	0.775
■ mAP@0.5	0.67	0.795

Fig. 10. Comparison between two different models using different epochs.

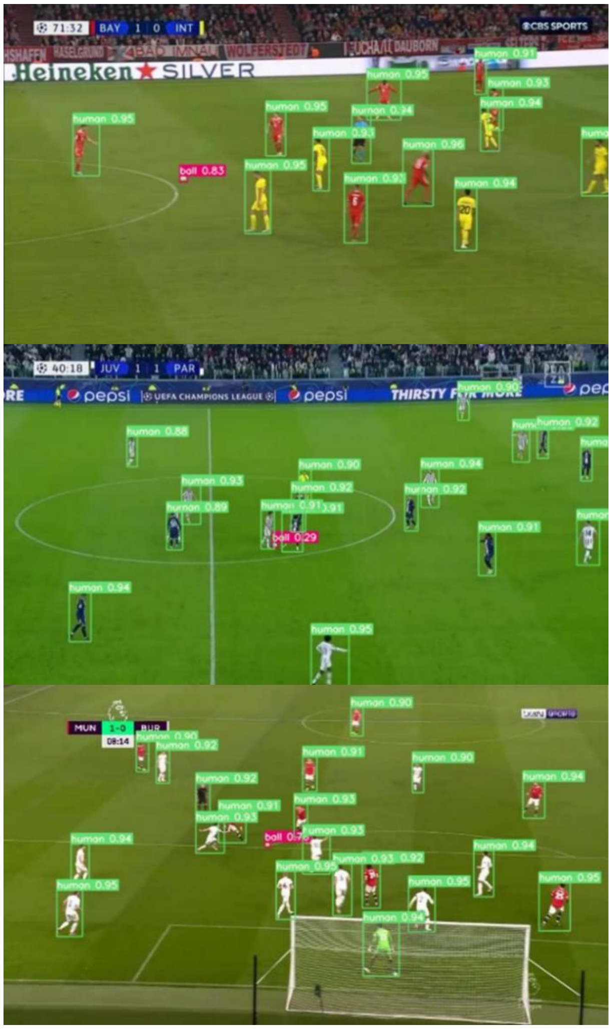


Fig. 11. Model Inference examples

B. Classification Results

The proposed human classification module has been tested on 5 different soccer matches to validate its performance for different color variances among humans' shirts on each single match. Fig. 12 shows cropped human images for each one of the five matches. These images are used in the discussions of the classification results. For each match, 11 to 14 frames have been selected to obtain 2 human templates for each one of the five human categories (team1, team2, goalkeeper1, goalkeeper2, and referee) and the rest detected humans are used to test the classification accuracy. A total number of 827 humans is existed among all selected soccer matches' frames. Table II shows the exact division of these images cross the five matches. Table III shows the division of the five tested human classes cross the five matches.

Fig. 13(a) to (e) shows the confusion matrices of the five matches. The classification accuracy for each human classes cross the five soccer matches is presented in Table IV.

In Match 1, the classification results exhibit strong accuracy with 94.8%. The model fails to predict correctly some players from Team2. This happens because of the similarity between the shirt color (green) and the background. The prediction of the referee fails in two cases because of the similarity between the Referee colors (black) and both Team1 and Goalkeeper2 colors.

Match 2 demonstrates consistent performance in classification, with an accuracy of 95.5%. The model succeeds in all cases where color variance between the five classes is high. It misclassifies some instances in Team2 and Goalkeeper2 because of the existence of black and dark colors in these different classes.

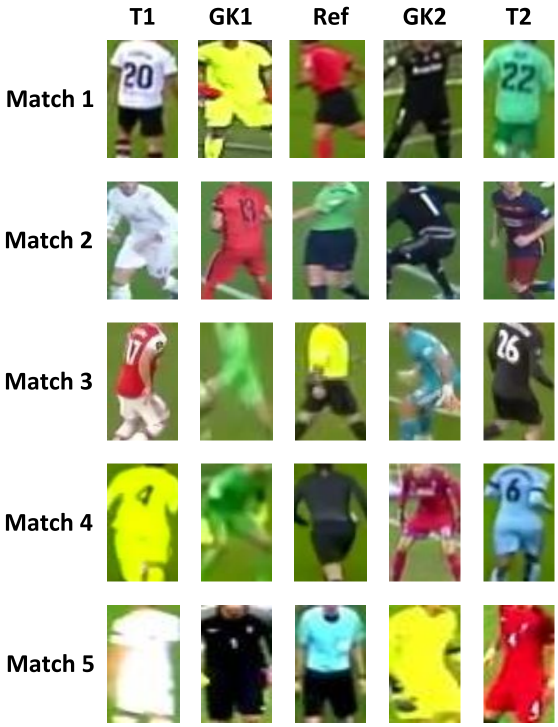


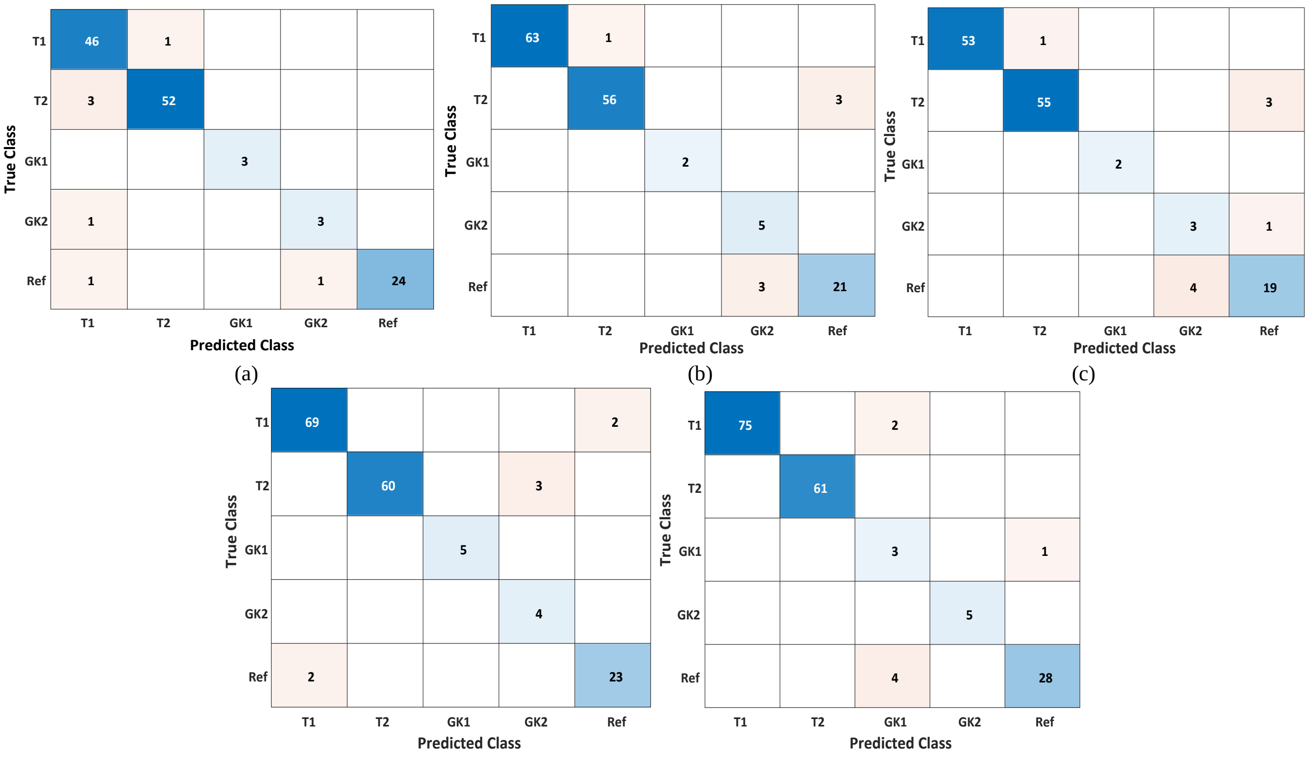
Fig. 12. Cropped images of all human classes in each Match.

TABLE II. HUMAN DATASET IMAGES' DIVISION CROSS THE FIVE SOCCER MATCHES

Soccer Match	No. of image frames	No. of human templates	No. of tested humans	No. of all existed humans
Match 1	11	10	135	145
Match 2	12	10	154	164
Match 3	11	10	141	151
Match 4	13	10	168	178
Match 5	14	10	179	189
	61	**50**	**777**	**827**

TABLE III. THE NUMBERS OF TESTED IMAGES FOR EACH HUMAN CLASS CROSS THE FIVE SOCCER MATCHES

	humans	Team1 (T1)	Team2 (T2)	Goalkeeper1 (GK1)	Goalkeeper2 (GK2)	Referee (Ref)
Match 1	135	47	55	3	4	26
Match 2	154	64	59	2	5	24
Match 3	141	54	58	2	4	23
Match 4	168	71	63	5	4	25
Match 5	179	77	61	4	5	32



	T1	T2	GK1	GK2	Ref
T1	46	1			
T2	3	52			
GK1			3		
GK2	1			3	
Ref	1			1	24

(a)

	T1	T2	GK1	GK2	Ref
T1	63	1			
T2		56			3
GK1			2		
GK2				5	
Ref				3	21

(b)

	T1	T2	GK1	GK2	Ref
T1	53	1			
T2		55			3
GK1			2		
GK2				3	1
Ref				4	19

(c)

	T1	T2	GK1	GK2	Ref
T1	69				2
T2		60		3	
GK1			5		
GK2				4	
Ref	2				23

(d)

	T1	T2	GK1	GK2	Ref
T1	75		2		
T2		61			
GK1			3		1
GK2				5	
Ref			4		28

(e)

Fig. 13. Confusion matrices of the classification module for: a) Match 1, b) Match 2, c) Match 3, d) Match 4, e) Match 5.

TABLE IV. THE CLASSIFICATION ACCURACY FOR EACH HUMAN CLASSES CROSS THE FIVE SOCCER MATCHES

Soccer Match	Team1 (T1)	Team2 (T2)	Goalkeeper1 (GK1)	Goalkeeper2 (GK2)	Referee (Ref)	**Overall Accuracy**
Match 1	97.9%	94.5%	100%	75.0%	92.3%	**94.8%**
Match 2	98.4%	94.9%	100%	100%	87.5%	**95.5%**
Match 3	98.1%	94.8%	100%	75.0%	82.6%	**93.6%**
Match 4	97.2%	95.2%	100%	100%	92.0%	**95.8%**
Match 5	97.4%	100%	75.0%	100%	87.5%	**96.1%**
	97.8%	**95.9%**	**95.0%**	**90.0%**	**88.4%**	

In Match 3, classification accuracy is slightly lower at 93.6%, but the model still maintains robust performance. Team2 and Referee classes have some misclassifications based on the color similarity. Referee classification is challenging.

Match 4 demonstrates outstanding classification accuracy at 95.8%. Goalkeeper1 and Goalkeeper2 classes perform exceptionally well while Team1 and Team2 show some misclassification due to the background of the cropped images.

In the final match, the model maintains high classification accuracy at 95.4%. The model has some misclassification results especially between Referee and Goalkeeper1 because of the similarity of colors between these two classes.

In general, the classification module performs exceptionally and has excellent classification accuracy with an overall accuracy of 95.2%. The model succeeds to classify 740 different instances from a total of 777 cross the five matches. For some instances, misclassification arises because of color similarities between shirt colors or the effect of background color (soccer field color). Low color variance may lower the classification performance since we use color feature descriptors (ICCV) in this module.

V. CONCLUSIONS

In this paper, we have proposed an automated system that has the ability to detect the soccer ball and classify players and referees on the soccer field using computer vision techniques. The proposed system implements a detection model using YOLOv7 to detect the ball and all humans on the field after building a labeled dataset of 1300 different soccer game frames. It also deploys Improved Color Coherence Vector (ICCV) features to classify all humans on the field to five classes (Team1, Team2, Goalkeeper1, Goalkeeper2, and Referee) using K-Nearest Neighbor algorithm. The proposed system has achieved high efficiency in both the detection and classification modules.

The proposed system can be considered the first phase of any computer vision application in soccer game matches. It can be deployed on game analysis, player performance evaluation, automatic offside detection and fan engagement. Furthermore, we delve into how it is poised to redefine the future of soccer as we know it. Through this exploration, it becomes increasingly evident that automated computer vision is not just a tool but a transformative force that is redefining the very essence of soccer analysis and appreciation.

REFERENCES

[1] B.T. Naik, M.F. Hashmi, and N.D. Bokde. "A comprehensive review of computer vision in sports: Open issues, future trends and research directions." *Applied Sciences* 12, no. 9 (2022): 4429.

[2] S. Kusmakar, S. Shelyag, Y. Zhu, D. Dwyer, P. Gastin, and M. Angelova. "Machine learning enabled team performance analysis in the dynamical environment of soccer." *IEEE access* 8 (2020): 90266-90279.

[3] FIFA Research Programme, available at https://www.fifa.com/technical/football-technology/research.

[4] P.R. Kamble, A.G. Keskar, K.M. Bhurchandi. "A deep learning ball tracking system in soccer videos." *Opto-Electronics Review* 27, no. 1 (2019): 58-69.

[5] B.T. Naik and M.F. Hashmi. "YOLOv3-SORT: detection and tracking player/ball in soccer sport." *Journal of Electronic Imaging* 32, no. 1 (2023): 011003-011003.

[6] B.T. Naik, M.F. Hashmi, ZW. Geem, ND. Bokde. "DeepPlayer-Track: player and referee tracking with jersey color recognition in soccer." *IEEE Access* 10 (2022): 32494-32509.

[7] G. Suzuki, S. Takahashi, T. Ogawa, M. Haseyama. "Team tactics estimation in soccer videos via deep extreme learning machine based on players formation." In *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, pp. 116-117. IEEE, 2018.

[8] Y. Ganesh, A. Sri Teja, SK. Munnangi, G. Rama Murthy. "A novel framework for fine grained action recognition in soccer." In Advances in Computational Intelligence: 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12-14, 2019, Proceedings, Part II 15, pp. 137-150. Springer International Publishing, 2019.

[9] J. Redmon and A. Farhadi. "YOLO9000: better, faster, stronger." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263-7271. 2017.

[10] J. Du. "Understanding of object detection based on CNN family and YOLO." In *Journal of Physics: Conference Series*, vol. 1004, p. 012029. IOP Publishing, 2018.

[11] J. Nelson, and J.Solawetz,"YOLOv5 is here: State-of-the-art object detection at 140 FPS" (2022). Available at: https://blog.roboflow.com/yolov5-is-here/.

[12] C.Y. Wang, I.H. Yeh, and HYM. Liao. "You only learn one representation: Unified network for multiple tasks." *arXiv preprint* arXiv:2105.04206 (2021).

[13] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464-7475. 2023.

[14] G. Pass, R. Zabih, and J. Miller. "Comparing images using color coherence vectors." In *Proceedings of the fourth ACM international conference on Multimedia*, pp. 65-73. 1997.

[15] X. Chen, X. Gu, and H. Xu. "An improved color coherence vector method for CBIR". *In Proceedings of the Graduate Students Symposium of Communication and Information Technology Conference*, Beijing. 2007.

[16] P. Cunningham and S. Delany. "k-Nearest neighbour classifiers-A Tutorial." *ACM computing surveys (CSUR)* 54, no. 6 (2021): 1-25.

[17] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Fourth Edition. Academic Press, 2008.

[18] N. Panse and A. Mahabaleshwarkar, "A Dataset & Methodology for Computer Vision Based Offside Detection in Soccer", *Association for Computing Machinery*, NY, USA, 2020, available at https://github.com/Neerajj9/Computer-Vision-based-Offside-Detection-in-Soccer.

[19] H. Rezatofighi, N. Tsoi, J.Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese. "Generalized intersection over union: A metric and a loss for bounding box regression." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 658-666. 2019.

[20] M. Everingham, L.V. Gool, C. Williams, J. Winn, and A. Zisserman. "The pascal visual object classes (voc) challenge." *International journal of computer vision* 88 (2010): 303-338.

[21] T.T. Nguyen, K. Vandevoorde, N. Wouters, E. Kayacan, J.G. De Baerdemaeker, and W. Saeys. "Detection of red and bicoloured apples on tree with an RGB-D camera." *Biosystems Engineering* 146 (2016): 33-44.