Social Distance Detection -Transfer Learning – YOLO

Aman Sharma
(Co-Author)
aman18csu015@ncuindia.edu
Dept. of Computer Science and Engineering
The Northcap University
Gurugram, India

Aditya Chaudhry
(Co-Author)
aditya18csu011@ncuindia.edu
Dept. of Computer Science and Engineering
The Northcap University
Gurugram, India

ABSTRACT

The ongoing COVID-19 corona virus outbreak has caused a global disaster with its deadly spreading. Due to the absence of effective remedial agents and the shortage of immunizations against the virus, population vulnerability increases. In the current situation, social distancing is thought to be an adequate precaution (norm) against the spread of the virus. The risks can be minimized by avoiding physical contact with people. The purpose of this work is, therefore, to provide a deep learning platform for social distance tracking using an frontal and side perspective. The framework uses the YOLOv3 object recognition paradigm to identify humans in video sequences. The transfer learning methodology is also implemented to increase the accuracy of the model. The detection model identifies peoples using bounding box methodology. Euclidean distance is used to measure distance from consecutive bounding box centroid and validate the appropriate distance. To estimate social distance violations between people, we used an approximation of physical distance to pixel value(70) and set a threshold. A violation list is established to monitor and store the violated values. In addition, a tracking algorithm is used to detect individuals in video sequences such that the person who violates/crosses the social distance threshold is tracked. Experiments are carried out on different video sequences to test the efficiency of the model. Our findings indicate that the developed framework successfully distinguishes individuals who walk too near and breach/violate social distance protocol. Also, to aid the overall functionality we implement transfer learning approach which boosts the overall efficiency of the model.

Keywords

Cv2framework, Blob, Yolo, Bounding Box, Centroid,
Coco, Euclidean Distance, Frame

1. Introduction

COVID-19 originated from Wuhan, China, has affected many countries worldwide. On 10-5-2021, the World Health Organization (WHO) gave the approx figure of 35,16,146 deaths and 16,50,48,459 number of total cases. On 18-5-2021, in India alone there were more than 4,67,890 active cases and roughly around 4,125 deaths confirmed. [1] The virus mainly spreads in those people, who are in close contact with each other (within 6 feet) for a long period. The virus spreads when an infected person sneezes, coughs, or talks, the droplets from their nose or mouth disperse through the air and affect nearby peoples. The droplets also transfer into the lungs through the respiratory system, where it starts killing lung cells. Social distancing associates with the measures that overcome the virus's spread, by minimizing the physical contacts of humans, such as the masses at public places (e.g., shopping malls, parks, schools, universities, airports, workplaces), evading crowd gatherings, and maintaining an adequate distance between people. [4] In the past decades, computer vision, machine learning, and deep learning has shown promising results in several daily life problems.

Recent improvement in deep learning allows object detection tasks to be carried out more effectively. New advancement in the field of deep learning helps us to determine the distancing between people, which can be achieved by clustering and distance-based methods. Most methods are developed using frontal or side view video sequences, which requires a camera to map pixels to distance, tracking and object detection. [2]

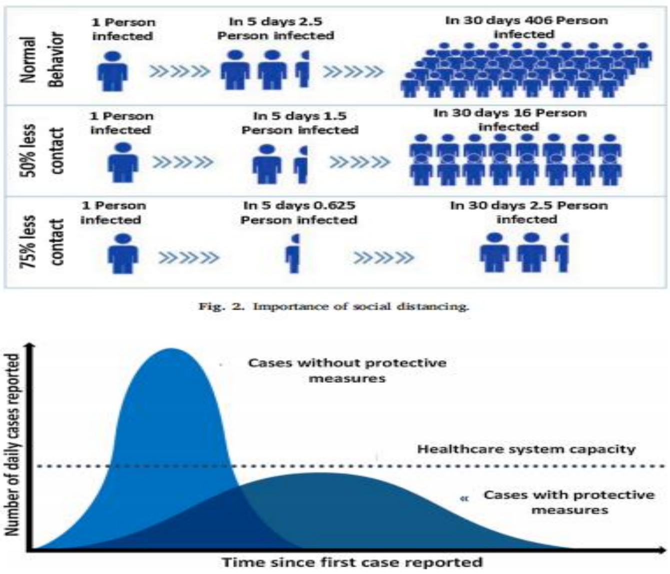

Fig 1 - Effect of social distancing.

This work aims to present a deep learning-based social distance monitoring framework using a deep learning model, i.e., YOLO (v3) which is applied for human detection. The current model (pre-trained on frontal and normal view data sets) has been already tested and trained several times and is available to use to give better results, saving time and effort and do task effectively. Transfer learning is also used to improve the efficiency of the detection model.[3] The detection model detects humans and gives bounding box information. After human detection, the Euclidean distance between each detected centroid pair is computed using the detected bounding box and its centroid information. A predefined minimum social distance violation threshold is specified using pixel to distance assumptions. To check, either the calculated distance comes under the violation set or not, the estimated information is matched with the violation threshold. The bounding box's color is formerly initialized as green; if the bounding box comes under the violation set, its color is updated to red.

2. Literature review

After the rise of the COVID-19 pandemic since late December 2019, Social distancing is deemed to be an utmost reliable practice to prevent the contagious virus transmission adopted as standard practice. The number of cases rises exceptionally, with two thousand to four thousand new confirmed cases reported per day.[1] Later, there has been a sign of relief. This is because of the social distance practice initiated by the people. The study revealed that moderate stages of exercise could be allowed for evading a large outbreak. So far, many countries have used technology based as well as biological advancements as a solution to overcome the pandemic loss. Several developed countries are employing GPS technology to monitor the movements of the infected and suspected individuals. Provides a survey of different emerging technologies, including Wi-fi, Bluetooth, smartphones, and GPS, positioning (localization), computer vision, and deep learning that can play a crucial role in several practical social distancing scenarios.[8]

Some researchers utilize drones and other surveillance cameras to detect crowd gatherings. Until now researchers have done considerable work for detection, some provides a smart healthcare system for pandemic using Internet of Medical Things, and findings ways to tackle the COVID-19 outbreak. The studies concluded that the early and immediate practice of social distancing could gradually reduce the peak of the virus attack.[7] As we all know, that although social distancing is crucial for flattening the infection curve.

Researchers provide effective solutions for social distance maintenance using surveillance videos along with computer vision, machine learning, and deep learning-based approaches.[9] we proposed framework using the YOLOv3 model to detect humans and the Deep sort of approach to track the detected people using bounding boxes and assigned IDs information. The work is also extended for the monitoring of facial masks. The drone camera and the YOLOv3 algorithm help identify the social distance and monitor people from the side or frontal view in public wearing masks.
The model is designed for individuals who do not obey a social distance restriction.
Stating the facts, we conclude that the researchers had done a considerable amount of work for monitoring of social distance in public environments.

3. Social distance monitoring

Researchers use a frontal or side perspective for social distance monitoring in public places through cctv and surveillance cameras. A deep learning-based detection paradigm is used to detect individuals in sequences. There are a variety of object detection models available, such as [Krizhevsky, Sutskever, and Hinton (2012), Simonyan and Zisserman (2014)] etc. Due to the best performance results for generic object detection, in this work, YOLOv3 (Redmon & Farhadi, 2018) is used.[11] The model used single-stage network architecture to estimate the bounding boxes and class probabilities. The model was originally trained on the COCO (Common objects in context) data set.

After detection, the bounding box information, mainly centroid information, is used to compute each bounding box centroid distance. We used Euclidean distance and calculated the distance between each detected bounding box of peoples. Following computing centroid distance, a predefined threshold is used to check either the distance among any two bounding box centroids is less than the configured number of pixels or not.[5] A centroid tracking algorithm is adopted for tracking so that it helps in tracking of those people who violate/breach the social distancing threshold. At the output, the model displays the information about the total number of social distancing violations along with detected people bounding boxes and centroids. In this work, YOLOv3 is used for human detection as it improves predictive accuracy, particularly for small-scale objects. The main advantage being adjusted network structure for multi-scale object detection.

The model's unique characteristic is that it performs detection at three separate scales, as depicted in Fig. 3. The convolutional layers with a given stride are practiced to down sample the feature map and transfer invariant-sized features (Redmon & Farhadi, 2018). Three feature maps, as shown in Fig. 3, are utilized for object detection. For that purpose, a transfer learning approach is adopted, that enhance the efficiency of the model. With transfer learning, the model is additionally trained without dropping the valuable information of the existing model. The architecture shown in Fig. 3 used a single-stage network for the entire input image to predict the bounding box and class probability of detected objects. For feature extraction, we use
convolution layers, and for class prediction, fully connected layers are used.[8] After this we store the values in defined lists and the model analyzes the video frame by frame and indicates if the current index[ID] falls in the violation set or not and represents it by green and red color outline boxes accompanied by warnings.

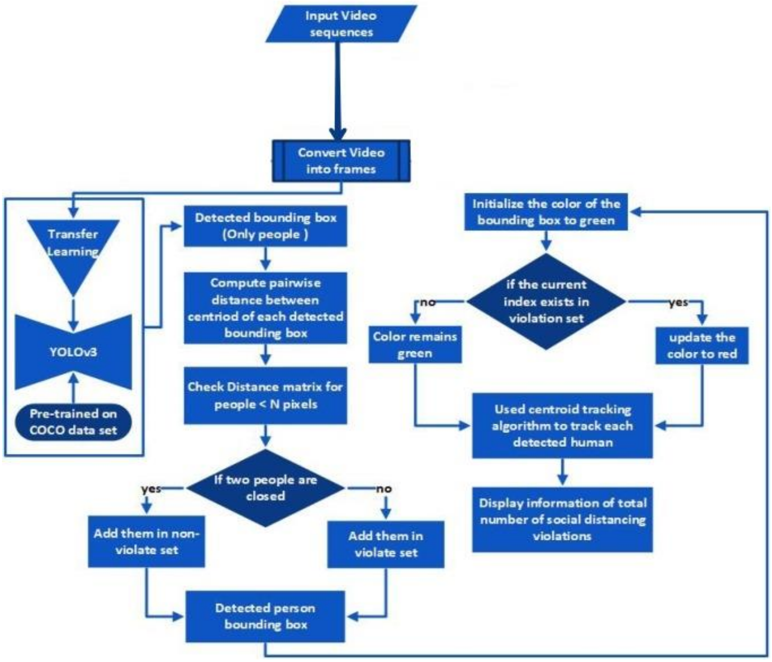

Fig 2- Flow diagram of overhead view social distance monitoring framework

4. PROPOSED METHODOLOGY

This project is based on 3 algorithm that are-
1. Object detection
2. Object tracking
3. Distance measurement between two objects

4.1 OBJECT DETECTION

In this project we have firstly used object detection technique, wherein we have first tried to classify object separately from one another. Object detection is very different from classification as classification involves assigning a class label to an image, whereas on the other hand object localization involves drawing a bounding box around one or more objects in an image. It is comparatively challenging as it involves combining two tasks and drawing a boundary.

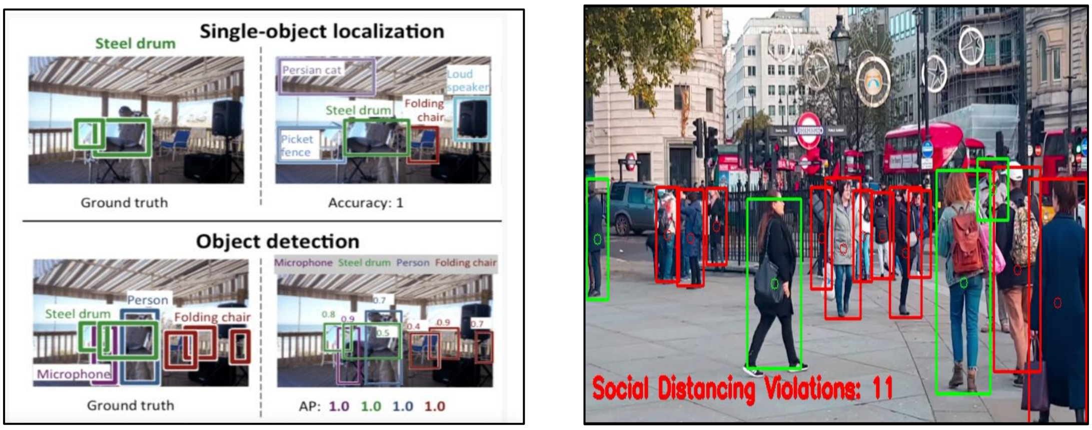

Fig 3- Classification vs Object Detection / Real Time object detection

4.1.1 YOLO

Here we have used YOLO technique, which stands for [YOU ONLY LOOK ONCE] which is a second family of techniques for object recognition designed for speed and real time use. During human identification, as seen in Fig 4 the input frame is divided into a region of S × S, also called grid cells. These cells are related to bounding box estimation and class probabilities.

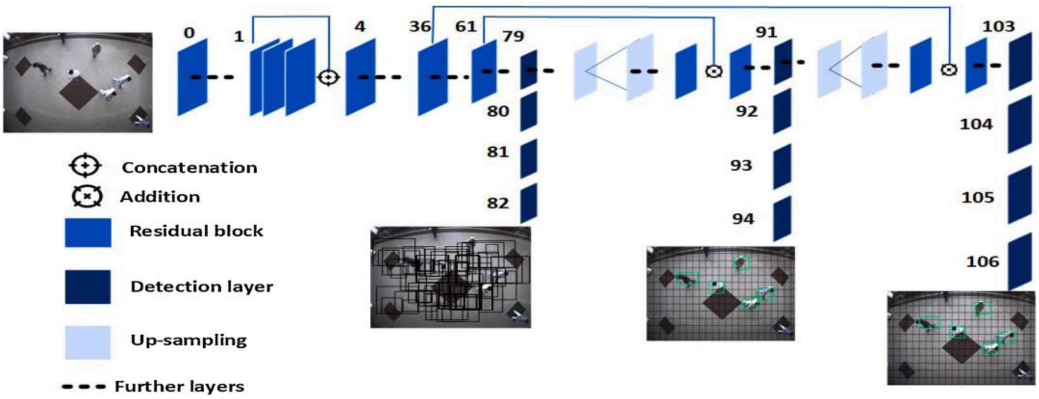

Fig 4- General architecture of YOLOv3 utilized for detecting person in the frame.

4.1.3 YOLO WORKING

The model works by first splitting the input image into a grid of cells, where each cell is responsible for predicting a bounding box if the center of a bounding box falls within the cell. Each grid cell predicts a bounding box involving the x, y coordinate and the width and height and the confidence.

4.1.2 R-CNN VS YOLO

The R-CNN models may be generally more accurate, yet the YOLO family of models are fast, much faster than R-CNN, achieving object detection in real-time. The YOLO model was first described by Joseph Redmon, et al. in the 2015 paper titled "You Only Look Once: Unified, Real-Time Object Detection. The approach involves a single neural network trained end to end that takes a photograph as input and predicts bounding boxes and class labels for each bounding box directly. The technique offers faster results but has a little less accuracy over R-CNN.

4.1.4 YOLOv2 vs YOLOv3

In this project we used v3 as it is more reliable, faster, gives better results per frame and also uses good techniques from v2, that is the darknet feature extraction which originally has 53 layers. For the detection task another 53 layers are stacked onto it, accumulating to a total of a 106-layer fully convolutional architecture.
YOLOv3 uses a new network for performing feature extraction. The new network is a hybrid approach **between** the network used in **YOLOv2** (Darknet-19), and the residual network, so it has some short cut connections for fast object detection.

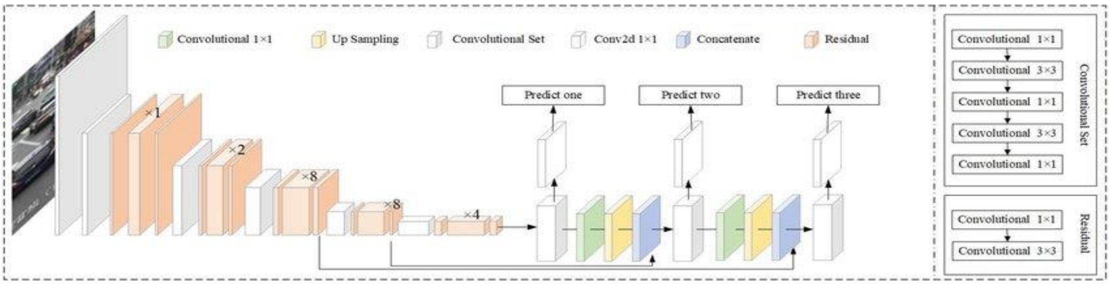

Fig 5- Structure detail of YOLO (v3)

4.2 OBJECT TRACKING

In this project we used object tracking technique, it is the task of taking an initial set of object detections, creating a unique ID for each of the initial detections, and then tracking each of the objects as they move around frames in a video, maintaining the ID assignment.

In this first to track the person with a box around them. And measure the centroid of the box and then measure the Euclidean Distance from each centroid, the closer centroid are the same person and it show it movement in the frame and we give new id to the further centroid and classify as a new person in the frame.

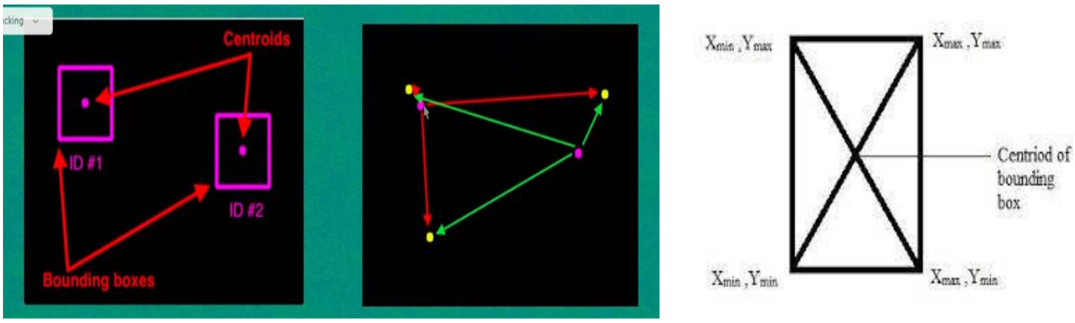

Fig 6- Bounding Box (ID's allocation)/ Centroid and distance from two consecutive ID's along with structural notation and co-ordinates.

4.2.1 BENEFITS OF USING OBJECT TRACKING

The key goals of tracking are as follows:

• To track the social distance between individuals, using Euclidean distance. In addition, a social distance violation threshold is specified using a pixel to get a distance estimation.

• A centroid tracking algorithm can keep track of the person who violates the social distance threshold.

• To assess the performance of pre-trained YOLOv3 along with added layer of transfer learning algorithm, which helps the tracking and detection to be easy and effortless.

4.3 DISTANCE MEASUREMENT BETWEEN TWO OBJECTS

After the second step that is object tracking comes the third and last step of the algorithm that is to measure the distance between two or more objects. In this we have already assigned ID's to the previous objects(persons) via a bounding box and a centroid which the algo stores and then takes the initial as a reference point to the other ID's. The distance between each detected centroid is calculated using Euclidean distance. For every subsequent frame in the video stream, we firstly compute bounding box centroids and then calculate the distance (highlighted with red lines) between each pair of detected bounding box centroids, The information of each centroid is stored in the form of a list. Based on distance values, a threshold is defined to check if any two people are less than N pixels apart or not. If the distance violates the minimum social distance, then the information is added into the violation set. The bounding box color is initialized as green. The information is checked in the violation set; if the current index exists violation set, the color is updated to red.

In simple terms if the gap between two objects is becoming less the system will turn the green indication to red, which will alert the people in the surroundings to maintain a safer distance from each other.

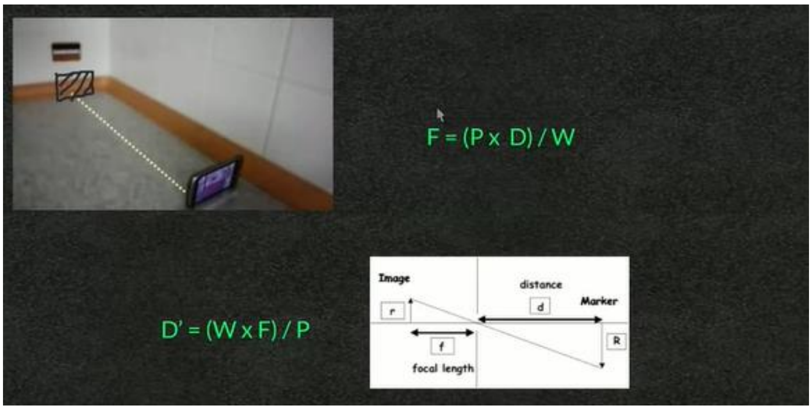

Fig 7- full backend processing that takes place to give us the distance.

4.3.1 FORMULA AND ARITHEMETIC

Here as it is represented, R is the marker (i.e. the object) whose distance is to be measured from the camera having F as it's focal length, P as the pixel size, D as the distance between the marker and camera, W as the width of the object.
The formula used to calculate the D is as follows –
f = (p x d)/w => d = (w x f)/p
Furthermore, the centroid tracking algorithm is used to track the detected people in the video sequence. The tracking algorithm also helps to keep track of people who are violating the social distance threshold. At the output, the model displays information about the total number of social distancing violations.

4.4 ARG-PARSE

ArgParse is the "recommended command-line parsing module in the Python standard library". The module makes it easy to write user-friendly command-line interfaces.
It is a good practice to use arg parse in coding as by changing the values we can see our perormance over multiple files.
Arguments passed in our project are as follows -
Input - to give the raw video footage to apply the model.
Output - to give the new file which will be generated by the algo after the full application of all the three techniques. ["my_output.avi"]
Display - The value of display has been set to "1" as we want to display the output, and "1" indicates true state.

5. Experimental Analysis and Results

The detailed descriptions of various experiments carried out in this work. There is no restriction on the mobility of persons throughout the scene. Peoples in the scene move freely; their visual appearance is affected by radial distance and camera position. From example frames, it can be observed that the human's visual appearance is not identical, and people's heights, poses, scales are varying in the coco dataset.

5.1 COCO

COCO is an advance state-of-the-art in object recognition Data Collection. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. The dataset contains photos of 91 objects types with a total of 2.5 million labeled instances in 328k images, which can be used by novel user interfaces for category detection, instance spotting and instance segmentation. This cuts down on the basic training time for any AI that needs to process images. Since here the model only considers human (person) class; therefore, only an object having an appearance like a human is detected.

The pre-trained model delivers good results and detects various size person bounding boxes, as shown with green rectangles. Multiple people are entering in the scene. In sample images, it can be seen that after person detection, the distance between each detected bounding box is measured to check whether the person in the scene follows the social distance or not.

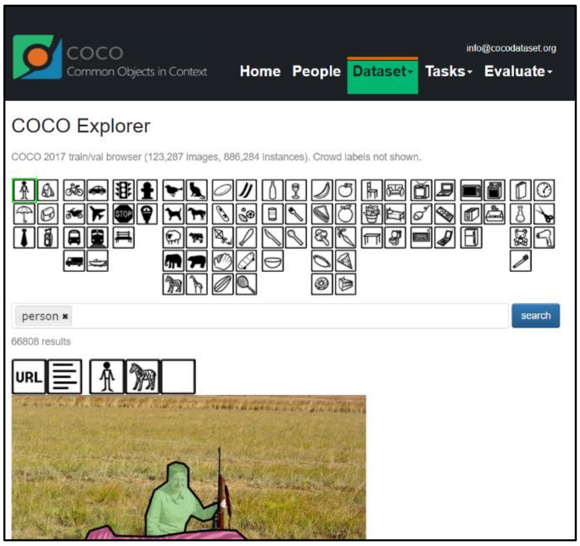

Fig 8- COCO Dataset

6.1 PICTORIAL UNDERSTANDING

A young blond woman walks in the frame and we can see that with red bounding boxes as she violates the social distancing threshold, but soon as she walks in the center of the frame the indication of the box is turned to green depicting that she is at an appropriate pixel length from other individuals and maintaining social distancing adequately.

NOTE

However, in some cases, the person's appearance is changing; therefore, the model gives miss detections. The reason for miss detection maybe, as the pre-trained model is applied, and an individual's appearance from an angle is changing, which may be misleading for the model.

Fig 9- blond woman at center maintaining appropriate social distance, so bounding box depicts "GREEN" colour box which tells us that proper social distancing is being followed.

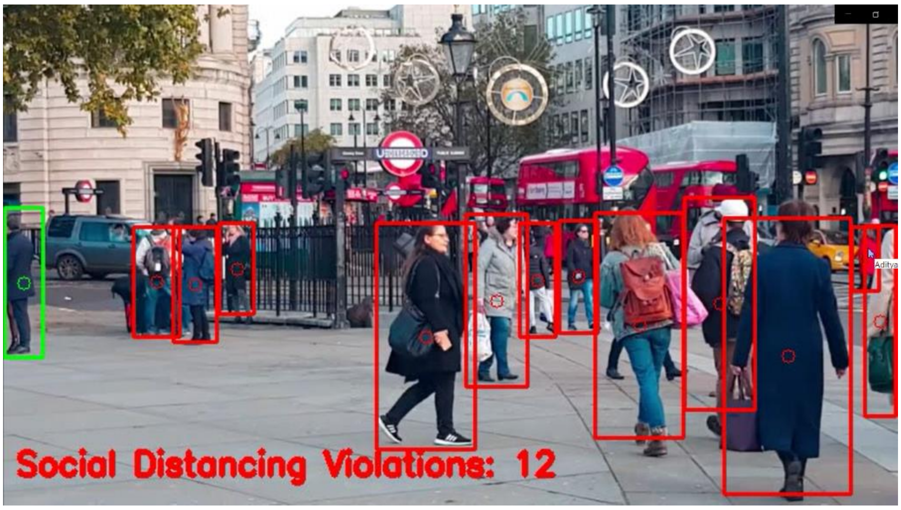

Fig 10- blond woman at going right side coming in close contact to people, violating social distancing norms and algorithm, as a result bounding box depicts "RED" colour box which tells us that proper social distancing is not being followed.

7. TRANSFER LEARNING

The transfer learning methodology is applied to improve the accuracy of the detection model. The model is now tested for other alternative methods that could have been used and experimental findings reveal that transfer learning significantly increases the detection results, as seen in it can be visualized that the model detects the individuals at various scene locations. People with various characteristics are effectively identified, and the social distance between people is also computed.

Table 1
Comparison results of YOLOv3 with other deep learning models.

S. no.	Model	True detection rate	False detection rate
1.	Fast-RCNN (pre-trained)	90%	0.7%
2.	Faster-RCNN (pre-trained)	92%	0.6%
3.	Mask-RCNN (pre-Trained)	92%	0.5%
4.	YOLOv3 (pre-trained)	92%	0. 4%
5	YOLOv3 (trained overhead data set)	95%	0.3%

Fig 11- Comparison results of YOLOv3 with other deep learning models.

8. Conclusion and future add-ons

In this work, a deep learning-based social distance monitoring framework is presented using a frontal-view perspective. The pre-trained YOLOv3 paradigm is used for human detection. As a person's appearance, visibility, scale, size, shape, and pose vary significantly from a complete view, the transfer learning method is adopted to improve the Coco(dataset) model's performance. The model consists of functions and parameters along with coordinates of frame, initializing new empty list, generating bounding box and forwarding the result to neural network and running loops on layer output and verifying minimum configuration index and appending the values to [BOX, CENTROID, CONFIDENCES].

The detection model gives bounding box information, containing centroid coordinates information. Using the Euclidean distance, the pairwise centroid distances between detected bounding boxes are measured. To check social distance violations between people, an approximation of physical distance to the pixel is used, and a threshold is defined. A violation threshold is used to check if the distance value violates the minimum social distance set or not. Furthermore, a centroid tracking algorithm is used for tracking peoples in the scene. Experimental results indicated that the framework efficiently identifies people walking too close and violates social distancing; also, the transfer learning methodology increases the detection model's overall efficiency and accuracy.

Also, as we've researched the internet and read related research papers for new techniques and ways to implement the code we've come across different procedure to do object detection, tracking, distance measurement and we'll try to implement all the best possible techniques in our project to further increase the outcome, usability in different fields, accuracy and performance. One such example of thing to append with this project is "OVERHEAD PERSPECTIVE" using the overhead dataset which may increase its accuracy to even more high reaching values.

9. References

1. Ruiz Estrada M.A. The uses of drones in case of massive Epidemics contagious diseases relief humanitarian aid: Wuhan-COVID-19 crisis. *SSRN Electron J.* 2020 doi: 10.2139/ssrn.3546547. (February) [CrossRef] [Google Scholar]

2. Nguyen T.T., Waurn G., Campus P. 2020. Artificial intelligence in the battle against coronavirus (COVID-19): A survey and future research directions. [CrossRef] [Google Scholar]

3. Maghdid H.S., Ghafoor K.Z., Sadiq A.S., Curran K., Rabie K. 2020. A novel AI-enabled framework to diagnose coronavirus COVID 19 using smartphone embedded sensors: design study; pp. 1–5.http://arxiv.org/abs/2003.07434 [Google Scholar]

4. Wang C.J., New T., Sun F. 2020. Response to COVID-19 in Taiwan big data analytics , new technology , and proactive testing; pp. 1–2. [PubMed] [CrossRef] [Google Scholar]

5. https://www.techuk.org/insights/news/item/17187-how-taiwan-used-tech-to-fight-covid-19

6. Bullock J., Alexandra Luccioni, Pham K.H., Lam C.S.N., Luengo-Oroz M. 2020. Mapping the landscape of artificial intelligence applications against COVID-19; pp. 1–14.http://arxiv.org/abs/2003.11336 [Google Scholar]

7. Wang S., Kang B., Ma J. 2020. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19) pp. 1–28. [Google Scholar]

8. Ali Narin, Ceren Kaya ZP. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks.

9. Jin C., Chen W., Cao Y., Xu Z., Zhang X., Deng L. 2020. Development and evaluation of an AI system for COVID-19 diagnosis; pp. 1–23. [Google Scholar]

10. Wang L., Wong A. 2020. COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images.http://arxiv.org/abs/2003.09871 [Google Scholar].