WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Implementation: Social Distancing Detection using YOLOv3 Object Detection and OpenCV

Pradeep Mondal P, Saloni Dhar, Prof. Asha M S, Rajath H R, Yogendra Kumar V R, Dr. Saravana Kumar

DSATM, Bangalore, India, 560082.

ABSTRACT

WHO recommends social distancing as a preventive measure to minimize the spread of COVID-19 and other infectious diseases. All government and healthcare authorities recommend 2 meters as the optimal distance to practice social distancing in crowded areas like schools, markets, etc. The Project uses a Deep-Learning BasedNeural Network model on common CCTV footage to automatically detect people, track them and estimate inter-people distances. We use YOLO object detection framework to accurately track people, even in situations of occlusion, partial visibility and variations in lighting. The output can be used to enforce social distancing in crowded public places to prevent the spread of infectious diseases and can have further applications like identifying areas to prioritize disinfecting and cleaning.

Keywords: "Social Distancing; COVID-19; People Detection and Tracking; Deep Learning; Convolutional Neural Networks"

I. INTRODUCTION

Social distancing refers to a precautionary action to prevent spread of infectious diseases, by minimizing the distance of people physical interactions in dense public places like schools, bus stations, etc. to reduce infection risk.

Artificial Intelligence (AI)has a major role in monitoring social distancing. Computer vision has become more important with various application like video surveillance, self-driving cars and more. It has beenused in solving various health related problemswhichhelps us to extract complicated features from the data. This information is used for understanding images by analysing the features.

As per WHO requirements, “the minimum distance between people must be at least 6 feet in order to effectively maintain social distancing”. This study helps in reducing the spread of the coronavirus by providing an automated solution for monitoring and detecting social distancingviolations among people using “Convolutional neural network (CNN) which is a neural network which has multiple convolutional layers which are used mainly for segmentation, image processing, classification”. The proposed model uses a YOLO based framework for bird’s eye transformation for accurate people detection and social distancing monitoring in various conditions.The model uses common CCTV security cameras for automated people detection, tracking, and inter-people distances estimation in the crowd.

II. LITERATURE SURVEY

“N. Dalal" [5] paper discusses methods used for detecting and tracking of people in a video footage by using the HOG featureshelps to train aSupport Vector Machine (SVM) classifier in a frame. Window of a particular size are cropped out from the image one by one in a Convolutional manner and Histogram of Oriented Gradients (HOG) features is used to train a Support Vector Machine (SVM) classifier for detecting human

beings in any frame. HOG is used to train the SVM for human detection.
Once the human in the frame are detected we need to track them. Initially we find the current position of each person in the frame and continuously compare them to the upcoming frames. k-means clustering is used to identify the location, the points are located for fixed iterations so that we can find a group which has least distance among them and path by connecting all points from previous frame to next frame.

"Ammar Abbas and Andrew Zisserman" [6] perform a bird eye view transformation which gives us a top view of a scene. We apply this method to an image in order to transform the perspective of the image taken to a top view of the image. To help in more accurately measuring the distance between people. This requires only four parameters to be specified (the vanishing point and vanishing line). We show that a CNN trained using our four scalar parameterisations to detect horizon and transform the image.
In any given applications that use a single camera to understand the 3D layout of the scene is always difficult.

Hence, we can use bird eye transformation which can be used for accurate distance prediction as the image is mapped to the top-view of the scene. This method is used in various computer vision applications like video surveillance for detection and tracking. First the image is tilted to get a horizontal view of the scene, then we identify the horizon and vanishing points.
The CNN is initially trained with pre-trained weights from ImageNet classification which contains different scene. Once the CNN is trained, we can use the model to convert the scene to top-view.

"Mahdi Rezaei, Mohsen Azarmi" [1] use a 3-stage model for social distancing which includes detection of people, tracking and distance estimation between the people. This can be used to monitor social distancing in any given place where any type of CCTV Camera is available with real time analysis. The model includes a DNN detector for extracting features and prediction and location of objects. The DNN model has to discover patterns in objects for faster prediction.The proposed method helps in reducing the spread of the coronavirus by providing an automated solution for monitoring and detecting social distancingviolations among individuals.

III. RESEARCH AREA

Artificial Intelligence (AI) has a major role in monitoring social distancing. Computer vision has become more important with various application like video surveillance, self-driving cars and more. It has beenused in solving various health related problems whichhelps us to extract complicated features from the data. This information is used for understanding images by analysing the features.
"Convolutional neural network (CNN) which is a neural network which has more than two convolutional layers which are used mainly for segmentation, image processing, classification**. "**

IV. METHODOLOGY

We used a four-stage model for social distancing which includes taking input video, bird's eye transformation, detection of people, distance estimation between the people and classification of people based on risk.

Working of YOLOv3 Object Detection model: YOLO divides the input image into an S × S grid. If the centre of an object falls into a grid cell, that grid cell is responsible for detecting that object.
To understand the YOLO algorithm, first we need to understand what is actually being predicted. Ultimately, YOLO aims to predict a class of an object and the bounding box specifying object location. Each bounding box can be described using four descriptors:

1. Centre of the box (bx, by)
2. Width (bw)
3. Height (bh)
4. Value c corresponding to the class of an object

Along with that we predict a real number pc, which is the probability that there is an object in the bounding box.

YOLO determines the probability that the cell contains a certain class.

The equation for the same is:

$$score_{c,i} = p_c \times c_i$$

Probability that there is an object of certain class 'c'
The class with the maximum probability is chosen and assigned to that particular grid cell.

YOLO model finds the bunding box with highest-class probabilities and repeats the same process, it is done until we are left with all the different bounding boxes.

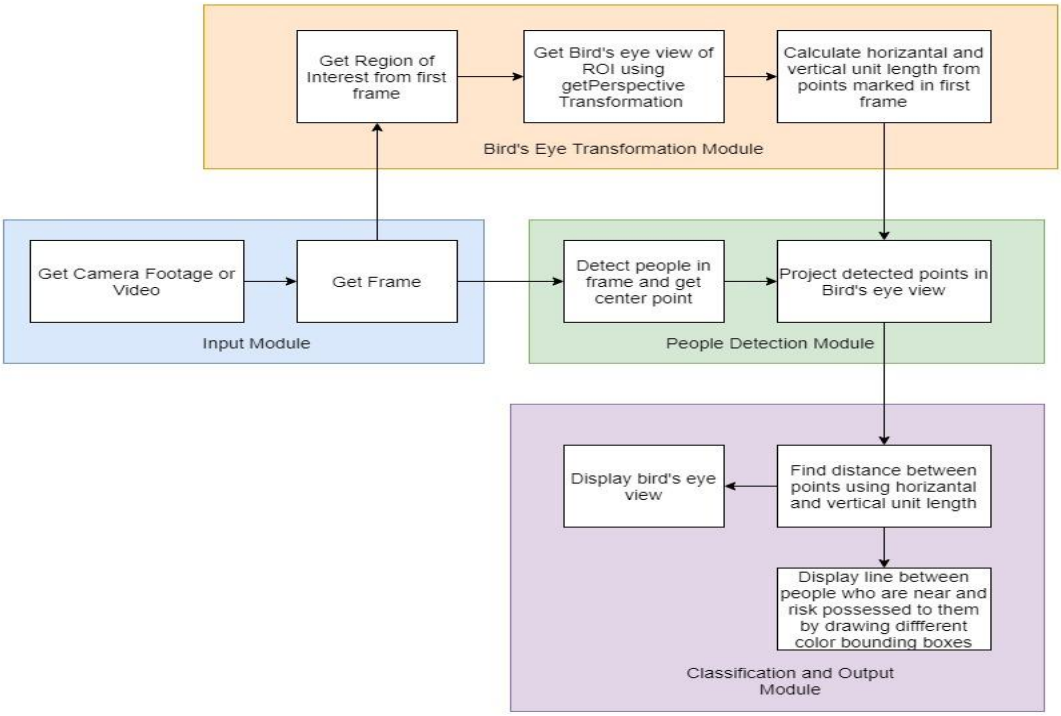

Fig 1. Modules

Input Module:

- In our project, we use numpy, argparse, openCV libraries to work on the videos.
- The argparse module makes it easy to write user-friendly command-line interfaces. The program defines what arguments it requires, using which we can give source and destination paths for input and output videos.
- We import video using openCV'svideoCapture() method which takes video path as an input. We calculate the height, width and framerate of the current video and then we use the object to iterate over the object to get each frame.
- The individual frames are processed for bird's eye view transformation and people detection.

Birds Eye View Transformation Module:

- Get Region of Interest from first frame.
- First the user has to select 4 points in the frame representing the region of interest. This is done using OpenCV's mouse event listener setMouseCallBack().
- Similarly, we take 3 more points to mark 6 feet distance horizontally and vertically.
- Get Bird's eye view of ROI using getPerspectiveTransform().
- Calculate horizontal and vertical unit length from points marked in first frame.

People Detection Module:

- Configure the network from the yolo model weights and configuration.
- Take each frame and preprocess it to feed it into the network.
- Feed the preprocessed frame as input to the deep neural network and the output.
- Compare the confidence scores against threshold to detect people in frame and get center point of each people's bounding box.
- Project detected points to Bird's eye view.

Classification and Output Module:

- Find distance between points using horizontal and vertical unit length.
- Classify the risk based on distance.

- Display bird's eye view.
- Display line between people who are near and risk possessed to them by drawing different color bounding boxes.

YOLOV3 model performance analysis

You only look once (YOLO) is a state-of-the-art, real-time object detection system. On a Pascal Titan X it processes images at 30 FPS and has a mAP of 57.9% on COCO test-dev.YOLOv3 is extremely fast and accurate. In mAP measured at .5 IOU YOLOv3 is on par with Focal Loss but about 4x faster. Moreover, you can easily tradeoff between speed and accuracy simply by changing the size of the model, no retraining required! This makes it ideal for our application as we require real time people detection even if the accuracy is moderate.

Snapshots:

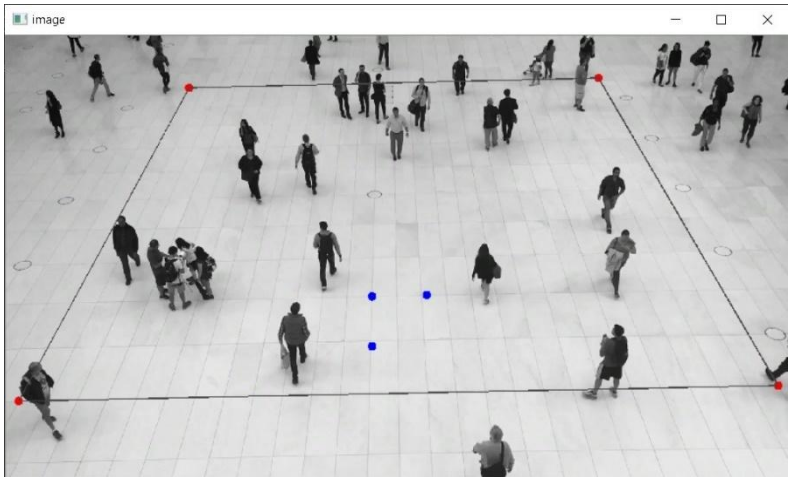

Figure 2 – Selecting Region of Interest (ROI) and distance scale.

Figure 3 – Detecting People in the frame and calculating inter-people distance

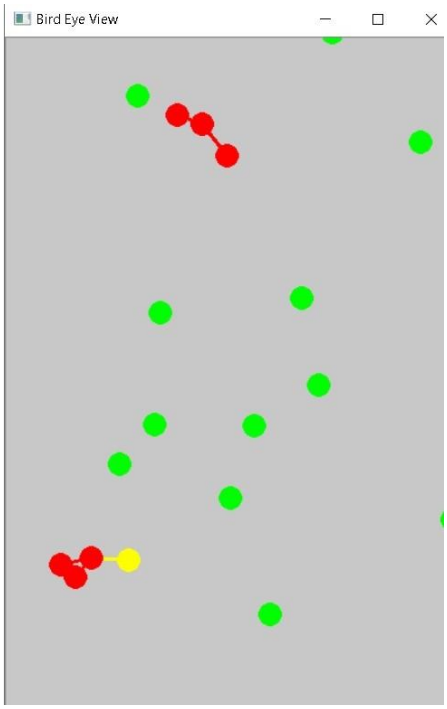

Figure 4 – Bird's Eye View Projection

V. CONCLUSION AND FUTURE WORK

The proposed method helps in reducing the spread of the coronavirus by providing an automated solution for monitoring and detecting breaking rules of social distancing among individuals. The proposed people detector model is able to detect and track stationary and moving people in public places and monitor and maintain social distancing standards in COVID-19 era and later. This Method is applicable in various placewhich uses any type of CCTV cameras.

This helps to identify zones based on risk, or bring more restrictions to certain areas and limited movement is allowed toparticular zones or re-design the scene of the environment to make it a saferplace. Based on the nature of environment and situation the restriction rules are varied. We can categorize people into various categories like safe, high-risk.

Future Enhancements: The proposed method can be used in bright environment where there is sufficient light for the camera to captured and therefore, we have more information in the image for image processing which results in more accurate results. In the night/low-light environment we can use thermal camera to capture the source of heat in the image. The speed and accuracy can also be increased from faster real-time processing.

ACKNOWLEDGEMENT

- We extend our sincere thanks to Dr. Saravana Kumar, Dept of CSE, DSATM
- We also would like to thank our project guide Assistant Prof. Asha M S, Dept of CSE, DSATM

REFERENCES

[1] Mahdi Rezaei, Mohsen Azarmi, "DeepSOCIAL: Social Distancing Monitoring and Infection Risk Assessment in COVID-19 Pandemic"
[2] World Health Organisation. WHO Corona-viruses Disease Dashboard (August 2020). Available at https://covid19.who.int/table.
[3] WHO Director, Generals. Opening remarks at the media briefing on COVID-19 (2020). WHO generals and Directors speeches.
[4] Nida M. Zaitoun, Musbah J. Aqel, "Survey on Image Segmentation Techniques"
[5] N. Dalal, B. Triggs and C. Schmid, "Human Detection and Tracking for Video Surveillance"
[6] Ammar Abbas and Andrew Zisserman, "A Geometric Approach to obtain a bird's eye view from an image"
[7] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement"
[8] https://pjreddie.com/darknet/yolo/
[9] https://towardsdatascience.com/yolo-you-only-look-once-3dbdbb608ec4
[10] https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006