Social Distance Detection Between Peoples in Video Using Deep Learning

Mishra Digvijay and Vishal Saini

Abstract **– This paper present object detection using deep learning in which we used regression based object detection model called YOLO and for identification of different objects we used Euclidean distance. Experiments have been taken to identify objects and measurement of Euclidean distance between moving peoples and we successfully implemented this method on actual footage of street and got results with good accuracy.**

I. INTRODUCTION

In this mean time we are going through corona pandemic and we have to maintain certain distance from others but in crowded places like railway stations, Bus stations and airports we can't put all the peoples under surveillance manually to obey social distancing rules, in such situation we have to develop some method which automatically take care of social distancing rules and report us if violation happens.

While doing this the major problem we have faced is how to identify weather it is Same person moved from one place to another or it is different person and distance between two persons moving at different speeds. It can be in same direction or it can be in different direction. We also have to differentiate between different kinds of objects and human in particular video.

We can use this method in different kinds of places like museum, Railway Stations, Airports, Shopping Malls etc. for maintenance of social distancing rules and can put fine for those who violate the rules. We have successfully implemented this method on one video and we are able to detect number of peoples violating social distancing rules.

II. BACKGROUND AND PREVIOS WORK

There are two types of object detection approaches 1. Traditional approach:The traditional approach of object detection usually has three stages: i) informative region selection, ii) feature extraction, and ii) classification of the object. In the first stage, we try to find the object location. In the second step, we do the feature extraction stage by using techniques like SIFT, HOG to extract the visual feature for recognizing the object. In the third classification stage we use Support Vector Machine(SVM) or Adaboost for classification of target objects from all the other categories and to make the representations more hierarchical, semantic, and informative for visual recognition.

The problem with the traditional approach is that the generation of candidate bounding boxes using the sliding window technique is computationally expensive and also the hand-engineered features are not always sufficient to perfectly describe all types of objects.

2. Modern approach: Deep learning is able to tackle some drawbacks of traditional approaches. The deep learning architectures are able to learn more complex features. There are two types of frameworks available in deep learning object detection models.

The first frame work is region proposal based and it consists of models like RCNN, SPP-NET, FRCNN, FasterRCNN and the second frame work is regression based and consists of MultiBox, AttentionNet, G-CNN, YOLO, SSD, YOLOV2.

Some of the previous work have been done for object detection and Measuring distance between objects for images with OpenCV and we have taken this idea further for videos.

III. DATASET

For training of model we used COCO dataset. The COCO dataset stands for Common Objects in Context, and is designed to represent a vast array of objects that we regularly encounter in everyday life. The COCO dataset is labelled, providing data to train supervised models that are able to identify the common objects in the dataset. If model is still far from perfection, so the COCO dataset provides a benchmark for evaluating the periodic improvement of models through computer vision research. Once the model is trained on the COCO dataset, it can be fine-tuned to learn other tasks, with a custom dataset.

Objects are annotated with a bounding box and class label. Humans are labelled with key points of interest (elbow, knee, etc.). we have only considered object detection of humans in our model training.

Some information about COCO data set : i) The COCO Dataset has 121,408 images. ii) The COCO Dataset has 883,331 object annotations. iii) The COCO Dataset has 80 classes. Iv)The COCO Dataset median image ratio is 640 x 480.
(https://cocodataset.org/#download)

IV. PROCEDURE AND EXPERIMENT

In this method there are mainly three steps : i)Object Detection ii) Object Tracking iii) Distance Measurements Between Detected Objects. We have started with using YOLO object detection technique which can detect up to 9000 classes and then we used COCO dataset to train model using only human label because in this method we are tracking humans and we ignore all other objects in particular video.

For tracking we assign new id to every new detected person, draw box over there and measure the centroid of that box. Lets say we have two centroids one is new and other is old and they are moving from one place to other in that case we calculate Euclidean distance from every old centroid to the new centroid and close pairs mark as the same person and different one assigned a new id. If any id disappear from the frame for consecutive 50 frames then we erase that id from our array.

For distance measurement we use the formula $D = (W*F)/P$ where D indicates distance, W indicates width of object, F indicates camera focal length and P stands for pixels covering in particular image.

Experiments have been taken with the images with different objects like human, cat, dog, horse etc. and we are able to detect humans in that particular image and also calculated distance between 2 or more humans in one particular image and then we used that technique on 25 FPS video of some street.

V. RESULTS

As we have mentioned earlier, we have used YOLO for human detection which is shown in figure 1. Once the object is detected how we do tracking and assign a ID to new person and also classify weather it is same person or new person is shown in figure 2, figure 3 and figure 4. In figure 5 and figure 6 we have shown frames of video which is indicating the

number of peoples who have violated social distancing rules at that particular time.

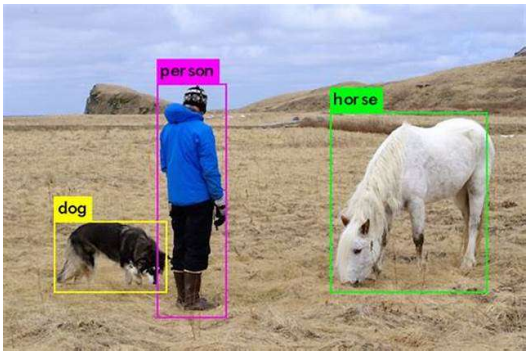

Figure 1

Here we have taken two centroids and assigned a two ID's to them and created box over them and also calculated their centroids.

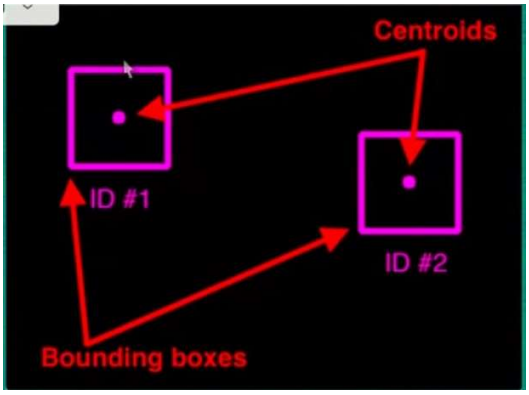

Figure 2

Here we have shown that old centroid is moving and we calculate Euclidean distance to identify old centroid and new centroid.

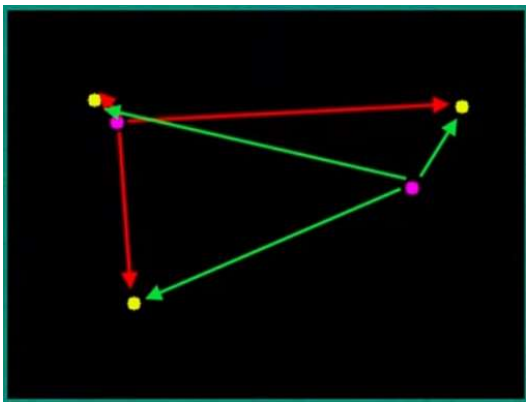

Figure 3

Here we can see, old centroids are classified and new centroid is marked as question mark. Now we will assign ID 3 to new centroid and process will repeat for upcoming frames. By doing this process we have achieved Mean Average Precision (MAP) of 37% and some of the examples have been shown here.

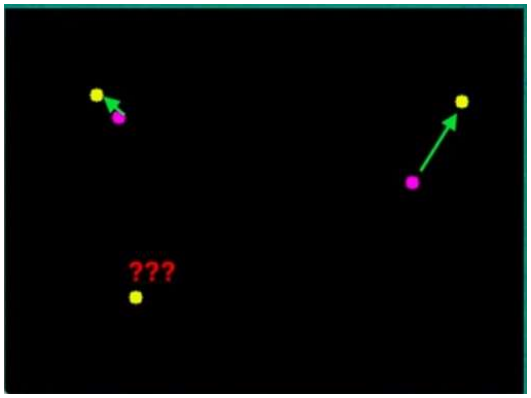

Figure 4

Here we can see 11 peoples have violated rules at that particular time.

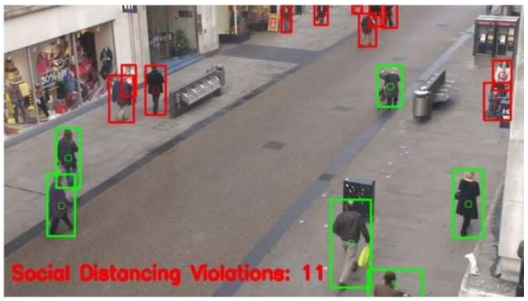

Figure 5

In the next instance we can see, 15 peoples had violated rules.

Figure 6

VI. CONCLUTION

We have used YOLO for human detection and Euclidean distance for tracking of humans. We have calculated distance between different centroids using formula D = (W*F)/P . in figure 5 and figure 6 we have shown that our method is working really well with real life situations where places are crowded. Given method can be used in different places like Railway Stations, Airports and Bus stands so people can follow social distancing rules in this pandemic time. We can combine this method with other methods like face detection and can punish those who violate the social distancing rules by matching these faces with the data collected while entering such places. We can manage to collect these types of data by using face recognition techniques at entrance.

VII. STATEMENT OF CONTRIBUTION

Project concept: Mishra Digvijay
Data Collection: Mishra Digvijay & Vishal Saini
Coding : Mishra Digvijay & Vishal Saini
Experimentation : Mishra Digvijay & Vishal Saini
Report Writing : Mishra Digvijay
Video Making : Vishal Saini