Real Time Human Tracking and Social Separation System Using Yolov5

A project report submitted in fulfillment of the requirements for B.Tech. Project

B.Tech.

by

Abhishek Pratap Singh (2018IMG-007)

विश्वजीवनामृतं ज्ञानम्

ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR-474 010

2021

CANDIDATES DECLARATION

We hereby certify that the work, which is being presented in the report, entitled **Real Time People Detection and Social Distancing Measure System for COVID-19**, in fulfillment of the requirement for the award of the Degree of **Bachelor of Technology** and submitted to the institution is an authentic record of our own work carried out during the period *June 2021* to *october 2021* under the supervision of **Dr. Sunil Kumar**. We also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

Date: 31-10-2021 Signatures of the Candidate

Abhishek

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date: 31-10-2021 Signatures of the Research Supervisors

Sunil Kumar
25/10/2021

ABSTRACT

COVID-19 is a respiratory disease that causes severe illness. It was discovered in December 2019 in Wuhan, China. It has resulted in a continuing pandemic with numerous cases of infection and deaths. Coronavirus is transmitted mostly through human-to-human contact. This study proposes a image-based artificial intelligence system for social distancing classification. The YOLOv5 approach is used to construct a revolutionary deep learning object detection technique for recognising and tracking individuals in indoor and outdoor environments. An algorithm is also used to automatically evaluate whether or not social distancing rules are being met, as well as to measure and evaluate the distance between people. As a result, the aim of the project is to discover if and how individuals adhere to social distancing norms in order to stop the COVID-19 virus from spreading. The suggested method is used to construct a comprehensive AI system for people tracking, social distance classification using camera photographs. Datasets obtained from distinct CCTV cameras are used in the training phase.

Keywords: YOLO, COVID, Object-detection, CCTV

ACKNOWLEDGEMENTS

I am highly indebted to **Dr. Sunil Kumar**, and obliged for giving me the autonomy of functioning and experimenting with ideas. I would like to take this opportunity to express my profound gratitude to him not only for his academic guidance but also for his personal interest in my project and constant support coupled with confidence boosting and motivating sessions which proved very fruitful and were instrumental in infusing self-assurance and trust within. The nurturing and blossoming of the present work is mainly due to his valuable guidance, suggestions, astute judgment, constructive criticism and an eye for perfection. He always answered myriad of our doubts with smiling graciousness and prodigious patience, never letting me feel that I am a novice by always lending an ear to my views, appreciating and improving them and by giving me a free hand in this project. It's only because of his overwhelming interest and helpful attitude, the present work has attained the stage it has.

Finally, I am grateful to our Institution and colleagues whose constant encouragement served to renew my spirit, refocus the attention and energy and helped in carrying out this work.

(Abhishek Pratap Singh)

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This section includes the an introduction about the COVID-19 situation and object detection.In this section a brief description of this project in which an Object detection model is Implemented by YOLOv5 Algorithm.

1.1 Introduction

COVID-19 pandemic has now spread to 188 nations worldwide. According to World Health Organisation (WHO), there have been 247,352,866 confirmed COVID-19 cases and 5,013,900 deaths worldwide as of October 31, 2021. It's symptoms include fever and chills. Researchers in China discovered that persons infected with the coronavirus had a high temperature in 99 percent of cases. The main goal is to create a comprehensive tool as well as effective technology for enforcing social separation. Technologies may play an essential role in facilitating the practise of social distancing. This project attempts to stop the virus from spreading throughout communities and save people's lives. In this project, we offer a deep learning object detection model for people tracking that is used in conjunction with an image based social distancing classification technique.

Social separation involves making adjustments to your usual life in order to decrease your close contact with others, such as:

- Being away from crowded areas and unimportant assemblies.
- Restricting interaction with persons who may be at high risk.
- Maintaining a safe distance of at least one metre from others.
- Avoiding common greeting like handshakes.

1.1.1 Object Detection

Object detection is a common Computer Vision task that involves identifying and locating specific classes of objects in an image. Object localization can be interpreted in a variety of ways, such as by drawing a bounding box around the object or by marking every pixel in the image that contains the object (called segmentation).

Object detection systems use a set of training samples to build a model for an object class. Although only one example may be required in the case of a fixed rigid object, many training instances are often required to capture some features of class variability.

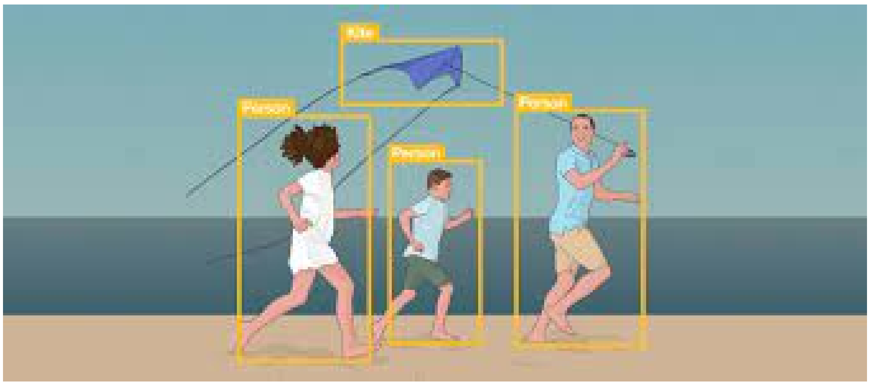

Figure 1.1: Object Detection [1].

1.2 Motivation

Corona virus is a serious virus that has resulted in the death of numerous people. Researchers in laboratory have discovered that the coronavirus can transfer from one person to another. Fever, cough, chills, shortness of breath, body aches, and loss of taste and smell are the most common COVID-19 symptoms. Coronavirus had a massive impact on numerous sectors of the world, including industrial, transportation, and agriculture. As a result of this impact, the world was forced to halt all activities and impose stringent guidelines for social distance and wearing a face mask on a priority basis [7]. Coronavirus's unpredictability, instability, and complexity have made it difficult to anticipate the pandemic's duration and spread. Nations suspended operations and closed the border as well as public spaces such as schools and offices to prevent people from coming into contact. According to reports, all infected countries that implemented the lock-down for their communities saw a decrease in the number of COVID-19 cases and deaths caused by the pandemic. According to the World Health Organization, social distancing is the recommended strategy for minimising physical contact with potential COVID-19 carriers by keeping a set distance between one person and another.

CHAPTER 2

LITERATURE SURVEY

2.1 Literature Review

In order to avoid the spread of disease, social distancing and temperature screening are helpful methods. Many organisations, including the World Health Organization, have recommended them. Russell et at [2] investigated the effects of social distancing tactics on coronavirus transmission. Using susceptible exposed infected removed (SEIR) methodologies, this paper offered scientific location contact patterns to produce the course of an outbreak. The authors also stated that a rapid easing of social barriers could accelerate the virus's infection and dissemination among humans.

The influence of social distancing tactics was underlined by Nabil Kahale [3]. The goal of the study was to come up with a rough estimate of how early social distancing techniques can significantly reduce economic loss and the number of new infections. Jennifer Berglund [3] proposed using GPS and built-in applications in smartphones to track a person infected with COVID-19. However, this technology has limits when it comes to tracking those who don't have access to Wi-Fi or phone service. Some authorities, on the other hand, use drones with video cameras installed on them to monitor crowds in the open air. Such technology is suitable for monitoring COVID-19 which could amid the coronavirus outbreak.

Thanks to advancements in computer vision and deep learning in general, the problem of identifying and recognising objects in an image has just been solved. As a result, computer vision research has concentrated on a variety of intriguing and difficult subjects, including neural style transfer, segmentation, and tracking, as well as, of course, object recognition [5].

Deep learning is an artifcial intelligence function (AI) that emulates the tasks of the human brain in data processing and object detection. It can be referred to as a neural network with a sophisticated algorithm. The history of the neural network dates to 1940 [6]. The neural network was created with the goal of ethically solving learning challenges. In deep learning models for object detection, a convolutional neural network (CNN) is commonly utilised. CNN is a deep learning method that takes an input image and assigns learnable weights and biases to multiple classes in the image, allowing them to be distinguished from one another. The evolution of a convolutional neural network that can be implemented on an embedded system with low-resolution input and low complexity [8]. There are various deep learning models such as R-CNN, Single-shot detector (SSD), and YOLO which are applied in diferent applications for object detection. These models are efcient algorithms for movement estimation in video scenes [7].

N. Singh Punn, S.K. Sonbhadra, S. Agarwal [10] has put together a system that uses the YOLO OR model to separate persons from the frame and the Deepsort technique to follow the discovered individuals using bounding boxes and assigned IDs. In terms of mAP, FPS, and loss values, the YOLO v3 model's results are compared to those of other well-known models such as quicker region-based CNN and SSD. Researchers proposed a method based on static crowds for a group of people who stayed in the same place for a long time in approach [11]. They used a support vector machine (SVM) to identify patches as essential crowds, and text features were used to extract these patches. Recent developments showed that the identifcation of individuals through video surveillance cameras can be achieved by face and a person's manner of walking. However, the detection of a person under crowds' technique is difcult and hard to optimize. [7].

The authors developed a method for detecting pedestrians with a low-resolution camera using background reduction and real-time classification of foreground silhouettes in method [12]. Sergio Saponara, Abdussalam Elhanasi and Alessio Gagliardi [7] proposed an artificial framework for the social distancing grouping of people utilizing thermal pictures. Using the YOLOv2 based approach, a detection procedure is produced which is deep learning based and is used for distinguishing and following individuals both in outdoor and indoor situations, But due to being developed on Yolov2 its exactness is very low on objects far away from camera. It performs somewhat similar when two or more object nearly identical to each other are in close vicinity.

Yolov5 received 140 FPS on Tesla P100, with the mean average precision (mAP) of 0.895 after training for just 100 epochs[16]. therefore in comparison to the other algorithms Yolov5 shows us thepossibility of receiving higher fps rates as well as accurate predictions.

2.2 Objectives

The goals in this project are:

1. Track Humans present in a video stream and calculate distance between them.

2. Improve the real-time performance of the previous system.
3. Improve the performance without increasing the memory size of the algorithm as this can be utilized in the CCTV systems.

CHAPTER 3

SYSTEM ARCHITECTURE AND METHODOLOGY

This Section discusses the architecture of the algorithm and other methodology that will be used to implement the project.

3.1 Neural Network Design

This System is centred on a solution for continually determining social separation using YOLO object discovery on video film and photographs. To recognise persons in a video, the system employs the YOLOv5[13] object detection algorithm. Using recognised bounding box data, the detection model separates people into groups.

The Euclidean distance is used to resolve the pairwise distances between the centroid of the distinct bounding boxes of the individuals. We use an estimate of the real distance to pixel and setting and edge to detect social distance infringement between persons. To determine whether the distance esteem exceeds the base social distance limit, an infringement edge is set up. The framework enables faster derivation times, making it suitable for delivering real-time results without sacrificing precision, even in very complicated setups.

3.1.1 Object detection using YOLOv5

YOLO v5 provides a faster surmising speed of up to 40fps with a one-third the model size and a 60-70 percent improvement in exactness in comparison to Yolov3. YOLO v5 also has a high level of accuracy when it comes to distinguishing smaller and farther away objects. There are 4 different network model for Yolov5 , they are Yolov5s (small network), Yolov5m (medium network), Yolov5l (large network), Yolov5x (extra large network). Out of the four organization strategies for Yolov5, we can use as light as required, for example, Yolov5s network design or Yolov5m.

YOLO v5 is a single-stage object detector, therefore it has three parts like any other single-stage object detector.

1. Model Backbone
2. Model Neck
3. Model Head

The basic purpose of Model Backbone is to extract key features from an input image. CSPNet is used as a backbone in Yolov5 to extract features from an input image. The model neck is used to make feature pyramids. Feature pyramids helps models in recognising the same object at various sizes and scales. PANet is utilised as the neck in Yolov5 to obtain feature pyramids. The final detection is carried out by the model Head. It creates final output vectors with class probabilities, objectness scores, and bounding boxes after applying anchor boxes to features. The head of the Yolov5 model is same as that of the Yolov3 and Yolov4 models. Leaky ReLU and sigmoid function are the activation functions employed. Leaky ReLU and sigmoid functions are used as activation functions. In the middle layers, the leaky ReLU is applied, and in the final detection layer, the sigmoid activation function is used. For the loss computation of class probability and object score, Yolov5 uses Binary Cross-Entropy with Logits Loss function.[14]

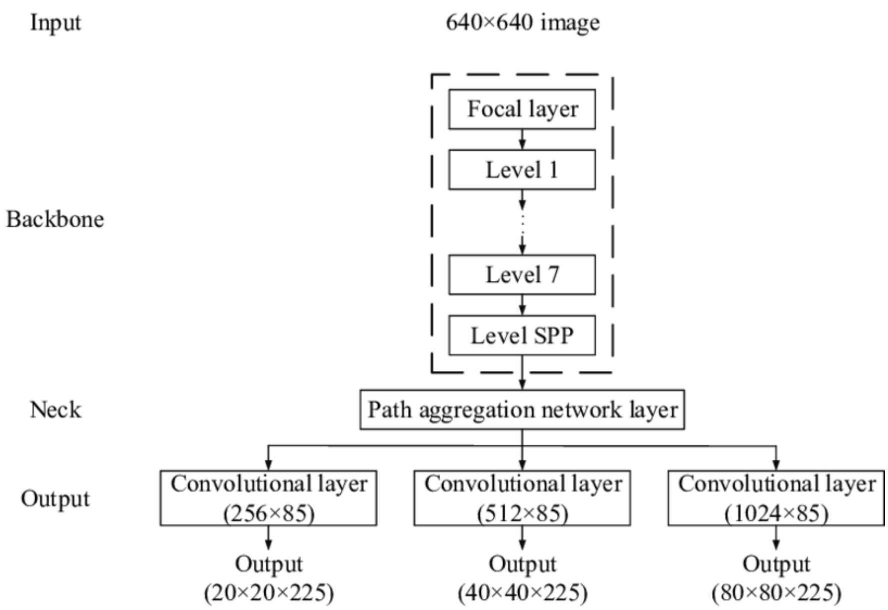

Figure 3.1: YOLOv5 Achitechture. [15]

3.2 Social Distancing Detector Steps

This section explains the necessary steps for establishing a workflow for monitoring social distancing on images, as seen in Figure below.

1. Prepare video from a CCTV camera which contains people.
2. Applying the deep learning object detector to detect people in images or video streams.
3. Check the number of persons that are in the images or video stream.
4. Compute the distance between the centroid of the bounding boxes which are enclosed to the detected people.
5. Finally, the algorithm will decide for safe or unsafe social distancing based on the number of persons and the measured distance between the centroid of bounding boxes.

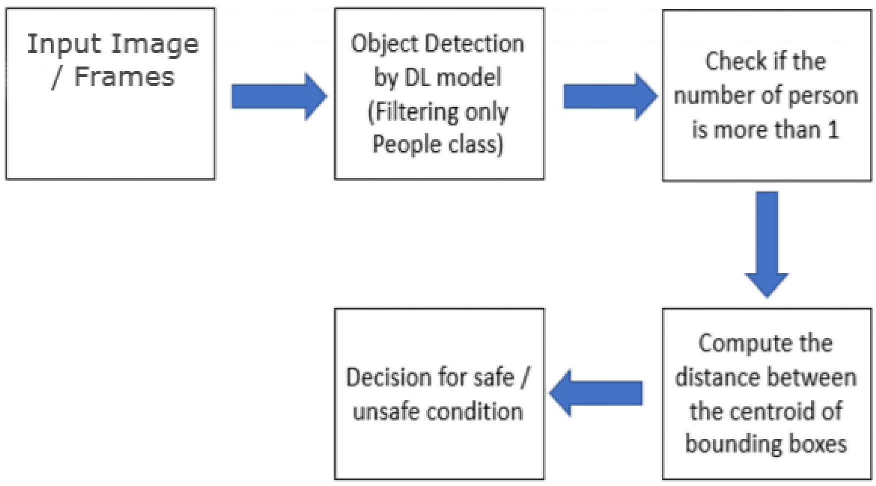

Figure 3.2: The steps involved for people detection and social distancing classifcation on thermal images. [7]

3.3 Algorithm For Distancing Classification

The code was implemented in Python to work with bounding boxes of a detected person in the images.This code categorises and determines whether or not the people in the image are within safe distances. The bounding boxes were given a green tint for safe social separation and a red colour for unsafe social distancing. First, we determine the number of people in each photograph.

A green colour is assigned to a bounding box of detected persons if it is just one. When there are two or more people. A function is implemented to find the color, this function determines whether the bounding box is two or more, as well as the distance between the centroid of the bounding boxes for the detected individual. The centre points of bounding boxes, C (x, y), are calculated using the equation shown in Eq.

$$C(X,Y) = \frac{X_{min} + X_{max}}{2}, \frac{Y_{min} + Y_{max}}{2} \tag{3.1}$$

where C is the centroid of the bounding box. The minimum and maximum values for the corresponding width of the bounding box are X_{min} and X_{max}. The minimum and maximum values for the corresponding height of the bounding box are Y_{min} and Y_{max}.

To measure the distance C1(X_{max}–X_{min}, and C2 (Y_{max}–Y_{min}), between the center of each bounding box, we used the Euclidean formula see Eq(2.2),where distance between pixels is converted to a metric distance (based on the camera's range and field of vision) and then compared to a threshold value. These boxes will be coloured red if the find color function discovers two bounding boxes with a distance less than the threshold value.

$$D(C1, C2) = \sqrt{(X_{min} - X_{max})^2 + (Y_{min} - Y_{max})^2} \tag{3.2}$$

where: D is the distance between the centers of bounding boxes.

When this method discovers two bounding boxes with a distance greater than the threshold value, the colour for these boxes will be green.A top view of the ground plane is required to correctly compute the distance between the people. By performing a homography transformation to the four point coordinates in the angled view, this can be accomplished. These four points can be transformed as shown in the Eq.(2.3)

$$\begin{bmatrix} X_{cr.top} \\ Y_{cr.top} \\ 1 \end{bmatrix} = M * \begin{bmatrix} X_{cr.ang} \\ Y_{cr.ang} \\ 1 \end{bmatrix} \tag{3.3}$$

where $X_{cr.ang}$ and $Y_{cr.ang}$ represent the pixel coordinates of one of the four points in the CCTV view image $X_{cr.top}$, $Y_{cr.top}$ represent the same point after transformed to the top view. M is the homography matrix.

To estimate the distance between persons in the real world, Eq. (3.3) and four point coordinates with homography matrix value are used to calculate the distance between individuals. The distance between the individuals is then scaled by factor S to obtain the real-world distance between them. The scaling factor S is calculated by counting the number of pixels in an image that corresponds to one metre in real life.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results

To evaluate the performance of an Object detection algorithm we must consider many parameters such as mAP (mean average precision), recall, precision, and inference time, below the statistics are presented with graphs.

1. The values of mAp for Yolov5s and Yolov5l are 0.925 and 0.930 respectively.

2. The precision parameter for Yolov5s and Yolov5l is 0.911 and 0.92 respectively.

3. The recall parameter for Yolov5s and Yolov5l are 0.850 and 0.874 respectively.

4. Yolov5s is giving us 31-36 FPS.

5. Yolov5l is giving us 29-34 FPS.

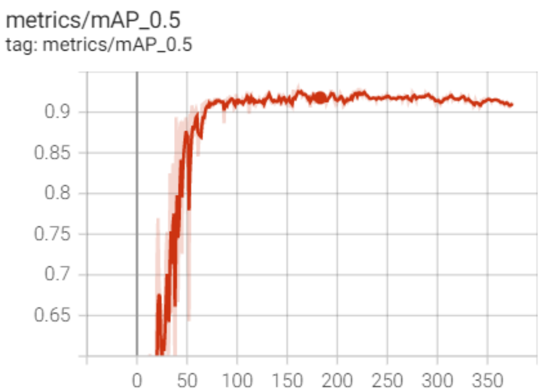

Figure 4.1: YOLOv5s mAp.

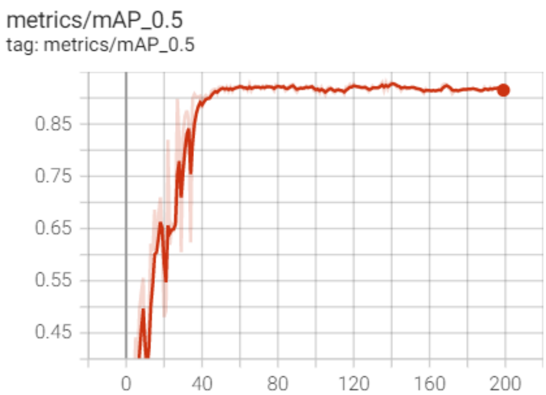

Figure 4.2: YOLOv5l mAp.

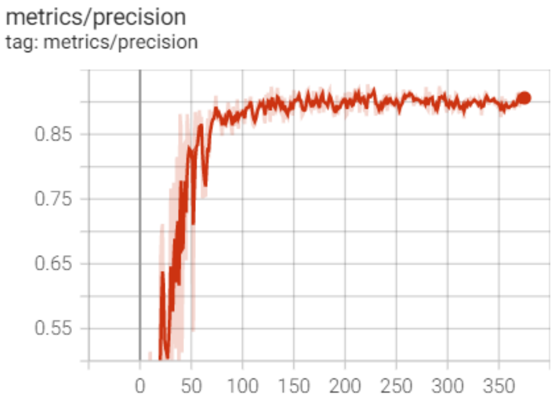

Figure 4.3: YOLOv5s precision.

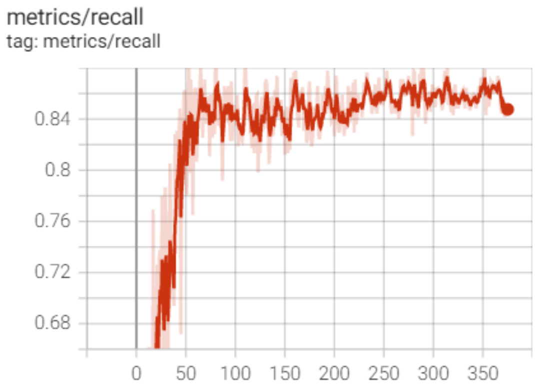

Figure 4.4: YOLOv5s Recall.

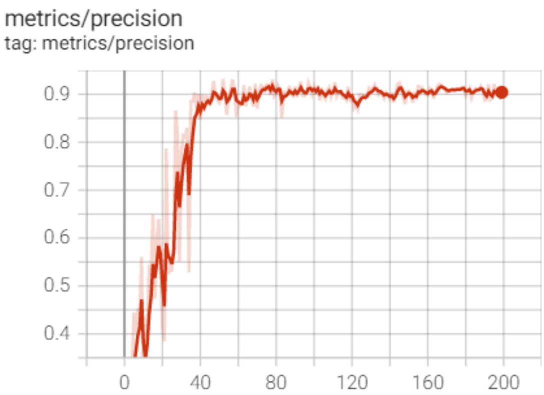

Figure 4.5: YOLOv5l precision.

Figure 4.6: YOLOv5l Recall.

Figure 4.7: Input Image

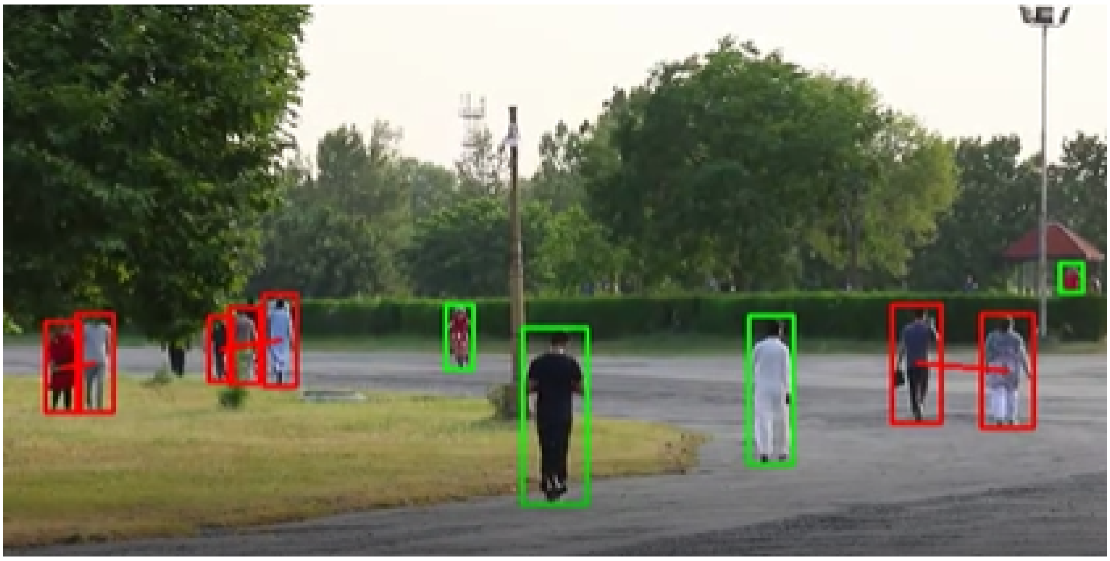

Figure 4.8: Resulting Image after processing through Yolov5 and distance detection steps.

4.2 Conclusion

This system, based on Yolov5, proposes a Social Distancing Tracker that may also be used with real-life cams. However, in this case, we used an input video to build our model's statistics. Yolov5 has four different network models, however We've trained on Yolov5s and Yolov5l only. Both of the above mentioned algorithm showed better results for real-time performance against the previous work. We have achieved upto 36 fps using the smaller version.

4.3 Future Scope

1. Can be extended to 3D through which we can eliminate the perspective effect.
2. The proposed approach can be implemented in a distributed video surveillance system and drones.

REFERENCES

[1] 2021. Computer Vision Object Detection: challenges faced. Exposit. <www.exposit.com/blog/computer-vision-object-detection-challenges-faced/>

[2] Prem, K., Liu, Y., Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N., Flasche, S., Clifford, S., Pearson, C.A., Munday, J.D. and Abbott, S., 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. The Lancet Public Health, 5(5), pp.e261-e270.

[3] Kahalé, N., 2020. On the economic impact of social distancing measures. Available at SSRN 3578415.

[4] Berglund, J., 2020. Tracking COVID-19: there's an App for that. IEEE pulse, 11(4), pp.14-17.

[5] Georgievski, Borijan, 2020. IObject Detection and Tracking in 2020. Medium.

[6] Pitts, W. and McCulloch, W.S., 1947. How we know universals the perception of auditory and visual forms. The Bulletin of mathematical biophysics, 9(3), pp.127-147.

[7] Saponara, S., Elhanashi, A. and Gagliardi, A., 2021. Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19. Journal of Real-Time Image Processing, pp.1-11.

[8] Wang, X., Ng, H.W. and Liang, J., 2017, November. Lapped convolutional neural networks for embedded systems. In 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 1135-1139). IEEE.

[9] Kajabad, E.N. and Ivanov, S.V., 2019. People detection and finding attractive areas by the use of movement detection analysis and deep learning approach. Procedia Computer Science, 156, pp.327-337.

[10] Punn, N.S., Sonbhadra, S.K., Agarwal, S. and Rai, G., 2020. Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. arXiv preprint arXiv:2005.01385.

[11] Manfredi, M., Vezzani, R., Calderara, S. and Cucchiara, R., 2014. Detection of static groups and crowds gathered in open spaces by texture classification. Pattern Recognition Letters, 44, pp.39-48.

[12] Alahi, A., Bierlaire, M. and Vandergheynst, P., 2014. Robust real-time pedestrians detection in urban environments with low-resolution cameras. Transportation research part C: emerging technologies, 39, pp.113-128.

[13] Jocher, G., 2020. <https://github.com/ultralytics/yolov5>

[14] Rajput, M., 2020. YOLO V5 — Explained and Demystified. Towards AI.

[15] Chen, W., Liqiang, Z., Tianpeng, Y., Tao, J., Yijing, J. and Zhihao, L., 2021, August. Research on the state detection of the secondary panel of the switchgear based on the YOLOv5 network model. In Journal of Physics: Conference Series (Vol. 1994, No. 1, p. 012030). IOP Publishing.

[16] Nelson,J., Solawetz,J., 2020, YOLOv5 is Here: State-of-the-Art Object Detection at 140 FPS. <blog.roboflow.com>