INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 16-09-2022

Accepted: 10-11-2022

Published: 21-12-2022

Citation: Bharathi G, Anandharaj G (2022) A Conceptual Real-Time Deep Learning Approach for Object Detection, Tracking and Monitoring Social Distance using Yolov5. Indian Journal of Science and Technology 15(47): 2628-2638. https://doi.org/10.17485/IJST/v15i47.1880

* **Corresponding author**.

jivibara@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

A Conceptual Real-Time Deep Learning Approach for Object Detection, Tracking and Monitoring Social Distance using Yolov5

G Bharathi[1]*, G Anandharaj[2]

1 Assistant Professor, Department of Master of Computer Application, Shanmuga Industries Arts and Science College, Thiruvalluvar University, Tiruvannamalai, 606603, Tamil Nadu, India
2 Assistant Professor & Head, Department of Computer Science and Application, Adhiparasakthi College of Arts and Science, Kalavai, Ranipet, Tamil Nadu, India

Abstract

Objectives: To develop a computer vision-based model that can detect, track and recognize individuals for the purpose of measuring social distance in road traffic videos using surveillance cameras. **Methods:** The real-time traffic surveillance webcam dataset was applied to validate the model, with better performance metrics outperformed to those of comparable cutting-edge models such as RetinaNet, Faster R-CNN and SSD. Our proposed methodology utilized object detection methods to recognize individuals followed by multiple objects tracking approach to track identified individuals using detected bounding boxes. Our research shows that the conventional method is successful in detecting persons who violate social distances. **Findings:** Our finding shows that our proposed object detection model successfully recognizes human and those who violating the social distancing measurements. For the purpose of detecting social distance, develop a highly accurate detection technique. Our YOLOv5 with multiple objects tracking algorithm delivered great outcomes with appropriate Precision of 93%, Recall of 94%, F1-Score of 96% and mAP of 95% measures given by object categorization and localization to measure social distancing in real-time traffic surveillance videos. The YOLOv5 model's results are then compared to many other prominent state-of-the-art models. **Novelty:** The YOLOv5 and MOTSORT is appropriate for finding whether people are maintaining social distancing or not, intended to identify, monitor and track those who are not following or violating social distances with overall accuracy and efficiency.

Keywords: Object Detection; Deep Learning; Social distance monitoring; multiple objects tracking; YOLOv5

1 Introduction

A surveillance monitoring network can range from a single camera to thousands of cameras scattered all over a city or area, all accessible through a single network. For this research purpose, videos are taken from the most visited tourist places or crowded areas in Tiruvannamalai, Tamil Nadu, India as shown in Figure 1. A deep neural network's convolution neural network is a type of deep neural network. CNNs are multi-layer networks that are fully connected. It is a strong image processing system that uses deep learning to accomplish certain image and video identification tasks. The input layer, convolution layers, pooling layers, and fully linked layers make up a CNN. Figure 2 shows a simplified illustration of CNN architecture. Multi-object detection and tracking of people to see if they are maintaining social distance is a critical task that is utilized in a variety of applications, including object detection, social distancing monitoring, and so on. Early deep learning neural network models are categorized into two stages, such as one-stage object detection algorithm and two-stage object detection algorithm. Faster Region based Convolution Neural Network (Faster R-CNN), Fast Region based Convolution Neural Network (Fast R-CNN) and Region based Convolution Neural Network (RCNN) are all two-stage object detection algorithm. The analysis of the literature used many methods, including computer vision and natural language processing. They gathered methods for literature mining in natural language processing to locate answers to questions. While using strategies for COVID-19 identification and diagnosis from X-ray pictures in computer vision, the ConVIRT model produced the best results. None of the reviews that have been published address the AI models that are used to recognize people who are close to one another. Since there is work being done on all sides, it is necessary to continue this tremendous effort with the same potential. The primary objective of this study is to apply the appropriate tracking and monitoring techniques for social distancing while maintaining everyday life routines and reducing the risk of infection. Whenever and wherever it is necessary, social distancing is carried out, as we have seen with COVID-19 over the previous year. DL technologies have shown potential in the image categorization and object identification fields. Therefore, it is necessary to apply various models to calculate the distance. In order to ensure strict adherence to the social distance standard, this research work is intended to provide a thorough evaluation of the AI methods utilized to measure social distance. The current real-time image detection models based on deep learning mainly uses one-stage target detection algorithms framed by frameworks such as the YOLO family. YOLOv5 is the simplest model in width and depth across all YOLO networks, with great speed, accuracy and precision in automatic object detection. Glenn Jocher of Ultralytics started YOLOv5 with Pytorch implementation. YOLOv5 includes CSPNet as the backbone, PANet for the heads, and Mosaic and CutMix for data augmentation. The authors of[(1)] provided a research study of the use of YOLOv5 to identify vehicles throughout the wintertime. YOLOv5 learns boundary box sizes automatically depending on the specified training dataset, allowing the detection scaling to be initialized for the current task's data. For various constraints, YOLOv5 supports all the sizes of models. As a result, for efficient individual detection, we used YOLOv5 as the backbone network and paired it with Multi Object Tracking DeepSORT. The authors[(2)] suggested a comprehensive strategy for spotting facial masks on people and keeping track of social distance breaches. Tracking-by-Detection, objects are first identified and localized in each frame, and then tracking is carried out by linking detections into itineraries using data associations. As a result, the tracking performance is greatly dependent on the detectors and based on the association method's performance. Simple Online and Real-time Monitoring (SORT) uses a Kalman Filter to display the correlation of frame-by-frame data, and a Hungarian technique effective for tracking multiple objects can be used to assess correlation. The Track estimator, which is part of the SORT method, uses the Kalman Filter to locate the position of an object's bounding box. The Kalman Filter is used here to determine an added feature between the old and calculated values in a recursive manner. Calculation based on the prior object velocity to make Kalman Filter predictions. The pictures are compressed using the Quantized Ternary Pattern based Singular Value Decomposition and Run Length Coding technique, which increases efficiency by computing the images block by block[(3)]. The recognized ROI's location is pixilated to generate Regions of Interest (ROIs) on fully connected layers of extracted features, which reflect confined bounding boxes all over detected objects with a strong possibility of presence of objects (people). RPN providing methods for the ROI pooling layer, which categorizes the target object as individual or even no individual as well as modifies the detected bounding box[(4)]. The South India Tourism Review provides an excellent opportunity for tourist planners to evaluate the interests, recommendations, and feelings of tourist's visitors. For the classification of tweet reviews, the Support Vector Machine algorithm is designed, which fulfills the specific needs of tourists and locates points of interest[(5)]. The term ”visual social distancing,” which is described as “Automatic measurement and characterization of interpersonal distance from a picture?” When established distance constraints are broken, VSD is utilized to offer statistics concerning the level of safety. Inverse perspective mapping (IPM) is a technique that uses real-world dimensions of an image being translated from a video source in combination with valuable information from of the mobile phone to produce a bird's eye view. Using top view pictures from remote sensing, a machine learning algorithm can determine the extent to which people maintain their social distance[(6)]. The authors[(7)] provided a research survey discusses and analyzes various state-of-arts of models that have been proposed for Deep Learning-based object detection. The author[(8)] suggested strategy to see whether pedestrians adhere to the norms of social distance in both indoor and outdoor scenarios.

Table 1 demonstrates the literature review based on techniques such as Faster R-CNN, YOLOv4 and Feature extraction modules based on merits and demerits of various techniques involved in detection, tracking and monitoring social distancing.

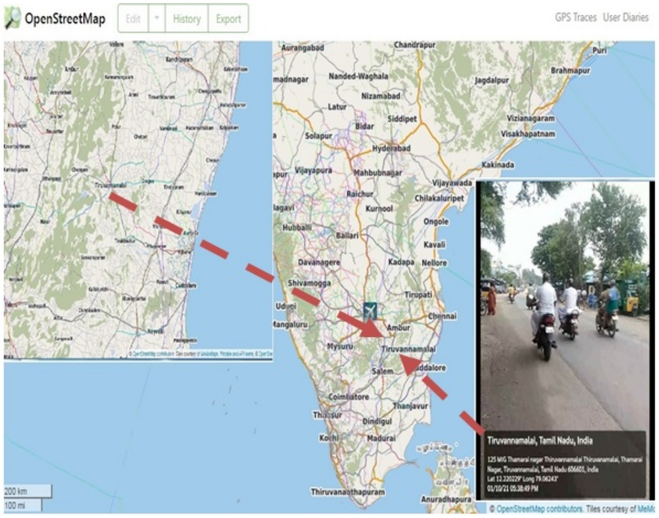

Fig 1. Study area of Open Street Map Image

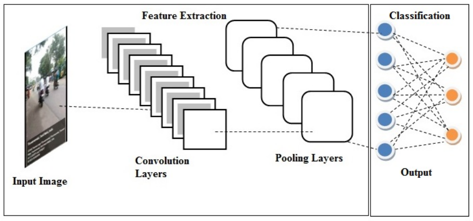

Fig 2. Architecture of CNN

Table 1. Comparison of various techniques involved in tracking and monitoring social distance

Reference	Proposed Scheme	Merits	Demerits
(9)	Faster R-CNN	One of the typical model designs for object detection to generate pyramidal features extracted.	Object detection is a difficult and time-consuming process. When there are numerous objects in a single frame, the method becomes more complicated.
(10)	YOLOv4 based Model	Object detection rate is significant, When compared to other algorithms; the computation time is quite short. The set of instructions is basic.	If the window size is increased, it consumes a lot of Memory. The algorithm is based on the assumption that the brightness of the image would remain constant throughout the tracking process.
(11)	Feature Extraction Modules	A simple method for tracking objects whose appearance is determined by their color.	It can only detect and track one moving object at a time. While there are interference issues, it is ineffective.

The effectiveness of our YOLOv5 and MOTSORT algorithms along with various comparisons has been demonstrated extensively over out dataset videos containing different objects. The contributions and novelty of the investigation are summarized below.

i) A new deep CNN model, named, YOLOv5 for multi object detection has been introduced. This is an improved version of the widely used YOLO family, providing superior performance.

ii) MOTSORT is a more efficient version of deep SORT in terms of both accuracy and speed for tracking multiple objects, as it involves frame-by-frame association only within the same-class objects, instead of taking all classes together as one class.

iii) YOLOv5 model was used to measure the distance between objects and process real-time images to solve the problem of a group of five or more people in the same space.

iv) Besides the investigation demonstrates a way of incorporating the recently emerged granular computing in deep learning network for on-line object tracking and classification directly from input video.

Therefore, our object detection framework model outperformed well to achieve the optimal balance between precision and time efficiency. Additionally, we suggest an efficient EMOR object identification that does not require additional metadata to eliminate redundant data and false positives. In Section 2 will present the proposed methodology, which includes the system architectures as well as our object detection algorithms, tracking, and monitoring social distancing. The experimental results and system performance will be compared to the state-of-the-art in Section 3, followed by discussions and conclusions in Section 4.

2 Methodology

The proposed methodology is a three-level module that integrates computer vision and convolution neural networks to focus on object or human identification or detection, tracking, and inter-distance measurement as a whole solution for social distance monitoring and tracking using image or video streaming datasets.

The dataset was first collected and preprocessed. Another issue with the dataset during preprocessing is that the quality of the image is not consistent. Some photos are taken in landscape format, while others are taken in portrait mode. As a result, adjusted the images into uniform shapes and generated training and testing folds. Whereas augmentation produced excellent outcomes in densely filled photos, it enhanced the outcomes overall. Throughout preprocessing, frames were resized and categorized before producing several folds of the dataset in order to generalize a prototype for object detection using deep learning architecture. These datasets were subjected to various augmentation techniques. The dataset's most annoying feature is that it includes all photographs of two-wheelers, four-wheelers, and persons from various angles of sight. There were pictures from the front, rear, and side of the traffic roads. Many photographs from the video were also included in the training to improve the collection of unusual category motor vehicles. These folds were trained autonomously with the YOLOv5 model with different phases during training. The new photos were then manually tagged with a tagging tool. The final state of our work is represented by images with bounding boxes surrounding the detected objects in an image with their own group.

Object detection is one of the most difficult problems in computer vision. An object detection network termed YOLOv5 is used throughout this work to allow the computer to automatically obtain and locate targets of interest from photos, balancing precision and speed. YOLOv5 has become the new state of the art for object recognition; it's mostly used for real-time picture detection. YOLOv5's main goal was to find the right balance between real-time performance and detection accuracy. Nevertheless, when compared to previous YOLO versions, the YOLOv5 produces much excellent performance. The main goal of this study is to detect objects and track whether people maintain social distance. Backbone, Neck, and Output are the main components of the standard YOLOv5. In order to do feature extraction from input photos, Backbone is used. Because of the pyramid structure used, Neck was able to obtain visual characteristics that were resistant to scale alterations. Assume that the input image is divided into S x S grids. Feature extraction from each grid cell is then used to classify objects. Full connection layers and many convolution layers make up the entire network structure. Following the completion of the different network, the tensor S x S (P x 5 + C) is output, where 'P' is the number of predicted targets in each grid and 'C' denotes the integer of grouping.

To emphasize the boldness of existence, locate and identify things in a photograph and label them with rectangular bounding-boxes. There are two types of object detectors: two stage detectors and one stage detectors. The conventional object detection pipeline, which includes object localization and classification, is followed by two stage detectors. To begin, object localization can be accomplished using either traditional methods or deep networks. The features retrieved from this proposed region have been used to complete the classification.

The YOLOv5 primarily consists of Bakbone, Neck, and Output. Focus could be a layer that splits the pictures into layers. Conv runs Conv, calculates the worth, and exports the result. Frames transforms the input file, calculates it at the bottleneck layer, adds the conv value from the initial input and thus the calculation value at the bottleneck layer in Concat, then converges and outputs it. The bottleneck uses Conv on the input value and outputs another Conv calculated value. MaxPooling values in Concat with the Conv value from the prevailing input value and sends them out after Conv after the Conv operation. Concat is responsible for combining input layers. Because final Conv values are currently an output result of detect.

Figure 3 shows the flow chart of overall algorithm for outputting the space between objects during this paper, and it explains so as from video or real-time data transmission to the method of boundary box. Initially the general structure receives video or real-time image data. In YoloV5, the video is processed until the top of the program. After training the YOLOv5 model, we expand its function to watch the social distancing of individuals; we all know that the model will obtain the position of the prediction boxes when detecting people during a picture. MOTSORT starts by receiving the important time object from YoloV5 as a Kalman Filter, and matches through the Hungarian algorithm. so as to detect the space between people, real-time image or video has two-dimensional coordinates and perspective, so there's a difference within the size of an object box under the influence of the y-axis compares the calculated value with the worth calculated from the Euclidean distance to calculate the degree of closeness between objects. When comparing with predicted points through IoU analysis, vector angle and distance are calculated using Euclidean distance.

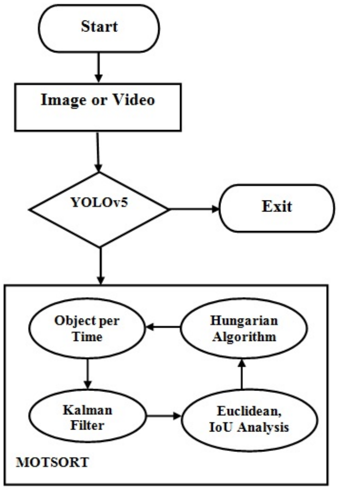

Fig 3. Structure of Proposed Model

After updating in Kalman Filter, it proceeds as object per time. MOTSORT works in a recursive manner, transmitting the coordinates and using the received coordinates to insert the bounding box into the OpenCV image. Observe the midpoint of the bounding box to examine the distance between objects throughout the positioning process. Calculate the Euclidean distance between the highest left vertices of the prediction boxes because the distance between people, if the space is bigger than a particular value as we set here as 100, it's judged that the 2 people are too close. Finally, draw a line to attach the highest left vertices on the prediction boxes which are too close, warning that the social distance between the 2 people is dangerous.

2.1 Exterior Maximal Optimum Repression

There are several overlapping detections after object identification on both videos and photographs, without delay continuous combining of all detection relating to almost the same objects into a particular detection. EMOR starts with the detections Bounding Boxes (BB) with scores S and rescores them using Equation (6). First, it chooses the detection with the greatest scorebb^{rs}_{hs}. EMOR concatenates it to the Absolute Detection (AD) set after eliminating it from the collection BB, and eliminates any outcomes with IoU scores greater than a threshold EMOR$_{th}$ from the collection BB. This procedure is done until the set BB is emptied. When re-scoring bounding boxes, however, EMOR uses a fixed threshold, which implies that if two predicted true positive have an overlapping bigger than EMOR$_{th}$, the one with the lower score is eliminated. To tackle this problem, EMOR uses a Gaussian penalty function as a re-scoring function in Eq. (7), assigning a lower score to the bounding box bb^{rs}_{j} if it has a massive overlap bb^{rs}_{hs}.

$$\check{\mathrm{E}}_{\mathrm{j}} = \int_0^{\check{\mathrm{E}}\mathrm{j}} \quad \begin{matrix} \mathrm{IoU}\left(bb^{rs}_{hs'}, bb^{rs}_{j}\right) < EMO\,Rth \\ \mathrm{IoU}\left(bb^{rs}_{hs^s}, bb^{rs}_{j}\right) \geq \ \mathrm{EMORth} \end{matrix} \,\mathrm{b} \tag{1}$$

Algorithm 1: Exterior Maximal Optimum Repression

```
    Input:
    $B = \{b_1, b_2, \ldots, b_n\}$ $DS = \{S_1, S_2, \ldots, S_n\}$ $EMOR_{th}$
    Where B = initial Detection boxes.
    DS = Detection Score.
```

```
$EMOR_{th}$ = Exterior Maximal Optimum Repression threshold.
Start:
D ← {}
While B # empty
do
 A ⟵ argmax S
 X ⟵ $B_m$
 D ⟵ DuX
B ← B − X
for $b_i$  in B
do
if IoU (X,$b_i$) ≥ $EMOR_{th}$
then
 B ← B − $b_i$;
  DS ← DS-S $S_i$;

end
end
end
return D,DS
End
```

There are many overlapping object detections after object identification on photos or videos, prompting the merging of all detected objects pertaining to almost the same object recognition into a single detection. Exterior Maximal Optimum Repression, the most frequent method, begins with the detection of boundary boxes with a total score. EMOR incorporates it into the final detection set and removes any results from the boundary box that have IoU scores larger than the threshold value. This operation is repeated until the boundary box set is no longer filled.

After object identification on pictures or videos, there are many overlapping object detections, leading the merging of all discovered objects belonging to nearly the same object recognition into a single detection. Exceptional case the most common method, Minimal Optimum Reduction, starts with the detection of border boxes with a total score.

2.2 Identifying the Centroid of a bounding box

$$\text{Cent} = \frac{(X_{\min}+X_{\max})}{2}, \frac{(Y_{\min}+Y_{\max})}{2} \tag{2}$$

The centre point of the bounding boxes is calculated from each of the minimum and maximum values for the corresponding width, Xmin and Xmax and height, Ymin and Ymax of the bounding boxes as shown in Figure 4.

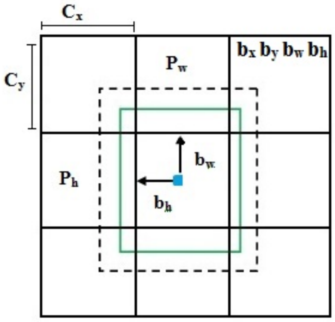

Fig 4. Location and Dimension priors of bounding boxes

Using clusters as anchor boxes, our algorithm predicts bounding boxes. The top, bottom, weight, and height of the detecting bounding box are specified by bx, by, bw, and bh. If the square is displaced from the top left corner of the image by Cx and Cy, and the bounding box cell has width and height of Pw and Ph, then there is an object. Convolution operations of 3 x 3, 5 x 5, or 7 x 7 can be converted from this procedure. The chance of detecting bounding boxes (B) of various ratios is assessed each time the screen location is modified.

$$\mathrm{B}=\left(\mathrm{b}_{\mathrm{x}},\mathrm{b}_{\mathrm{y}},\mathrm{b}_{\mathrm{w}},\mathrm{b}_{\mathrm{h}}\right) \tag{3}$$

2.3 Identifying the Distance between two bounding boxes

$$D(\text{ Cent }_1,\text{ Cent }_2)=\left((X_{\min}+X_{\max})^2+(Y_{\min}+Y_{\max})^2\right)^{1/2} \tag{4}$$

To calculate the distance between the two detected object bounding boxes by using the centroid of the detected bounding boxes. The above formula is used to measure the distance between the bounding box Cent$_1$ and Cent$_{2.}$ After obtaining that centroid, the algorithm calculates the distance using the code's minimum pixel range. According to the obtained prediction box and the ground truth box, Next, we need to calculate the positioning accuracy of the model i.e., Intersection over Union (IoU). We first detect the entire test set pictures to obtain the position of the prediction boxes, and then get the position of the ground truth boxes from the annotation files.

$$Io\,U=\frac{US_u}{\cap S_i} \tag{5}$$

S$_u$ is the total area occupied by the prediction box and the ground truth box and S$_{i\ is}$ the overlapping area between the prediction box and the ground truth box, and. Determine whether the sample is correctly classified according to whether the IoU is greater than the confidence threshold.

2.4 Performance Metrics

We also obtain outcomes, by reviving the social distance prediction problem as a binary classification, Precision, Recall, and F1-Score are the usual evaluation metrics for binary classification issues. The average precision (AP) and mean average precision (mAP) metrics are now used to assess the effectiveness of object detection algorithms. For safe distances, we established various social distance boundaries. We consider a distance between two people to be problematic if it is less than the threshold, and safe otherwise. The unsafe case is classified as a positive class.

The samples predicted correctly as positive samples are TP (True Positives), the samples predicted correctly as negative samples are TN (True Negatives), the samples labeled improperly as positive samples are FP (False Positives), and the samples categorized improperly as negative samples are FN (False Negatives). Precision measures the percentage of positive samples that contain TP, while recall corresponds to the percentage of positive samples that contain TP. Precision, Recall and F1-Score formulas are defined as follows:

$$\text{Precision }=\frac{TP}{TP+FP} \tag{6}$$

$$\text{Recall }=\frac{TP}{TP+FN} \tag{7}$$

$$F1\text{ Score }=2*\left(\frac{\text{Precision }*\text{ Recall}}{\text{Precision }+\text{ Recall}}\right) \tag{8}$$

2.5 Experimental Setup

Google Colab is a free Integrated Development Environment from Google that supports Artificial Intelligence training and research. Google Colab comes pre-installed with major Deep Learning packages like Pytorch, Tensor Flow, GPU, and OpenCV, and it's completely free to use. Normal PCs do not have GPUs because machine learning or deep learning algorithms demand a system with high speed and computational resources. Train the models and put them together for the final output during training. Google Colab was used to train the entire model. Our model was trained at a 640x640 pixel image resolution.

2.6 Dataset

For the analysis of the proposed framework, a specific dataset is used. The data was gathered at Tiruvannamalai, Tamil Nadu, India, which is a popular tourist destination. The analysis of the proposed framework in such a densely populated area would assist us in better analyzing the model's performance. The test dataset consists of 346 RGB frames. Frames are gathered using a Samsung galaxy note 10+ motionless ToF camera mounted 4.5 feet above the ground with regular camera view calibration. Of the 5450 images in our dataset, 3855 images were used as the testing dataset and the remaining ones were used as the testing dataset. Tiny objects in the data set are defined as objects with a size less than 32*32. A moderate object has a dimension between 32*32 and 96*96, while other objects are referred to as big objects, having a dimension greater than 96*96.

2.7 Training Process

The test was done out along with the Yolov5 control all aspects in the execution of this article in the analytical outcomes. The proposed framework YOLOv5 is explained in this article by giving a brief overview of object detection utilizing deep CNN architecture to construct the RoI map that depicts multiple object regions in videos and pictures, hence improving detection accuracy. For accurate and quick searching, the method incorporates an enhanced version of the deep SORT algorithm known as MOTSORT. In this section, we illustrate these features over a wide range of fine films, as well as compare them against a number of prospective detection and tracking technologies.

The approach begins by using the Boundary Box label generator for training to create a boundary box for each indicator. Some types of marks are used in the labeling process. A picture can be hosted by more than one bounding box. In this step, a single class detector model is utilized, with a symbol representing a single training model. The resulting value of the bounding box drawing tool is object dimensions in the form (x1, y1, x2, y2).

3 Results and Discussion

The Pytorch YOLOv5 repository is used to generate our code. Figures 5 and 6 shows the input and output of our model under various scenarios. For a particular image or video captured from a different view of the road, our model was able to identify and recognize the majority of the human and vehicular objects. Figure 5 depicts our effectiveness of the algorithm on images from the videos. In both heavily crowded and less densely populated photos, it was able to locate and categorize the majority of the humans and vehicles.

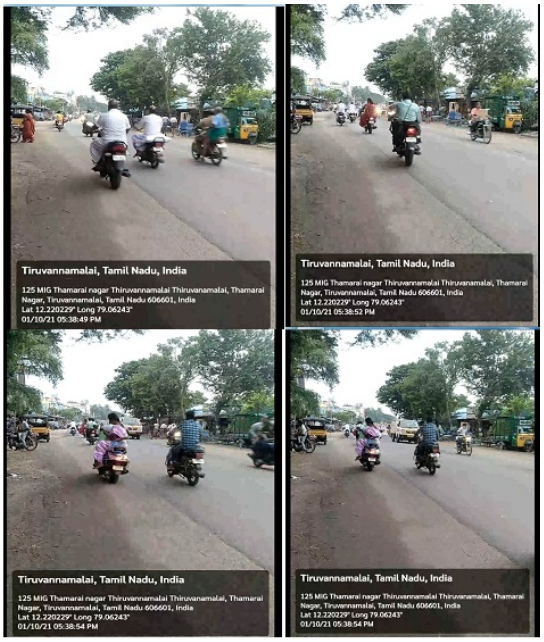
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:49 PM
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:52 PM
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:54 PM
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:54 PM

Fig 5. Input images from videos

There is no requirement for human involvement when using deep learning techniques to identify complex connections in raw data at different levels of abstraction. Machine learning techniques were used for the object detection task. One group of the above-mentioned approaches used a two-stage deep learning method called Faster-RCNN; the other group of solutions used YOLOv1, YOLOv3, and SSD. Yolov5 is used more commonly than other techniques to calculate a human's distance from another person during both stages of social distancing detection[12].

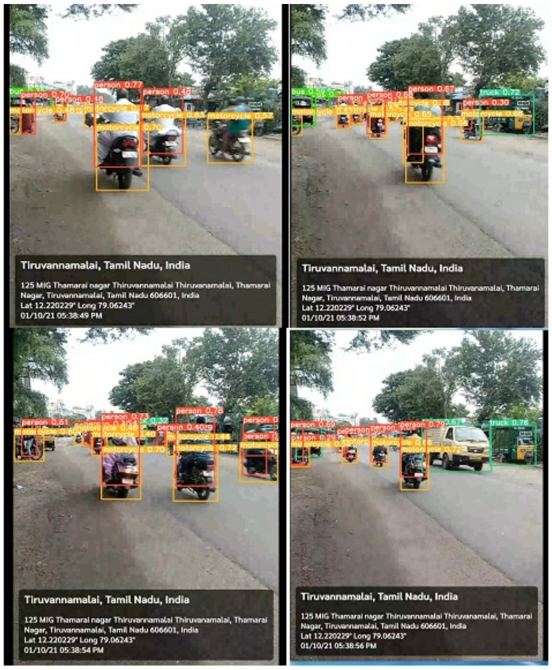
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:49 PM
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:52 PM
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:54 PM
Tiruvannamalai, Tamil Nadu, India
125 MIG Thamarai nagar Thiruvannamalai Thiruvanamalai, Thamarai Nagar, Tiruvannamalai, Tamil Nadu 606601, India
Lat 12.220229° Long 79.06243°
01/10/21 05:38:56 PM

Fig 6. Output images of object detection

3.1 Evaluation Metrics

Precision is used to measure how accurate the model is in making predictions, while recall measures how well the trained model anticipates positive samples. The F1-score, which is always between 0 and 1, is a statistic for overall binary classification performance. An F1-score of 1 indicates a perfect categorization result. These metrics aids in determining the performance of a model that has already been trained in terms of prediction. To evaluate our performance, we used mean Average Precision (mAP) across the training period. The average precision (AP) technique combines the precision-recall gradient into a single figure that indicates the average of all precision. It's a generally used metric for evaluating object detection models' performance. The following is the AP formula:

$$AP_C = \int_{c=0}^{c=n-1} [\text{Recall}(c) - \text{Recall}(c+1)] * \text{Precision}(c) \tag{9}$$

The number of thresholds is n, and the average precision for a class is c. For object detection, the formula for computing mean Average Precision is as follows:

$$mAP = \frac{1}{n}\int_{c=1}^{c=n} AP_C \tag{10}$$

The number of classes is n, and the average precision for class c is APc. During the ensemble, we employed Non-Maximum Suppression and adjusted the confidence threshold for every predicted bounding box.

We compare our algorithm with other existing object detection models such as RetinaNet[13], Faster R-CNN and YOLOv4[14], Visual Social Distancing[15]. We compare our performance outcomes with the existing models using different threshold values. Finally, we report the performance outcomes of our proposed work with high accuracy and efficiency, other state-of-arts of object detection models. The performance outcomes are demonstrated in Figure 7 indicates that our proposed model obtains the highest precision, recall, F1-Score and mAP values.

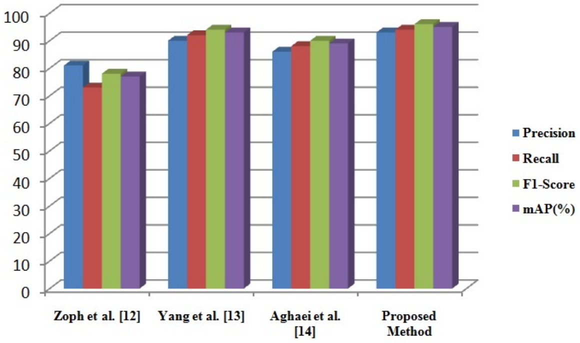

Fig 7. Comparison of Performance outcomes

4 Conclusion

This study provides an effective real-time deep learning-based system for computerizing of social distancing monitoring using object detection and tracking techniques. The proposed framework easily identifies individuals going too close together or vehicles moving close together. Detection and recognizing minimum social distancing in traffic area using surveillance camera, the outcomes demonstrate the applicability of this approach in COVID-19 mitigation in crowded places. This method might be used to identify infected people and infectious locations. It will assist in stopping future respiratory diseases that harm our vocal cords, such as Covid-19. Finally, our methodology could be combined with others for social distance tracking and monitoring individuals. According to the outcomes of the research, the detection model's overall efficiency and accuracy are perfect. In the future, the research could be expanded for various indoor and outdoor surroundings. The priority of future research will be on enhancing object detection techniques to correct adverse information discrepancies and increase geo-location precision, which may be used to promote exact and accurate maximum distance calculation. Various detection and tracking techniques may be used to monitor the individuals or group who really breaking or violating the social distance threshold.

References

1) Sharma T, Debaque B, Duclos N, Chehri A, Kinder B, Fortier P. Deep Learning-Based Object Detection and Scene Perception under Bad Weather Conditions. *Electronics*;11(4):563–563. Available from: https://doi.org/10.3390/electronics11040563.
2) Walia IS, Kumar D, Sharma K, Hemanth JD, Popescu DE. An Integrated Approach for Monitoring Social Distancing and Face Mask Detection Using Stacked ResNet-50 and YOLOv5. *Electronics*;10(23):2996. Available from: https://doi.org/10.3390/electronics10232996.
3) Preethi RA, Anandharaj G. Quantized ternary pattern and singular value decomposition for the efficient mining of sequences in SRSI images. *SN Applied Sciences*. 2020;2(10):1697. Available from: https://doi.org/10.1007/s42452-020-03474-8.
4) Ottakath N, Elharrouss O, Almaadeed N, Al-Maadeed S, Mohamed A, Khattab T, et al. ViDMASK dataset for face mask detection with social distance measurement. *Displays*. 2022;73:102235. Available from: https://doi.org/10.1016/j.displa.2022.102235.
5) Bharathi G, Anandharaj G. Sentiment Classification of Tourist's Opinion on Tourist Places of Interest in South India using Tweet Reviews. *The Computer Journal*. 2021. Available from: https://doi.org/10.1093/comjnl/bxab197.
6) Bharathi G, Anandharaj D. Real time deep learning framework to monitor social distancing using improved single shot detector based on overhead position. *Earth Science Informatics*. 2022;15:585–602. Available from: https://doi.org/10.1007/s12145-021-00758-4.
7) Saeed HH, Alazzawi A. Social distance in object detection: Survey based on cutting-edge deep learning approach. *International Journal of Nonlinear Analysis and Applications*. 2022;13:2865–2880. Available from: https://doi.org/10.22075/IJNAA.2022.27431.3598.
8) Saponara S, Elhanashi A, Zheng Q. Developing a real-time social distancing detection system based on YOLOv4-tiny and bird-eye view for COVID-19. *Journal of Real-Time Image Processing*. 2022;19(3):551–563. Available from: https://doi.org/10.1007/s11554-022-01203-5.
9) Shorfuzzaman M, Hossain M, Alhamid MF. Towards the sustainable development of smart cities through mass video surveillance: A response to the COVID-19 pandemic. *Sustainable Cities and Society*. 2021;64. Available from: https://doi.org/10.1016/j.scs.2020.102582.
10) Yadav S, Gulia P, Gill NS, Chatterjee JM. A Real-Time Crowd Monitoring and Management System for Social Distance Classification and Healthcare Using Deep Learning. *Journal of Healthcare Engineering*. 2022;2022:1–11. Available from: https://doi.org/10.1155/2022/2130172.
11) Ilyas N, Ahmad Z, Lee B, Kim K. An effective modular approach for crowd counting in an image using convolutional neural networks. *Scientific Reports*. 2022;12(1). Available from: https://doi.org/10.1038/s41598-022-09685-w.
12) Sriharsha M, Jindam S, Gandla A, Allani LS. Social Distancing Detector using Deep Learning. *International Journal of Recent Technology and Engineering (IJRTE)*;10(5):146–149. Available from: https://doi.org/10.35940/ijrte.E6710.0110522.
13) Zoph B, Cubuk ED, Ghiasi G, Lin TY, Shlens J, Le QV. Learning Data Augmentation Strategies for Object Detection. *Computer Vision – ECCV 2020*. 2020;p. 566–583. Available from: https://doi.org/10.48550/arXiv.1906.11172.
14) Yurtsever YD, Renganathan E, Redmill V, Ozgüner KA, U. A Vision-Based Social Distancing and Critical Density Detection System for COVID-19. *Sensors*. 2021;21(13):4608. Available from: https://doi.org/10.3390/s21134608.

15) Aghaei M, Bustreo M, Wang Y, Bailo G, Morerio P, Bue AD. Single Image Human Proxemics Estimation for Visual Social Distancing. In: IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE. 2021;p. 2784–2794. Available from: https://doi.org/10.3390/s22020418.