2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT 2022)

Indore, India
23-24 April 2022

IEEE Catalog Number: CFP2218P-POD
ISBN: 978-1-6654-8039-0

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP2218P-POD
ISBN (Print-On-Demand):	978-1-6654-8039-0
ISBN (Online):	978-1-6654-8038-3

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC. proceedings.com

2022 11th IEEE International Conference on Communication Systems and Network Technologies

CSNT 2022

Table of Contents

Track-B: Power Systems, Energy and Control System

Track C: Digital Implementation and Application

Track D: Machine Learning, Deep Learing and Optimisation techniques

Track E: Data Analytics, Social Networking, Neural Networks and Automation

Track F: Network Protocols, Design, Algorithm and IOT Applications

Track -G: Network Security and Cryptography

Track -H: Signal and Image Processing

2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT) | 978-1-6654-8038-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/CSNT54456.2022.9787660

Monitoring Social Distancing based on Regression Object detector for reducing Covid-19

Ruchi Jayaswal
Department of CSE
MITS
Gwalior, M.P., India
ruchi.jayaswal23@gmail.com

Manish Dixit
Department of CSE
MITS
Gwalior, M.P., India
dixitmits@mitsgwalior.in

***Abstract*— To deal with the worldwide coronavirus epidemic situation, the procedure of pulling down the Covid-19 cases will be tough to handle, if people do not take steps to thwart the virus from spreading. One of the most significant strategies in this pandemic is to keep a safe distance among the persons in public areas. This paper aims to detect persons with social distance monitoring as a preventive technique in minimizing physical connection between persons. The workflow of this paper is to detect the people in the area of interest using the YOLOv5 model. The model is trained on the Open Images V6 dataset and takes categories of people and human faces. Further, a social distancing algorithm is applied to check the distance between two persons and mark the red boundary box if any person violates the rules. The network provides inference speed capable of offering real-time results with maximum accuracy. The accuracy is achieved 99% in real-time scenarios using the proposed work. In various contexts, the suggested social distancing strategy produces promising outcomes.**

***Keywords*— Social Distancing, Covid-19, Deep Learning, YOLOV5, CNN models, Human Detection**

I. Introduction

Coronavirus is a contagious disease that can be turned out to be an pandemic, proclaimed by the WHO (World Health Organization). In late 2019, Wuhan, China, was the first place where this virus was reported. But then the COVID-19 global pandemic hits the world. As of 30th November 2021, there have been more than 261 million confirmed cases globally. According to the WHO report, some vaccines have been invented, and approximately 7 billion vaccine doses are administered until 28th November 2021[5]. Still, the second and third waves of COVID hit worldwide with new coronavirus variants, so all people must take preventive measures. The WHO has issued rules to limit people's exposure to the virus. Therefore, the government has taken strong and necessary steps to control the spread worldwide. To address the situation of Covid-19, the citizens must follow the guidelines of government such as wearing a mask and maintaining the social distancing in public areas like the shopping mall, public road, hospitals, schools, colleges, workplaces, metro stations, airports. This paper aims to make a framework for keeping the social distancing rules in public places, i.e., WHO's recommended solution to minimize the virus among people.

In other words, keeping a distance of at least six feet between persons is an effective strategy for avoiding physical contact with possible new variants of coronavirus carriers. Figures 1 (a) and (b) depict the social distancing preventive measure taken in open areas with some signs on the roads that prevent people from gathering [1].

Artificial Intelligence(AI) can be useful in facilitating the monitoring of social distancing. Computer vision is a subclass of AI that has demonstrated its promise in CT-Scan, or X-ray-based Covid-19 detection[1] can also assist in social distance monitoring. Moreover, Deep Neural Networks (DNN) help us to obtain perplexing features from images, allowing us to analyze and classify these features to provide a more accurate interpretation of the images. Social distance monitoring uses the person's detection algorithm to locate the persons accurately in the region of interest and then evaluate the distance among the detected people. Deep learning models such as Single Shot Detector, Mobilenet, Resnet, YOLO models, etc., can train the dataset to detect the objects in the real-time frames.

(a) (b)

Figure 1. Social distancing enforced in public place a) Marked circle on road b) Marked line on the road

The study of this paper is to present a method of social distancing monitoring using YOLOv5 and demonstrate the analysis with CNN models and compare with the existing methods.

The organization of the paper is described in four sections. Section I presents the introduction part about social distance monitoring. Section II describes the research background of this work. The research methodology is described in section III. The experimental and result analysis with other methods is demonstrated in section IV. At last, Section V discusses the conclusion part.

978-1-6654-8038-3/22/$31.00 ©2022 IEEE
DOI: 10.1109/CSNT.2022.110

II. Related Research

A. Object Detectors Based on Regression

Object Detectors based on Regression, such as Single Shot Detector (SSD) and You Only Look Once (YOLO), be substantially faster than region-based object detectors [3,6]. However, YOLO has been more widespread and produced more accurate recognition results among other similar object detectors. New versions of YOLO such as YOLOv3, YOLOv4, YOLOv5, and many more have been released with more detection accuracy. YOLO takes the complete image as input, unlike area-based detectors, which generate region predictions that are forwarded to the classifier. As a result, it is much faster than other detectors. The model workflow starts with an RGB image that has been split into G×G grid cells. Each grid cell is crucial to determining B boundary boxes. Each bounding box is given five values: a, b, h, w, and k. The height and weight of a bounding box are h and w, respectively, whereas the coordinates of a bounding box's centre point relative to a grid cell are a, b. The presence of an object in a bounding box has a confidence score of k[6]. The object detector's output for class probabilities C is a shape tensor.

$$(G \times G \times (B \times 5 + C)) \quad (1)$$

B. Object Detection based on Machine Learning

Viola and Jones [7] proposed a very common solution for object detection in 2001. Haar features are used extract features, and for categorization purposes, they used cascade classifiers with the Adaboost learning method. This technique is fifeteen times faster than the traditional methods. By combining kinematic and visual characteristics, the author Fr-Chun Hsu et al [8] suggested a hybrid method for detecting the body parts(Head and Shoulders). They tracked down that the Histogram of Oriented Optical Flow (HOOF) descriptor is better for sectioning moving items in video streams. It can effectively control cluttered and obstructed contexts.
The author Vijay et al. [9] presented a real-time pedestrian detection system for better driver support. The methods are Edgelet features and the ‘k-means’ clustering technique as a classifier used to boost accuracy and reduce system complexity. The authors [10] have proposed a human detection framework for city-wide surveillance via CCTV cameras. They used the background subtraction approach to segment moving objects, and HOG descriptor is used for features extraction while the Support Vector Machine method is employed for object categorization.

C. Object Detection based on CNN models

Object detection through deep convolutional neural networks (CNN) has made substantial progress in recent years. CNN is a type of intense, feed-forward ANN that has proven effective in computer vision applications like object detection and classification. With the help of the convolution process, CNN can extract robust features. Its high attribute representation capability [11]. Pre-trained CNN-based networks are utilized for several object recognition applications. The authors of a related study [12] used the COCO dataset as a benchmark input. The result found that the Inception-resnet feature extractor model has a maximum Mean Average Precision (mAP) score of thirty and a minimum mAP score of twenty. SSD Mobilenet is a faster model compared to the Inception Resnet model with a mAP score of nineteen but the Faster R-CNN Inception-resnet101 model produces an accurate result with a score of 34.2.

D. Monitoring Social Distancing in real-time based on past research

Using a computer vision-based detection model, Khandelwal et al[13] used surveillance camera video to monitor masked faces and breach social distancing rules. MobileNet model is used for mask detection and SSD detector for monitoring social distancing. The authors [1] suggested a Mobilenet SSD object tracking model and OpenCV library for detecting the people in the region of interest. The distance between the people detected in the video footage will be calculated and compared to a set of pre-determined pixel values. The output is achieved between 56.5% to 68% for outdoor locations. Aquib Ansari[2] presented a human detection method using CNN-based detector and then evaluated the distances among each pair of the person detected. The validation accuracy of model 1 and model 2 have achieved up to 97% and 98%.
The author [4] discussed an automatic human detection using a YOLOv4-based object detector model in crowds in both inside and outside areas by surveillance cameras. When combined with a modified Inverse Perspective Mapping(IPM) method and the SORT tracking technique, the suggested deep-neural network model produces reliable human detection and observes people's distance. The experimentation was carried out in occlusion, partial vision, and varying lighting conditions with an mAP of 99.80% and a speed of 24.11 frames per second in real time

III. Research Methodology

This study proposes a framework for monitoring social distancing in real-time environment. The idea is developed on the regression-based object detector, which uses Python3, Torch, Cuda, Tensorflow-keras, OpenCV frameworks. These are used to gain the visibility of image processing methods with faster computation. The methodology of this proposed architecture is to take a video from CCTV footage from public areas for human detection and compute the distance between the two persons and raise a red alert for violation the norms of social distancing. The terminology we used in this workflow has been discussed in this section.

A. Dataset

The dataset used in this methodology is Open Images Dataset V6 for training the model. We have taken 15000 specific images of persons and human faces so that the model is trained only with the human category. Due to this, computation time and space complexity are balanced. Annotations of each associated image file are also downloaded from the database library. The bounding box co-ordinates and labels for each image are then take out from XML files and standardized to the image’s width and height. For testing the algorithm of social distancing, some images are crawled from the internet and tested the algorithm's robustness.

Figure 1. Sample images of Open Images Dataset V6

B. *YOLO V5 Architecture and its Functioning*

YOLO's object detection method, which stands for "You Only Look Once," separates images into a grid structure. Each grid cell is incapable of detecting objects within itself. Due to its speed and precision, YOLO is one of the most well-known object detection frameworks. YOLOv5 version was released after YOLOv4, and it is written in the Ultralytics Pytorch framework in 18^{th} May 2020. In a comparison of YOLOv4, YOLOv5s[18] is simpler to use and very effective in training the custom dataset. It is a lightweight model and trains quickly, inferences quickly, and performs well.

The architecture of YOLOv5 is shown in figure 2. It is a single-stage object detector and comprises three important fragments. First is Model Backbone, and the backbones can be running on CPU or GPU. The purpose of the backbone is to extract the prominent features from the input image. YOLOv5 incorporated the CSPDNet (Cross Partial Network) into Darknet framework, ensuing in creating CSPDarknet as the spine of a network. CSPNet fixes the issue of recurrent gradient info in large-scale backbones by as well as gradient modifications into the mapping of features, lowering parameters, and the Floating-point Operations per second (FLOP), which assures the inference speed and accuracy, also reducing the size of the model. Speed and accuracy are crucial terms in social distance monitoring tasks, and the model's size dictates its inference efficiency.

The second is Model Neck which is specially used to produce feature Pyramids. The Feature pyramids network (FPN) aids in the simplification of models on an object ruler. It assists in recognizing the same object in various scales and sizes. The YOLOv5 used a Path Aggregation network (PANet) to optimize the information flow, and PANet uses a new FPN topology with an improved bottom-up approach to improving low-level feature propagation. Simultaneously, Adaptive Feature Pooling ensures that beneficial information from each feature level reaches the following subnetwork by connecting the feature lattice to all feature levels.

Lastly, for the final stage detection, the model head is applied. After applying anchor boxes to features, it creates final resultant vectors with probabilities of class, objectness values, and several bounding boxes.

The activation functions are Leaky ReLU and Sigmoid used in YOLOv5. The Leaky ReLU function is applied in the middle layers, and in the final detecting layer, the Sigmoid function is applied. In this work, we have used an ADAM optimizer for training purposes.

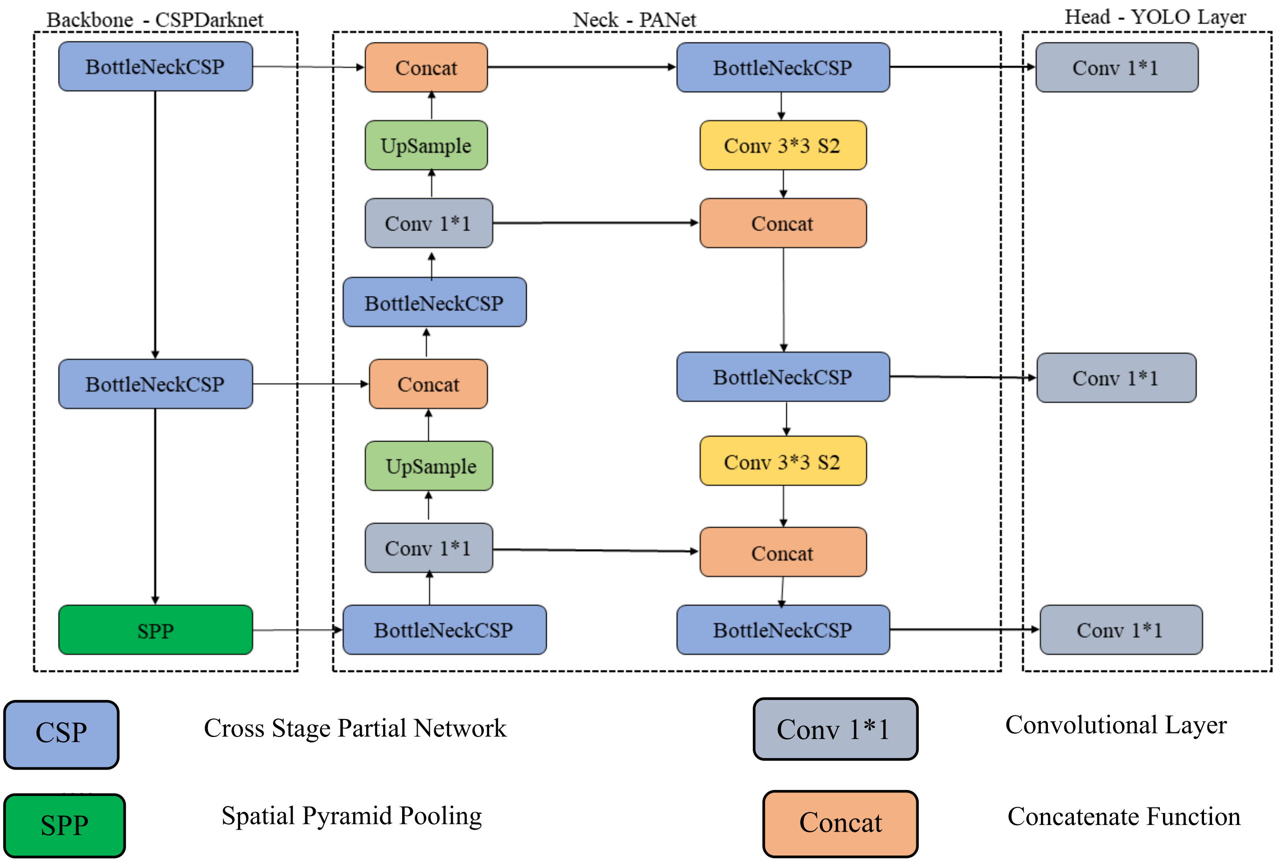

Figure 2. The architecture of YOLOv5 Model [18]

Identify applicable funding agency here. If none, delete this text box.

C. Monitoring Social Distance

Monitoring social distance is the second stage of the working methodology. The algorithm is carried out with two foremost functions. The first function assists in determining the locations of persons available in an image. It employs a people detection method and offers coordinates of people location such as X_A(left), X_B(right), Y_A(top), Y_B(bottom). The centroid points (X and Y) of different objects are evaluated from these coordinate values. Moreover, these points are forwarded to the following function to compute distance of people.

$$X = (X_A + X_B) * 0.5 \quad (1)$$

$$Y = (Y_A + Y_B) * 0.5 \quad (2)$$

The next function is utilized to measure the closeness among two objects through Euclidean metric. The distance is based on comparing this distance vector to a previously defined threshold value. Suppose the Euclidean distance [2] between two persons is fewer than a certain threshold or remarkable value. In that case, it is expected that persons are neither adhering to the rules of social distancing or nor created maximum distance among people. This study set 2 m to 2.5 m as a threshold limit and classified it in three categories. If the distance is less than 2m, it marks as a High Risk; if the value of the distance is in between the threshold limit, it marks as a Moderate Risk otherwise Safe Distance.

$$\{d = \sqrt{((X_2 - X_1)^2 + (Y_2 - Y_1)^2} \} \quad (3)$$

Here, (X_1, X_2) and (Y_1, Y_2) are centroid values of two persons.

The workflow of the proposed methodology is described in a stepwise manner:

Input: V_D : Number of frames of size 512*512 contained in the video.

Output: D: Mark the rectangles using distance vector d around the person.

Step 1: Collect the database of persons and human face categories.
Step 2: Train the image database with YOLOv5 architecture with changes in the configuration file.
Step 3: Load the trained model for the testing phase.
Step 4: Apply a social distancing monitoring algorithm to detect the distance between two persons.
Step 5: Draw the red rectangles around the person if the distance is less than the assigned threshold value.

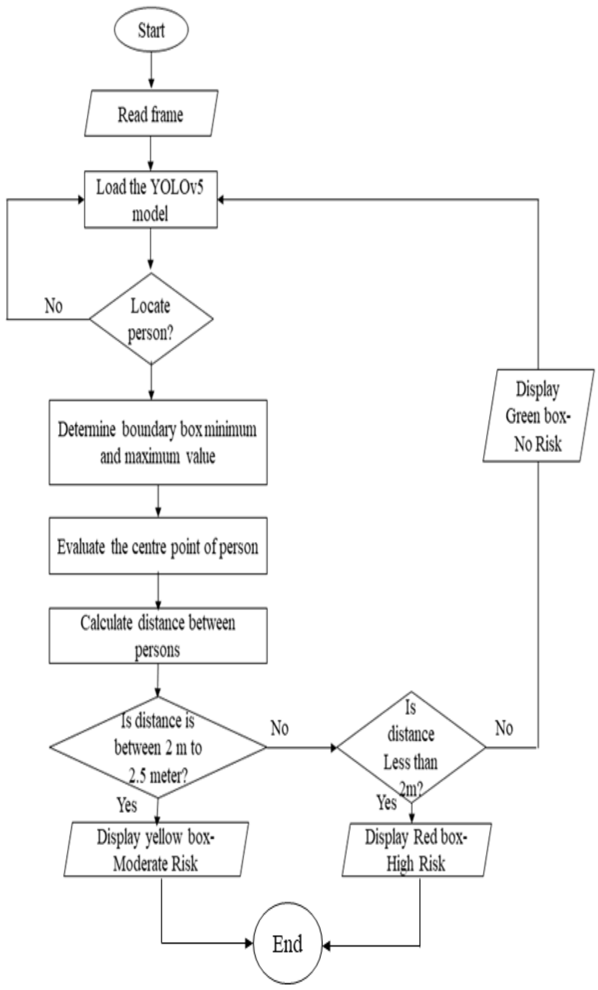

Figure 3. Flowchart of the proposed methodology

IV. Experiments and Results

A. Experimental Setup

The model YOLOv4 was constructed on the PyTorch framework and trained the model on i7 processor with 4GB RAM GPU and utilizing the Cuda framework. Webcam and standard dataset videos are used for testing the system. The frame is set to a 521*521 ratio.

B. Metrics

Several measures are used to assess the system's performance. Here two metrics will be used to judge the efficiency of our solution, which are discussed below.

1) Intersection over Union(IoU): The overlap between the two bounding boxes is determined by IoU. A result of 1 represents perfect prediction accuracy, while lower values indicate the least prediction accuracy[3].

$$\frac{area(P_B \cap G_B)}{area(P_B \cup G_B)} \quad (4)$$

Where P_B is the area or region of the Predicted Boundary Box, and G_B is the area of the Ground Truth Boundary Box.

2) Accuracy: The system's effectiveness is measured by accuracy(A)[14]. It can also be estimated in terms of positive and negatives. It can also be defined as the proportion of the number of truthful predictions and the total amount of predictions.

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

C. Results Analysis

The accuracy of the social distancing system is tested on four distinct videos, including the TownCentre, VIRAT_S, PETS2009[15-17], and videos captured from the webcam. Video footage from the webcam is captured in indoor and outdoor environments.

TABLE 1. PARAMETERS ESTIMATION DURING TRAINING THE MODEL

Model	Dataset	Training Accuracy	Validation Accuracy	Validation Loss	Training Time
YOLOv5	Open Image V6	0.99	0.98	0.02	23 h

TABLE 2. OUTCOMES DURING TESTING PHASE ON A DIFFERENT DATASET

Video Dataset for testing	Testing Accuracy	Length of video	Testing Time
Self taken	98%	Continuous frames	0.02sec
TownCentre	98%	0.04sec	9.6s
VIRAT_S	92%	24 sec	32.9s
PETS2009	95%	13 sec	42.21s
Epfl-Mpv	98%	1.58 min	136.54s

TABLE 3. COMPARISON OF EXISTING APPROACH WITH PROPOSED METHODS ON TOWNCENTRE VIDEO.

	Existing Approach		Proposed Approach
	Author[1]	Author [4]	
Training model	MobileNet SSD	YOLOv4	YOLOv5
System Accuracy	62.5%	98%	99%

D. Outcomes on a different dataset

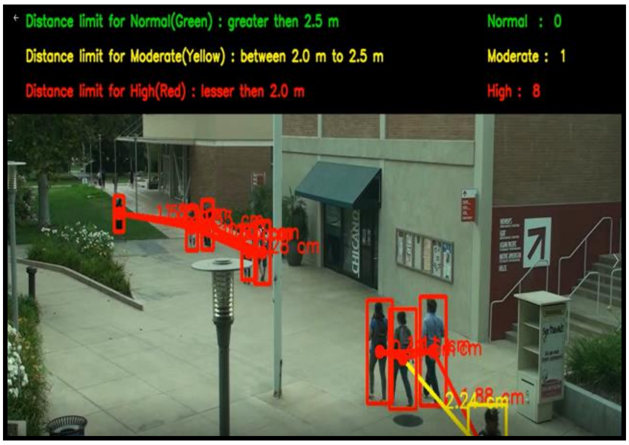

Figure 4. Persons violating the social distancing norms represent in red boundary box on dataset Virat_S(Outdoor scene)

Figure 5. Persons violating the social distancing norms represent in red boundary box on dataset of TownCenter(outdoor scene)

Figure 6. Persons violating the norms represent in red line through webcam stream(indoor scene)

Figure 7. Persons violating the social distancing norms represent in red line on dataset of EPFL-Mpv(indoor scene)

The above outcomes represent the violation of social distancing norms in outdoor and indoor environments on different datasets and live webcam. The red line or boundary shows the high risk (i.e. less than 2m distance), yellow mark represents the moderate risk (the distance between the 2.0m to 2.5m) and the green mark shows the safe distance i.e., greater than 2.5m.

V. Conclusion

Social distancing is one of the most significant safety measures in avoiding physical contact that could lead to the spread of coronavirus. Virus transmission rates would increase if WHO's norms were not considered. This study proposed a Deep learning-based object detection technique to observe the distance among people in a real-time situation. The system has been designed in Python and torch library to implement the functions. YOLOv5 is used to train the model on Open Images V6 dataset for persons and human face categories because the YOLO framework is the best architecture to detect the objects. Further, this model is used for maintaining the social distancing algorithm. This algorithm detects the stages of risk from the input frame and draws the red rectangle if a high-risk violation is found. This paper has seen to meet all of its objectives based on the overall results. The system effectively and efficiently detects the social distance between persons and generates an alert that can be controlled and monitored.

The future scope of this proposed model is to apply in advanced applications such as monitoring the human location and observing people's gestures. Human tracking is the ardent area for researchers like human action recognition, abnormal behavior recognition, and many more in the surveillance field.

References

[1] Ahamad, Afiq Harith, Norliza Zaini, and Mohd Fuad Abdul Latip. "Person Detection for Social Distancing and Safety Violation Alert based on Segmented ROI." 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), pp. 113-118. IEEE, 2020.

[2] Ansari, Mohd Aquib, and Dushyant Kumar Singh. "Monitoring social distancing through human detection for preventing/reducing COVID spread." International Journal of Information Technology 13, no. 3 (2021): 1255-1264.

[3] Bhambani, Krisha, Tanmay Jain, and Kavita A. Sultanpure. "Real-time Face Mask and Social Distancing Violation Detection System using YOLO." In 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), pp. 1-6. IEEE, 2020.

[4] Anil Dnyandev Sapkal, Chaitanya Rathod and Ajmal Shaikh,"OpenCV Social Distancing Detector", Journal of Emerging Technologies and Innovative Research, Volume 8, Issue 1, pp: 21-46, January 2021.

[5] World Health Organization Coronavirus disease (Coivid-19) advice for the public. In: WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data(accessed 2 Dec 2021).

[6] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[7] Viola P, Jones M (2001) Robust real-time object detection using a boosted cascade of simple features. IJCV 12:18–24

[8] Hsu FC, Gubbi J, Palaniswami M (2013) Human head detection using histograms of oriented optical low in low quality videos with occlusion. In: 2013, 7th International conference on signal processing and communication systems (ICSPCS), IEEE, pp 1–6

[9] Gaikwad V, Lokhande S (2015) Vision based pedestrian detection for advanced driver assistance. Proc Comput Sci 46:321–328

[10] Singh DK, Paroothi S, Rusia MK, Ansari MA (2020) Human crowd detection for city wide surveillance. Proc Comput Sci 171:350–359. https://doi.org/10.1016/j.procs.2020.04.036.

[11] Jiao L et al (2019) A survey of deep learning-based object detection. IEEE Access 7:128837–128868. https://doi.org/10.1109/ACCESS.2019.2939201

[12] N. Yadav and U. Binay, "Comparative Study of Object Detection Algorithms," Int. Res. J. Eng. Technol., pp. 586–591, 2017, [Online]. Available: www.irjet.net.

[13] Prateek Khandelwal, Anuj Khandelwal, Snigdha Agarwal, Deep Thomas, Naveen Xavier and Arun Raghuraman, "Using Computer Vision to enhance Safety of Workforce in Manufacturing in a Post COVID World", 2020, arxiv:2005.05287.

[14] R. Jayaswal and M. Dixit, "Comparative Analysis of Human Face Recognition by Traditional Methods and Deep Learning in Real-Time Environment," 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), 2020, pp. 66-71, doi: 10.1109/CSNT48778.2020.9115779.

[15] S. Oh et al., "AVSS 2011 demo session: A large-scale benchmark dataset for event recognition in surveillance video," 2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011, no. 2, pp. 527–528, 2011, doi: 10.1109/AVSS.2011.6027400..

[16] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," Proc. 12th IEEE Int. Work. Perform. Eval. Track. Surveillance, PETSWinter 2009, pp. 0–5, 2009, doi: 10.1109/PETSWINTER 2009.5399556

[17] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3457–3464, 2011, doi: 10.1109/CVPR.2011.5995667.

[18] YOLOv5 study Introduction to YOLOv5 Object Detection with Tutorial - MLK - Machine Learning Knowledge(online accessed on 5th December 2021).