JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

COVID-19 SAFTEY PROTOCOL VOILATION DETECTION USING YOLO

Tushant Yadav, Vivek Kumar & Tarun Sharma, Bhavnesh Jaint
Department of Electrical Engineering
Delhi Technological University
Delhi, India

Abstract *:* In this current ongoing situation of a Worldwide Pandemic due to the COVID-19 Virus, Social Distancing and wearing a Face Mask are two of the most important and viable means in prevention of this Virus. As per the guidelines issued by WHO, Social Distancing should be practiced with a distance of at least 1m between 2 individuals. You Only Look Once (YOLO) is one of the best available CNN Architecture for detecting various Objects. For Object detection, bounding boxes for different objects placed under different classes deem to be the best option. YOLO architecture utilizes the concept of Transfer Learning. Transfer learning is a machine learning method in which a model which has already been developed for a particular task is reused as the base to perform another task. We have implemented YOLO in our project using the Darknet framework. In this paper, we will be using a pre-trained YOLO model and train it for Social Distancing detection and Face Mask detection.

Keywords—COVID-19, Social Distancing, CNN, You Only Look Once (YOLO), Transfer Learning, Darknet.

I. Introduction

COVID-19 also known as the Coronavirus Disease first originated in 2019 in Wuhan, China. Since then, it has become a worldwide crisis in 2020, but the situation has improved a little in 2021 due to the production of vaccines. It is a highly infectious virus which is caused by a virus called Sars-Coronavirus(SARS-CoV).[1] Many of the people who have been infected with the COVID-19 virus might experience moderate to high difficulty in breathing and will also be able to recover without the need for any medical attention. Only else people and persons with any underlying chronic medical diseases are highly likely to develop serious illnesses. But the guidelines issued by WHO. Maintaining Social distance and wearing a Face Mask in all public places are two of the best strategies for the prevention of this virus. So in this paper, we have utilized Darknet[2] & YOLO[3], [4] models to measure Social distancing in a group of people inside a video and Face mask detection for the same. We have used different versions of YOLO for the implementation of both these tasks..

II. Yolo and object detection

Object detection is a computer technology and it deals with the task of detecting different instances of an object belonging to a particular class inside an image. It is related to computer vision and is highly used in today's day and age. Object detection becomes highly efficient when the objects to be identified or detected are labelled using bounding boxes inside an image. Object detection combines Image Classification and Object-Localization tasks. Now, image classification includes designating class labels in an image and object localization includes constructing a bounding box over one or more objects inside an image and identifying its location.[5] And after both these tasks are done object detection begins by drawing bounding boxes around the object to be detected and assigning them to the detectable class label. Object detection tasks are mainly tackled using:
i) R-CNN Model Family - It is referred to as "Region-Based Convolutional Neural Network. It revolves around using Faster R-CNN techniques, Fast R-CNN, and R-CNN.
ii) YOLO Model Family - It refers to as "You Only Look Once" or YOLO as already mentioned earlier.

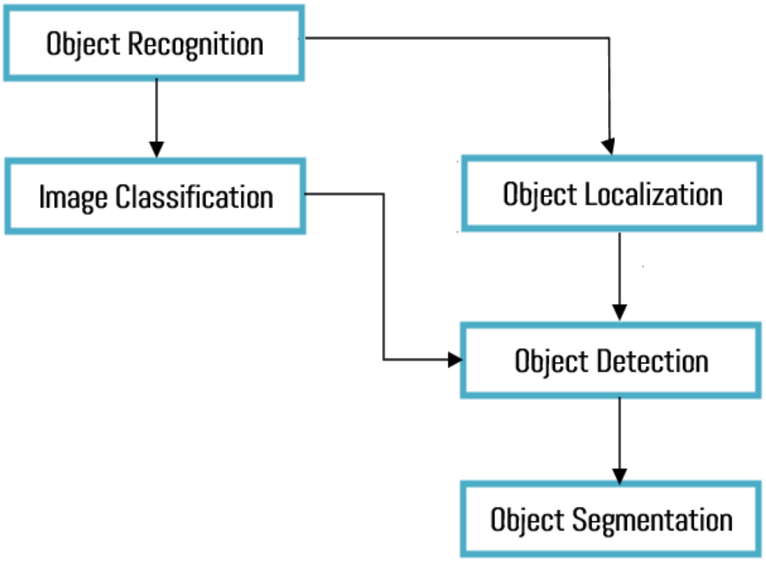

Fig. 1: Basic working flow chart of an object detector algorithm

A. About YOLO and it's different versions

YOLO Model family are included in many of the one-step object detection architectures proposed over the last few years. One-step object detection specifies that YOLO only needs a single forward-propagation pass through the model to make accurate predictions. YOLO is a regression-based algorithm, which means that unlike other architectures that uses feature extraction techniques for object detection, YOLO can predict bounding boxes for the objects in an image and it looks at the entire image as a whole during the entire process. In this paper we decided to use the YOLO framework due to its incredible speed and unique approach towards object detection problems.[3], [6] YOLO is also well known as it achieves high accuracy even while being run in real-time. YOLO trains over a complete image and also directly optimizes object detection performance.

Merits of YOLO over other object detection models:

a. YOLO has superb speed.

b. YOLO processes whole image when passes through network so it also encodes information about all classes and how they look.

c. YOLO learns object generalization so it can easily outperform other detection methods.

YOLO performs the best using the Darknet framework as backbone. Darknet is widely used as an open-source neural network framework.

In this paper, we have used YOLOv3[4] for social-distancing detection and YOLOv5 for face mask detection.

YOLO has many different variants such as:

i) YOLO
ii) YOLOv2
iii) YOLOv3
iv) YOLOv4
v) YOLOv5
vi) PP-YOLO

Each new version of YOLO is an upgrade of the previous version in all aspects such as speed, accuracy etc.

.

III. Dataset Used

Datasets are a group of data with a definite structure. In this paper we have used COCO dataset, Face Mask Detection dataset from Kaggle and a custom labelled dataset for training and testing purposes.

1. COCO Dataset
2. Kaggle Face Mask Detection Dataset
3. Custom labelled dataset for face mask detection

A. COCO Dataset - It is a large scale widely used object detection dataset consisting of 80 classes in total. YOLOv3 has been pre-trained on this dataset.

B. Kaggle Face Mask Detection Dataset - This is a dataset that came into the spotlight after the COVID-19 pandemic began. It is a dataset consisting of 2 classes (With Mask, Without Mask) and contains 7553 images divided among these 2 classes.

C. Custom labelled dataset for facemask detection - a custom dataset was used for both training and testing for the proper facemask detection model using YOLOv5.

We used a labelling tool to label the images inside 2 classes (With mask, Without Mask).

IV. Methodology

YOLO undertakes the task of multiple object detection through the method of fixed grid regression. YOLO uses only a single CNN to predict many bounding boundaries and category probabilities for each classes at the same exact time.[6], [7] With YOLO our aim is to predict a class of an object and the bounding boxes which specifies the location of that object.

Steps taken by YOLO for object detection:

1. Firstly, an input image is entered into the framework.

2. Then, the framework speratees that input image into grids of different sizes.
3. At Last, object classification and object localization to each sub image. Then, YOLO draws all the bounding boxes with class probabilities.

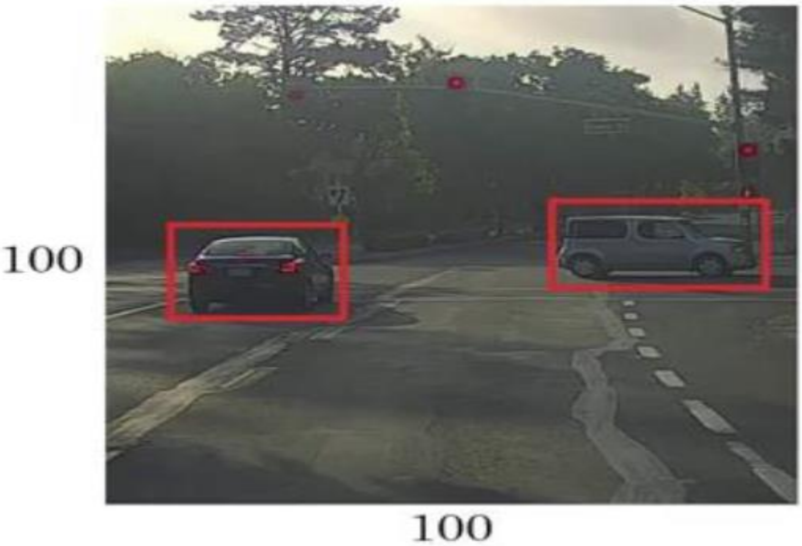

Fig. 2: Step 1-Input image is entered

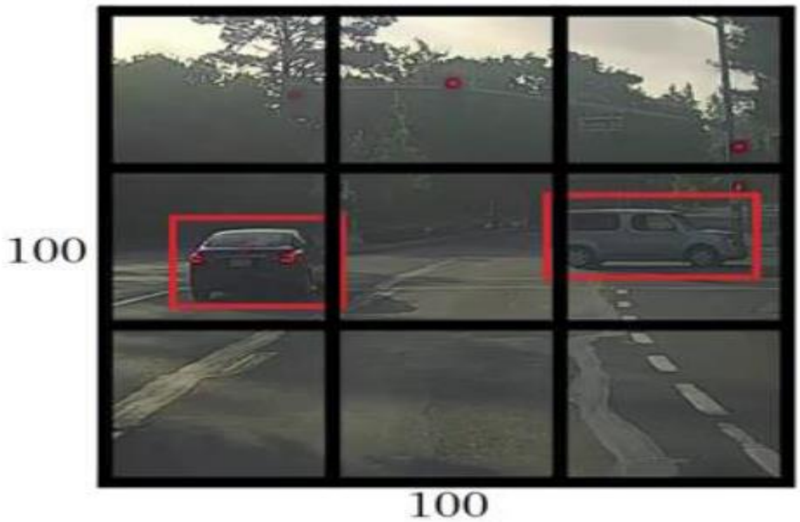

Fig. 3: Step 2- Processed image after grid division

Now, each bounding box can be represented using 4 parameters:

a. Centre of a bounding box (bx,by)
b. Height of a bounding box (bh)
c. Width of a bounding box (bw)
d. And a value 'c' which represents the class of an object.

A parameter 'pc' also used represents the probability of an object located in that bounding-box. Now, for each grid cell of an image, a 'y label' consisting of all the different parameters mentioned above which is a dimensional vector. And after all this encoding of bounding boxes is done. YOLO predicts numerous bounding boxes for every sub image. So, if the center/centroid of an object is found to be inside a subimage, then that particular grid cell is held selected for detecting that object.[8]

Table 1: Representation of dimensional vector 'y' used in bounding box encoding

y =	**pc**
	bx
	by
	bh
	bw
	c

Now, overall a bounding box is selected and held accountable for detecting an object using parameters such as, Intersection over Union (IoU) & Non-Max Suppression.
In this paper, we have used YOLOv3 for social distancing detection over the Darknet open-source framework and YOLOv5 for face mask detection using PyTorch[9] framework.

A. Physical Social Distancing Detection

We have used a pre-trained YOLOv3 for Social Distancing Detection.[10] YOLOv3 utilizes a version of Darknet (Darknet-53). As the name suggests it uses standard network train on ImageNet with 53-layer. Furthermore, the task of object localization another stack of 53 layers is added before it resulting in 106 layer CNN.[4]

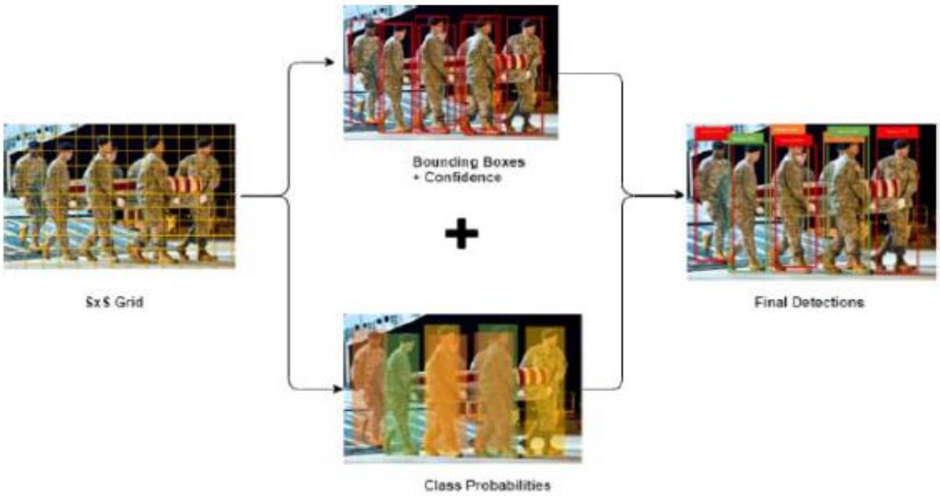

Fig. 5: Overview of YOLOv3 algorithm

The developed architecture separates the image into numerous sections, each representing a bounding box, with class likelihood scores for each component to consider as an item in this project. This method provides tremendous increases in terms of speed while sacrificing efficiency. The detector module has strong generalization capabilities, allowing it to represent a complete image. Many research discoveries have been reported in recent years based on the above notions. Crowd counting has developed as a promising field of study with numerous societal implications. We pridected using one dataset with blended annotations for object recognition and classification, sample segmentation, and multichanneled picture details would allow investigate object identification tasks more effectively, as well as encourage progress toward true scene understanding.[11] All of the reviewed literature and related research work paints a clear picture of how human being detection can have variety of utilizations to address current situation, such as checking prescribed hygiene standards, social distancing, work practices, and so on.[8]
Now, the framework continues as follows:
1.To find and trace a people in a video, fine-tune the trained object identification model.
2. Surveillance footage is fed to the trained model. For each person detected, the model generates a set of bounding cuboid and an ID.
3. Each individual has a three-dimensional feature space (x, y, d), where (x, y) are the centroid coordinates of the bounding cuboid and d is the depth of the individual as seen by the camera.

$$d = ((2 * 3.14 * 180)/(w + h * 360) * 1000 + 3) \quad (1)$$

where the enclosing cuboid's width is w and its height is h.
4. The pairwise L2 norm is computed for the set of bounding cuboids.
5. The dense matrix of the L2 norm is then used to assign neighbours to each people who meets the minimum distance sensitivity requirement. The minimum distance threshold is dynamically adjusted in extended trials based on the spatial location of the person in a particular frame ranging between (90, 170) pixels.
6. Any person who meets the closeness property is allocated a neighbour or neighbours, producing a group that is colour-coded differently than the rest of the population.
7. The creation of groups implies a breach of the practice of social separation, which can be measured using the following formula:
- Take as the number of identified groups or clusters,
- and as the total strength of people detected in close proximity.
- **Vi = (np)/(ng)**, where Vi stands for the variable index of violation

Table. 1: YOLOv3 Network Architecture

Type	Filters	Size	Output
Convolutional	32	3 x 3	256 x 256
Convolutional	64	3 x 3 / 2	128 x 128
Convolutional	32	1 x 1	
Convolutional	64	3 x 3	
Residual			128 x 128
Convolutional	128	3 x 3 / 2	64 x 64
Convolutional	64	1 x 1	
Convolutional	128	3 x 3	
Residual			64 x 64
Convolutional	256	3 x 3 / 2	32 x 32
Convolutional	128	1 x 1	
Convolutional	256	3 x 3	
Residual			32 x 32
Convolutional	512	3 x 3 / 2	16 x 16
Convolutional	256	1 x 1	
Convolutional	512	3 x 3	
Residual			16 x 16
Convolutional	1024	3 x 3 / 2	8 x 8
Convolutional	512	1 x 1	
Convolutional	1024	3 x 3	
Residual			8 x 8
Avgpool		Global	
Connected		1000	
Softmax			

Face Mask Detection

In this paper, we have been using YOLOv5 for the purpose of face mask detection. We implemented YOLOv5 for face mask detection using the PyTorch[9] framework. YOLOv5 is an upgrade over all the previous versions of YOLO due to it's performance speed and it's the ability to retain its organization even on our own custom dataset. Detecting a face mask on a person's face was fairly easier than the social distancing detection model mentioned above. This was because YOLOv5 is pre-trained over different classes and the concept of Transfer learning makes it easier for the model to make accurate predictions over newly added classes. The dataset we used was divided into only 2 classes. So, we used the pre-trained weight of the YOLOv5 and changed the name of the classes to be predicted based on our datasets. All this made our work easier and we were able to get good results.

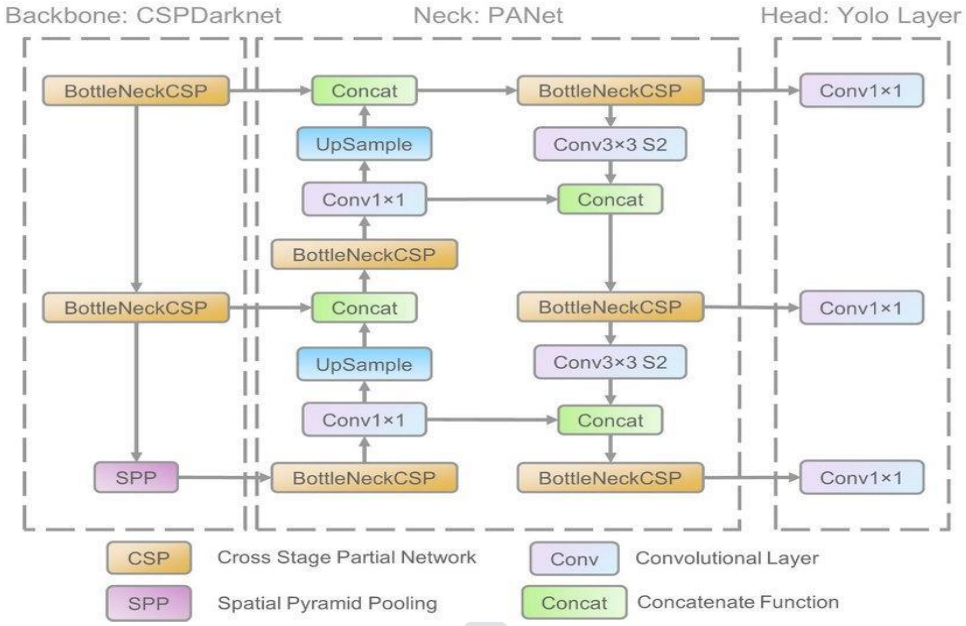

Fig. 7: YOLOv5 Network Architecture

V. Examination Results

Various parameters are considered when measuring the overall performance of a model. Some of these parameters are IoU, Accuracy, Precision, Recall, mAp etc.[12] The parameters used for YOLO are mentioned as:

a. IoU (Intersection over Union) - It is an evaluation metric and is used to measure the accuracy of an object detector. It is widely used in models such as YOLO, R-CNN based object detectors. It helps in measuring how accurate the object identified in an image/frame is.

$IoU = \text{area}\ A \cap B A \cup B$ (2)

where A is the actual bounding cuboid labelled and B is the predicted bounding cuboid by the model.

b. Accuracy = (TP+TN)(TP+TN+FP+FN) (3)

c. Precision = TPTP+FP (4)

d. Recall = (TP)(TP+FN) (5)

Now, as our YOLOv3 model was pre-trained for the 'person' class, so the results obtained are the same as that of the actual model. While testing we achieved an accuracy of 91.25% due to the instability of the camera frames.

Fig. 8: Test image before detection

Fig. 9: Test image after detection

As it can be seen in Fig. 9 that the model is detecting the persons in the image and accurately calculating the the number of social distancing violations.

Fig. 10: Before face mask detection

Fig. 11: After face mask detection

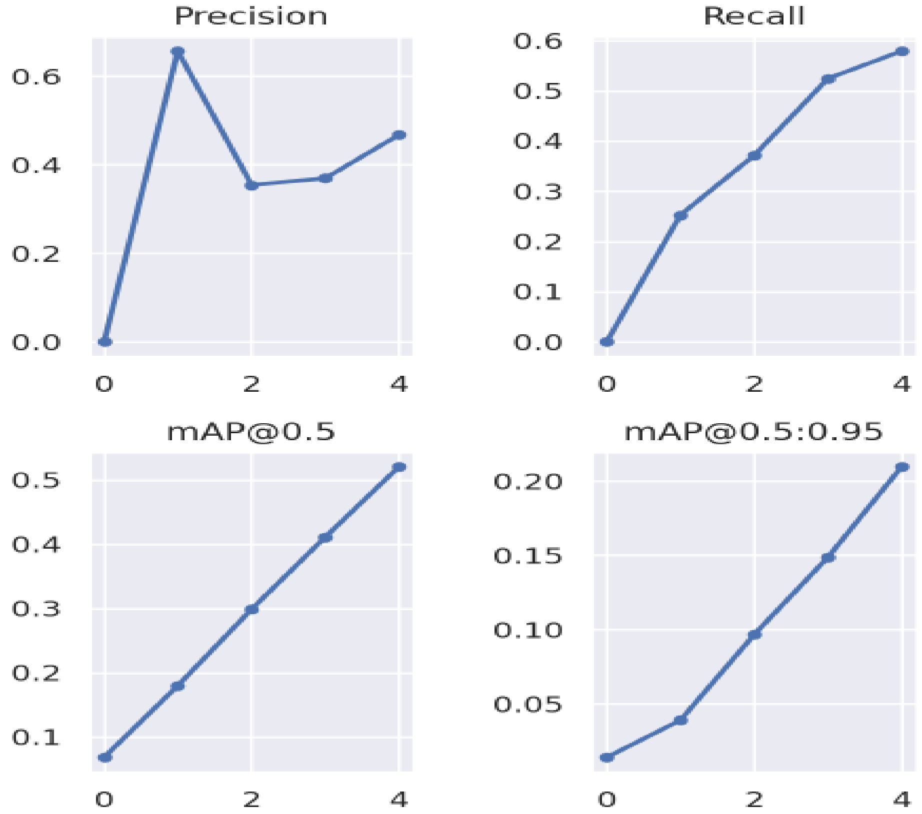

Fig. 12: Performance parameters of YOLOv5 after training and testing

VI. Conclusion and future scope

In this paper, social distancing detection is applied on videos using the model pre-trained on 1 class('person') from the COCO dataset. Also, for the face mask detection model, the results we achieved were satisfactory using YOLOv5. Both YOLOv3 & YOLOv5 are excellent choices for object detector models mainly because of their speed and efficiency in implementing them for various applications.

For the future work of our project, we will be improving on both our models and implementing them on Live Surveillance cameras. This is a necessary step in the time of an ongoing global pandemic. We will also be using other object detector models and comparing them with the models we have developed. Furthermore, we will be working on a path tracking model as well.

References

[1] "W. H. Organization, 'WHO corona-viruses (COVID-19),' https://www. who.int/emergencies/diseases/novel-corona-virus-2019, 2020, [Online; accessed May 02, 2020]." .

[2] "@misc{darknet13, author = {Joseph Redmon}, title = {Darknet: Open Source Neural Networks in C}, howpublished = {\url{http://pjreddie.com/darknet/}}, year = {2013--2016} }." .

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[4] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv*, 2018.

[5] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[6] "https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006." .

[7] N. Mittal, A. Vaidya, and A. Shreya Kapoor, "Object Detection and Classification Using Yolo," vol. 5, no. 2, 2019.

[8] I. Ahmed, M. Ahmad, J. J. P. C. Rodrigues, G. Jeon, and S. Din, "A deep learning-based social distance monitoring framework for COVID-19," *Sustain. Cities Soc.*, vol. 65, no. November, p. 102571, 2021, doi: 10.1016/j.scs.2020.102571.

[9] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," *arXiv*, no. NeurIPS, 2019.

[10] A. Bathija, "Visual Object Detection and Tracking using YOLO and SORT," *Int. J. Eng. Res. Technol.*, vol. 8, no. 11, pp. 705–708, 2019, [Online]. Available: https://www.ijert.org.

[11] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques," *arXiv*, pp. 1–10, 2020.

[12] S. Saponara, A. Elhanashi, and A. Gagliardi, "Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19," *J. Real-Time Image Process.*, no. 0123456789, 2021, doi: 10.1007/s11554-021-01070-6.