IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 26, 2022, accepted March 11, 2022, date of publication March 22, 2022, date of current version March 29, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3161441

DeepPlayer-Track: Player and Referee Tracking With Jersey Color Recognition in Soccer

BANOTH THULASYA NAIK[1], MOHAMMAD FARUKH HASHMI[1], (Senior Member, IEEE), ZONG WOO GEEM[2], (Senior Member, IEEE), AND NEERAJ DHANRAJ BOKDE[3]

[1]Department of Electronics and Communication Engineering, National Institute of Technology, Warangal, Warangal 506004, India
[2]College of IT Convergence, Gachon University, Seongnam 13120, South Korea
[3]Department of Civil and Architectural Engineering, Aarhus University, 8000 Aarhus, Denmark

Corresponding author: Zong Woo Geem (geem@gachon.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government through MSIT under Grant 2020R1A2C1A01011131.

ABSTRACT In real-world sports video analysis, identity switching caused by inter-object interactions is still a major difficulty for multi-player tracking. Due to similar appearances of players on the same squad, existing methodologies make it difficult to correlate detections and retain identities. In this paper, a novel approach (DeepPlayer-Track) is proposed to track the players and referees, by representing the deep features to retain the tracking identity. To provide identity-coherent trajectories, a sophisticated multi-player tracker is being developed further, encompassing deep features of player and referee identification. The proposed methodology consists of two parts: (i) the You Only Look Once (YOLOv4) can detect and classify players, soccer balls, referees, and background; (ii) Applying a modified deep feature association with a simple online real-time (SORT) tracking model which connects nodes from frame to frame using cosine distance and deep appearance descriptor to correlate the coefficient of the player identity (ID) which improved tracking performance by distinct identities. The proposed model achieved a tracking accuracy of 96% and 60% on MOTA and GMOTA metrics respectively with a detection speed of 23 frames per second (FPS).

INDEX TERMS Soccer, referee detection, player tracking, identity switch, JCD-SORT.

I. INTRODUCTION

Visual object tracking is a popular and well-explored research domain. In the face of various challenges, any object, its motion, and appearance have its level of tracking difficulties. With an increase in demand for video analysis of sports, multiple object tracking becomes necessary for executing advanced operations in sports. For instance, the team coach might use the movement and locations of players in the soccer game and the position of the ball with the aid of the automatic tracking system to strategically enhance the team's performance and precisely assess each player. It also enhances the viewing experience by providing information such as statistical data and activity of the player, to assist the viewers in better understanding the game they are watching. During fast-paced sports events, continuous, robust, precise athlete and ball tracking help in making critical decisions and detecting violations of game rules.

The associate editor coordinating the review of this manuscript and approving it for publication was Wai-Keung Fung.

Soccer is the world's most famous and well-liked sport [1]. Its popularity drew in a slew of researchers to work on player and football tracking for a wide range of applications. They include player performance and strategy analysis, highlight extraction, shot classification, incident detection, referee gesture recognition, material insertion, commentary assistance, and so on. All of these applications need real-time player and football position data, as well as their tracks as a reference. As a result, accurate athlete identification, football detection, and tracking of them are critical activities.

Even though the field of multiple object tracking has been gaining popularity among computer vision researchers, most approaches described in the literature are not specifically applicable to sports video analysis for the following reasons: The foremost is, the active movements of football players and the referee on the field differ significantly from those in actual videos of people walking along the street, vehicles driving down the lane, etc. In the soccer video, the motions of the ball and players produce nonlinear motion patterns, as a result of the football's location changing irregularly even between frames, the orientations and velocities of foreground objects

are erratic. Second, extreme occlusions often occur between players when they compete for possession of the football across the entire field. Finally, comparable shapes and the color of the players' jerseys make the tracking problem more difficult to solve.

This paper introduces a unique approach for multiple object recognition and tracking in soccer. The main concept of the method recommended in the paper is to detect all objects in the soccer field (players, soccer ball, referee, and background) to classify and exclude the background from the playfield. Since boundaries of the players and referee are similar to the persons (or) objects moving outside the playfield, the proposed methodology effectively eliminates objects detected outside the playfield as the background, which makes it robust to detect and classify the fast-moving objects in the playfield by removing irrelevant background features whereas traditional techniques still face significant challenges to identify moving object. Rather than using conventional methods for identifying tracking models, which have been extensively used in previous approaches, the proposed method Jersey color + Cosine Distance (JCD) based SORT tracking model has a good ability to effectively deal with object characteristics under complex scenes of the soccer game by preventing identity switching of the players even under severe occlusions. Experiments show that the proposed system can track multiple objects in a soccer video accurately in a variety of conditions. This paper's main contribution can be summarized as follows:

- Using the YOLOv4 object detection and classification model, foreground regions such as players, soccer balls, and referees in/out of the playfield are correctly identified and categorized in a given frame, and the background is successfully omitted. This technique is very useful for distinguishing between objects (players, referee, and ball) in the playfield and objects outside the playfield (peoples moving outside of the playfield), which is also a challenging task for conventional moving object tracking approaches.
- Based on convolutional neural network (CNN) features, Jersey color of the players/referee, and Cosine distance, a novel approach for tracking multiple objects is suggested, which is terribly beneficial in soccer videos. Moving objects in the playfield are clearly identified using the YOLOv4 object detection and classification method, which aids in effectively updating the tracking model of each object even in the presence of extreme occlusions and removes objects detected outside the playfield as background after classification. Furthermore, the JCD-SORT tracking model successfully prevents the players' identity switching even under severe occlusions, which efficiently improves the tracking performance.

The rest of the paper is organized as follows. Section 2 provides a review of the existing literature related to the soccer ball and player detection and tracking. Methods and materials used to achieve the desired results are discussed in Section 3. Section 4 presents the proposed methodology in detail. The experimental results and performance metrics related to the work are presented in Section 5. Finally, the conclusion and scope of future work are discussed in Section 6.

II. RELATED WORK

The topic of multi-object detection and tracking is quickly expanding due to its diverse possibilities. This section provides a concise summary of significant studies for soccer video analysis.

A. PLAYER AND BALL DETECTION

The identification of players and soccer balls is an essential phase in soccer video analysis. As a low-level image processing activity, player detection provides measurements that can be used to initiate the subsequent tracking step. Because of its direct impact on the player tracking step, player detection precision is extremely important. Following player detection, a bounding box can be used to approximate each player, and the left corner or center of the bounding box may be used to identify the player's position in the frame. Halbinger *et al.* [2] presented an approach for soccer ball detection in difficult situations such as severe occlusions. It can also handle low-resolution images from a single static camera system. Pallavi *et al.* [3] utilized circular Hough Transform and modified several versions to detect the soccer ball. Mazzeo *et al.* [4] used various feature extraction techniques to detect the soccer ball. Murthy and Hashmi [5] proposed a network using tiny-YOLOv3 to detect pedestrians in real-time. The method fails to detect severely occluded pedestrians in real-time. A recent studies on object detection architectures is covered in [6]. Motion-based player detection by background subtraction was proposed in [7]–[10] and texture-based player detection in [11], template-based player detection by classifying each window in the image into a player or non-player in [12] are the other approaches proposed to detect the players. The drawback of this approaches was that they were failed to detect the players when the players were captured using moving cameras. A survey on ball/player detection and tracking [13], [14] presents the depth of this challenging topic. Multi-player detection in soccer broadcast videos using a blob-guided particle swarm optimization (PSO) method proposed in [15], a two-step blob detection step was combined with an efficient search mechanism based on PSO for player detection. This approach failed to detect players in severe occlusion conditions [16]. Yang *et al.* [17], [18] developed Bayesian based multi-dimensional model to create probabilistic and identified occupancy map to detect the player and identify the position of the player.

B. PLAYER TRACKING

First of all, early approaches centered on distinguishing each target from the background and other tracked objects. Xing *et al.* [19] used player detection and playfield segmentation results to perform progressive observation modeling.

They created a coherent tracking framework by using the dual-mode Bayesian inference scheme, which combines forward filtering and backward smoothing to effectively overcome the problem caused by isolated single objects as well as multi occlusions. Bewley *et al.* [20] attempted to correlate a high-performing detector with a simple tracking system. Their work was further extended with the discriminative appearance model to provide a structure for data association metrics, which significantly enhances tracking accuracy [21]. Liu *et al.* [22] used a visual codebook developed using the widely used color learning approach to represent various soccer teams in two sorts of tracking models for video. Even when occlusions and shape similarities were present, Lee *et al.* [23] used a partial least square technique to develop relationships between them. For alleviating the influence of rapid camera movements, Zhang *et al.* [24] introduced a particle filter that takes into account both presence and cross-domain context, such as trajectories predicted using the homography transform. Even if these approaches have a fairly reliable tracking outcome, they always fail to correctly understand trajectories when more than two objects were occluded. The correlation filter and its variations [25]–[27], which were widely employed in recent years for single object tracking, are now being used for multiple objects tracking as well. Their technique is simple and effective: an individual correlation filter is added to each item independently while effectively maintaining the processing speed advantage. For example, as compared to representative approaches specifically developed for multiple object tracking [28], textcolorredthe kernelized correlation filter-based approach [26] worked quickly and produced comparable results. Nonetheless, the uncertainty created by extreme occlusions between players and complex motion patterns seen in soccer games continues to impact most approaches. The context conditioned motion models were implemented in [29] and has been used hierarchical data association to indirectly integrate complicated inter-object associations. Lu *et al.* [30] introduced a new tracking system that can locate, classify, and track several players by using the conditional random field to exploit both temporal and collective exclusion constraints. Shitrit *et al.* [31] proposed a multi-person tracking global optimization framework that considers image-appearance inputs, even if they are only accessible at various time intervals. Hurault *et al.* [32] proposed a soccer player recognition and tracking system based on deep learning that is robust to tiny players and a variety of situations. For reliably tracking multiple players in the dynamic environments of a soccer match, Kim [33] presented the topographic surface approach to extract the foreground region. For the identity switch problem, the k-shortest paths tracking framework is introduced in [34], [35].

C. BALL TRACKING

Yu *et al.* [36] proposed a novel paradigm with two procedures: one for recognizing the ball in broadcast soccer video based on its attributes (such as shape, color, and height), and another for tracking the ball using the Kalman filter and they achieved an accuracy of about 89.85%. Liang *et al.* [37] suggested a similar schema, with the exception that instead of using just Kalman filter to track the ball, they used template matching and Kalman filter. This schema demonstrated that it can perform well even in poor playfield conditions, with an accuracy of 89.7%. Yu *et al.* [38] suggested a two-phased approach for offline ball recognition and tracking in soccer video: first, the system creates a series of candidate balls, and second, the algorithm leverages ball trajectories to aid in ball recognition and tracking. It achieved an accuracy rate of about 81%. Huang *et al.* [39] proposed a method for tracking a small ball in a soccer video that employs an adaptive particle filter to manage occlusion and motion prediction. Because this method identifies the ball within a frame, identifying the incorrect ball causes tracking loss. Sanyal *et al.* [40] resolved the challenge of [39] by developing a modern adaptive particle filter algorithm that includes three steps: predicting the area of measuring around a football ball, assigning weights to predicted particles to measure the space points, and resampling winner points depending on their weight from measurement space. In soccer videos, Kamble *et al.* [41] proposed deep learning-based ball detection and tracking (DLBT) algorithm. Using a deep learning algorithm, this approach classifies the image into three categories: background, players, and ball, and it achieved high precision of around 87.45% in predicting the position of the ball when it went missing. Komorowski *et al.* [42] presented a player and ball identification system based on deep neural networks.

Players in the playfield and people roaming outside the playfield look alike. So, a court mask was used by earlier methods to avoid showing people detected outside the playfield, which cannot detect an official (referee) who watches a game closely for ensuring whether the rules are followed and matters arising from the play are arbitrated. But, the proposed method efficiently detects the players and referee in the playfield and eliminate people outside the playfield by detecting as the background without using a court mask. The earlier methods used for tracking resolved the problem of ID switching, but cannot track by classifying the team players and the referee. The proposed approach utilized a color mask to classify the players and referees to improve tracking efficiency to overcome the problem of ID switching.

III. PROPOSED METHODOLOGY

To detect the players, soccer ball, referee, and to eliminate those moving outside the playfield as background, YOLOv4 [43] a SOTA method was utilized. JCD-SORT was used for robust and accurate tracking of players, soccer balls, and referees. To overcome the problem of identity switching among the players and referees, color representation was introduced to classify detected bounding boxes in terms of players and referees based on the color of the jersey, which can make the tracking module track the players and referee accurately without switching their identity. The complete flowchart of detecting the players, referee,

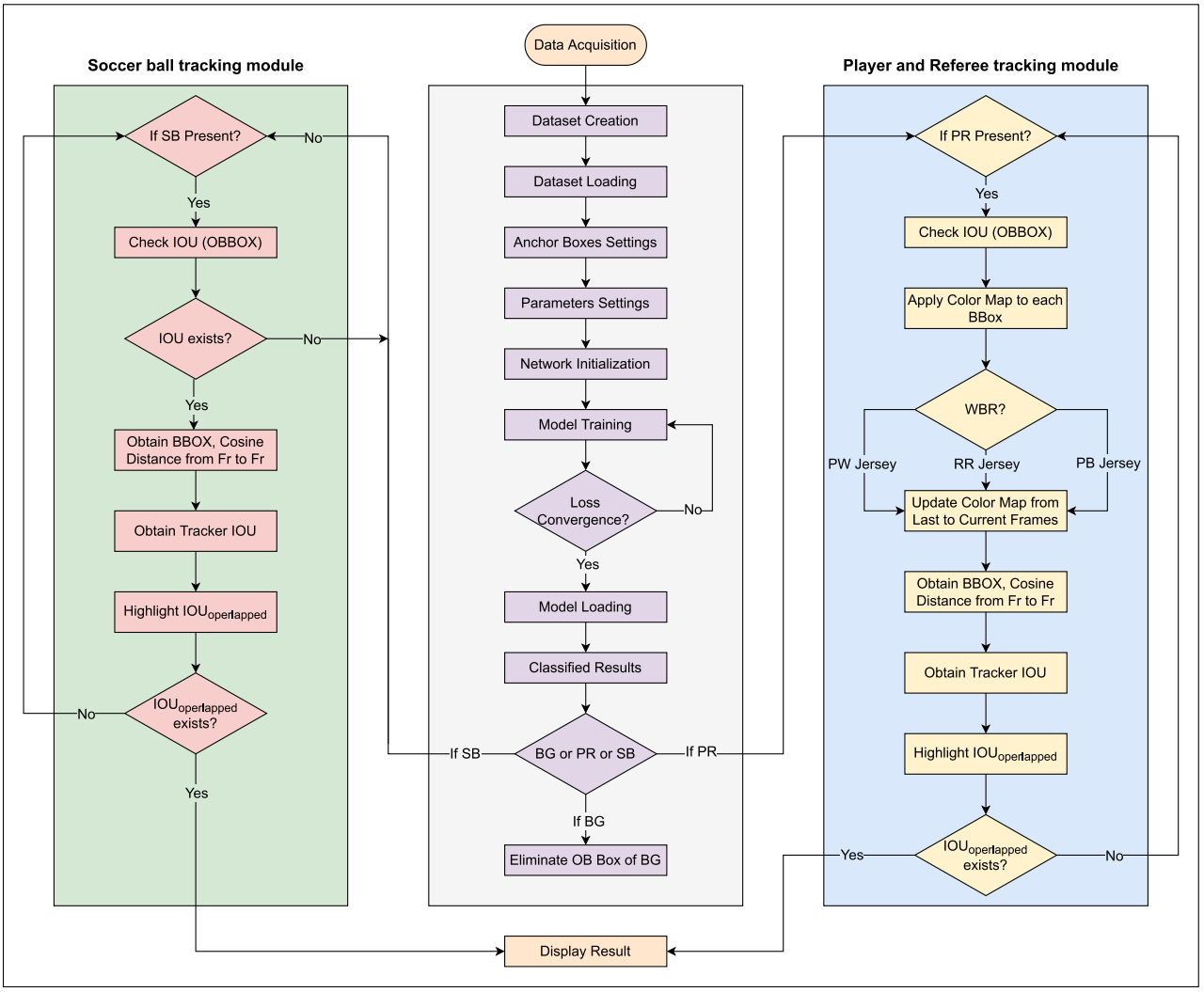


FIGURE 1. Detailed flowchart of the proposed algorithm. (PW, PB, RR represents player with white jersey, player with a blue jersey, and referee with red jersey, respectively. OBBOX represents Object Bounding BOXes).

and soccer ball, followed by the tracking module of the soccer ball, players, and referee after classifying them using the color mask is shown in Figure 1. After obtaining the detected and classified Object Bounding BOXes (OBBOX) from YOLOv4 [43] detection model, the color mask was applied to player OBBOX and referee OBBOX for further classification as a Player with White jersey (PW jersey), a Player with a Blue jersey (PB jersey) and Referee with Red jersey (RR jersey). By concatenating the color mask and cosine distance to the obtained tracker IoU from frame to frame through the whole video sequence, YOLOv4 + JCD-SORT detection and tracking algorithm successfully tracks the players without mismatch and switching the identity under various soccer game situations, which will be discussed in the subsections that follow.

A. PLAYER/BALL/REFEREE/BACKGROUND DETECTION AND CLASSIFICATION USING OPTIMIZED YOLOv4

Player/ball/referee and background detection is the main task in the whole system, which gives the bounding boxes and category probabilities for each object in the soccer play-field. Among the SOTA object detectors, YOLOv4 [43] outperformed all other approaches in terms of detection speed (FPS) and detection accuracy (mAP). The structure of YOLOv4 is composed of CSPDarknet-53 [44], Spatial Pyramid Pooling in Deep Convolutional networks (SPPnet), Path Aggregation Network (PANet) [45], and YOLO head as shown in Figure 2. As the backbone of YOLOv4, CSPDarknet-53 is responsible for extracting deep features of the input image through 5 Residual block bodies (CSP1-CSP5). The network contains 53 convolution layers with the sizes of 1×1 and 3×3, and each convolution layer is connected with a batch normalization (Bn) layer and a Mish activation layer. Furthermore, all activation functions in YOLOv4 are replaced with leaky-ReLU that requires less computation.

The neck of YOLOv4 is responsible for collecting feature maps from various phases, as shown in Figure 2. Usually, a neck is composed of several bottom-up paths using Spatial Pyramid Pooling Layer (SPP), which can overcome one of the issues caused by the constant size image presented by a

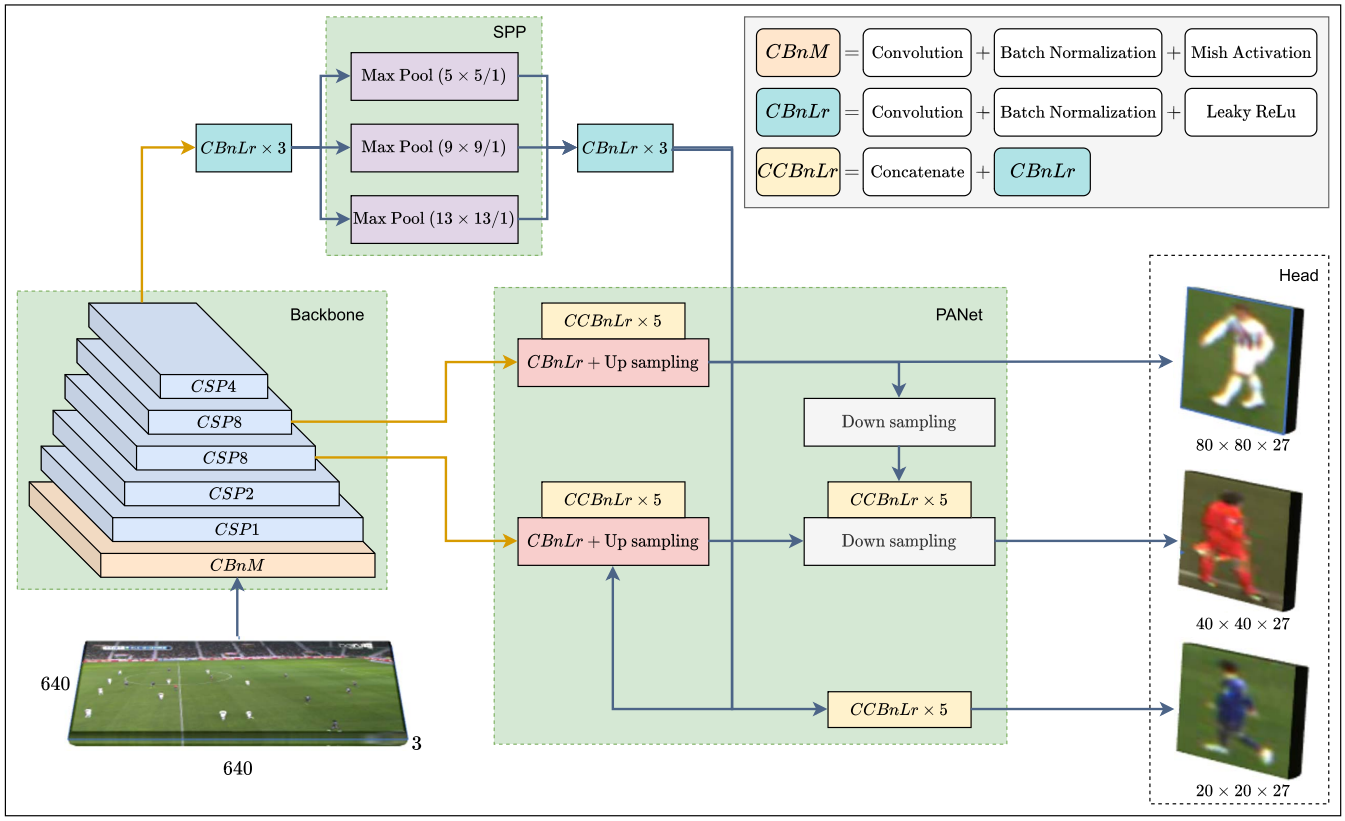


FIGURE 2. YOLOv4 architecture for payer/referee/ball detection in soccer sports.

fully connected network. The other issue caused by CNN is the fixed size of the sliding window. The result of CNN is the feature map, which is generated by features of different filters. To make it simpler, a filter capable of detecting circular geometric shapes is used; this filter generates a feature map highlighting the shapes while maintaining the shape's position in the image. Whatever the size of the specified feature maps, the SPP layer allows for the generation of fixed-size features. It utilizes pooling layers, such as Max Pooling, to produce various representations of the specified feature maps to create a fixed size. The SPP is changed in YOLO to keep the output spatial dimension. A max pool is applied to a sliding kernel of the size of 5×5, 9×9, and 13×13. The spatial dimension is preserved, and the result is a concatenation of feature maps from multiple kernel sizes. YOLO heads with sizes of 20×20, 40×40, and 80×80 are used to fuse and interact with feature maps of different scales to detect objects of different sizes. The complete YOLOv4 network architecture shown in Table 1, contains 162 layers and pre-trained weights of 137 layers were utilized to avoid the training of the model from the initial level because these 137 layers contain low and medium-level features of the images. The remaining 25 layers are the part of the head, which were trained on the sports dataset as the head is the part where actual detection is taking place. The model trained on all on the samples such as background (BG), players, and referee (PR), a small object which is moving with high velocity such as soccer ball (SB) to avoid biased training.

The model was trained on 55000 iteration steps with the hyper-parameter values such as input frame size is 640×640, weight decay of 0.0005, momentum is 0.949, the batch size is 32, by considering the initial learning rate as 0.001. After training, the following metric was considered to measure the training and validation loss of the model, i.e. mAP at IoU = 50% for each saved step and the best result was used for further evaluation. For evaluation of the detector performance, FPS, mAP, average mAP, Average IoU, Re-call, Precision, and F1-score metrics were calibrated. Also, True Positive (TP), False Positive (FP), and False Negative (FN) rates were determined, at 50% of the IoU threshold.

B. PLAYER AND REFEREE TRACKING WITH DEEP FEATURES ASSOCIATION

The detection module classifies the stakeholders as players and referees, but it will not classify the players based on the team, to which they belong. This results in identity switching between players, even identity mismatch, sometimes when they are occluded. So, this can be resolved by applying the color mask to each player and referee bounding boxes, because the major difference between teams and the referee is jersey color: one team is white, the other is blue and the referee is red. After detecting and classifying players and

TABLE 1. Network architecture of optimized YOLOv4.

	Input	Type	Filters	Stride	Output
	640×640×3	Convolutional	32	3×3/1	640×640×3
	640×640×32	Convolutional	64	3×3/2	640×640×64
5×	320×320×64	Convolutional	64	1×1/1	320×320×64
	320×320×32	Convolutional	64	3×3/1	640×640×64
		Convolutional			320×320×64
	320×320×64	Shortcut L4	128	3×3/2	
		Route L1, L2, L8			320×320×128
		Convolutional			160×160×64
5×	160×160×128	Shortcut L14, L17	64	1×1/1	
		Route L11, L12, L21			160×160×256
2×	160×160×64	Convolutional	64	3×3/1	160×160×64
	160×160×128	Convolutional	128	1×1/1	160×160×128
	160×160×128	Convolutional	256	3×3/2	80×80×256
2×	80×80×256	Convolutional	128	1×1/1	80×80×128
		Route L24			80×80×256
		Convolutional			80×80×128
9×	80×80×128	Shortcut L27, L30, L33, L36, L39, L42, L45, L48	128	1×1/1	
		Route L25, L52			80×80×256
8×	80×80×128	Convolutional	128	3×3/1	80×80×128
	80×80×256	Convolutional	256	1×1/1	80×80×256
	80×80×256	Convolutional	512	3×3/2	40×40×512
2×	40×40×512	Convolutional	256	1×1/1	40×40×256
		Route L55			40×40×512
		Convolutional			40x40x256
9×	40×40×256	Shortcut L58, L61, L64, L67, L70, L73, L76, L79	256	1×1/1	
		Route L56, L83			40×40×512
8×	40×40×256	Convolutional	256	3×3/1	40×40×256
	40×40×512	Convolutional	512	1×1/1	40×40×512
	40×40×512	Convolutional	1024	3×3/2	20×20×1024
2×	20×20×1024	Convolutional	512	1×1/1	20×20×512
		Route L86			20×20×1024
		Convolutional			20×20×512
9×	20×20×512	Shortcut L89, L92, L95, L98	512	1×1/1	
		Route L87, L102, L107, L108, L110, L112			20x20x1024
6×	20×20×512	Convoltional	1024	3×3/1	20×20×1024

the referee, three-color masks were applied to classify the players and the referee based on their jersey color. Jersey color + Cosine Distance-based Simple Online Real-Time Tracking (JCD-SORT) as shown in Figure 3, in comparison to Deep Association metric based SORT [21] process, attempts to enhance the detection of objects under occlusion by re-identifying objects that emerge in the scene after being untracked in a few frames utilizing existing data. To get the minimum cosine distance of the tracked objects, a Kalman filter was used and color masks were updated from frame to frame to the tracker module to assign tracker ID to players and referees based on the jersey color.

In particular, detected bounding box color feature and an appearance distance metric based on feature descriptors extracted using a CNN were added to the assignment cost. The cost of j^{th} detection is assigned to the i^{th} track ($C_{i,j}$) as follows:

$$C_{i,j} = d^{v}(i,j) + (PI^{p}_{OBBoxj}(A) + PI^{r}_{OBBox(j+1)}(A)) \quad (1)$$

where PI^{p}_{OBBoxj} represents pixel intensity of j^{th} detected player bounding box, and $PI^{r}_{OBBox(j+1)}$ represents pixel intensity of $(j+1)^{th}$ detected referee bounding box respectively. Where $A \in (R, G, B)$ denotes red, green, and blue channels of the bounding boxes as shown in (2).

$$PI^{p}_{OBBoxj}(x, y) = OBBOX((x_1, x_2, x_3), (y_1, y_2, y_3)) \quad (2)$$

where (x_1, y_1), (x_2, y_2), and (x_3, y_3) are the range of intensity values of red, green, and blue channels. To be specific, detected bounding boxes were cropped from the detector model output frame and pixel intensities were extracted. Thereafter pixel intensities were picked manually as shown in algorithm1 from j^{th} bounding box of the player and $(j+1)^{th}$ bounding box of the referee to concatenate the color mask with cosine distance as shown in 1. Therefore, to retain the identity of players and referees, it is necessary to compare each bounding box intensities from frame to frame as shown in Algorithm 1 (Figure 4).

The d^{v} represents the visual appearance metric which is calculated as the cosine distance between the j^{th} detection and the i^{th} track, and to get the min-cosine distance of the tracked

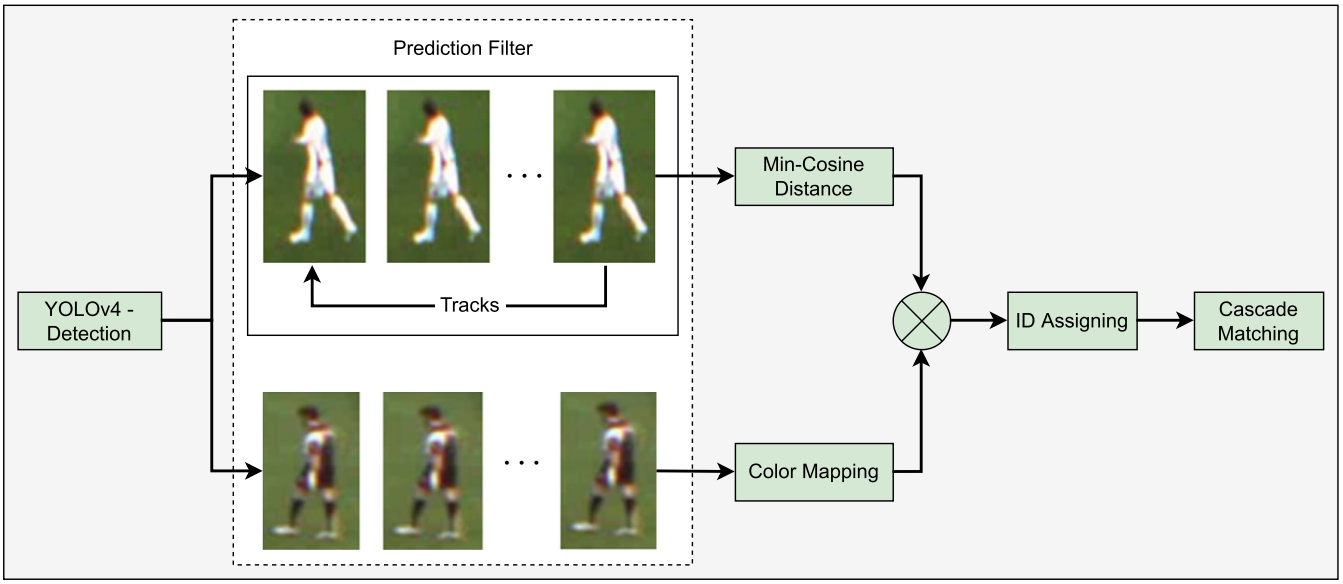


FIGURE 3. Proposed block diagram (JCD-SORT) of payer/referee/ball tracking in soccer sports.

objects, Kalman filter was used which is given as shown in 3:

$$d^{v}(i,j) = \min \frac{1 - r_j^T r_k^i}{r_k^i} \in \mathbb{R}_i \tag{3}$$

Here r_j represents the appearance descriptor with $|r_j| = 1$ extracted from a portion of the image j^{th} bounding box, and $\mathbb{R}_i$ is the set of last n appearance descriptors r_k^i associated with the track i.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The performance of the proposed methods for multiple object detection and tracking is illustrated with ISSIA and SoccerNet datasets in this section. The proposed technique is compared to the benchmark using two datasets (ISSIA and SoccerNet), which are the most commonly used to track players in soccer video. To demonstrate the proposed method's efficiency and robustness, it is compared to previous methods often used for multiple object detection and tracking, which is illustrated in the following subsections.

A. EXPERIMENTAL SETUP

The detection and classification model was trained using Google Colab with the experimental configuration of RAM 27.3 gigabytes, Tesla P100-PCIE-16GB GPU, and the tracking model was trained and tested using the workstation with the configurations RAM 64 gigabytes, NVIDIA Quadro P4000/8 GB/1792 Cuda cores GPU.

B. DATASET

Due to the unavailability of a dedicated dataset, the ISSIA dataset [46] and SoccerNet [47] were modified by annotating it with the classes of the player, soccer ball, referee, and playfield background. Around 7500 frames, which contain soccer balls from six videos of ISSIA and SoccerNet datasets, were considered to generate the required dataset. Figure 5 shows a few samples of all four classes. Among 10217 frames, 7327 frames were considered for training the model and 2040 frames for validating the trained model, and 1000 frames for testing the model.

C. QUALITATIVE EVALUATION

In this sub-section, detection and tracking results are shown in obtained video frames using ISSIA and SoccerNet Datasets.

1) DETECTION AND CLASSIFICATION OF OBJECTS IN SOCCER

For efficient tracking, all the players, referees, and soccer balls in the playfield had to be detected accurately. People roaming outside the playing field were detected, for which a class was included i.e. background. It can detect and classify people as background and referee moving outside the playfield as is shown in Figure 6.

2) TRACKING PAYERS/REFEREE AND SOCCER BALL

All the players, soccer balls, and the two referees were tracked by the proposed method and some of the tracking results from five videos of the ISSIA dataset are shown in Figure 7. YOLOv4 + Deep-SORT simply tracks the detected players; as all the detected players had similar features, it did not work in the case of identity switching, which is shown in Figure 9 (Players with ID-1 and ID-4 are switched). In particular, it can be observed that the position of each player is accurately detected, classified based on their jersey color, and tracked even under dynamic scenes of soccer sport using the proposed YOLOv4 + JCD-SORT algorithm, as seen in Figure 8. The

Algorithm: Proposed Algorithm to avoid identity switch and identity mismatch
Input: Player and Referee Bounding Boxes of all the Frames
Output: Assigning the ID based on the color of the player jersey

```
Proposed Algorithm()
  While true:
  Iterating for every frame
  Predict the object position
  Update the object position
  Store the last location
    pos_last = self.to_tlbr()
    last_mean = self.mean
    last_covariance = self.covariance
    get the present prediction
  For each object id, both for last and present position, apply red, white, blue color masks to each object
    Def get_Color(img, bbox):
  Crop object from the frame
    img = get_crop(img, bbox)
  Get the Color mask that gives maximum positive pixels
    red = get_per(img, (x1, x2, x3), (y1, y2, y3))
    white = get_per(img, (x1, x2, x3), (y1, y2, y3))
    blue = get_per(img, (x1, x2, x3), (y1, y2, y3))
    if max(red, white, blue) == blue:
      return "blue"
    if max(red, white, blue) == white:
      return "white"
    if max(red, white, blue) == red:
      return "red"
  Comparing first and present prediction Color to detect identity switch
    if first_Color == present_Color:
      self.features.append(detection.feature)
    else:
  Reset to the last object if last and current identity mismatches
    if last_Color != present_Color:
      self.mean = last_mean
      self.covariance = last_covariance
      self.features.append(detection.feature)
```

FIGURE 4. Proposed Algorithm to avoid identity switch and identity mismatch.

proposed approach is robust in the case of identity switching between two players.

All the players, soccer balls, and referees were tracked by the proposed method and some of the tracking results from two different videos of the SoccerNet dataset are shown in Figures 10 and 11. When the jersey color of the referee is other than red color then the proposed YOLOv4 + JCD-SORT method fails to detect the referee, and it detects as player and tracks as a player as shown in Figure 10. In particular, it can be observed that the position of each player is accurately detected, and classified based on their jersey color and tracked even under dynamic scenes of soccer sport using the proposed YOLOv4 + JCD-SORT algorithm, (Players with ID-5, ID-7, and ID-31 identities were retained) as seen in Figure 10. The proposed approach is robust in the case of identity switching between two players.

Previous techniques, in particular, often miss target objects and suffer from tracking drift due to complex motion patterns and occlusion conditions, and cannot identify the referee, the position of the ball, or differentiate the players based on the team the player belongs to. The proposed approach is capable of dealing with such challenging situations. It is important to note that the proposed approach is capable of providing accurate results of the position of the player based on their team since it works on the basis of the jersey color of the player instead of the segmentation approach or the rectangular window, two commonly used approaches in tracking methods earlier.

D. QUANTITATIVE EVALUATION

In this sub-section, detected, classified and tracking results were measured using various performance metrics.

1) PERFORMANCE OF DETECTION AND CLASSIFICATION OF THE OBJECTS IN SOCCER

To evaluate the detection and classification of the objects, the following metrics were considered. The results of detection and classification of players, referee, soccer ball, and background were compared with other SOTA detection algorithms and tabulated which is shown in Table 2. The proposed detection methodology detects four classes (player, soccer ball, referee, and background) and the compared methodologies detected only one class (i.e. player or ball).

Precision is defined as (4)

$$P = \frac{T_P}{T_P + F_P} \tag{4}$$

Recall/True Positive Rate (TPR) is defined as (5)

$$R/TPR = \frac{T_P}{T_P + F_N} \tag{5}$$

The harmonic mean of precision and recall can be defined as F1-score as (6)

$$F1 - score = 2 \cdot \frac{P \cdot R}{P + R} \tag{6}$$

Average Precision (AP) is given by (7)

$$AP = \sum_{t} (R_n - R_{n-1}) P_n \tag{7}$$

2) EVALUATION OF TRACKING PERFORMANCE ON SOCCER VIDEO

Sub-videos comprising 500 frames were manually annotated with thirteen players, one soccer ball, and one referee from the original videos in ISSIA and SoccerNet datasets, and sub-videos comprising 500 frames were manually annotated with sixteen players, one soccer ball, and one referee from the original video in SoccerNet dataset to quantitatively analyze the efficacy of tracking algorithms. The proposed tracking methodology detects three classes (player, soccer

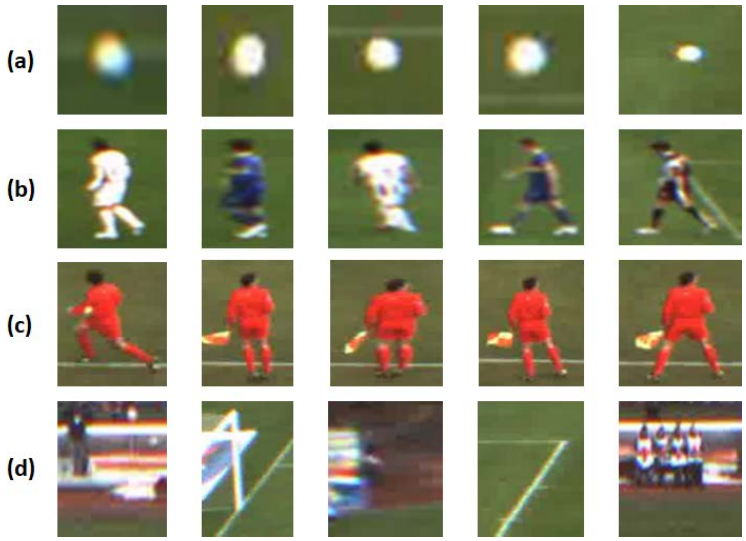


FIGURE 5. Sample images from modified ISSIA dataset (a) soccer ball (b) player (c) referee (d) background.

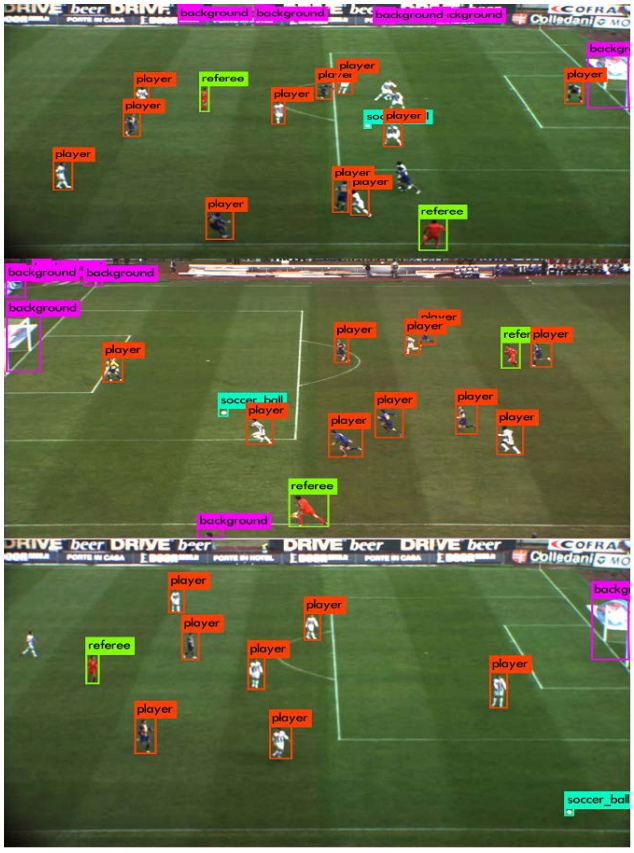


FIGURE 6. Player, soccer ball, referee, background detection and classified results on ISSIA and SoccerNet datasets.

ball, and referee) and the compared methodologies detected only one class (i.e. player or ball). To start with, the completeness of the tracking method was evaluated using seven metrics that are commonly used in this field and described as follows:

The performance of Multi-object tracking (MOT) can be measured by Multiple Object Tracking Accuracy (MOTA) as shown in (8).

$$MOTA = 1 - \frac{\sum_t (C_m \cdot m_t + C_m \cdot f_{p_t} + C_s \cdot mme_t)}{\sum_t g_t} \quad (8)$$

In order to evaluate the identity switch influence, '*gmme*' was used to measure the proportion of identity switches in a global manner. To get '*gmme*' greater than '*mme*' it must be selected that $C_s = C_m = 1$. Global Multiple Object Tracking Accuracy (GMOTA) [48] is shown in (9).

$$GMOTA = 1 - \frac{\sum_t (C_m \cdot m_t + C_m \cdot f_{p_t} + C_s \cdot gmme_t)}{\sum_t g_t} \quad (9)$$

where $m_t = \sum_n \frac{L_n}{g_{tn}}, f_{p_t} = \sum_n \frac{f_{pn}}{g_{tn}}, gmme_t = \sum_n \frac{mme_t}{g_{tn}}$.

Here, m_t is the number of missing tracks (Player with ID-8 is not detected in few frames), f_{p_t} is the number of mismatches (as the player with ID-8 mismatch with ground truth track), mme_t is the number of instantaneous identity switches, and g_t is the number of detection ground truth (total number of ground truth tracks over 500 frames are 55).

a: TRACKING PERFORMANCE ON ISSIA DATASET

The weight factors were set to $C_m = 1$ and $C_s = ln10$ with reference to [48]. Therefore, for $m_t = 1, f_{p_t} = 1, mme_t = 0$;

$$MOTA = 1 - \frac{1+1}{55} = 0.96 \quad (10)$$

where $m_t = \sum_n \frac{L_n}{g_{tn}} = (\frac{1}{9} \times 17), f_{p_t} = \sum_n \frac{f_{pn}}{g_{tn}} = (\frac{1}{9} \times 17)$, $gmme_t = \sum_n \frac{mme_t}{g_{tn}} = 0$.

Here, g_{tn} denotes number of actual players at n^{th} frame corresponding to ground truths (actually ground truth players

FIGURE 7. Tracking results from five film roles of ISSIA dataset.

are 9), Ln denotes the number of lost players (detected and tracked 8 players out of 9 ground truth players in 17 frames, therefore $L_n = 1$), and f_{p_n} denotes the number of players missed the tracks in the n^{th} frame (i.e. 1 player last the track in 17 frames). Table 3 shows the comparison of tracking performance depending on GMOTA metrics.

$$GMOTA = 1 - \frac{1.8 + 1.8}{9} = 0.60 \tag{11}$$

b: TRACKING PERFORMANCE ON SOCCERNET DATASET

Where the number of missing tracks are 7 (five players not detected in a few frames and after a few frames only two players were not detected), the number of mismatches (referee is tracked as player and ID-8 and ID-16 were tracked as ID-72 and ID-75 after few frames which mismatch with ground truth track), mme_t is the number of instantaneous identity switch, and gt is the number of detection ground truth (total number of ground truth tracks over 500 frames are 33) as shown in Figure 10. Therefore, for $m_t = 7, f_{p_t} = 3, mme_t = 0$,

$$MOTA = 1 - \frac{7 + 3}{33} = 0.69 \tag{12}$$

where $m_t = \sum_n \frac{L_n}{g_{tn}} = ((\frac{5}{33} \times 32) + (\frac{2}{33} \times 113))$, $f_{p_t} = \sum_n \frac{f_{pn}}{g_{tn}} = ((\frac{5}{33} \times 32) + (\frac{2}{33} \times 113))$, $gmme_t = \sum_n \frac{mme_t}{g_{tn}} = 0$.

Here, g_{tn} denotes the number of actual players at n^{th} frame corresponding to ground truths (actually ground truth players are 16), L_n denotes the number of lost players (detected and tracked 11 players out of 16 ground truth players in first 32 frames, detected and tracked 14 players out of 16 ground truth players in remaining frames, therefore, $L_n = 7$), and f_{p_n} denotes the number of players missed the tracks in the n^{th} frame (i.e. five player last the track in 32 frames and two players lost in 113 frames among the total frames). Table 3 shows the comparison of tracking performance depending on GMOTA metrics.

$$GMOTA = 1 - \frac{11.6 + 11.6}{33} = 0.29 \tag{13}$$

It indicates the capacity of the tracker to predict specific object positions, regardless of the proficiency in detecting object configurations, keeping consistent trajectories, etc. The proposed approach was compared to four representative tracking algorithms in order to evaluate tracking performance qualitatively, i.e., Tracklet-based Multi-Commodity Network Flow (T-MCNF) [31], Topography [33], Small-Soccer

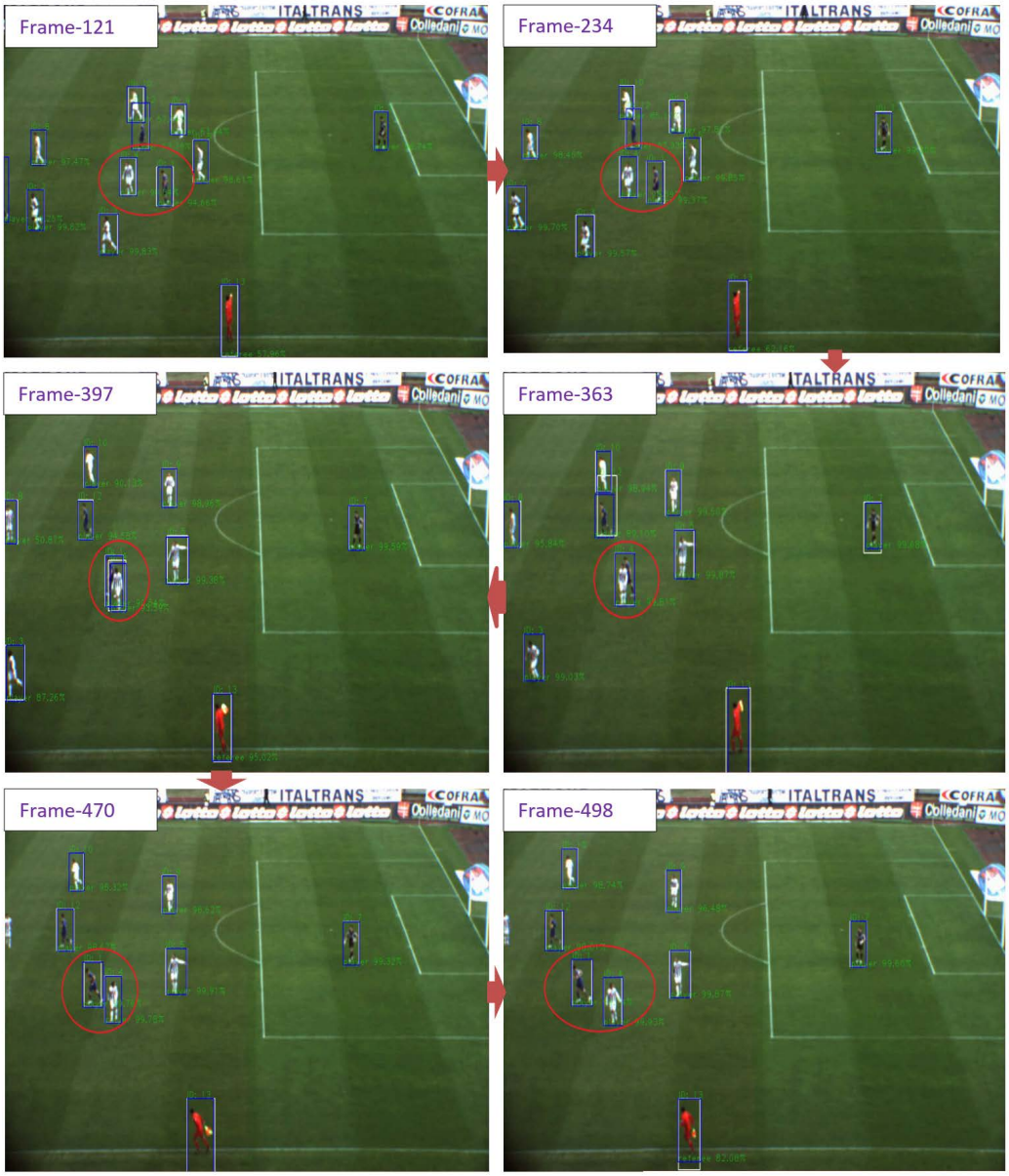


FIGURE 8. Tracking results in the case of identity switching on ISSIA dataset Using YOLOv4 + JCD-SORT, which can be seen in the red circle (Frame-121 to Frame-498).

TABLE 2. Comparative analysis of proposed detection/tracking methodology on soccer videos.

Method	Dataset	Class	Precision (%)	Recall (%)	F1-Score (%)	Average IOU (%)	AP (%)	mAP (%)	Accuracy (%)	FPS
DLBT [41]	ISSIA	Soccer Ball	93.25	73.25	-	52	-	-	87.45	10
FootAndBall [42]	ISSIA	Player	-	-	-	-	92.1	-	-	-
Small-Soccer Player [32]	ISSIA	Payer	-	-	–		97.3	-	-	-
Proposed	ISSIA	Player, Soccer Ball, Referee, and background	93	95	94	76.62	-	92.47	-	23
	SoccerNet	Player, Soccer Ball, Referee, and background	94	95	94	77.5	-	91.76	-	21

Player [32], and MCMPTI [34] on ISSIA dataset depending on MOTA/GMOTA metrics as shown in Table 3. In some of the frames player with ID-8 was frequently not detected due to which track was missed (m_t) in those frames and again frequently detected which causes the mismatch (f_{p_t}) between the ground truth track and

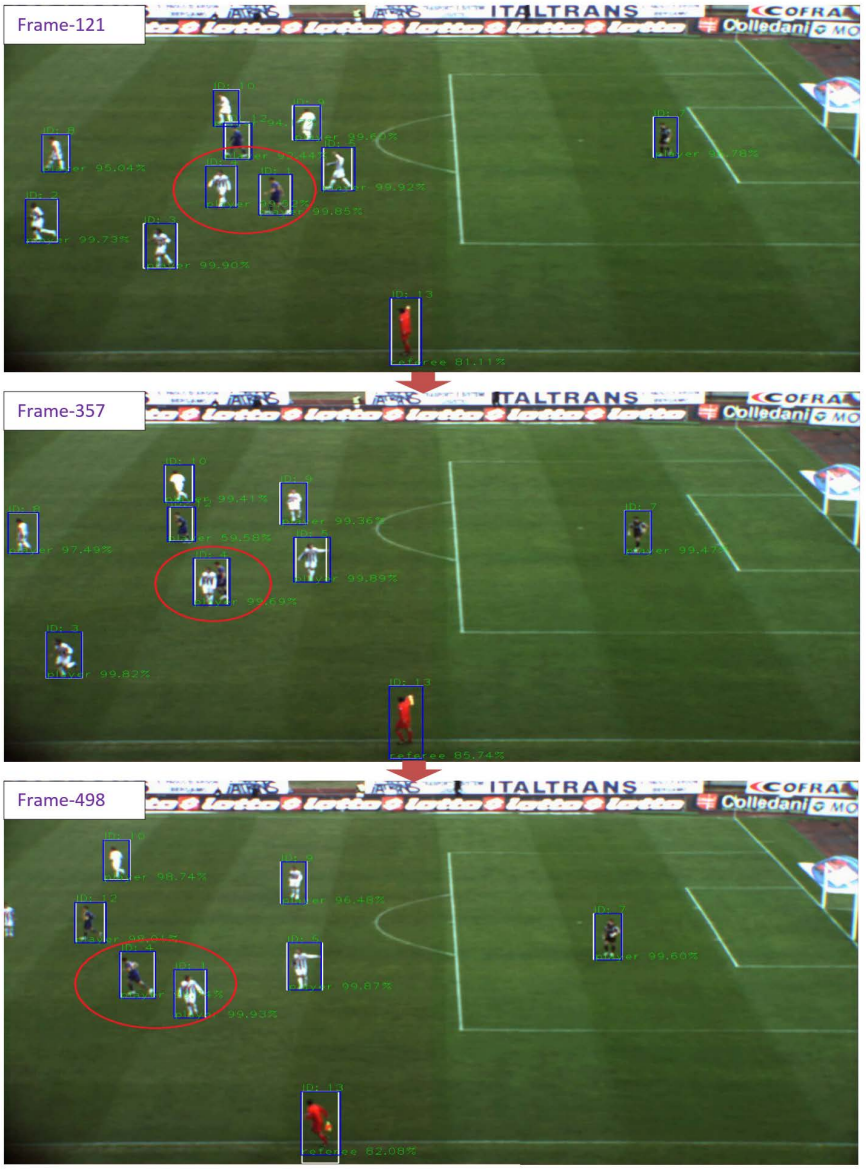


FIGURE 9. Tracking results in the case of identity switching Using YOLOv4 + Deep-SORT, which can be seen in the red circle (Frame-121 to Frame-498).

TABLE 3. Comparative analysis of the proposed tracking algorithm with SOTA methodologies on ISSIA dataset using MOTA/GMOTA metrics.

Method	T-MCNF [31]	Topographic [33]	Small-Soccer Player [32]	K-SP MCMPTI [34]	Proposed (ISSIA)	Proposed (SoccerNet)
Class	Player	Player	Player	-	Player, Soccer Ball, Referee	Player, Soccer Ball, Referee
MOTA ↑	0.53	0.92	0.811	92.9	0.96	0.69
mme_t ↓	0.1	2	-	4	0	0
m_t ↓	-	-	-	-	1	7
fp_t ↓	-	-	-	-	1	3
FPS	187.5	-	6.5	-	8.7	11.3
GMOTA ↑	0.38	-	-	0.723	0.60	0.29
gmme ↓	0.05	0	-	4	0	0
m_t ↓	0.18	0.03	-	-	1.8	11.6
fp_t ↓	0.18	0.056	-	-	1.8	11.6

predicted track. The number of players and referees was tracked accurately and verified, which is shown in Figure 7. Based on the above results, it is apparent that the proposed approach has more accurate players/referee and soccer ball tracking than previous approaches in the soccer game.

The accuracy is reduced owing to the non-uniformity of the number of players in different frames- as in, for the first

FIGURE 10. Tracking results in the case of identity switching on SoccerNet dataset Using YOLOv4 + JCD-SORT which can be seen in red circle (players with ID-5, ID-7 and ID-31 identities were retain) (Frame-98 to Frame-325).

few frames one player and in subsequent frames few more player starts disappearing resulting in an incongruence in the analysis as shown in Figure 10.

E. ABLATION STUDY

To evaluate the effect of identity switching, the model was implemented in two variants and tested on the ISSIA and SoccerNet datasets. The ablation baseline included the following: (i) YOLOv4 + DEEP-SORT: the proposed model without the color mask, (ii) YOLOv4 + JCD-SORT: the proposed model with the color mask. The performances of both the models were evaluated on MOTA and GMOTA metrics.

TABLE 4. Comparative analysis of proposed methodology using MOTA/GMOTA metrics on ablation study.

Method	YOLOv4 + Deep-SORT	YOLOv4 + JCD + SORT
MOTA ↑	0.94	0.96
$mme_t \downarrow$	1	0
$m_t \downarrow$	1	1
$fp_t \downarrow$	1	1
GMOTA ↑	0.51	0.60
gmme ↓	1	0
$m_t \downarrow$	1.8	1.8
$fp_t \downarrow$	1.8	1.8

From Table 4, it is clear that the YOLOv4 + JCD-SORT model is significant for multi-player tracking. Specifically, identity switching occurs frequently, if the color mask

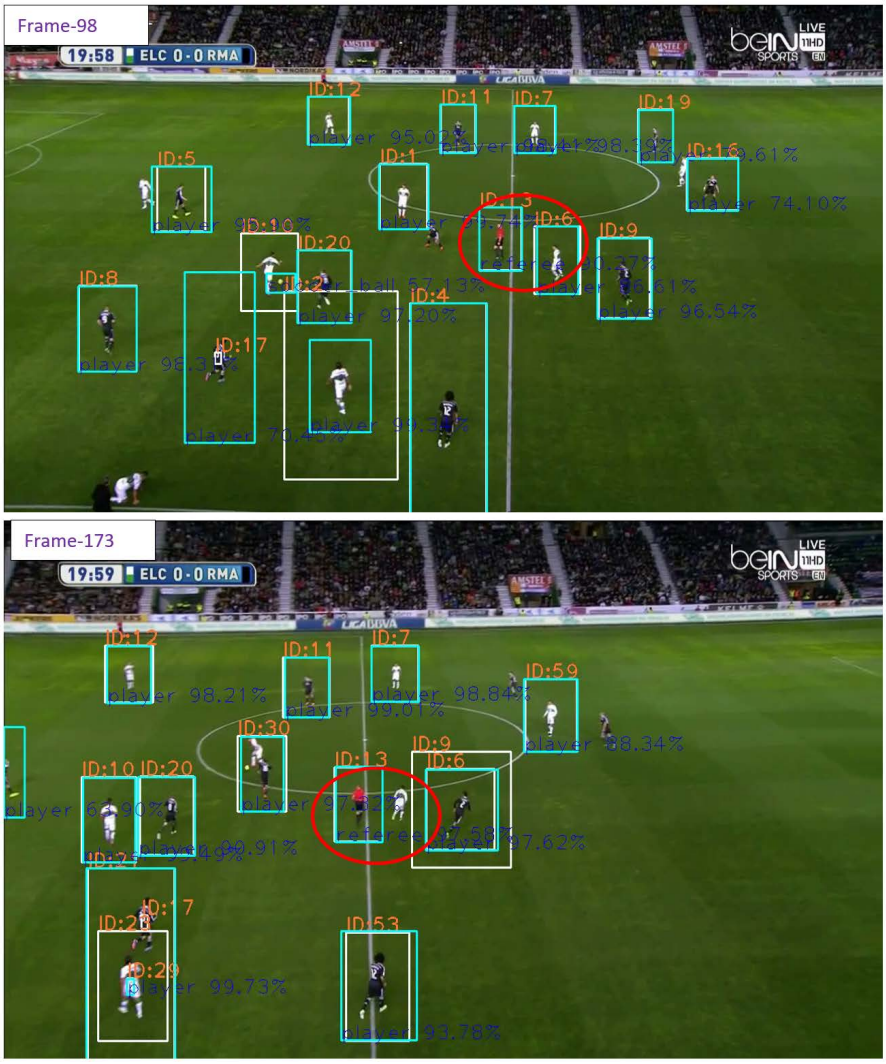


FIGURE 11. Tracking results SoccerNet dataset video which has referee jersey colour red using YOLOv4 + JCD-SORT which can be seen in red circle.

is not applied to classify the players based on jersey color.

V. CONCLUSION AND SCOPE FOR FUTURE WORK

A robust multi-object tracking methodology is proposed in this paper. In this approach, the objects detected outside the playfield are successfully eliminated as background, which improved the performance of detecting the objects (such as players, soccer balls, and referees) on the playfield. Specifically, by applying the color mask, team payers and referees were successfully classified based on their jersey color and different IDs were assigned to each payer and referee to handle ID switches, which improved tracking performance. The results are discussed in three sections; (1) qualitative evaluations presented the results in pictorial formats; (2) quantitative evaluations measured the results in various performance metrics; (3) ablation experiments were done to verify that the ID switch problem was effectively tackled. The results of the proposed methodology effectively handle identity switches and surpass SOTA trackers on soccer video. The limitation of the proposed method is that when the player with the same jersey color is occluded, the ID of the player is switched. The major drawback of the proposed methodology is higher computation cost, which reduces FPS compared to those of existing methods.

In the future, to improve the tracking performance of real-time vs precision, jersey number recognition will be considered to track players from the same squad. The approach suggested can also be used in other real-world scenarios, such as basketball and rugby sports.

REFERENCES

[1] *World Cup 2018 Breaks Viewing Records Across Streaming Platforms as Soccer Fans Tune*. Accessed: Nov. 23, 2018. [Online]. Available: https://www.cnbc.com/2018/07/14/world-cup-2018-breaks-viewing-records-across-streaming-platforms-as-so.html

[2] J. Halbinger and J. Metzler, "Video-based soccer ball detection in difficult situations," in *Proc. Int. Congr. Sports Sci. Res. Technol. Support*, Cham, Switzerland: Springer, 2013, pp. 17–24.

[3] V. Pallavi, J. Mukherjee, A. K. Majumdar, and S. Sural, "Ball detection from broadcast soccer videos using static and dynamic features," *J. Vis. Commun. Image Represent.*, vol. 19, no. 7, pp. 426–436, Oct. 2008.

[4] P. L. Mazzeo, M. Leo, P. Spagnolo, and M. Nitti, "Soccer ball detection by comparing different feature extraction methodologies," *Adv. Artif. Intell.*, vol. 2012, pp. 1–12, Oct. 2012.

[5] C. B. Murthy and M. F. Hashmi, "Real time pedestrian detection using robust enhanced YOLOv3+," in *Proc. 21st Int. Arab Conf. Inf. Technol. (ACIT)*, Nov. 2020, pp. 1–5.

[6] C. B. Murthy, M. F. Hashmi, N. D. Bokde, and Z. W. Geem, "Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review," *Appl. Sci.*, vol. 10, no. 9, p. 3280, May 2020.

[7] K. Choi and Y. Seo, "Automatic initialization for 3D soccer player tracking," *Pattern Recognit. Lett.*, vol. 32, no. 9, pp. 1274–1282, Jul. 2011.

[8] R. Martín and J. M. Martínez, "A semi-supervised system for players detection and tracking in multi-camera soccer videos," *Multimedia Tools Appl.*, vol. 73, no. 3, pp. 1617–1642, 2013.

[9] M. A. M. Laborda, E. F. T. Moreno, J. M. del Rincón, and J. E. H. Jaraba, "Real-time GPU color-based segmentation of football players," *J. Real-Time Image Process.*, vol. 7, no. 4, pp. 267–279, Dec. 2012.

[10] R. Hamid, R. K. Kumar, M. Grundmann, K. Kim, I. Essa, and J. Hodgins, "Player localization using multiple static cameras for sports visualization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 731–738.

[11] T. Mochizuki, M. Fujii, M. Shibata, and Y. Sakai, "Flast identification of player position in soccer broadcast video by block-based camera view angle search," in *Proc. 6th Int. Symp. Image Signal Process. Anal.*, Sep. 2009, pp. 408–413.

[12] M. Heydari and A. M. E. Moghadam, "An MLP-based player detection and tracking in broadcast soccer video," in *Proc. Int. Conf. Robot. Artif. Intell.*, Oct. 2012, pp. 195–199.

[13] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, "Ball tracking in sports: A survey," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1655–1705, Oct. 2019.

[14] M. Manafifard, H. Ebadi, and H. A. Moghaddam, "A survey on player tracking in soccer videos," *Comput. Vis. Image Understand.*, vol. 159, pp. 19–46, Jun. 2017.

[15] M. Manafifard, H. Ebadi, and H. A. Moghaddam, "Appearance-based multiple hypothesis tracking: Application to soccer broadcast videos analysis," *Signal Process., Image Commun.*, vol. 55, pp. 157–170, Jul. 2017.

[16] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.

[17] Y. Yang, R. Zhang, W. Wu, Y. Peng, and M. Xu, "Multi-camera sports players 3D localization with identification reasoning," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 4497–4504.

[18] Y. Yang, M. Xu, W. Wu, R. Zhang, and Y. Peng, "3D multiview basketball players detection and localization based on probabilistic occupancy," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Dec. 2018, pp. 1–8.

[19] J. Xing, H. Ai, L. Liu, and S. Lao, "Multiple player tracking in sports video: A dual-mode two-way Bayesian inference approach with progressive observation modeling," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1652–1667, Jun. 2011.

[20] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3464–3468.

[21] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[22] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, "Automatic player detection, labeling and tracking in broadcast soccer video," *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 103–113, 2009.

[23] S.-H. Lee, M.-Y. Kim, and S.-H. Bae, "Learning discriminative appearance models for online multi-object tracking with appearance discriminability measures," *IEEE Access*, vol. 6, pp. 67316–67328, 2018.

[24] T. Zhang, B. Ghanem, and N. Ahuja, "Robust multi-object tracking via cross-domain contextual information for sports video analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 985–988.

[25] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550.

[26] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[27] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2016.

[28] J. Kwon, K. Kim, and K. Cho, "Multi-target tracking by enhancing the kernelised correlation filter-based tracker," *Electron. Lett.*, vol. 53, no. 20, pp. 1358–1360, Sep. 2017.

[29] J. Liu, P. Carr, R. T. Collins, and Y. Liu, "Tracking sports players with context-conditioned motion models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1830–1837.

[30] W.-L. Lu, J.-A. Ting, J. J. Little, and K. P. Murphy, "Learning to track and identify players from broadcast sports videos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1704–1716, Jul. 2013.

[31] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Multi-commodity network flow for tracking multiple people," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1614–1627, Aug. 2014.

[32] S. Hurault, C. Ballester, and G. Haro, "Self-supervised small soccer player detection and tracking," in *Proc. 3rd Int. Workshop Multimedia Content Anal. Sports*, Oct. 2020, pp. 9–18.

[33] W. Kim, "Multiple object tracking in soccer videos using topographic surface analysis," *J. Vis. Commun. Image Represent.*, vol. 65, Dec. 2019, Art. no. 102683.

[34] R. Zhang, L. Wu, Y. Yang, W. Wu, Y. Chen, and M. Xu, "Multi-camera multi-player tracking with deep player identification in sports video," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107260.

[35] Q. Liang, W. Wu, Y. Yang, R. Zhang, Y. Peng, and M. Xu, "Multi-player tracking for multi-view sports videos with improved K-shortest path algorithm," *Appl. Sci.*, vol. 10, no. 3, p. 864, Jan. 2020.

[36] X. Yu, C. Xu, Q. Tian, and H. Wai Leong, "A ball tracking framework for broadcast soccer video," in *Proc. Int. Conf. Multimedia Expo. (ICME)*, Jul. 2003, p. 273.

[37] D. Liang, Y. Liu, Q. Huang, and W. Gao, "A scheme for ball detection and tracking in broadcast soccer video," in *Proc. Pacific-Rim Conf. Multimedia*. Cham, Switzerland: Springer, 2005, pp. 864–875.

[38] X. Yu, H. W. Leong, C. Xu, and Q. Tian, "Trajectory-based ball detection and tracking in broadcast soccer video," *IEEE Trans. Multimedia*, vol. 8, no. 6, pp. 1164–1178, Dec. 2006.

[39] Y. Huang and J. Llach, "Tracking the small object through clutter with adaptive particle filter," in *Proc. Int. Conf. Audio, Lang. Image Process.*, Jul. 2008, pp. 357–362.

[40] S. Sanyal, A. Kundu, and D. P. Mukherjee, "On the (soccer) ball," in *Proc. 10th Indian Conf. Comput. Vis., Graph. Image Process.*, Apr. 2016, pp. 1–8.

[41] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, "A deep learning ball tracking system in soccer videos," *Opto-Electron. Rev.*, vol. 27, no. 1, pp. 58–69, Mar. 2019.

[42] J. Komorowski, G. Kurzejamski, and G. Sarwas, "FootAndBall: Integrated player and ball detector," 2019, *arXiv:1912.05445*.

[43] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

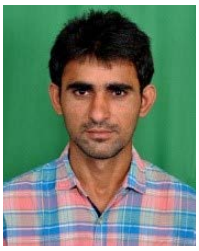
[44] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.

[45] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.

[46] T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo, and P. L. Mazzeo, "A semi-automatic system for ground truth generation of soccer video sequences," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 559–564.

[47] S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, "SoccerNet: A scalable dataset for action spotting in soccer videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1711–1721.

[48] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 17–35.



BANOTH THULASYA NAIK received the B.Tech. degree in electronics and communication engineering from the JNTUK University College of Engineering Vizianagaram and the M.Tech. degree in electronic instrumentation from NIT Warangal, in 2019, where he is currently pursuing the Ph.D. degree under the supervision of Dr. Md. Farukh Hashmi. His research interests include computer vision, object detection, pattern recognition and classification, and motion analysis using deep learning.

MOHAMMAD FARUKH HASHMI (Senior Member, IEEE) received the B.E. degree in electronics and communication engineering from MIT Mandsaur/RGPV Bhopal University, in 2007, the M.E. degree in digital techniques and instrumentation from SGSITS (Autonomous State Government) Indore/RGPV Bhopal University, in 2010, and the Ph.D. degree from the Visvesvaraya National Institute of Technology (VNIT), Nagpur, in 2015, under the supervision of Dr. Avinash G. Keskar. He is currently working as an Assistant Professor with the Department of Electronics and Communication Engineering, National Institute of Technology (NIT), Warangal. He has published up to 75 articles including 25 SCI indexed research papers in international/national journals/conferences of publishers like IEEE, Elsevier, and Springer. He has published one patent to his credit. He was a principal investigator of one research project of worth five lakhs funded by the Institute Seed Grant through TEQIP III. He has a teaching and research experience of 13 years. He is also serving as an Active and Potential Technical Reviewer for IEEE ACCESS, *IET Image Processing*, *IET Computer Vision*, *Wireless Personal Communications*, IEEE SYSTEMS JOURNAL, *Sensors* (MDPI), *Electronics* (MDPI), *Diagnostic* (MDPI), *The Visual Computer*, *Applied Soft Computing* (Elsevier), *Color Research and Application*, *The Journal of Supercomputing*, and various other journals like Elsevier/Springer/IEEE TRANSACTIONS publishers of repute. He has supervised two Ph.D. scholars. He is presently guiding four Ph.D. scholars. His current research interests include computer vision, machine vision, machine learning, deep learning, embedded systems, the Internet of Things, digital signal processing, image processing, and digital IC design. He is a Life Member of IETE, ISTE, and IAENG societies.

ZONG WOO GEEM (Senior Member, IEEE) received the B.Eng. degree from Chung-Ang University, the M.Sc. degree from Johns Hopkins University, and the Ph.D. degree from Korea University. He researched at Virginia Tech, University of Maryland, College Park, and Johns Hopkins University. He is currently an Associate Professor with the College of IT Convergence, Gachon University, South Korea. He invented a music-inspired optimization algorithm, and harmony search, which has been applied to various scientific and engineering problems. His research interests include phenomenon-mimicking algorithms and their applications to energy, environment, and water fields. He has served for various journals as an Editor (an Associate Editor for *Engineering Optimization* and a Guest Editor for *Swarm and Evolutionary Computation*, *International Journal of Bio-Inspired Computation*, *Journal of Applied Mathematics*, *Applied Sciences*, *Complexity*, and *Sustainability*).

NEERAJ DHANRAJ BOKDE received the M.E. degree in embedded systems from the EEE Department, BITS Pilani, Pilani Campus, India, and the Ph.D. degree in data science from the Visvesvaraya National Institute of Technology, Nagpur, India. Then, he has worked as a Postdoctoral Researcher at different departments at Aarhus University, Aarhus, Denmark. His research interests include the domain of data science topics, focused majorly on time series analysis, software package development, and prediction applications in renewable energy. He is serving in editorial positions in *Data in Brief*, *Frontiers in Energy Research*, *Energies*, and *Information* journals.