ORIGINAL ARTICLE

OBB detector: occluded object detection based on geometric modeling of video frames

Supriya Agrawal[1] · **Prachi Natu**[1]

Abstract

Object detection is an important research area in video surveillance systems, aimed at identifying and locating target objects within recorded scenes. Various object detectors fail when partial occlusion occurs in which only some features of the objects are visible due to overlapped bounding boxes. This situation can result in miscounting of the objects and misaligning the bounding boxes leading to localization loss. To address these problems, we have proposed a geometric-based axis-aligned bounding box method with occlusion prior conditions to estimate the location of overlapped bounding boxes with a single viewpoint. Firstly, the proposed method detects the closest points of the detected bounding boxes by extracting geometric features namely the width, height, and area of the detected objects. Secondly, occlusion prior condition is used to detect the partial occlusion and compute the overlapped area under different levels of occlusions such as (i) 20–40% and (ii) 40–70%. The performance of the proposed method has been tested on two benchmark datasets: Highway and PETS 2006, both containing outdoor video frames. The experimental results show that the proposed method can detect the objects under partial occlusion which are approximately 65% occluded with 92.7% precision in the Highway dataset and 85.1% precision in the PETS2006 dataset. Also, it has been observed that the bounding box localization loss of the proposed method has been improved by 1.76% in the Highway dataset and 2% in the PETS2006 dataset by generating the correct aligned bounding boxes on the detected objects, especially in the case of partial occlusion.

Keywords Bounding box · Geometric modeling · Localization error · Occlusion detection · Overlapping objects · Security · Video surveillance

1 Introduction

Video surveillance is an essential application of computer vision to provide real-time information by analyzing video footage for security purposes. These systems are helpful to detect, track, and recognize objects of interest, such as people, vehicles, animals, and many more. Object detection is a crucial aspect of video surveillance, which involves identifying objects in a video frame and localizing them in the scene. Significant research has been conducted in the field of object detection and classification using several frameworks to improve the accuracy of object detection algorithms [1, 2]. However, object detection in video surveillance is challenging due to occlusion which occurs when objects are partially or fully obstructed from view, making them challenging to detect and track [3, 4]. Occlusion condition occurs when some part of the object is hidden by another object or the objects are overlapped. Due to occlusion, object detectors fail in the correct counting of the objects or the correct classification of the objects. Occlusion can be caused by objects moving in front of each other or objects being partially hidden by other obstructions. Occlusion can be found in a variety of real-world situations as shown in Fig. 1.

In Fig. 1a, the car's window is obscured by the person due to the same color leads to the miscounting of the objects. While Fig. 1b represents the front side of the car's feature is obscured by the pedestrian which shows a loss of feature visibility. Partial occlusion is particularly challenging, as it is difficult to accurately estimate the size, position, and

✉ Supriya Agrawal
   supriya.agrawal@nmims.edu

   Prachi Natu
   prachi.natu@yahoo.com

1 Department of Computer Engineering, Mukesh Patel School of Technology Management & Engineering, SVKM's Narsee Monjee Institute of Management Studies (NMIMS) University, Mumbai, India

Fig. 1 Different scenarios of occlusion occurrences [7]

(a) (b)
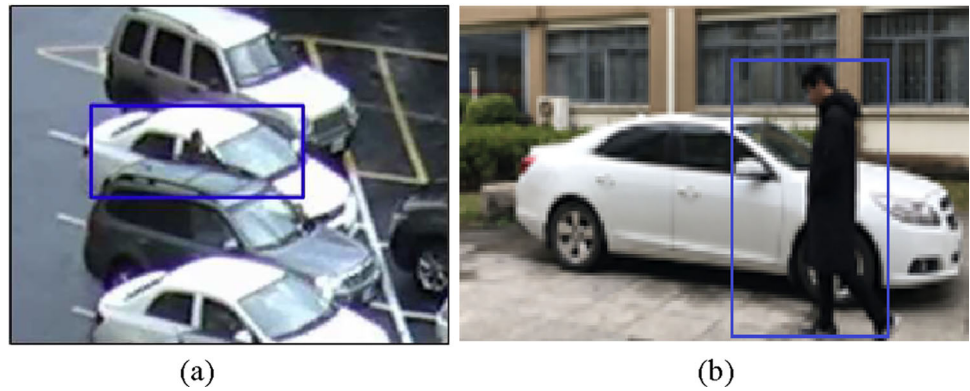
shape of the occluded object. Occlusion detection is the process of determining the interested objects in an image or video sequence where some parts of the detected objects are overlapped by another object in the recorded scene. Video surveillance systems do not perform well when interacting objects have similar appearance traits, it is more difficult to track them efficiently [5]. Occlusion can lead the system to lose track of the object being monitored and the erroneous object to be tracked after overlapping. Over the decade, various research has been published on the subject of vehicle and person detection under occlusion condition [6, 7].

There are two types of occlusions: (i) Full Occlusion and (ii) Partial Occlusion. Full occlusion occurs when an object is completely hidden by other objects and it is difficult to detect the objects without knowing the object's visibility. Detection of full occlusion is not the scope of our research. In areas with sophisticated traffic patterns, the effectiveness of vehicle and pedestrian detection methods may suffer significantly from partial occlusion. The partial occlusion condition merges the objects when they are nearby and causes incorrect counting and misclassification of objects. The performance of an object detector slows down to recognize partial occlusion due to visibility features, such as the same shape and colors. Also, when multiple targets in the scene produce frequent occlusions, the handling becomes much more challenging [8]. As a result, an appropriate occlusion reasoning method is required.

Existing methods only predict bounding boxes that need to be closed to ground truth without considering their relationships. As a result, they make the detectors sensitive to the non-maximum suppression (NMS) threshold in situations that are filled with occlusions. To that aim, Wang et al. [9] proposed a repulsion loss that not only drives each proposal toward its intended target but also keeps it away from the other ground truth objects and their corresponding designated proposals. To handle overlapping, it is challenging to control the balance between the repulsion and attraction elements in the loss function. In the literature, various bounding box regression techniques were proposed to localize the

objects efficiently [10, 11] but estimating the orientation of objects has become challenging at the time of occlusion state estimation. A method has been proposed based on the ability of occlusion handling that could be classified into high-accuracy and high-speed types [12]. In high-speed algorithms, architecture is quite simple which makes them efficient for detecting objects without occlusion. A method proposed in [13] discusses the relative distance and occlusion relationship from viewpoint. Relative distance determines the closeness of the two objects and relative occlusion provides the intersection of two overlapped objects. The proposed complicated architecture of occlusion handling is built for high-accuracy methods, which makes them robust for occlusion but unavoidably takes greater computational resources, even for objects without occlusion. However, there is no single solution that works well in all occlusion instances; the optimum approach is determined by the specific occlusion condition and the desired trade-off between object detection accuracy and bounding box localization loss.

With this motivation, we solved the partial occlusion detection challenge by proposing an OBB Detector (Occluded Bounding Box Detector). To identify the orientation of the object is a great challenge. A method has been proposed in the paper [14] to segment the objects with bounding boxes but it merges similar types of objects in the presence of partial occlusion. Our method solves this issue by integrating the region proposal technique with Axis Aligned Bounding Box method which generates the enclosed bounding boxes separately on the detected objects. Moreover, the proposed method can detect the orientation of the detected objects precisely using occlusion prior condition. Many object detectors are not able to analyze the overlapped area of the occluded objects which effects the system's performance [16, 17]. We also explored the functionality of the IoU (Intersection over Union) similarity indexed-based evaluation metric [18] which is used to identify overlapping between two objects with scale-invariant features. The research findings reveal that our method provides good precision under partial occlusion conditions.

The main contributions of this research work are as follows:

- We propose a bounding box generation method by integration of region proposal and Axis Aligned bounding box approach to locate the position of the detected objects. The proposed method is versatile in the generation of enclosed bounding boxes on detected objects precisely for improving localization performance. This method helps in calculating geometric features such as area, size, and centroid of the enclosed bounding box.
- We propose an occlusion prior condition to detect the overlapped area of bounding boxes with vertical and horizontal alignment axis.
- We provide occlusion analysis of partially detected objects by calculating the degree of overlap and analyzing bounding box localization loss (when a misaligned bounding box is detected).
- We compare the performance of the proposed method with state-the-art methods on benchmark datasets with different levels of occlusions.

The proposed research work can generate the bounding boxes under occlusion conditions and reduce the bounding box localization loss. The rest of the paper is organized as follows: Sect. 2 discusses related work. Section 3 presents a detailed description of the proposed methodology. The evaluation of the proposed work and comparative analysis are discussed in Sect. 4. The results discussion is shown in Sect. 5. Section 6 discusses the conclusion with future work.

2 Related work

There are many challenges faced by object detector systems, such as localization errors, counting and classifying wrong objects, illumination changes, occlusion, and many more. A lot of research has been done to solve occlusion problems automatically comparable with the human visual system to achieve high precision of detected objects. In this section, we discuss object detection under occlusion conditions, challenges, and existing solutions. As per the literature, two types of partial occlusion occur: (i) low partial occlusion—less part of the object is overlapped, and (ii) heavy occlusion—larger part of the object is overlapped.

Occlusion occurs when a few parts of the object are not visible due to the presence of another object which cause more false positives. To handle occlusion issues, image segmentation techniques have been implemented in the literature [4, 19] where pixel-wise masks were generated for interested objects, resulting in accurate localization of the objects along with their shape. In image segmentation techniques, feature matters. Various visual representation segmentation

tasks have been considered over the past decade based on shape-based features and appearance-based features [2]. The shape modeling method was proposed for object recognition which is used to describe shapes and predict the topology of an object in advance but fails to identify curved lines under occluded conditions [20]. To overcome this issue, a method has been proposed to detect curved objects precisely with hand-crafted features even if it is occluded [21]. The model-based technique was also used by the authors to address occlusion [22]. Their approach was built on the geometric hashing of the 2D and 3D occluded objects. Multiple objects in obscured scenes could be recognized by the shape-based recognition method. They created their suggested technique using spin image representation for matching surfaces point by point.

On the other side, researchers proposed various methods using appearance-based models to handle occlusion conditions. Scale-invariant feature transform (SIFT) is one of the appearance-based models that generate interest points of the segmented objects and produces scale-invariant features to detect overlapped pixels [23]. Many improvements have been made in the SIFT feature extractor method. The authors used SIFT features with histograms to extract color information for occluded face recognition [24]. The occlusion condition has been detected using similarity-based features by extracting histogram features using the local binary pattern. This method is not able to detect match facial features if appropriate key points are not extracted. To overcome this issue, a structure-aware key point tracking method was proposed which is based on the search space method to generate key points for segmenting interested regions [25]. This method helps in detecting small view parts in crowded scenes. Many researchers have adopted SURF feature descriptors to detect occluded objects. This descriptor is featured with scale invariant and rotation capability to search the objects from the database. A method has been proposed using a symmetric SURF feature to detect the face occlusion parts [26]. A local matching technique called Metric Learned Extended Robust Point Set Matching (MLERPM) uses feature set matching with SIFT and SURF features to solve the partial occlusion problem [27].

Other researchers have developed models to detect pedestrians using Histogram of Gradient (HOG) features. The purpose of this method is to analyze the distribution of edges orientation in the object for shape recognition. A method has been proposed in which the HOG feature descriptor has been combined with the local binary pattern (LBP) feature to detect pedestrians under partial occlusion conditions [28]. Another article has been presented to detect pedestrians in urban driving environments using HOG with the support vector machine (SVM) classifier [29]. The deformable part-based model (DPM) was introduced to detect hidden features of the objects [30]. Based on DPM, a two-person detector

was introduced to handle occlusion [31]. The DPM detector is based on sliding window protocol approach in which filter is applied at all positions of the object. They view the opacity between persons as a quirk rather than as interference. Though, part detectors are now frequently employed to address occlusion problems but have the disadvantage that parts are manually designed, which may not be ideal. Even with heavy or partial occlusion, these detectors can accurately estimate the bounding boxes of two individuals but do not work well in a complex environment.

Identifying the area of the bounding box along with the rotation has become a great advantage in the object detection field [32]. The bounding box helps in finding an overlapping area to calculate the degree of occlusion on two or more detected objects. The method proposed in the paper [33] is an Iterative Bounding Box expansion approach that predicts the amodal mask by repeatedly extending the amodal segmentation approach. Amodal mask provides the complete information about the object included occluded part. Here, the segmentation is done on the basis of threshold value where pixels intensities in the heatmap are greater than the threshold. They computed the area ratio to calculate occlusion. In the literature, Axis Aligned Bounding Box and Oriented Bounding Box techniques have also been discussed for the detecting intersection point of the objects along with their orientation [17, 34, 35]. These methods are useful for collision detection in large amounts of datasets. Table 1 represents the various methods and challenges solved under the occlusion condition.

In the literature, deep learning models have been developed for object detection under occlusion conditions [24, 36–39]. The work presented in [36] shows that deep learning models do not perform well if the dataset is small because the efficiency of these models is increased by more images in the training set. Typically an object detector model includes the whole image rather than its part. Since there is no supervision for locating an object, segmentation results fail to detect the boundary of the objects, leading to localization errors. With this motivation, this research study proposed a model to detect partial occlusion efficiently by reducing bounding box localization loss and false positives to make object detection methods accurate and efficient. Section 3 discusses the proposed method of this research work to solve the partial occlusion problem.

3 Proposed method

In this section, the detailed methodology of the proposed work has been discussed. Occlusion occurs when a few features of one object are not visible due to the existence of another object. The detection of the occluded region depends on the overlapping parts of the generated bounding boxes of multiple objects. Occlusion conditions can be categorized as non-occlusion, partial occlusion, and full occlusion as shown in Fig. 2. If there is no occlusion occurs between the objects, then bounding boxes are not overlapped (refer to Fig. 2a) otherwise bounding boxes are overlapped (refer to Fig. 2b and c) when an occlusion exists.

Partial occlusion generates more error at the time of object detection as most parts of the objects are not visible which creates confusion for the object detector to classify the correct class of the objects if either object is of the same shape or objects are identical. This research work focuses on solving the issue of detecting correct objects under partial occlusion conditions.

This proposed work is divided into two stages: (i) generation of enclosed bounding box for computing geometric features of objects and (ii) detection of occlusion to estimate the percentage (%) of the overlapped area of the bounding boxes. The general framework of the proposed work has been illustrated in Fig. 3. The bounding box generation stage is used to generate an enclosed bounding box of the input frame to detect the region of interest called objects. Firstly, the adaptive background modeling method [40] has been used to generate a binary mask of a given sequence and then the region proposal method with the axis-aligned bounding box (AABB) method are applied to generate the enclosed bounding box on the detected objects. The next step is to check the overlapped bounding boxes for occlusion detection. If multiple bounding boxes are overlapped, they can be identified as occluded objects by using various levels of occlusions otherwise detected objects are considered as non-occluded objects. We also calculate the percentage % of the overlapping area of bounding boxes to understand the visible features of occluded objects.

3.1 Generating enclosed bounding box

Object detection is an important task in computer vision. One of the challenges of object detection is dealing with occlusion—when part of the object is blocked from view by another object or part of the background. In the object detection field, the bounding box plays an important role in detecting the foreground objects in input sequences. The two-dimensional bounding box generates the four coordinates (x_1, y_1, x_2, y_2) based on the geometric features of the objects, such as height and width (refer to Fig. 4). The dataset is associated with a ground truth bounding box whose coordinates are represented by $(x_{G1}, y_{G1}, x_{G2}, y_{G2})$.

The enclosed (tightness) bounding box in geometry specifies the smallest rectangle which covers the whole object in 2D dimension. When objects are partially occluded, sometimes misaligned bounding boxes are formed, resulting in bounding box localization loss and increased false positives.

Table 1 Various methods to detect different types of occlusion

| Occlusion detection models	Approaches used	Detected objects	Type of occlusion
Shape-based model [20, 25]	Pixel-wise mask generation	Persons and Vehicles	Partial occlusion
Appearance-based model [24, 26]	Scale-invariant and similarity-mapping features	Face and Human	Partial and heavy occlusion
Part-based model [7,0]	Deformable part-based feature	Face, human, and vehicles	Partial and heavy occlusion
Bounding box regression model [15, 18, 34]	Intersection over Union-based overlapping detection	Different categories of objects	Partial occlusion

Fig. 2 Various occlusion conditions

(a) Non-occlusion   (b) Partial occlusion   (c) Full occlusion
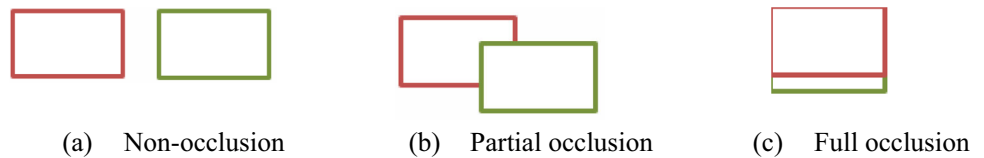
Fig. 3 Block diagram illustrates the working of the proposed work
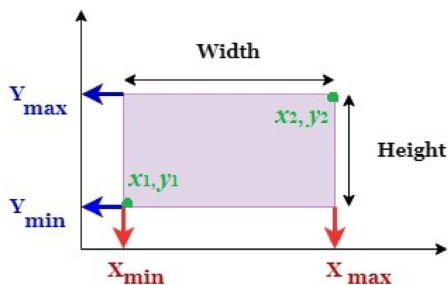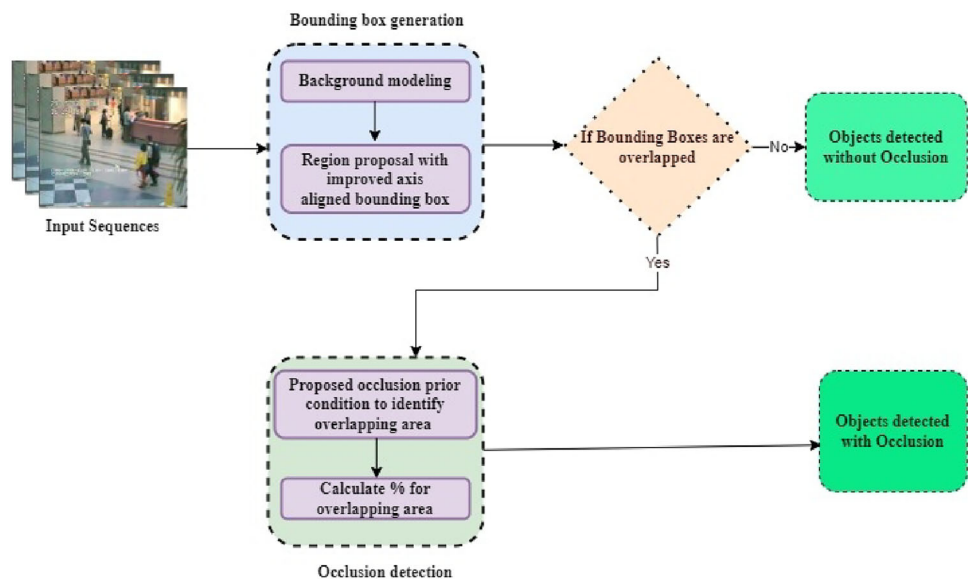
Fig. 4 Generation of coordinates of bounding box

To deal with this issue, we have proposed an integrating approach using region proposal with an axis-aligned bounding box (AABB) to generate enclosed (tight) aligned bounding boxes even in occlusion conditions. The axis-aligned bounding box (or AABB) for a given point set is the minimum bounding box under the condition that the box's edges are parallel to the coordinate axis. The AABB method generates the position vectors that help in computing the size and relative distance of bounding boxes. In the proposed method, the coordinates of the predicted bounding box (x_{P1}, y_{P1}, x_{P2}, y_{P2}) have been generated and then geometric features of the bounding boxes have been computed such as width, height, area, and centroid of the generated bounding box using the following equations where W_{BB} represents the width, H_{BB} represents height, A_{BB} represents the area, and

C_{BB} represents the centroid of the bounding box coordinates.

$$W_{BB} = x_{p2} - x_{p1} \qquad (1)$$

$$H_{BB} = y_{p2} - y_{p1} \qquad (2)$$

$$A_{BB} = W_{BB} \times H_{BB} \qquad (3)$$

$$C_{BB}(x_P, y_P) = \frac{x_{P1} + x_{P2}}{2}, \frac{y_{P1} + y_{P2}}{2} \qquad (4)$$

Similarly, the geometric features of the ground truth bounding box have been computed. After that, the relative distance (R_d) has been computed between the generated and ground truth bounding boxes to find the closest points for generating the tightened bounding box which is near to the ground truth bounding box as shown in Fig. 5. Here, the center-to-center distance method based on Euclidian distance has been used to measure the relative distance between the generated (predicted) bounding box and the ground truth bounding box using Eq. (5).

$$(R_d) \leftarrow \sqrt{(x_P - x_G)^2 + (y_P - y_G)^2} \qquad (5)$$

where $(x_P, y_P) \,\epsilon$ center points of the predicted bounding box and $(x_G, y_G) \,\epsilon$ center points of ground truth bounding box.

To tighten the bounding box around a set of points, we started with the initial coordinates of the predicted bounding box and divided the width and height into half to compute the closest points. Similarly, the midpoints of the ground truth bounding box have been computed. After that, the absolute difference between the center point (x_P) and midpoint $\left(\frac{W_P}{2}\right)$ of predicted bounding boxes have been computed. The absolute difference between center point (x_G) and midpoint $\left(\frac{W_G}{2}\right)$ of ground truth bounding boxes have been computed. Then, the minimum value of the predicted and ground truth result was computed by using Eq. (6) to get the closest points $\left(X_{CP_1}\right)$ of tightening the bounding box.

$$\left(X_{CP_1}\right) = min\{P\left(|x_P - \frac{W_P}{2}|\right), G\left(|x_G - \frac{W_G}{2}|\right)\} \qquad (6)$$

Also, the maximum coordinate value $\left(X_{CP_2}\right)$ of x axis was computed using Eq. (7).

$$\left(X_{CP_2}\right) = max\{P\left(|x_P + \frac{W_P}{2}|\right), G\left(|x_G + \frac{W_G}{2}|\right)\} \qquad (7)$$

Later, the minimal closest point $\left(Y_{CP_1}\right)$ for y-axis was calculated by taking the minimum value of the absolute difference between y_P and $\left(\frac{H_P}{2}\right)$ of the predicted bounding box and the absolute difference between y_G and $\left(\frac{H_G}{2}\right)$ of ground truth bounding box as shown in Eq. (8). Here, the y-axis is related to the height of the bounding box.
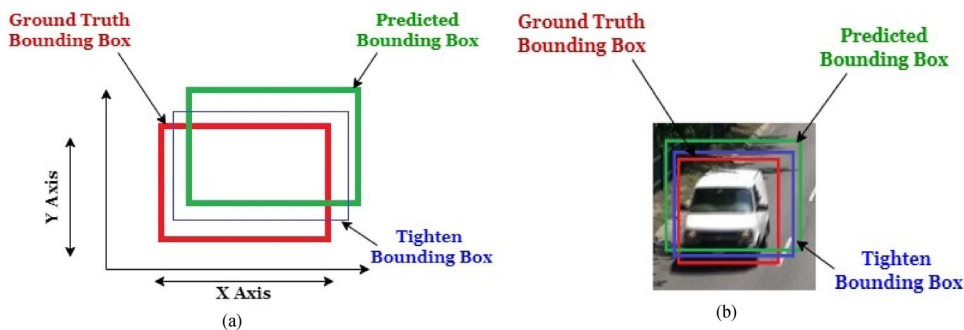
$$\left(Y_{CP_1}\right) = min\left\{P\left(\left|y_P - \frac{H_P}{2}\right|\right), G\left(\left|y_G - \frac{H_G}{2}\right|\right)\right\} \qquad (8)$$

The maximum coordinate value $\left(Y_{CP_2}\right)$ of y axis was calculated using Eq. (9).

$$\left(Y_{CP_2}\right) = max\left\{P\left(\left|y_P + \frac{H_P}{2}\right|\right), G\left(\left|y_G + \frac{H_G}{2}\right|\right)\right\} \qquad (9)$$

Algorithm 1 discusses the generation of the bounding box that is enclosed using the proposed method.

Fig. 5 Formation of enclosed (tightened) bounding box

(a)

(b)

Algorithm 1 Generating enclosed bounding box

Input: Coordinates of predicted and ground truth bounding box
Output: Coordinates of the enclosed bounding box
Procedure:

1. Read the coordinates of predicted and ground truth bounding boxes
2. Calculate the geometric features: width (W_P), height (H_P) and area (A_P) of predicted bounding box
3. Calculate the geometric features: width (W_G), height (H_G) and area (A_G), of ground truth bounding box
4. Calculate the centroid of the predicted bounding box

$$C_{BB}(x_P, y_P) \leftarrow \frac{x_{P1} + x_{P2}}{2}, \frac{y_{P1} + y_{P2}}{2}$$

5. Calculate the centroid of the ground truth bounding box

$$C_{BB}(x_G, y_G) \leftarrow \frac{x_{G1} + x_{G2}}{2}, \frac{y_{G1} + y_{G2}}{2}$$

6. Calculate the relative distance between the predicted and ground truth bounding box

$$R_d \leftarrow \sqrt{(x_P - x_G)^2 + (y_P - y_G)^2}$$

7. Calculate the coordinates of the enclosed bounding box:

   7.1. Find the closest points (CP) between the predicted and ground truth bounding box

$$(X_{CP_1}) \leftarrow min\{P\left(|x_P - \frac{W_P}{2}|\right), G\left(|x_G - \frac{W_G}{2}|\right)\}$$
$$(X_{CP_2}) \leftarrow max\{P\left(|x_P + \frac{W_P}{2}|\right), G\left(|x_G + \frac{W_G}{2}|\right)\}$$
$$(Y_{CP_1}) \leftarrow min\{P\left(|y_P - \frac{H_P}{2}|\right), G\left(|y_G - \frac{H_G}{2}|\right)\}$$
$$(Y_{CP_2}) \leftarrow max\{P\left(|y_P + \frac{H_P}{2}|\right), G\left(|y_G + \frac{H_G}{2}|\right)\}$$

   7.2. Compute the new coordinates using closest points

$$(x_{\min\_new}, x_{\max\_new}) \leftarrow \{X_{CP_1}, X_{CP_2}\}$$
$$(y_{\min\_new}, y_{\max\_new}) \leftarrow \{Y_{CP_1}, Y_{CP_2}\}$$

8. Calculate the area of new coordinates:

   8.1. Compute the width (W_{new_BB}) and height (H_{new_BB})   //using new coordinates

   8.2. Compute area (A_{new_BB}) $\leftarrow W_{new\_BB} \times H_{new\_BB}$ //using new coordinates

   8.3. **if** ($A_{new\_BB} < A_P$) $\leftarrow$ ***True***

   RETURN ($x_{\min\_new}, x_{\max\_new}, y_{\min\_new}, y_{\max\_new}$) // Coordinates of ground truth bounding box

   else   RETURN ($x_{P1}, y_{P1}, x_{P2}, y_{P2}$)   // Coordinates of Predicted bounding box

---

As per Algorithm 1, first, select the coordinates of predicted and ground truth bounding boxes. Then, the geometric features such as width, height, area, and center points of predicted and ground truth bounding boxes were computed. After that, the relative distances have been calculated to find the closest points (as shown in steps 4 and 5). Later, the area of new primitive coordinates was computed. The area of the closest points has been compared with the area of the predicted bounding box and the lesser value has been considered for the closest points (as shown in step 7). After computing the enclosed bounding box, the next section discusses the detection of the overlapped area of the objects under occlusion condition.

3.2 Detection of the occluded area

In this research work, the static camera with a single view point is considered to initialize the object detection system. The direction from which an object is viewed plays an important role in the detection of occlusion state. An object might be partially or fully occluded depending on the viewer's perspective. Once the object is detected, the algorithm for
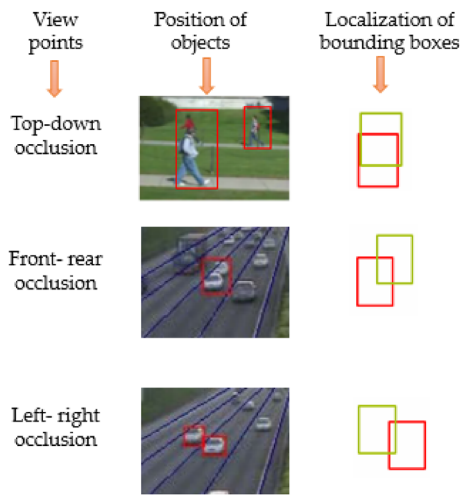
Fig. 6 Different viewpoints of the occlusion detection model

detecting occlusion begins to start under the assumption that the detection method works well. The main challenge is how frequent occlusion conditions occur and are identified in different views. In the literature, many methods are available to detect occlusion states in different views but these methods are not scalable and require different setups for each view point.

Viewpoint is an important factor in occlusion, as it can significantly impact the level and type of occlusion that occurs. When an object is partially or fully occluded, the degree of occlusion can depend on the viewer's perspective or viewpoint. An object may appear fully occluded from one viewpoint, but only partially occluded from another viewpoint. Many methods learn the occlusion pattern from the captured data and require a separate model to detect occlusion for each view point which increases computation time[41, 42]. To overcome this issue, the proposed method detects different levels of occlusion for three view points such as (i) Top-down view, (ii) Front-rear view, and (iii) Left–right view as shown in Fig. 6, considering a single viewpoint at a time using a single static camera. The top-down view represents the positions of the detected objects in the vertical direction and similarly, bounding boxes are formed. In front-rear view, one object is behind the other object while in left–right view, objects are viewed in left–right direction. Our method does not require information on camera's statistics such as position, angle, and type of camera. It requires only geometric information about objects, such as height, width, and area.

Before applying the occlusion detection method, an assumption has been made that the static camera is considered with a particular view point. After generating the coordinates of the enclosed bounding box (using Algorithm 1), the next step is to identify the occluded region if exists. The occluded area of two bounding boxes is calculated using the intersection area. To achieve this aim, we incorporated

a diagonal direction feature in the proposed Axis Aligned Bounding Box to compute the intersection area for occlusion detection (considering low diagonal orientation). Let $x_i = \{x_1, x_2, x_3 \ldots \ldots x_n\}$ be a set of bounding box pixels considered visible parts of detected objects. The occlusion prior condition based on occlusion ratio $P(x_i|O_c)$ is calculated by identifying the pixels x_i on detected bounding boxes that are being occluded O_c To measure the condition of occlusion, first, we get the geometric features of enclosed bounding boxes such as height H_{Obj} and width W_{Obj} from Algorithm 1 then compute the projected width $\widehat{w}$, relative distance d from the closest to the farthest edge of the object, and projected height $\widehat{h}$ of the occluded region as illustrated in Eqs. (10), and (11), respectively.

$$\widehat{w} = W_{Obj}.\cos(\theta) + l.\sin(\theta) \tag{10}$$

$$\widehat{h} = H_{Obj}.\cos(\psi) + d(\theta).sin(\psi) \tag{11}$$

where $\theta \in [0, 2\pi]$ assumes all rotations around the vertical axis and l is the length of the object. H_{Obj} is the height of the object and ψ is the elevation angle. The elevation angle is calculated between two center points ($C_{B_{ix}}$, $C_{B_{iy}}$) of generated bounding boxes based on the horizontal axis. The elevation angle helps in deriving the orientation of bounding boxes.

$$\psi = \arctan(C_{B_{ix}}, C_{B_{iy}}) \tag{12}$$

Then, the projected dimensions $(\widehat{w}, \widehat{h})$ are used to derive occlusion prior conditions to check whether the bounding boxes are overlapped or not. The formation of the occluded bounding box and their orientation have been shown in Fig. 7 where the red color area represents the occluded part between two bounding boxes.

The proposed method generates the occlusion priori conditions to detect the overlapped part of the enclosed bounding boxes. To generate the bottom-left coordinates (x_left, y_bottom), the minimum value of the x and y coordinates of the two tighten bounding boxes have been calculated using Eqs. (13) and (14).

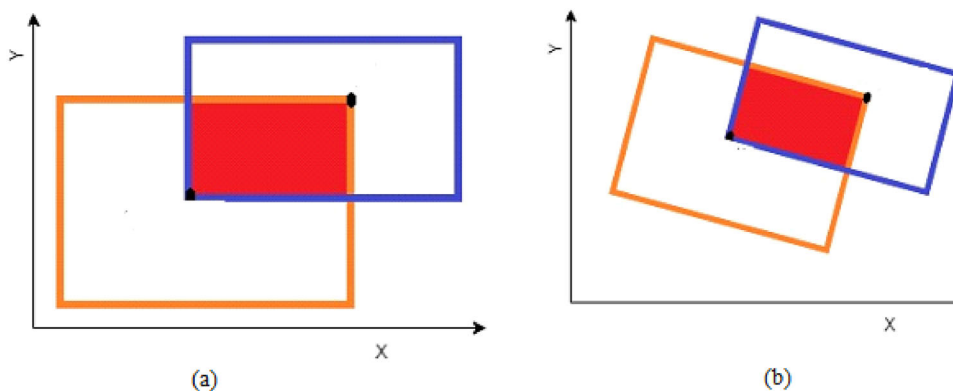$$x\_left = \min(B_1(x), B_2(x)) \tag{13}$$

$$y\_bottom = \min(B_1(y), B_2(y)) \tag{14}$$

Similarly, the maximum values of the x and y coordinates of the two tighten bounding boxes have been considered to generate top-right coordinates (x_right, y_top) by using Eqs. (15) and (16).

$$x\_right = \max(B_1(x), B_2(x)) \tag{15}$$

Fig. 7 Formation of overlapped bounding boxes and their orientation

(a)   (b)

$$y\_top = \max(B_1(y), \, B_2(y)) \tag{16}$$

Algorithm 2 represents the prediction of an occluded area of intersecting bounding boxes.

3.3 Different levels of occlusion and prior likelihood

In this section, we discuss different levels of partial occlusions of detected bounding boxes and occlusion prior likeli-

---

Algorithm 2 Prediction of the intersection area of two bounding boxes

---

Input: List of bounding boxes $B_1 \ and \ B_2$
Output: Coordinates and area of overlapped bounding box
Procedure:

1. Read the geometric features of tightened bounding boxes   //computed from Algorithm 1
2. Calculate the rotation angle of tightened bounding box

$$\psi = \arctan(C_{B_1}, C_{B_2}) \cdot \frac{180}{\pi}$$

3. Compute the coordinates of the overlapped area of bounding boxes
   a. x_left = $\min(B_1(x), B_2(x))$
   b. y_bottom = $\min(B_1(y), B_2(y))$
   c. x_right= $\max(B_1(x), B_2(x))$
   d. y_top = $\max(B_1(y), B_2(y))$
4. RETURN overlapped rectangle coordinates: (x_left, y_top, x_right, y_bottom)
5. Compute the occlusion condition
   If (x_right < x_left || y_bottom < y_top) **then**
      Intersection occurs (Occlusion exists)
   else No Occlusion exists
6. Calculate the intersection area
   $A_I = $ (x_left $-$ x_right) $\times$ (y_top $-$ y_bottom)
7. Calculate occlusion ratio $P(x_i|O_c)$

   Occlusion ratio $= \frac{A_I}{A_1 + A_2 - A_I}$ // where A_1 = Area of predicted bounding box 1 and
   A_2 = Area of predicted bounding box 2

In Algorithm 2, the geometric features of bounding boxes have been stored and then the rotation of bounding boxes has been stored. From step 3, the occlusion prior condition was calculated to know the pixel values that come under the occlusion condition by calculating the coordinates of bounding boxes. In step 5, the occlusion prior condition was checked for occlusion existence, and then the area of an overlapped rectangle for occlusion analysis was calculated. This computation helps in identifying the various levels of occlusion.

hood for analyzing the performance of the proposed method. When objects are highly occluded, it is difficult to get the correct information about objects that are not in the scope of this research. More false positive rates are increased when some parts of the objects are not visible which leads to the misclassification of objects. In partial occlusion condition, the visible features must be specified which helps an object detection method to recognize the objects efficiently. The proposed method is enriched with occlusion prior likelihood

Table 2 Ranges to identify different levels of occlusion [14]

Occlusion levels	Overlapping range	Visibility criteria
No occlusion	0–10%	Objects are visible
Low-level occlusion	10–20%	Objects are visible
Partial occlusion	20–40% and 40–70%	Object visibility is less and leads to more false positives and localization error
Heavy occlusion	> 70%	Object visibility is very less

Table 3 Ground truth frames and test frames under occlusion of datasets

Datasets	Ground truth frames	Test frames under different ranges of occlusion level	
		20–40%	40–70%
Highway	100	30	25
PETS2006	1200	200	256

to check the visibility part and intersection part of bounding boxes. Equation (17) is used to generate a conditional likelihood function

$$P(x_i|\lambda, O_c) = \int P(x_i|\lambda, O_c, Z)P(Z|\lambda, O_c)dZ \qquad (17)$$

Here, $P(Z|\lambda, O_c)dZ$ is the conditional probability of visible data x_i for occlusion condition parameters λ and Z is an occluded variable parameter. The integral over Z represents the process of marginalizing over the occluded variables, which allows us to compute the likelihood of the observed data given the model parameters and occlusion condition.

To estimate the visibility of objects in an occlusion state, we used the variable x_i which gives all the information about visibility regions under occlusion conditions. Then, the approximation likelihood $P(V_i|\lambda, O_c)$ were computed using Eq. (18) for the entire visibility region V_i where $x_i \in V_i$

$$P(V_i|\lambda, O_c) \simeq P\big(V_i|(V_j)^*, O_c\big) \qquad (18)$$

$$V_j^* = argmax\, P\big(V_i|V_j, O_c\big) \qquad (19)$$

In this research work, different levels of occlusion were considered based on the overlapping area of bounding boxes[14]. Various levels of occlusion help in identifying how much an object is occluded that information can be associated with an object detection model to understand the object visibility. Table 2 discusses the visibility parts and different levels of occlusion. In our research work, we consider the range of overlapping areas of detected bounding boxes from 20 to 40% and 40% to 70% under partial occlusion conditions to improve localization error of bounding boxes. The overlapping area of (10–20%) can be omitted as it does not impact the performance of the object detection model.

The experimental results of the proposed work based on different levels of occlusion have been discussed in the next section.

4 Experimental setup and results analysis

In this section, we present the results of our experiments evaluating the proposed method to detect partially occluded objects in video sequences. We first describe the experimental setup used, followed by an analysis of the results. The method was implemented on MATLAB 2018Rb version, Intel Core i6 processor with 16 GB RAM.

4.1 Datasets used

We have used two benchmark datasets (i) Highway and (ii) PETS 2006 to detect overlapped areas under partial occlusion conditions [43]. The Highway dataset is a widely used dataset that contains a sequence of 1700 frames, which was captured from a camera mounted on a highway. The frames contain a continuous traffic flow, and the goal of change detection is to identify any changes or anomalies in the traffic flow. Another dataset is PETS 2006 which contains video sequences of walking pedestrians, captured from a camera-mounted outdoor environment. The frames contain a group of people walking, and the goal is to identify any unusual behavior or events. The ground truth annotations of both datasets for a few frames provide the position of the object in the captured scene. The ground truth values help in evaluating the performance of object detection approaches based on various metrics, such as accuracy, robustness, and efficiency. In both Highway and PETS 2006 datasets, ground truth data of 100 frames and 1200 frames are available with bounding box coordinates values, respectively. We considered a few frames of both datasets for experimentation purposes under different levels of occlusion (as shown in Table 3). The availability and standardization of discussed datasets make them a valuable resource for researchers and practitioners in the field of computer vision for surveillance applications. In our experimentation, we have considered occlusion ranges from 20–40% to 40–70% for the given datasets. The occlusion ranges between 0 and 20% and does not impact the performance of object detectors.

4.2 Occlusion evaluation metrics

Occlusion evaluation metrics are used to measure the effectiveness of algorithms designed to detect and handle occlusions in images or videos. We used the following evaluation matrices for the experimentation as follows:

(i) *Intersection over Union (IoU)* in an evaluation metric to calculate the amount of overlap or intersection area between two bounding boxes B_1 and B_2. The intersection area is calculated by finding the coordinates of the intersection region and computing its area using a geometric formula using Eq. 20

$$IoU = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|} \tag{20}$$

where B_1 and B_2 are bounding boxes. The IoU range should lie from 0 to 1 for detected bounding boxes where 0 represents no occlusion and 1 represents complete occlusion. IoU loss (L_{IoU}) is also calculated to understand the correct prediction of bounding boxes.

$$L_{IoU} = 1 - IoU \tag{21}$$

(ii) *Generalized Intersection over Union (GIoU)* is an extension of the IoU metric that takes into account the size and location of the bounding boxes. GIoU measures the distance between the bounding boxes B_1 and B_2 as well as the overlap between them, and then normalizes the result.

$$GIoU = IoU - \frac{|C - (B_1 \cup B_2)|}{|C|} \tag{22}$$

where C is an enclosed area of bounding boxes. The range of GIoU becomes -1 to 1 from non-occlusion to occlusion respectively. Then, GIoU loss (L_{GIoU}) is calculated to know the correct localization of bounding boxes.

$$L_{GIoU} = 1 - IoU + \frac{|C - (B_1 \cup B_2)|}{|C|} \tag{23}$$

(iii) *Complete Intersection over Union (CIoU)* is another evaluation metric that is used to identify the correct localization and size of the bounding boxes by considering three geometric factors such as aspect ratio (V), overlapped area, and central point distance. This metric also helps in evaluating the diagonal position of bounding boxes.

$$CIoU = IoU + \frac{\rho^2(B_1, B_2)}{c^2} + \alpha V \tag{24}$$

where c is the central point of an enclosed bounding box, ρ is the Euclidean distance, α is a trade-off parameter and V is

Occlusion detection at Front-rear view

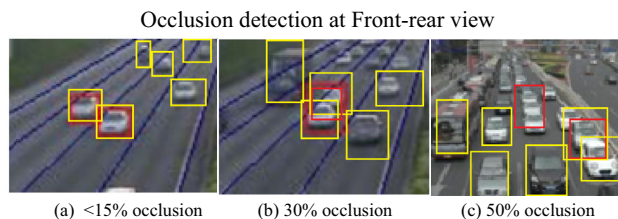(a) <15% occlusion   (b) 30% occlusion   (c) 50% occlusion

Fig. 8 Qualitative representations of the proposed method at the front-rear view. **a** represents 423rd frame of the Highway dataset at left-right view with < 15% occlusion, **b** represents 396th frame of the Highway dataset at front-rear with 30% occlusion **c** represents 200th frame of the PETS 2006 dataset captured at top-down view with 50% occlusion
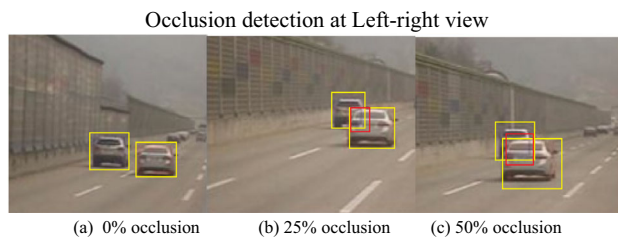
Occlusion detection at Left-right view

(a) 0% occlusion   (b) 25% occlusion   (c) 50% occlusion

Fig. 9 Qualitative results of the proposed method on the Highway dataset at left side view along with different levels of occlusion. **a** represents 350th frame with 0% occlusion **b** represents 387th frame with 25% occlusion **c** represents 398th frame with 50% occlusion

the aspect ratio.

$$V = \frac{4}{\pi^2}(archtan\frac{w1}{h1} - archtan\frac{w2}{h2})^2 \tag{25}$$

$$\alpha = \begin{cases} 0, \ if \ IoU < 0.5 \\ \frac{V}{(1-IoU)+V}, \ if \ IoU \geq 0.5 \end{cases} \tag{26}$$

CIoU metric minimizes the distance between two bounding boxes and converges faster than GIoU. The range of CIoU is 0 to 1 from non-occlusion to occlusion conditions, respectively. Then, CIoU loss (L_{CIoU}) is calculated using the following equation

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(B_1, B_2)}{c^2} + \alpha V \tag{27}$$

4.3 Qualitative results of the proposed method

The proposed method for detecting occlusion is evaluated in this section, considering different views and levels of occlusion. The results are presented through visual representations in Figs. 8, 9, and 10. The detected objects are highlighted with a yellow bounding box, while the occluded areas are marked with a red bounding box. The levels of occlusion discussed in Sect. 3.3 are used as a basis for the evaluation.

Fig. 10 Qualitative results of the proposed method on PETS2006 dataset at Top-down view along with different levels of occlusion. **a** represents 56th frame with 0% occlusion, **b** represents 90th frame with 22% occlusion, **c** represents 128th frame with 65% occlusion
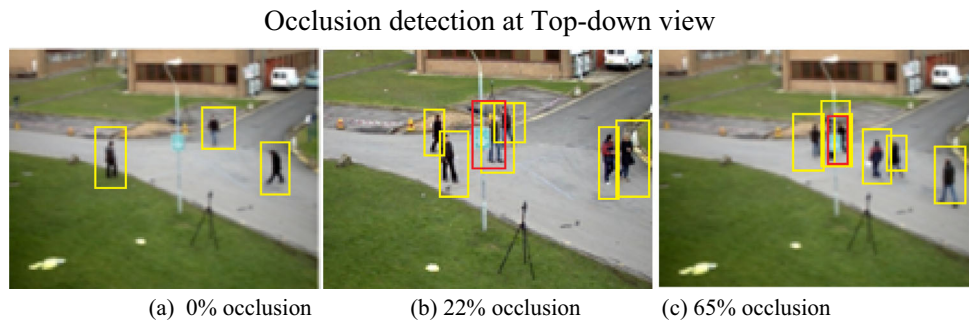
Occlusion detection at Top-down view

(a) 0% occlusion   (b) 22% occlusion   (c) 65% occlusion

Fig. 11 Qualitative representations of the proposed method on the Highway dataset at front-rear viewpoint along under no occlusion and occlusion case. **a** represents 200th frame with no occlusion, **b** represents 256th frame with no occlusion, **c** represents 287th frame with occlusion
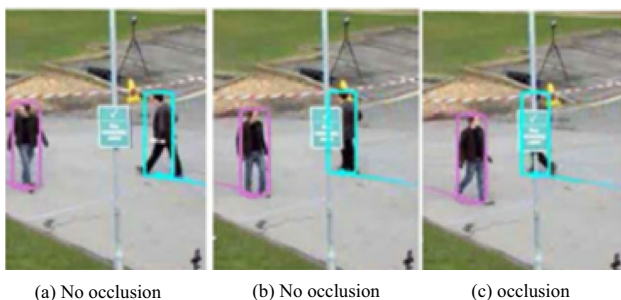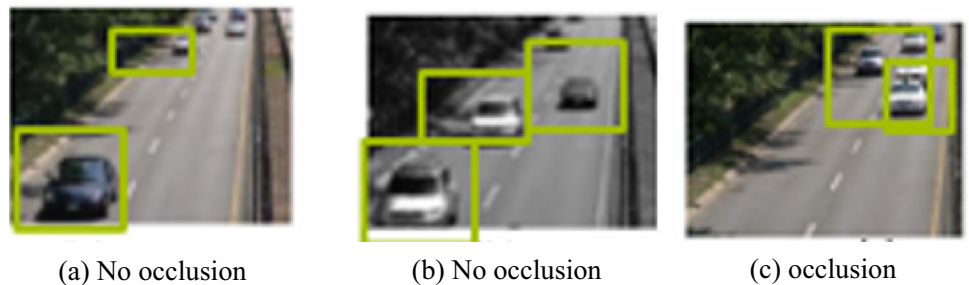

(a) No occlusion   (b) No occlusion   (c) occlusion

(a) No occlusion   (b) No occlusion   (c) occlusion

Fig. 12 Qualitative representations of the proposed method on the Highway dataset at the front-rear viewpoint along under no occlusion and occlusion case. **a** represents 102nd frame with no occlusion, **b** represents 106th frame with no occlusion, **c** represents 115th frame with occlusion

4.4 Quantitative analysis of the proposed method

In this section, we have employed three evaluation metrics, namely IoU, GIoU, and CIoU to assess the statistics of pixel values under partial occlusion conditions. Although IoU may not be adequate in determining the intersection area of two arbitrary shapes' bounding boxes, we have also considered GIoU and CIoU metrics to address this issue. We have presented the results of our evaluation in Figs. 11 and 12 for Highway and PETS 2006 datasets, respectively, showing varying levels of occlusion. The quantitative analysis has been analyzed for the same and shown in Table 4. In the Highway dataset, the object is not occluded in the 200th frame so the quantitative value of the IoU metric is 0. Based on IoU value, GIoU, and CIoU evaluation metrics have been computed. In the 256th frame, the boundaries of detected

bounding boxes are near due to the IoU value being minimal at 0.01 but GIoU and CIoU values are in the range of non-occlusion. While there is occlusion occurring in frame no. 287th with an IoU value of 0.551 which shows a 55.1% occlusion rate. Based on the IoU metric, GIoU, and CIoU values have been calculated. The analysis reveals that GIoU and CIoU are scale-invariant, and hence, the generated values are lower than IoU to identify the precise location of bounding boxes. In the PETS2006 dataset, there are no overlapped objects in the 102nd frames while the 106th frame occurs with 46% occlusion area and the 115th frame occurs with 65.7% occlusion. Overall, this evaluation highlights the effectiveness of GIoU and CIoU metrics in accurately aligning bounding boxes under partial occlusion conditions.

The experimental results show that the Intersection over Union (IoU) metric for occlusion detection has a zero gradient in non-overlapping cases. This zero gradient adversely affects the convergence rate of object detection methods. On the other hand, the Generalized Intersection over Union (GIoU) and Complete Intersection over Union (CIoU) have gradients in both overlapping and non-overlapping cases. Therefore, GIoU and CIoU are the lower bounds of IoU, as observed in Table 4 for the occlusion state, due to their ability to handle overlapping and non-overlapping cases with nonzero gradients.

4.5 Analysis of bounding box localization loss of the proposed method

The evaluation of the accuracy of object detection systems involves the use of Bounding Box Localization loss (BBL)

Table 4 Occlusion analysis on Highway dataset and PETS2006 dataset of the proposed method

Datasets	Frame sequences	Occlusion			Non-occlusion		
		IoU	GIoU	CIoU	IoU	GIoU	CIoU
Highway Dataset	200th frame	---	---	---	0	− 1	− 1
	256th frame	---	---	---	0.01	− 1	− 1
	287th frame	0.551	0.549	0.547	---	---	---
PETS 2006 Dataset	102nd frame	---	---	− −	0	− 1	− 1
	106th frame	0.460	0.457	0.458	---	---	---
	115th frame	0.657	0.624	0.621	---	---	− −

--- represents no value for that frame

metrics. The BBL loss helps in identifying the correct localization of objects by comparing predicted and ground truth bounding boxes to detect true positives and false positives. In this study, three types of loss functions, namely L_{IoU}, L_{GIoU}, and L_{CIoU}, were considered to evaluate the accuracy of the proposed system [44]. The BBL loss is directly influenced by the threshold values (δ) of overlapped detection, which is a measure of detecting the overlapped amount of bounding boxes. This value is determined by calculating the loss of overlaps at different IoU threshold values, allowing for an accurate assessment of the system's performance. Here, 10 different IoU threshold values ranging from 0.25 to 0.70 have been used for both the Highway and PETS 2006 datasets to calculate the bounding box prediction loss for overlapped detection. These measurements were taken under various levels of occlusion based on different threshold values and are presented in Tables 6 and 7, respectively. The relative improvement in loss is evaluated using the L_{IoU} value, which represents the percentage reduction in the loss value. A positive value indicates that L_{GIoU} and L_{CIoU} are better than L_{IoU}, while a negative value indicates that L_{IoU} is better than the other losses. A larger positive value suggests a greater relative improvement of L_{GIoU} and L_{CIoU} over L_{IoU}.

Table 5 shows that under 65% occlusion, the proposed method generates a relative loss improvement of 14.2% in GIoU loss and 17.6% in CIoU loss. However, at threshold value δ = 0.35, the proposed method generates a negative relative loss result indicating misalignment of the detected bounding box. Additionally, the percentage of relative loss at different threshold values was compared and then the average loss values were evaluated. It has been observed that the average loss value for L_{GIoU} is 17.1% which indicates a good prediction of bounding boxes. In the Highway dataset, some frame sequences were positioned diagonally, and the enclosed bounding box area and diagonal distance were calculated to compute the L_{CIoU}. We observed a

relative improvement for L_{CIoU} value from 50 to 70% occlusion, demonstrating the ability to detect optimized bounding boxes.

Table 6 represents the outcomes of evaluating the proposed approach with diverse threshold values across different occlusion levels. The results indicate that the proposed method generates a relative improvement in CIoU loss by an average of 17.3% when 65% occlusion is present. To estimate the loss prediction accurately, we used the Average loss and observed that CIoU loss has been improved by 17.6% for the PETS2006 dataset. We arrived at this value by comparing the percentage of relative loss across various threshold values.

We have also computed the correlation between different occlusion threshold values and relative loss improvement as shown in Fig. 13 for the Highway dataset and PETS 2006 dataset.

The graphical representation demonstrates that loss has been improved when an object is 60% to70% occluded. The research finding reveals that the proposed method can detect objects correctly for up to 65% occlusion. The BBL loss for the Highway dataset generates a negative value at the occlusion threshold of 0.35 which states that the predicted bounding box is not properly aligned at the time of occlusion detection. In this research work, we have compared the proposed method with existing methods to test the performance of the proposed system.

4.6 Comparison with state-of-the-art methods

We tested the number of frames (as discussed in Table 3) under different levels of occlusion and compared the average occlusion analysis of the proposed method with four state-of-the-art methods (i) CBOT[19], (ii) OIMOD[29], (iii) SAOD[3], and (iv) AWOD[17]. The chosen state-of-the-art methods are based on the latest research and utilize the most advanced algorithms to improve the accuracy of the occlusion detection method. The comparison has been done based on the qualitative and quantitative results analysis.

Table 5 Quantitative comparison of BBL loss on different overlapped threshold values (δ) for proposed method on Highway dataset

BBL Loss	$\delta = 0.25$	$\delta = 0.30$	$\delta = 0.35$	$\delta = 0.40$	$\delta = 0.45$	$\delta = 0.50$	$\delta = 0.55$	$\delta = 0.60$	$\delta = 0.65$	$\delta = 0.70$	Avg_{Loss}
IoU	0.231	0.286	0.342	0.385	0.438	0.492	0.537	0.583	0.629	0.673	0.45
L_{IoU}	0.019	0.014	0.008	0.015	0.012	0.008	0.013	0.017	0.021	0.027	0.015
L_{GIoU}	0.015	0.013	0.006	0.012	0.009	0.007	0.010	0.014	0.018	0.022	**0.012**
$RI\%$	0.163	0.071	0.250	0.20	0.250	0.050	0.230	0.176	0.142	0.185	**0.171**
L_{CIoU}	0.016	0.012	0.010	0.011	0.010	0.006	0.009	0.012	0.017	0.021	0.012
$RI\%$	0.157	0.135	− 0.25	0.266	0.166	0.187	0.253	0.247	0.176	0.222	0.156

RI Relative Improvement, δ overlapped threshold value. Bold values depict better results. A negative value represents a misalignment of the bounding box

Table 6 Quantitative comparison of BBL loss on different overlapped threshold values (δ) for proposed method on PETS 2006 dataset

BBL Loss	$\delta = 0.25$	$\delta = 0.30$	$\delta = 0.35$	$\delta = 0.40$	$\delta = 0.45$	$\delta = 0.50$	$\delta = 0.55$	$\delta = 0.60$	$\delta = 0.65$	$\delta = 0.70$	Avg_{Loss}
IoU	0.229	0.277	0.327	0.375	0.427	0.479	0.531	0.588	0.627	0.674	0.453
L_{IoU}	0.021	0.023	0.023	0.025	0.023	0.021	0.019	0.012	0.023	0.026	0.021
L_{GIoU}	0.019	0.020	0.018	0.024	0.019	0.017	0.018	0.009	0.020	0.023	0.018
$RI\%$	0.095	0.100	0.217	0.032	0.173	0.161	0.021	0.250	0.130	0.088	0.124
L_{CIoU}	0.018	0.020	0.017	0.021	0.017	0.016	0.016	0.010	0.019	0.022	**0.017**
$RI\%$	0.142	0.130	0.239	0.160	0.243	0.219	0.131	0.167	0.173	0.153	**0.176**

RI Relative Improvement. Bold values depict better results
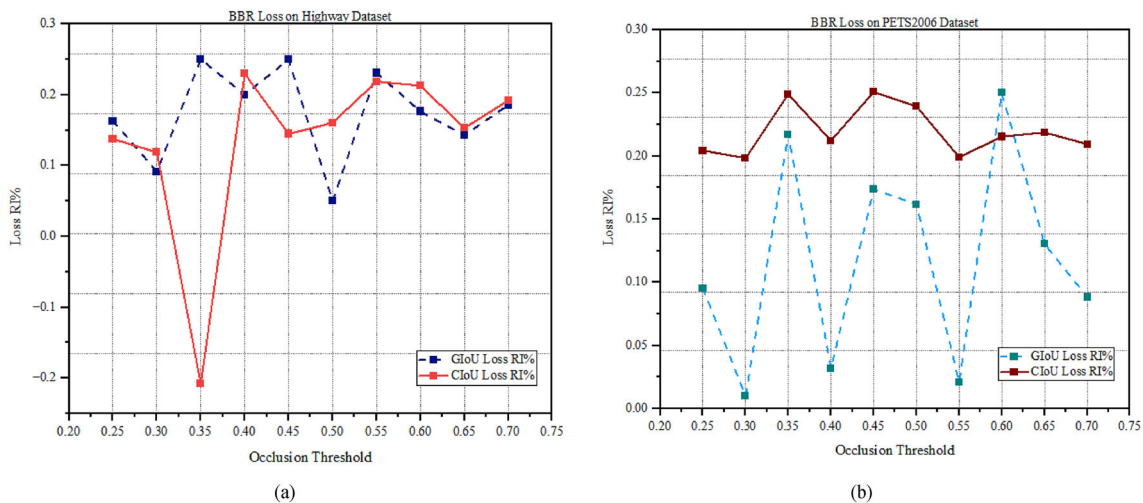
Fig. 13 Relative loss improvement (RI) of the proposed method on different occlusion threshold values. **a** Highway dataset **b** PETS2006 dataset

4.6.1 Qualitative results

The qualitative comparative results of the Highway dataset and the PETS 2006 dataset have been shown in Figs 14 and 15, respectively. The results were compared based on the ground truth frame. (See the Appendix section). The red color bounding box represents the occluded part as given in the ground truth frame. After experimentation, it has been observed that the CBOT and OIMOD methods generate only an occluded bounding box (purple color box) which is not properly aligned and generates a large level of occlusion. The OIMOD method detects only one object if the occluded area is 45% to 50% while the SAOD method generates the bounding boxes (purple color) on detected objects. This method calculates the amount of overlap without overlapped bounding box area which sometimes causes the occurrence of false
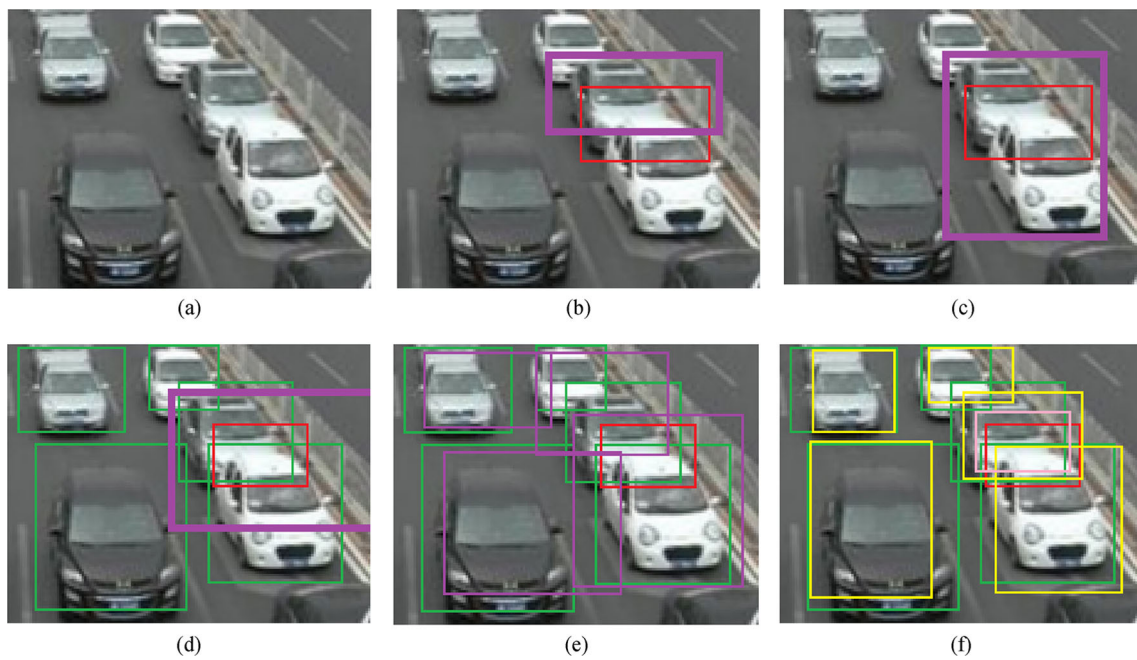
Fig. 14 Qualitative comparison of the proposed method with state-of-the-art methods on the 209th frame of the Highway dataset. **a** Input frame, **b** CBOT method, **c** OIMOD method, **d** SAOD method, **e** AWOD method, **f** Proposed method. Appendix section shows ground truth image
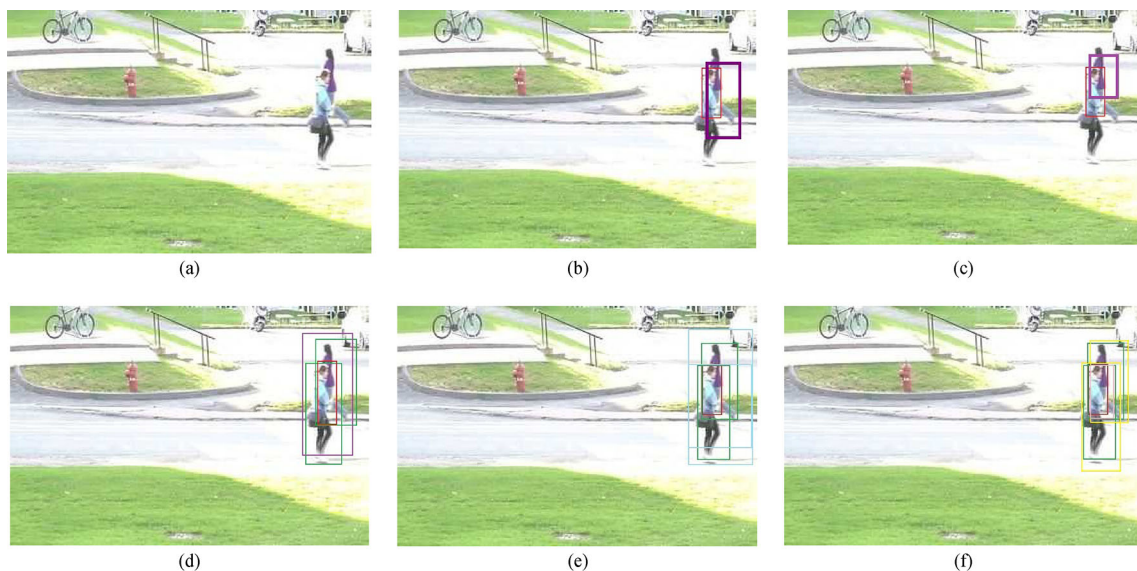

Fig. 15 Qualitative comparison of the proposed method with state-of-the-art methods on the 429th frame of the PETS2006 dataset. **a** Input frame, **b** CBOT method, **c** OIMOD method, **d** SAOD method, **e** AWOD method, **f** Proposed method. The appendix section shows the ground truth image

positives. The proposed method can generate the bounding boxes (yellow color) on detected objects as well as the bounding box (pink color) on the occluded area.

4.6.2 Quantitative results

This section discusses quantitative comparative results that examine object localization and occlusion analysis based on bounding boxes. The experiment assumes that the occluded area and enclosed bounding boxes have non-negative values to produce the experimental results. This study compares the

Table 7 Comparative results of average occlusion analysis

| Datasets | Methods | Average Occlusion (%) | | | | | |
| | | 20–40% | | | 40–70% | | |
		Avg_{IoU}	Avg_{GIoU}	Avg_{CIoU}	Avg_{IoU}	Avg_{GIoU}	Avg_{CIoU}
Highway	CBOT [19]	47.80	38.07	40.64	58.11	43.27	45.37
	OIMOD [29]	32.65	35.32	33.8	57.49	49.1	48.5
	SAOD [3]	33.21	26.04	20.7	52.3	55.2	48.9
	AWOD [17]	39.28	25.41	30.21	56.2	49.64	49.23
	Proposed	**37.84**	**36.33**	**36.05**	**60.8**	**57.06**	**56.45**
PETS2006	CBOT [19]	40.32	35.01	48.9	55.36	43.28	47.7
	OIMOD [29]	42.6	29.56	37.6	62.1	57.1	58.5
	SAOD [3]	37.9	29.0	38.9	54.5	45.8	**52.7**
	AWOD [17]	*40.7*	*38.1*	*37.6*	*62.1*	*59.4*	*57.5*
	Proposed	**39.5**	**39.02**	**38.9**	**65.7**	**65.2**	**64.7**

Bold represents the best result. Italics represents the second best result

Average Occlusion Analysis of Highway Dataset
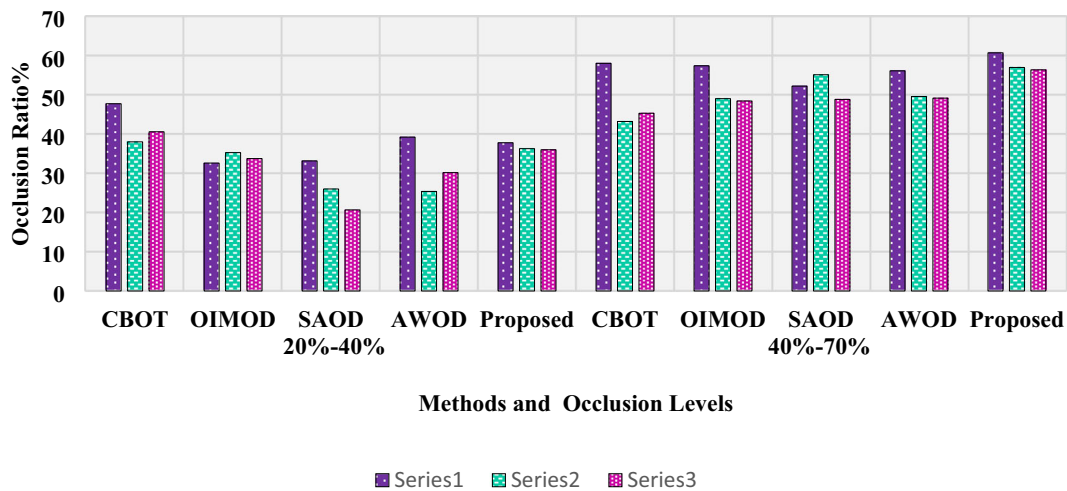
Fig. 16 Comparative representation of average occlusion analysis on Highway dataset

proposed method to state-of-the-art techniques using different levels of occlusion and shows that the proposed method can detect an average overlapped amount of 37.84% for the Highway dataset and 39.5% for the PETS 2006 dataset when occlusion occurs in between 20 and 40%. The proposed method detects a 60.8% overlapped amount for the Highway dataset and 65.7% for the PETS2006 dataset under the occlusion range of 40% to 70%. We have used GIoU and CIoU occlusion evaluation metrics and observed that the Avg GIoU and Avg CIoU values are closed to the Avg IoU value, which indicates the effectiveness of the proposed method. Other comparative methods produce wide variations in Avg GIoU and Avg CIoU compared to Avg IoU values, indicating misalignment of bounding boxes on the occluded

area. In Table 7, the Avg CIoU value is similar to the Avg IoU of the proposed method because of the similar geometry of the occluded bounding box. The proposed method can detect 65% occlusion and the number of objects that lead to correct localization of the bounding boxes. Although the AWOD method generates better occlusion values in terms of Avg GIoU and Avg CIoU, it is still lower than the proposed method due to a less enclosed occluded bounding box. The graphical comparison for occlusion analysis of the proposed method and other methods are shown in Figs. 16 and 17.

Table 8 presents a comparison between the proposed method and state-of-the-art methods regarding the average precision (AP) based on the loss functions used for bounding box localization, specifically under occlusion conditions.
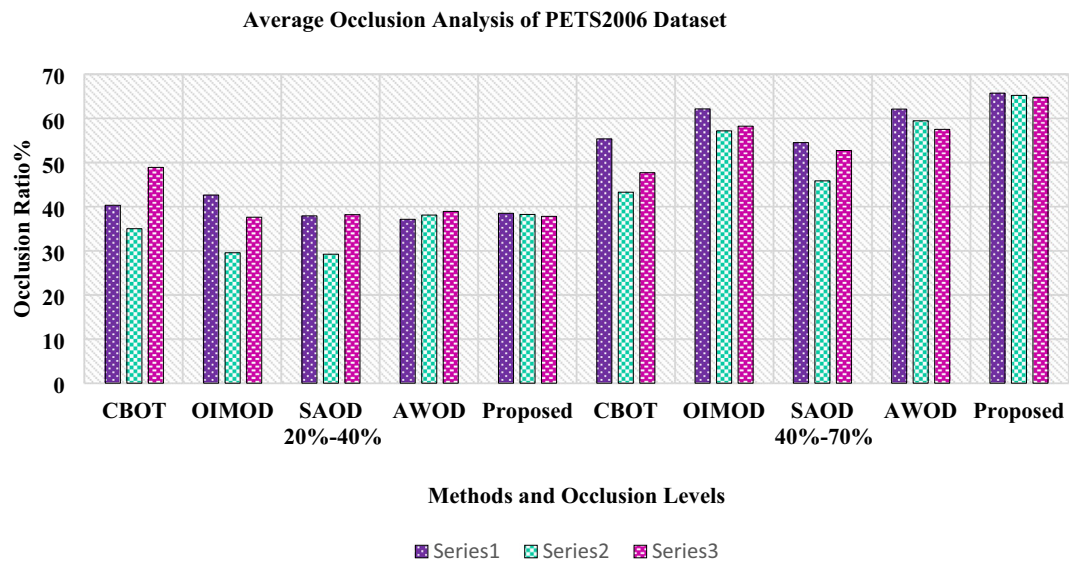
Average Occlusion Analysis of PETS2006 Dataset

Fig. 17 Comparative representation of average occlusion analysis on PETS2006 dataset

Table 8 Quantitative comparison of Average Precisions under occlusion conditions using $\mathbf{AP_{IoU}}$(baseline)

| Average Precision (AP) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Datasets | Average Precision | $\delta = 0.35$ | | | | | $\delta = 0.65$ | | | | |
| | | CBOT | OIMOD | SAOD | AWOD | Proposed | CBOT | OIMOD | SAOD | AWOD | Proposed |
| **Highway** | AP_{IoU} | 0.388 | 0.396 | 0.410 | 0.392 | 0.438 | 0.386 | 0.341 | 0.397 | 0.405 | 0.463 |
| | AP_{GIoU} | 0.392 | 0.404 | 0.419 | 0.403 | 0.451 | 0.392 | 0.353 | 0.416 | 0.427 | 0.492 |
| | RI% | 1.030 | 2.020 | 2.195 | 2.806 | **2.968** | 1.554 | 3.519 | 4.785 | 5.432 | **6.263** |
| | AP_{CIoU} | 0.397 | 0.409 | 0.425 | 0.409 | 0.462 | 0.395 | 0.361 | 0.421 | 0.429 | 0.502 |
| | RI% | 2.319 | 3.178 | 3.658 | 4.336 | **5.479** | 2.331 | 5.865 | 5.700 | 5.925 | **8.423** |
| **PETS2006** | AP_{IoU} | 0.354 | 0.386 | 0.372 | 0.408 | 0.447 | 0.401 | 0.428 | 0.452 | 0.515 | 0.524 |
| | AP_{GIoU} | 0.361 | 0.395 | 0.383 | 0.423 | 0.463 | 0.408 | 0.435 | 0.461 | 0.53 | 0.539 |
| | RI% | 1.977 | 2.331 | 2.956 | 3.579 | **3.619** | 1.745 | 1.635 | 1.991 | 2.901 | **2.905** |
| | AP_{CIoU} | 0.368 | 0.398 | 0.387 | 0.430 | 0.471 | 0.411 | 0.442 | 0.471 | 0.548 | 0.556 |
| | RI% | 3.954 | 3.108 | 4.032 | 5.392 | **5.362** | 2.493 | 3.271 | 4.203 | 6.357 | **6.407** |

RI Relative Improvement. Bold values depict better results

To evaluate the performance of the system under different occlusion ranges, we computed the average IoU loss at two threshold values: $\delta = 0.35$ for occlusion ranges between 20 and 40%, and $\delta = 0.65$ for occlusion ranges between 40 and 70%. To evaluate the performance of the methods, we selected approximately 10 frames from each of the PETS 2006 and Highway datasets for each occlusion range. The choice of these threshold values is used to determine the effectiveness of the proposed method for detecting occlusion in the tested datasets. The bounding box localization loss depends on the correct prediction of detected and ground truth bounding boxes of occluded areas. Then, the average precision values have been computed for benchmarked

datasets using IoU loss as a baseline evaluation metric. We have computed average precision for GIoU loss and CIoU loss to achieve a certain degree of performance.

Based on the experimentations, it has been observed that the proposed method using GIoU loss can effectively reduce the object localization error by achieving significant improvements in the average precision (AP) metric for the Highway and PETS2006 datasets under different occlusion ranges. Specifically, for the Highway dataset, the proposed method achieved 2.968% AP and 5.479% AP for $\delta = 0.35$ and $\delta = 0.65$, respectively, under occlusion ranges of 20%-40% and 40%-70%. For the PETS2006 dataset, the proposed method attained 3.579% AP and 6.374% AP for $\delta = 0.35$

Fig. 18 Graphical comparison of AP for different occlusion ranges using GIoU loss and CIoU loss for Highway and PETS2006 datasets

(a)

(b)

(c)

(d)
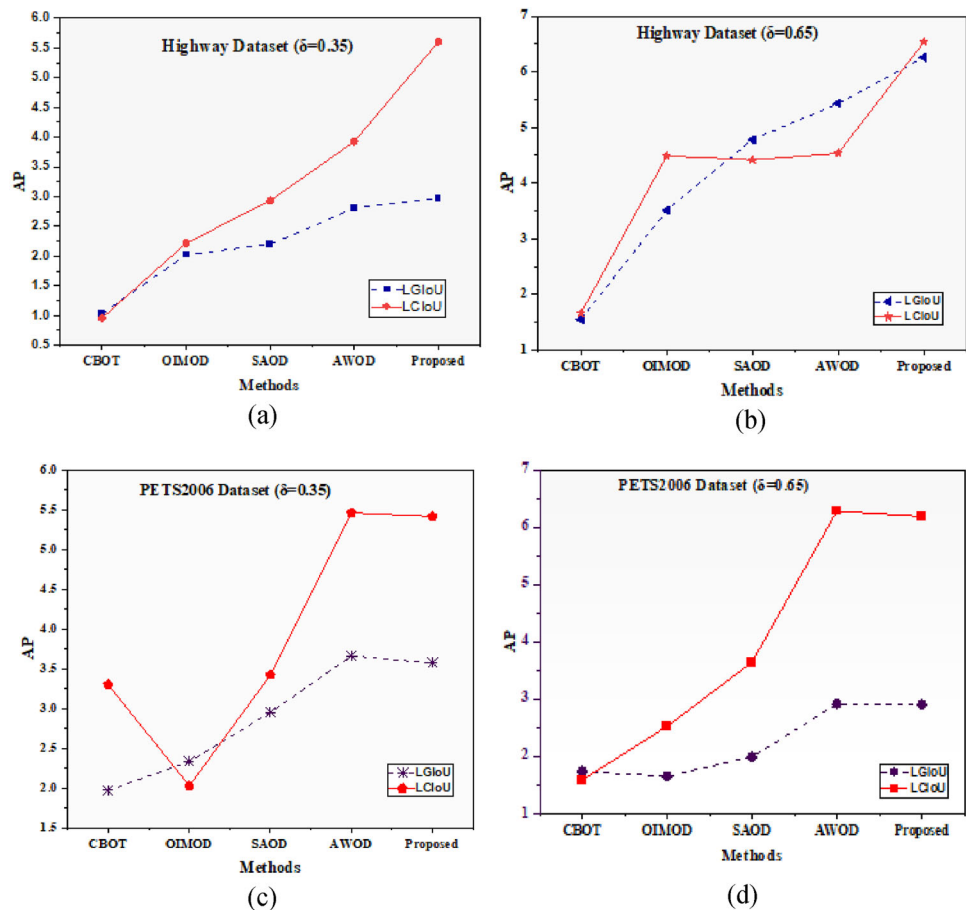
and $\delta = 0.65$, respectively. Moreover, the CIoU loss, which considers the three geometric factors of overlapped bounding boxes, i.e., center points, Euclidian distance, and aspect ratios, has shown better performance than the GIoU loss in the same datasets. The CIoU loss achieved 6.263% AP and 8.423% AP in the Highway dataset for different occlusion ranges, and 2.905% and 6.374% AP in the PETS2006 dataset for $\delta = 0.65$. Furthermore, the AWOD method has also shown better performance in the PETS2006 dataset due to its property of correlation between foreground objects and background scenes with sub-patch feature extraction, which reduces the size of the bounding box for precise detection of the occluded area. The graphical comparison of average precision values under occlusion conditions based on GIoU loss and CIoU loss shown in Fig. 18 indicates that the CIoU loss generates more correct predictions of occluded bounding boxes than GIoU-based prediction. Although the GIoU loss has a small gain, the CIoU loss outperforms due to its geometric properties, where consistency of aspect ratios along with central point distance between predicted bounding boxes affects the prediction rate.

4.6.3 Performance evaluation

We assessed the efficiency and effectiveness of the proposed method by comparing it with state-of-the-art methods using bounding box prediction accuracy. To evaluate our method's performance in detecting partially occluded objects, we used three additional metrics: Precision, Recall, and F1-Score. The precision determines the accuracy of object identification, Recall measures the ability to detect objects correctly, and F1-Score incursions a balance between Precision and Recall. To optimize our method's Precision and Recall values, we set the threshold value $\delta = 0.60$ and compared it with other methods on the Highway and PETS2006 datasets. The results are illustrated in Table 9, which includes an average analysis of Precision, Recall, F1-Score, Intersection over Union (IoU), and Bounding Box Localization Loss (BBL) to measure performance under occlusion conditions. Our proposed method outperformed other methods in terms of F1-Score by accurately predicting bounding boxes. Specifically, our method achieved 65.8% and 64.7% overlapping amounts with 0.035 and 0.041 localization losses in the Highway and PETS2006 datasets, respectively (Fig. 19).

Table 9 The comparison results of performance measures under occlusion condition

Methods	Highway dataset					PETS2006 dataset				
	Precision ↑	Recall ↑	F1 ↑	IoU ↑	BBL ↓	Precision ↑	Recall ↑	F1 ↑	IoU ↑	BBL ↓
CBOT [19]	0.773	0.785	0.778	0.589	0.054	0.767	0.743	0.754	0.574	0.093
OIMOD [29]	0.815	0.804	0.809	0.625	0.048	0.742	0.688	0.713	0.562	0.131
SAOD [3]	0.863	**0.898**	0.880	0.619	0.051	0.785	0.726	0.752	0.594	0.105
AWOD [17]	*0.915*	*0.881*	*0.897*	*0.644*	*0.041*	*0.816*	*0.825*	*0.820*	*0.625*	*0.072*
Proposed	**0.927**	0.875	**0.900**	**0.658**	**0.035**	**0.847**	**0.851**	**0.848**	**0.647**	**0.041**

Bold represents the best results. Italic represents the second best result
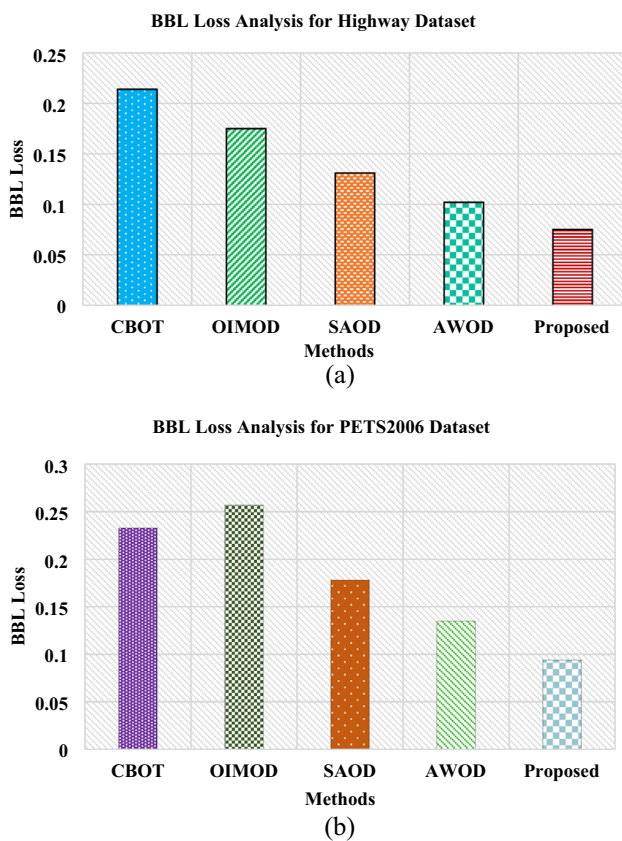
Fig. 19 Comparison of Bounding box localization loss of proposed method with state-of-the-art methods on Highway and PETS2006 datasets

The following graphical representations show that the proposed method generates less bounding box localization loss as compared to other methods.

5 Discussion

The quantitative results demonstrate the performance of the proposed method can detect objects under partial occlusion conditions. The occlusion pattern helps in detecting various levels of occlusion and our method achieves good results by computing localization loss at different levels of occlusion for different threshold values. The negative loss value reflects the misalignment of the bounding boxes. After experimentation, we have found that the proposed method accurately detects objects under occlusion conditions with a 65–70% occlusion rate. The Avg GIoU and Avg CIoU values play an important role in identifying the amount of overlapping. This study reveals that GIoU and CIoU values should be similar to IoU due to tightening lower bound which shows more similarity and proximity of bounding boxes. Table 8 illustrates the results of average IoU, GIoU, and CIoU to understand the occlusion reasoning. Then, the localization loss has been evaluated for different threshold values to know the effectiveness of the proposed method. The proposed method needs only the intersection area to detect the occlusion relationship rather than the large enclosed bounding box. Then, localization loss has been computed for the correct prediction of bounding boxes. The efficiency of the proposed method is measured using average precision (AP) metric based on localization loss. Sometimes AP value is low because CIoU loss is a little lesser than IoU loss as the aspect ratio consistently does not contribute to the prediction of accuracy. Overall, the results suggest that the CIoU loss outperforms the GIoU loss in terms of AP for both datasets, particularly for the Highway dataset where the improvement is significant. However, the GIoU loss still achieves reasonable performance in both datasets. We also considered the occluded part of the objects to observe the effects of occlusion for evaluation. Our datasets contain similar class objects which help us to quantify the different levels of occlusion for accuracy prediction and improving the performance of object detection methods.

Our method relies on the geometric features of the bounding box which are directly extracted using a blob which helps in reducing the computational cost. The proposed method takes around 30 ms to process one frame. The proposed method can detect occluded objects with three different views. However, continuous improvements are necessary

for occluded object detection techniques to minimize localization errors. The accurate detection of occluded objects is vital for a variety of real-world applications, including autonomous driving, robotics, video surveillance, and object recognition. Precise occluded object detection can assist autonomous vehicles in identifying and avoiding obstacles, allow robots to handle objects in cluttered settings, enhance the accuracy of surveillance systems, and improve object recognition in natural environments.

6 Conclusion and future scope

Occlusion detection is a difficult challenge for object detectors, as many methods struggle to locate bounding boxes accurately when the object's features are less visible. To address this issue, we proposed a geometric feature-based axis-aligned bounding box method that generates enclosed bounding boxes surrounding objects to ensure their proper localization. We then introduced an occlusion prior condition to check the statistics of pixels and determine whether they are occluded by calculating the overlapped area of bounding boxes. The proposed OBB Detector is capable of detecting different levels of occlusions ranging from 20 to 70%. We compared our approach with two state-of-the-art methods on benchmark datasets, including PETS 2009 and Highway, using similarity measures based on evaluation metrics such as IoU, GIoU, and CIoU for evaluating bounding box localization loss and found that our method generates BBL 0.039 for the Highway dataset and 0.051 for the PETS2006 dataset as compared to state-of-the-art methods. The efficiency of the proposed method is evaluated using average precision (AP) for occluded part detection. It has been observed that at threshold value (δ) 0.65, the AP of the proposed method has increased by 8.4% in the Highway dataset and 6.3% for the PETS2006 dataset. The results demonstrated that the proposed method performed better and achieved higher accuracy in object detection under partial occlusion conditions. The effectiveness of the proposed method can be seen using the F1-Score factor which demonstrates significant improvements in the correctness of the detected bounding boxes under occlusion conditions.

The proposed work provides a good impact on real-world applications in areas such as surveillance systems, robots, and driverless cars, where objects are frequently partially obscured. The proposed method can help make these systems safer and more reliable by enhancing their object-detecting capabilities. Object recognition, tracking, and classification are only some examples of applications that could benefit from using the proposed method, all of which depend on the precise location of the items While our method is limited to heavy occlusion detection, we believe that it can be extended by reconstructing the occluded parts from multiple viewpoints.
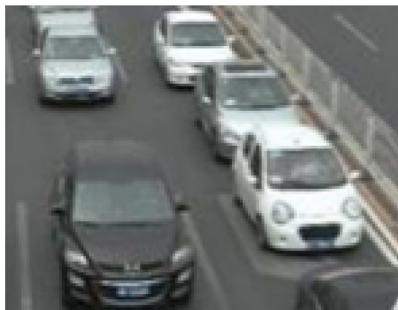
Declarations

Conflict interest The authors declare that they have no competing interests.
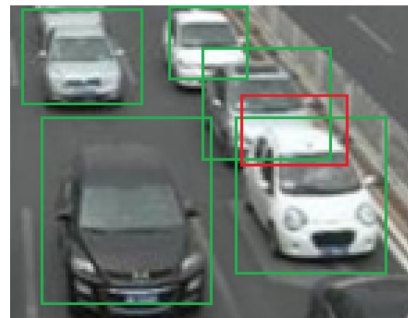
Ethics approval Not applicable.

Consent for publication Not applicable.

Appendix

This section contains the input frame and ground truth frame of Highway and PETS 2006 datasets. The red color bounding box depicts an occluded area.

(a)　　Input frame of Highway dataset

(b)　　Ground truth frame of Highway dataset

(c) Input frame of PETS2006 dataset

(d) Ground truth frame of PETS2006 dataset

References

1. Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J.: Object Detection in 20 Years: A Survey. In: Proc. IEEE, no. June, pp. 1–20, 2023, doi: https://doi.org/10.1109/jproc.2023.3238524.
2. Amit, Y., Felzenszwalb, P., Girshick, R.: Object Detection. In: Computer Vision: A Reference Guide, Cham: Springer International Publishing, 2020, pp. 1–9. doi: https://doi.org/10.1007/978-3-030-03243-2_660-1
3. Yuan, Y., Chu, J., Leng, L., Miao, J., Kim, B.-G.: A scale-adaptive object-tracking algorithm with occlusion detection. EURASIP J. Image Video Process. **2020**(1), 7 (2020). https://doi.org/10.1186/s13640-020-0496-6
4. Shi, P., Hou, B., Chen, J., Zu, Y.: An algorithm of occlusion detection for the surveillance camera. Sci. Program., (2021) https://doi.org/10.1155/2021/6698160
5. Li, Y.Y., et al.: Occlusion. Comput. Vis. Image Underst. **17**(3), 346–359 (2008). https://doi.org/10.1016/j.cviu.2007.09.014
6. Gholamhosseinian, A., Seitz, J.: Vehicle classification in intelligent transport systems: an overview, methods and software perspective. IEEE Open J. Intell. Transp. Syst. **2**, 173–194 (2021). https://doi.org/10.1109/OJITS.2021.3096756
7. Xie, H., Zheng, W., Shin, H.: Occluded pedestrian detection techniques by deformable attention-guided network (Dagn). Appl. Sci. (2021). https://doi.org/10.3390/app11136025
8. Fantacci, C., Vo, B.-N., Vo, B.-T., Battistelli, G., Chisci, L.: Robust fusion for multisensor multiobject tracking. IEEE Signal Process. Lett. **25**(5), 640–644 (2018). https://doi.org/10.1109/LSP.2018.2811750
9. Wang, C. Xinlong and Xiao, Tete and Jiang, Yuning and Shao, Shuai and Sun, Jian and Shen, Repulsion loss: Detecting pedestrians in a crowd.," *{Proceedings IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)}*, 2018
10. Meshgi, S., Ishii, K.: The state-of-the-art in handling occlusions for visual object tracking. IEICE Trans. Inf. Syst. **98**(7), 1260 (2015)
11. C. Zhou and J. Yuan, "Bi-box Regression for Pedestrian Detection and Occlusion Estimation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11205 LNCS, pp. 138–154, 2018, doi: https://doi.org/10.1007/978-3-030-01246-5_9.
12. E. Bochinski, V. Eiselein, and T. Sikora, "High-Speed tracking-by-detection without using image information," in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017, pp. 1–6. doi: https://doi.org/10.1109/AVSS.2017.8078516.
13. Li, Y., Tu, Y., Chen, X., Zhao, H., Zhou, G.: Distance-Aware Occlusion Detection with Focused Attention. IEEE Trans. Image Process. **31**, 5661–5676 (2022). https://doi.org/10.1109/TIP.2022.3197984
14. B. Pepikj, Bojan and Stark, Michael and Gehler, Peter and Schiele, "Occlusion Patterns for Object Class Detection," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013.
15. P. Cai *et al.*, "Collision Detection Using Axis Aligned Bounding Boxes," in *Simulations, Serious Games and Their Applications*, Y. Cai and S. L. Goei, Eds. Singapore: Springer Singapore, 2014, pp. 1–14. doi: https://doi.org/10.1007/978-981-4560-32-0_1.
16. K. Saleh and Z. Vámossy, "BBBD: Bounding Box Based Detector for Occlusion Detection and Order Recovery," pp. 78–84, 2022, doi: https://doi.org/10.5220/0011146600003209.
17. X. Tian, H. Li, and H. Deng, "An improved object tracking algorithm based on adaptive weighted strategy and occlusion detection mechanism," *J. Algorithms Comput. Technol.*, vol. 15, 2021, doi: https://doi.org/10.1177/1748302620973536.

18. H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 658–666, 2019, doi: https://doi.org/10.1109/CVPR.2019.00075.

19. Moutakki, Z., Ouloul, I.M., Afdel, K., Amghar, A.: Real-time video surveillance system for traffic management with background subtraction using codebook model and occlusion handling. Transp. Telecommun. **18**(4), 297–306 (2017). https://doi.org/10.1515/ttj-2017-0027

20. Ferrari, V., Jurie, F., Schmid, C.: From Images to Shape Models for Object Detection. Int. J. Comput. Vis. **87**(3), 284–303 (2010). https://doi.org/10.1007/s11263-009-0270-9

21. Nevatia, R., Binford, T.O.: Description and recognition of curved objects. Artif. Intell. **8**(1), 77–98 (1977). https://doi.org/10.1016/0004-3702(77)90006-6

22. Mao, W., Zheng, J., Li, B.: "Patch-based object tracking using corner and color with partial occlusion handling", in. IEEE International Conference on Progress in Informatics and Computing **2014**, 269–274 (2014). https://doi.org/10.1109/PIC.2014.6972339

23. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94

24. Park, S., Lee, H., Yoo, J.-H., Kim, G., Kim, S.: Partially Occluded Facial Image Retrieval Based on a Similarity Measurement. Math. Probl. Eng. **2015**, 217568 (2015). https://doi.org/10.1155/2015/217568

25. Li, X., et al.: Multi-Task Structure-Aware Context Modeling for Robust Keypoint-Based Object Tracking. IEEE Trans. Pattern Anal. Mach. Intell. **41**(4), 915–927 (2019). https://doi.org/10.1109/TPAMI.2018.2818132

26. Hu, K., Huang, G., Yang, Y., Pun, C.M., Ling, W.K., Cheng, L.: Rapid facial expression recognition under part occlusion based on symmetric SURF and heterogeneous soft partition network. Multimed. Tools Appl. **79**(41–42), 30861–30881 (2020). https://doi.org/10.1007/s11042-020-09566-2

27. Weng, R., Lu, J., Tan, Y.-P.: Robust Point Set Matching for Partial Face Recognition. IEEE Trans. Image Process. **25**(3), 1163–1176 (2016). https://doi.org/10.1109/TIP.2016.2515987

28. X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 32–39. doi: https://doi.org/10.1109/ICCV.2009.5459207.

29. Mishra, P.K., Saroha, G.P.: Occlusion handling strategies for multiple moving object classification. Int. J. Comput. Vis. Robot. **10**(2), 167–184 (2020). https://doi.org/10.1504/IJCVR.2020.105683

30. C. C. Ghiasi, Golnaz and Fowlkes, "Occlusion Coherence: Localizing Occluded Faces with a Hierarchical Deformable Part Model," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014.

31. A. Y. Stewart, Russell and Andriluka, Mykhaylo and Ng, "End-To-End People Detection in Crowded Scenes," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016.

32. Jia, Q., Chen, X., Wang, Y., Fan, X., Ling, H., Latecki, L.J.: A rotation robust shape transformer for cartoon character recognition. Vis. Comput. (2023). https://doi.org/10.1007/s00371-023-03123-2

33. L. Qi, L. Jiang, S. Liu, X. Shen, and J. Jia, "Amodal instance segmentation with kins dataset," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 3009–3018, 2019, doi: https://doi.org/10.1109/CVPR.2019.00313.

34. H. A. Bin Sulaiman, M. A. Othman, M. Z. A. A. Aziz, and A. Bade, "Implementation of axis-aligned bounding box for opengl 3D virtual environment," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 2, pp. 701–708, 2015.

35. D. Bartz, Dirk and Klosowski, James T and Staneker, "Tighter bounding volumes for better occlusion culling performance," 2005.

36. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031

37. Heo, J., Wang, Y., Park, J.: Occlusion-aware spatial attention transformer for occluded object recognition. Pattern Recognit. Lett. **159**, 70–76 (2022). https://doi.org/10.1016/j.patrec.2022.05.006

38. T. Mostafa, S. J. Chowdhury, M. K. Rhaman, and M. G. R. Alam, "Occluded Object Detection for Autonomous Vehicles Employing YOLOv5, YOLOX and Faster R-CNN," in *2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2022, pp. 405–410. doi: https://doi.org/10.1109/IEMCON56893.2022.9946565

39. Wang, M., Du, H., Mei, W., Wang, S., Yuan, D.: Material-aware Cross-channel Interaction Attention (MCIA) for occluded prohibited item detection. Vis. Comput. **39**(7), 2865–2877 (2023). https://doi.org/10.1007/s00371-022-02498-y

40. Agrawal, S., Natu, P.: ABGS Segmenter: pixel wise adaptive background subtraction and intensity ratio based shadow removal approach for moving object detection. J. Supercomput. (2022). https://doi.org/10.1007/s11227-022-04972-9

41. B. Pepik, M. Stark, and P. Gehler, "Occlusion Patterns for Object Class Detection," *CVPR*, pp. 3286–3293, 2013, doi: https://doi.org/10.1109/CVPR.2013.422.

42. Zhou, C., Yuan, J.: Occlusion Pattern Discovery for Object Detection and Occlusion Reasoning. IEEE Trans. Circuits Syst. Video Technol. **30**(7), 2067–2080 (2020). https://doi.org/10.1109/TCSVT.2019.2909982

43. "Changedetection.net." http://jacarini.dinf.usherbrooke.ca/dataset2012

44. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D.: "Distance-IoU loss: Faster and better learning for bounding box regression", AAAI 2020–34th AAAI Conf. Artif. Intell. **2**, 12993–13000 (2020). https://doi.org/10.1609/aaai.v34i07.6999

Supriya Agrawal has earned her Bachelor of Technology in Computer Science & Engineering from Uttar Pradesh Technical University, Lucknow in 2007 and Master of Technology in Computer Science from Banasthali University, Rajasthan in 2013. Currently, she is pursuing PhD in Computer Engineering from NMIMS University, Mumbai. She has 2 years of industrial experience in IT Company. She is currently working as Assistant Professor in Computer Engineering department at Mukesh Patel School of Technology Management & Engineering, NMIMS University, Mumbai, India. Her research interest is Image Processing and Computer Vision. She has published many research papers in various conferences and journals. Ms.

Agrawal is a Technical Program Committee member of ICACDS conference, Springer. She is also a member of IAENG society.

Prachi Natu has completed her B.E. in Electronics in 2004 and Telecommunication and M.E. in 2011 in Computer Engineering from Mumbai University. She has earned her PhD in Image Processing from NMIMS University, Mumbai. She has 13 years of experience in academics and was working as Assistant Professor in Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai. Her area of interest are Image processing, Computer Vision, Machine learning and Deep learning. She has published many research papers in various journals and conferences.