Noname manuscript No.
(will be inserted by the editor)

Real-time social distancing detector using SocialdistancingNet-19 deep learning network

Rinkal Keniya · Ninad Mehendale

Received: date / Accepted: date

Abstract With no doubt, the COVID-19 pandemic has put the world to a halt. The world we lived in a few months prior is completely different than what it is now. The virus is spreading quickly and is a danger to the human race. Seeing the necessity of the hour one must always take certain precautions of which one being social distancing. Maintaining social distancing during COVID-19 is a must to ensure a slowdown in the growth rate of new cases. Our manuscript focuses on detecting if the people around are maintaining social distancing or not. Using our own self developed model named SocialdistancingNet-19 for detecting the frame of a person and displaying labels, they are marked as safe or unsafe if the distance is less than a certain value. This system can be used for monitoring people via video surveillance in CCTV. Our model achieved an accuracy of 92.8 %.

Keywords Social distancing · Object detection · COVID

1 Introduction

Coronavirus is an infectious disease caused by the corona virus-2 extreme acute respiratory syndrome. The disease was first detected in Wuhan, China in December, which has contributed to a spread across the world. When in close contact, the virus spreads mainly between individuals, including by tiny droplets formed when sneezing or coughing. Droplets falling on the ground will pass through the air through the body of a human. For the first three days, the infection is the most infectious. Many typical symptoms include nausea, dry cough, and fatigue. Severe and harmful human consequences have contributed to a worldwide halt. Many such signs may include sore throat and headache. It takes a fortnight for a person with mild symptoms to get healed. The duration of recovery for individuals with severe symptoms depends on the extent, along with an individual's immune capability. The main diagnostic approach is from a nasopharyngeal swab by a real-time reverse transcription-polymerase chain reaction (RRT-PCR). Chest CT imaging is also useful for the diagnosis of people with an elevated probability of infection based on signs and risk factors. Seeing the devastating spread of the disease, the World Health Organization (WHO) suggested favoring the term social distancing. To slow down the rate of spread of the disease it is necessary to maintain physical distance. Maintaining a distance of two meters between two individuals is a must to remain safe and get back to the world we lived a few months back. After the COVID-19 pandemic, the CDC changed the concept of social distancing as keeping out of congregate environments, preventing public meetings, and preserving, when appropriate, a gap of around six feet or two meters from everyone. Recent findings have shown that droplets from a sneeze or a deep breath will fly more than six meters during exercise. And hence maintaining the norm of social distancing is a necessity and also in our benefit to live a safer and healthier life. Our work proposes to determine whether or not an individual is following the rule of social distancing. The findings are verified using both a live stream as well as a video feed. By measuring the gap of two frames of people from the centroids, we can understand whether or not a person is maintaining social distancing. Also, they are labelled as safe and unsafe.

* Corresponding author
N. Mehendale
B-412, K. J. Somaiya College of Engineering, Mumbai, India
Tel.: +91-9820805405
E-mail: ninad@somaiya.edu

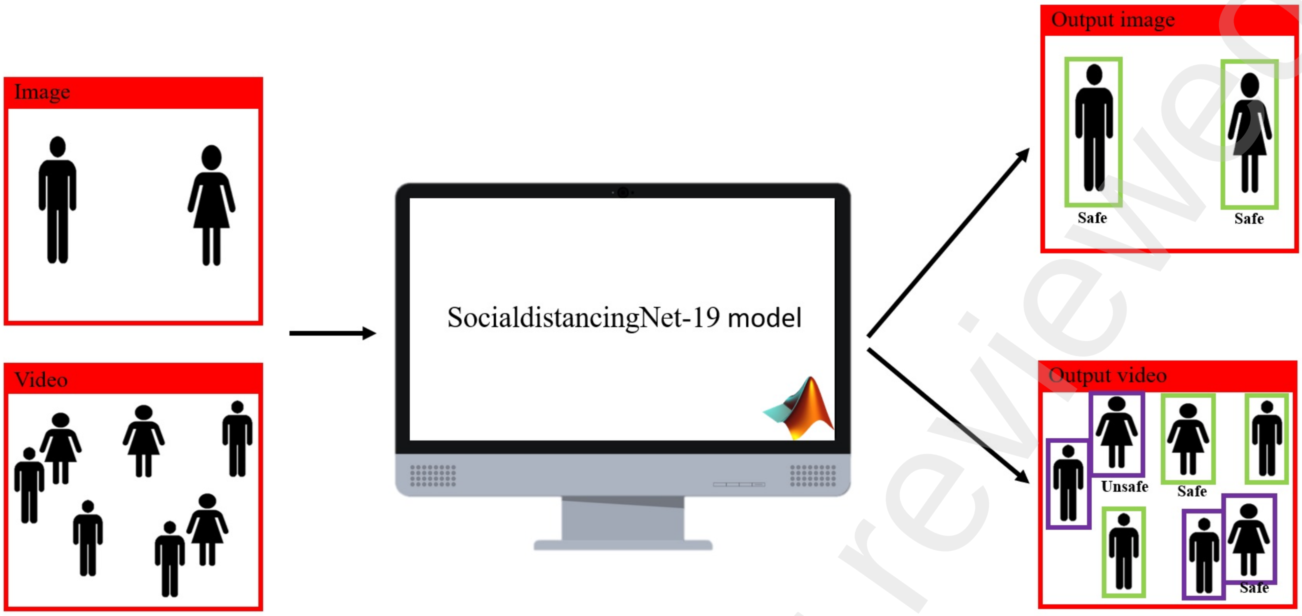

Fig. 1 A video stream or an image is fed as an input to our self developed model named SocialdistancingNet-19. The people are detected as maintaining social distancing or not depending on the distance maintained between two individuals. They are marked in frames of different colours and also labels are marked for each of them.

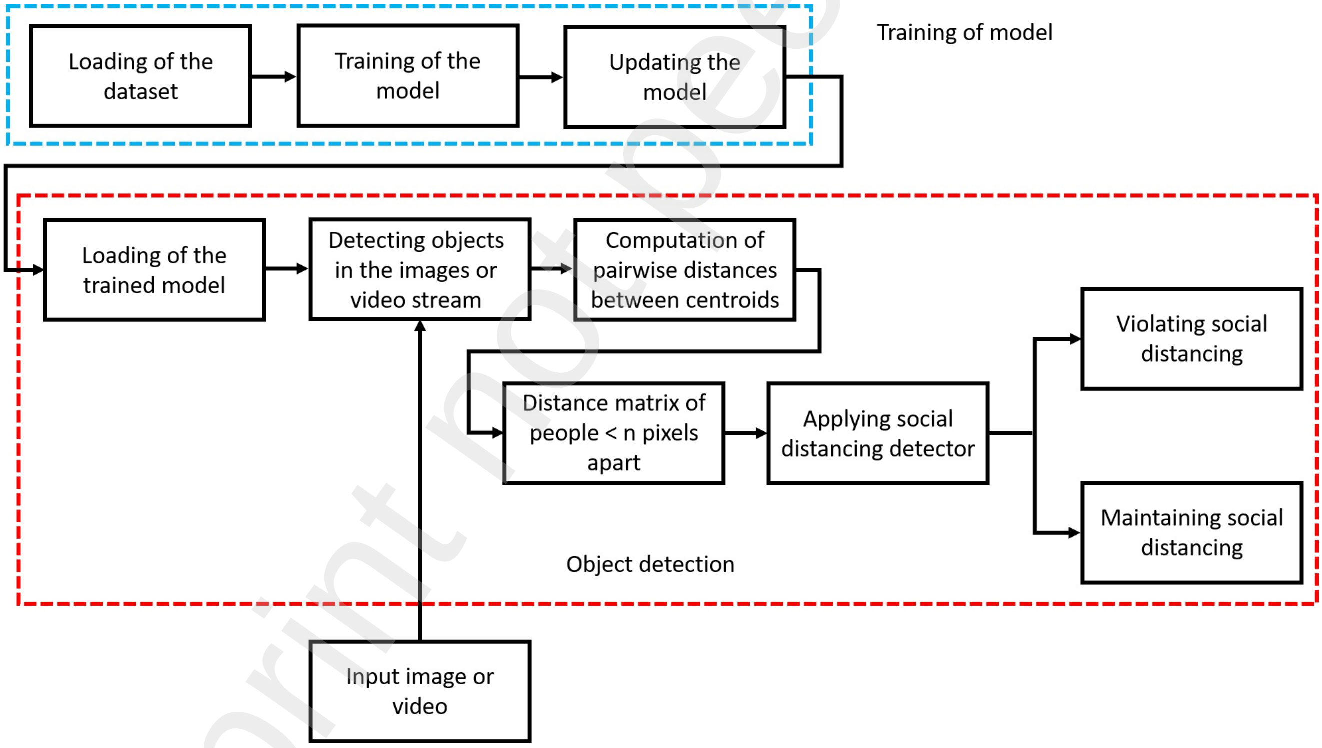

Fig. 2 The training of the model is first carried out by loading the dataset into the model and then trained. Later, the model is loaded and then objects are detected in the image and video stream. Further depending on the distance frames are marked on the people along with labels indicating the marking as maintaining or violating social distancing.

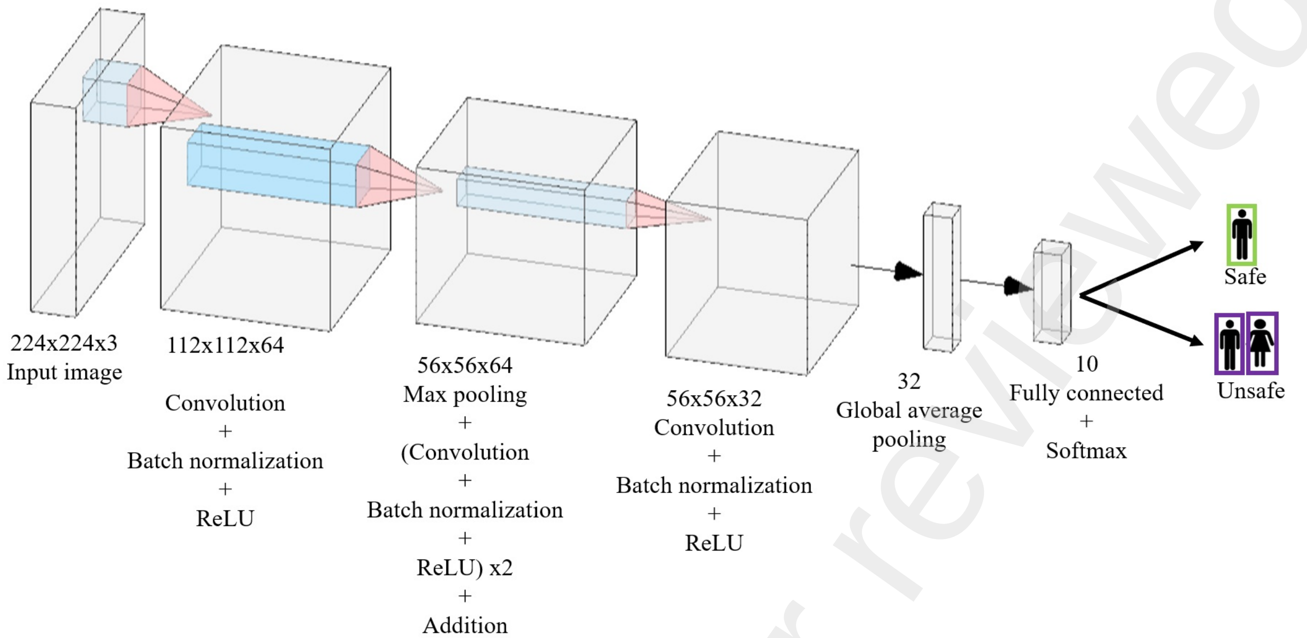

Fig. 3 SocialdistancingNet-19 has an architecture of 19 layers. The network is fed with an input image. Then it is further passed through a convolution, batch normalization and ReLU (Rectification Linear Unit) layers. After that it is passed through a single max pooling layer, two convolution layers, two batch normalization layers, two ReLU layers and a single addition layer. It is then passed through single convolution, batch normalization and ReLU (Rectification Linear Unit) layers. And at the end was finally passed through a fully connected and a softmax layer. And then we received the classification output.

Model	Accuracy (%)
Yadav *et al.* [1]	91
Sener *et al.* [2]	93.3
Liu *et al.* [3] (SSD300)	74.3
Liu *et al.* [3] (SSD512)	76.8
ResNet-50	86.5
ResNet-18	85.3
SocialdistancingNet-19 (Proposed method)	92.8

Table 1 Comparison of the accuracy values of the different methodologies. The SocialdistancingNet-19 model gave the highest accuracy as compared to the other models.

2 Literature review

Various research work has been carried out on social distancing using different techniques. Yadav *et al.* [1] proposed a system that used raspberry pi4 with a camera to automatically track public spaces in real-time to prevent the spread of Covid-19. The trained model with the custom data set was installed in the raspberry pi4, and the camera was attached to it. The camera is fed with real-time videos of public places to the model in the raspberry pi4, which continuously and automatically monitors public places and detects whether people keep safe social distances and also checks whether or not those people wear masks. Their method operates in two stages: first, when a person identified without a mask his photo was taken and sent to a control center at the State Police Headquarters; and second, when the detection of a social distance violation by individuals was detected continuously in threshold time, there rings an alarm that instructs people to maintain social distance and a critical alert is sent to the control center of the State Police Headquarters for further action. They achieved an accuracy of 91 %. Singh Punn *et al.* [4] proposed a real-time based deep learning to monitor social distancing using object detection and tracking approaches. The number of violations was given by computing the number of groups formed and the violation index term computed as the ratio of the number of people to the number of groups. Different object detection models were used like Faster RCNN, SSD, and YOLO v3, where YOLO v3 with balanced performance of FPS and mAP score. An AI monocular camera-based real-time system to monitor social distancing was proposed by Yang *et al.* [5]. The proposed method uses a

Fig. 4 Results when the input video and images were given to the model. In (a) and (b), the people were detected as maintaining social distancing or not depending on the distance maintained between two individuals. They were marked in frames of different colours. Green colour is marked for violating social distancing and labelled as unsafe. The purple frame was marked for those maintaining social distancing and labelled as safe.In (c) and (d), the people are detected and frames were marked as per the distance between two individuals. Along with this the number of violations was also counted.

critical social density to avoid overcrowding by modulating inflow to the region of interest. The method was verified using 3 different pedestrian crowd datasets. But there were some missing detections in the train station dataset, as in some areas the density of pedestrians is very high and occlusion happens. However, after some analysis, they concluded that the maximum pedestrians were captured and the idea of social density is valid. In the proposed method by Sener *et al.* [2] the motion of the communicating people was extracted from each region of the detected individual. Then, visual descriptors for two persons are created. As the relative spatial positions of communicating people are likely to complement the visual descriptors, we propose to use embedding of spatial multiple instances, which implicitly integrates the distances between people into the learning process of multiple instances. Experimental findings on two benchmark datasets validate that the use of two-person visual descriptors along with multiple-instance spatial learning provides an efficient way to infer the form of interaction. They achieved an accuracy of 93.3 %. Bielecki *et al.* [6] did a study of 508 male soldiers with average age of 21years. They followed the number of soldiers into two groups. For the 354 soldiers affected before social distancing was introduced, COVID-19 caused 30 % to become sick. While no soldier in a population of 154, in which infections occurred after social distancing had been introduced. An innovative localization method was proposed to by Nadikattu *et al.* [7] to track humans' positions in the surrounding based on sensors. This AI smart device is not only handy for maintaining social distancing but also detects symptoms of COVID in and person if any. The system will warn the user if anyone is near him within the vital six-foot radius. Ghorai *et al.* [8] proposed a deep learning solution that would alert the person as soon as on violates social distancing. A video stream is captured from the CCTV camera and with the PoseNet model the people are detected and then kept a rack of the number of people present in the video stream. If the distance between 2 frames of people is less then the authorities in-charge are alerted. Using deep learning techniques a drone was proposed by Ramadass *et al.* [9] for inspection of social distancing and also to check if a person is wearing a mask or not. In the camera of the drone is installed the qualified yolov3 algorithm with the custom data collection. The drone camera runs the yolov3 algorithm and determines whether or not social space is preserved and whether the individuals wearing masks are in the crowd. The drone is made fit to operate automatically. Reluga *et al.* [10] proposed a differential-game for determining whether persons during an outbreak can use social distancing and associated self-protective behaviors. The differential game is used as a mitigating tool to research the possible utility of social distancing by measuring the equilibrium actions under several cost functions. Following outbreak detection, computational techniques are used to measure the cumulative expense of an infection under equilibrium practices as a result of the period until mass vaccination. The main parameters in the study are the specific number of reproductions and the underlying efficacy of social distancing. To slow the spread of the COVID-19 virus via airborne transmission, a "social distancing" approach of around 1.83 m (6 feet) was recommended in the proposed method by Feng *et al.* [11]. It was also found that the wind effect on droplet transport and deposition is dynamic and highly dependent and localized on the wake flow patterns. Secondary flow intensities between the two simulated beings, and calm currents. High RH=99.5 % leads to higher deposition fractions on both human bodies and the ground, which is not necessarily related to higher exposure risks. High RH=99.5 % can enhance the condensation effect, and the cough droplet sizes keep growing during their transport in the air until the partial pressure at the droplet surface is equal to the saturation pressure of water vapor. In contrast, RH=40 % triggers the evaporation of the water in cough droplets, thereby leading to droplet size reduction, which may lead to a long time suspended in the air. High RH=99.5 % results in higher percentages of deposition on both human bodies and the environment, which are not generally correlated with a higher risk of radiation. Venkateswaran *et al.* [12] proposed a System Dynamics (SD) model of the Covid-19 pandemic spread in India. The model is an age-structured compartment based approach to explore different modes of disease propagation, greatly extending from the traditional SEIR approach. The model was adapted for India using the correct population ladder, matrices for touch levels, external arrivals. They also specifically monitored the results of models like touch recording, segregation of COVID-positive patients, quarantining, use of masks, better grooming procedures, social distancing by and touch levels in various places of home, college, school, and other locations. Results of the simulation suggest that any non-trivial number of pathogens will be left even after a prolonged lockout and the pandemic will resurface. Liu *et al.* [3, ?] presented a method for detecting objects in images using a single deep neural network. The model named single shot multibox detector (SSD), discretized the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. The results on the PASCAL VOC, COCO, and ILSVRC datasets showed that SSD has competitive accuracy to methods that utilize an additional object proposal step and is much faster, while providing a unified framework for both training and inference. The accuracy for SSD300 was 74.3 % and for SSD512 was 76.8 %.

3 Methodology

We loaded 295 images from the dataset, where each image had single or multiple labels inside it which were used for training the model. Further, more images and labels were generated using an auxiliary dataset. The auxiliary dataset is a variation of the images in terms of rotation(+5,-5), scaling (0.95 to 1), and cropping(0.95 to 1). The dataset was then stored into two different columns. First, the image file path and the second is the corresponding label. Later the dataset is split into training and testing for validation and 60 % of the dataset is selected for training, 10 % for validation, and the remaining 30 % for testing of trained detectors. We used the SocialdistancingNet-19 architecture for the training purpose. Box labels were used to create the data for training and evaluation purposes. A rectangular box was used to mark the object. This network comprises of 2 subnetworks- feature extraction and feature detection. The feature extraction was carried out by a pre-trained convolutional neural network (CNN) model. We also used a reduced ResNet-50, MobileNet-V2 and ResNet-18 network. The detection of sub-networks of small CNN is compared to feature extraction and is composed of a few convolutional layers specific to the YOLO object detection model. The YOLO detection model is similar to the single-stage detector model. This algorithm views object recognition as a problem of regression, taking a given input image or video stream and concurrently knowing the bounding box coordinates and the corresponding labels of class probabilities. YOLO has three tuning parameters, network input sizes, anchored box, and feature extraction network. First, the frame is detected. We then compute bounding box coordinates and then derived the center of the bounding box. Using the box coordinates the top-left coordinates are derived. Afterwhich the frame is pre-processed giving three results which are confidence, bounding box, and centroids of each person. The euclidean distance is calculated and used to find the distance between centroids. After the comparison of the distance between the centroids of two individuals, it is compared with the minimum distance in terms of pixels. The pairs are marked as red or green depending on if they have violated social distancing or not. The user specifies the input size and number of classes while choosing a network. With the minimum size for a network, the size of the training image and the computational cost was optimized. We tried to find the best model as per input size and set of training images and optimize it to handle larger data sets than the current dataset. SocialdistancingNet-19 has an architecture of 19 layers. The network is fed with an input image of dimension 224x224x3. Then it is further passed through a convolution, batch normalization and ReLU (Rectification Linear Unit) layer each of dimension 112x112x64. After that it is passed through a single max pooling layer, two convolution layers, two batch normalization layers, two ReLU layers and a single addition layer. Each of these layers were of dimension 56x56x64. Further it was passed through single convolution, batch normalization and ReLU (Rectification Linear Unit) layers, each of dimension 56x56x32. Then it was passed through a global average pooling layer of dimension 1x1x32. And at the end was finally passed through a fully connected and a softmax layer each of dimension 1x1x10. And then we received the classification output. The reduced computational cost was having 224x224x3 which was the bare minimum size required to run any network. Image resizing was the only pre-processing operation required before training. Then, the estimated anchor boxes were used for object training to account for resizing before the training. Also, the estimated anchor resizes. This was done to transform the process with the number of anchor boxes estimated in the resized images. And later stored in the processed data directly. Activation layer 40 of ReLU (Rectification Linear Unit) is generally selected for the feature extraction layer and we refresh the activation layer with the detected sub-network. The feature extraction layer outputs the feature maps and down samples it by the factor of 16. The amount of downsampling was good to maintain the tread between the special resolution and strength of the extracted feature. This feature extracted downs to the encoder with a stronger image feature that was used to estimate the cost of the special resolution. Data augmentation was carried out to improve the accuracy by randomly transforming the data while training. Data augmentation added more variety during training. And actually, increases the number of labels in the training data samples. The use of transform augmentation during the training allows random keeping of images. The associated box labels are also flipped horizontally. Augmentation is not performed for the validation and test data and hence evaluation can be carried out unbiasedly since the data is unmodified. NVIDIA GPU- 1660, 1408 Cuda core with 6GB DDR5 RAM and 192 bits memory bus was used to train the network.

4 Results and discussion

The accuracy of developed model SocialdistancingNet-19 was 92.8 %. The accuracy of the ResNet-50 network was 86.5 %. For ResNet-18 the accuracy was 85.3 %. We tested our model using a video stream and images. Of which, we could see the proper detection of people according to the distance between a pair. The frames were also labelled as safe and unsafe accordingly. Also, the count of the violations made were counted and were constantly updating. While using the webcam, it is necessary to have people moving continuously else the detection goes incorrect. This could happen due to the detection method, wherein the entire frame is detected, and further, the distance calculation and comparison between the centroids takes place. The results obtained by the model are displayed in fig 4. The purple and green coloured images displayed along with the labels indicate if the person is maintaining social distancing or not. The table 1 shows the comparison with different models tested and found in the reviews and their respective accuracies. The maximum accuracy was 93.3 % and 74.3 % was the minimum accuracy.

5 Conclusions

Our work distinguishes the social distancing pattern and classifies them as a violation of social distancing or maintaining the social distancing norm. Additionally, it also displays labels as per the object detection. The classifier was then implemented for live video streams and images also. This system can be used in CCTV for surveillance of people during pandemics. Mass screening is possible and hence can be used in crowded places like railway stations, bus stops, markets, streets, mall entrances, schools, colleges, etc. By monitoring the distance between two individuals, we can make sure that an individual is maintaining social distancing in the right way which will enable us to curb the virus.

6 Acknowledgement

Authors would like to thank all colleagues from COVID research group.

Compliance with Ethical Standards

Conflicts of interest

Authors R. Keniya, and N. Mehendale, declare that he has no conflict of interest.

Involvement of human participant and animals

This article does not contain any studies with animals or Humans performed by any of the authors. All the necessary permissions were obtained from the Institute Ethical Committee and concerned authorities.

Information about informed consent

No informed consent was required as the studies does not involve any human participant.

Funding information

No funding was involved in the present work.

References

1. S. Yadav, Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence
2. F. Sener, N. Ikizler-Cinbis, Two-person interaction recognition via spatial multiple instance embedding, Journal of Visual Communication and Image Representation **32**, 63 (2015)
3. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, in *European conference on computer vision* (Springer, 2016), pp. 21–37
4. N. Singh Punn, S.K. Sonbhadra, S. Agarwal, Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques, arXiv pp. arXiv–2005 (2020)
5. D. Yang, E. Yurtsever, V. Renganathan, K.A. Redmill, U. Ozgüner, A vision-based social distancing and critical density detection system for covid-19, arXiv e-prints pp. arXiv–2007 (2020)
6. M. Bielecki, R. Züst, D. Siegrist, D. Meyerhofer, G.A.G. Crameri, Z.G. Stanga, A. Stettbacher, T.W. Buehrer, J.W. Deuel, Social distancing alters the clinical course of covid-19 in young adults: A comparative cohort study, Clinical Infectious Diseases (2020)
7. R.R. Nadikattu, S.M. Mohammad, P. Whig, Novel economical social distancing smart device for covid-19, International Journal of Electrical Engineering and Technology (IJEET) (2020)
8. A. Ghorai, S. Gawde, D. Kalbande, Digital solution for enforcing social distancing, Available at SSRN 3614898 (2020)
9. L. Ramadass, S. Arunachalam, Z. Sagayasree, Applying deep learning algorithm to maintain social distance in public place through drone technology, International Journal of Pervasive Computing and Communications (2020)
10. T.C. Reluga, Game theory of social distancing in response to an epidemic, PLoS Comput Biol **6**(5), e1000793 (2010)
11. Y. Feng, T. Marchal, T. Sperry, H. Yi, Influence of wind and relative humidity on the social distancing effectiveness to prevent covid-19 airborne transmission: A numerical study, Journal of aerosol science p. 105585 (2020)
12. J. Venkateswaran, O. Damani, Effectiveness of testing, tracing, social distancing and hygiene in tackling covid-19 in india: A system dynamics model, arXiv preprint arXiv:2004.08859 (2020)