Deep Learning and Computer Vision-based Social Distancing Detection System

Radhika Kumari[1], Ms. Upasna Joshi[2]

[1]Student, Department of Computer Science and Engineering, Delhi Technical Campus, Greater Noida, U.P., India
[2]Professor, Department of Computer Science and Engineering, Delhi Technical Campus, Greater Noida, U.P., India

---***---

Abstract - *The COVID-19 pandemic has certainly put a stop to the world. In a few months, the world we lived in has totally changed. The virus spreads rapidly and poses a risk for mankind. When one sees the need for the hour, certain precautions must always be taken that are socially distant. In order to ensure a slowdown in the rate of growth of new cases, it is important to maintain social distancing during COVID-19. The video surveillance system in CCTV can be used to track people. With its lethal spread in more than 190 countries, RCDs 2019 (COVID-19) have resulted in a global epidemic, where nearly 3,519,901 confirmed cases and a total of 247,630 deaths worldwide, compared to 4 May 2020. The shortage of therapeutic active ingredients and tolerance to COVID19 increase public susceptibility. As no vaccines are available, the only effective way to tackle this pandemic is to social distance.*

***Key Words*: Social Distancing, Epidemic, COVID-19, Pandemic, Surveillance.**

1. INTRODUCTION

Coronavirus is a corona-virus-2 severe acute respiratory syndrome infectious disease. In December, in Wuhan, China, the disease was first identified which helped spread throughout the world. The virus spreads mainly among people in close contact, including minute droplets produced while sneezing or coughing. Droplets that fall to the earth move the air into a human body.

The infection is the most contagious for the first three days. Nausea, dry cough, and fatigue are two of the common symptoms. The worldwide halt also led to severe and dangerous human consequences. Many of these symptoms may include headache and sore throat. An individual with mild symptoms needs a couple of weeks to recover. The recovery period depends on the degree and immune function of individuals who are suffering from serious symptoms.

The principal diagnostic technique is the reverse transcription/polymerase chain reaction of a nasopharyngeal swab (RRTPCR). CT imaging of the chest is also useful for detecting people with high infection likelihood based on signs and risks. The World Health Organization (WHO) proposed that the word social distancing should be used to understand the destructive spread of the disease. It is necessary to maintain physical distance to slow the spread rate of the disease. Holding two meters away from two people is a must to be safe and to return to the world which we lived a few months ago. Following the pandemic in COVID-19, the CDC modified the idea of the social distance to avoid public meetings and to establish separation from everyone, which was about six or two feet apart. Latest studies have shown that sneeze or deep breath droplets are moving for more than six meters.

The maintenance of the social distance level is therefore a requirement to live a healthy and safer life and in our interest. Our job is to decide whether a person follows the social dividing rule or not. Both a live stream and a video feed check the results. By measuring the distance between two frames of people in the middle, we can understand whether or not an individual holds social distance. They are often called healthy and uncertain.

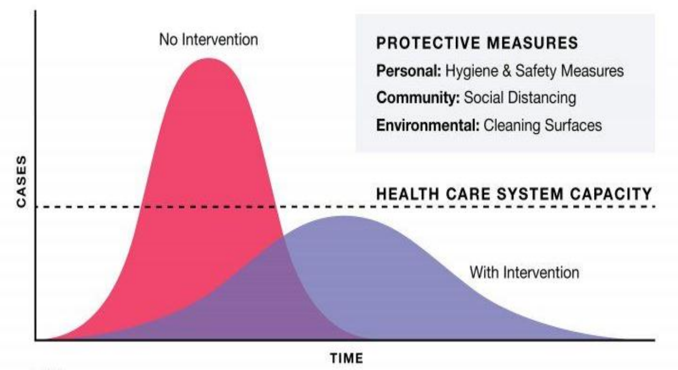

Figure-1-: Graph showing the result of social distancing as a reduced peak of the epidemic matched with available healthcare capacity

In this article, the system proposed uses the YOLO v3 object detection model to isolate a person from the surroundings and the Deep Learning method to track detected individuals with bounding boxes and assigned identities.

2. LITERATURE SURVEY

Several research projects on social distancing have been performed using various techniques. The social gap for COVID-19. Since December 2019, COVID-19 has induced extreme acute respiratory syndrome worldwide [1]. Recent work has shown that public distancing is a significant action for slackening and fall off of the expansion of COVID-19 [2]. Further analysis [5] also shows that social distance has important economic benefits. COVID-19 could not be completely removed in the short run, but an automated system to track and interpret measures of social distance can substantially support our community. Another survey [4] explores both general object detection and pedestrian detection methods in deep learning. Detectors of state-of-the-art objects use deep learning techniques, which can be typically split into two categories. The initial one is called a two-tier detector, typically based on Region-Based Convolutional Neural Networks [20, 22, 11, 7], which begins with regional bids and then form a division and regression of the bounding box. The following is referred to as a single level predictor with YOLOv1-v4 as the prevalent models. [13, 16, 25, 26], SSD [17], RetinaNet[18], and Effective Det[19].

Pedestrian detection can be used either as the part of a general detection issue or just as a basic activity to detect humans. *Singh Punn et al.* [10] gave a real-time deep learning approach using target detection and tracking. The numbers of groups created and the infringement index are determined as a proportion of the population to the number of groups.

Different models like Faster RCNN, SSD, and YOLOv3 were used for object detection. *Yadav et al.* [3] proposed an automatic real-time system to track public areas with the help of a raspberry pi4 with a camera installed with it to avoid the spread of the COVID virus. The raspberry pi4 was installed with the custom data and the camera was fed with real-time video streams to make sure if people are following social distancing or not. It even verifies whether the individuals wear masks or not. This was the two-step method in which the photo of the persons not wearing the masks was taken and sent to the Headquarters of government police and also when the field of concern is overcrowded. This method was carried out using three different datasets of pedestrians but there were a few incomplete train station dataset detections. In the proposed procedure, *Sener et al.* [8] the motion of the communicators was extracted from each region of the person detected. Visual descriptors for two people are then established. Because the relative position of spatial experiences of people is likely to complement the visual descriptors, they proposed to integrate many spatial instances into the learning process of multiple instances, which meant that distances between people are included. Experimental findings from two baseline datasets support the efficient way to deduce the form of contact by using two-person visual descriptors together with several instances of spatial learning. They obtained a precision of 93.3 percent. A precise social distance monitoring approach using YOLO version 3 and Deepsort was suggested for the identification and tracking of a person, accompanied by the estimation of the infringement index for comportments of non-social distance. The method is stimulating, but there is no statistical analysis of the findings. Moreover, no other subject than the infringement index exists for implementation of privacy. *Nadikattu et al.* [6] proposed a groundbreaking localization approach to monitor human positions in the sensor-based world. This AI smart device is not only useful for maintaining social distance but also detects signs of COVID in and in person if any. The machine will alert the consumer if anyone is close to him within a critical six-foot radius.

Bielecki et al. [9] examined 508 military men, aged 21 years on average. The number of soldiers in the two groups was followed. For 354 soldiers affected until the social distance is enforced, COVID-19 caused 30% of them to become ill. Although there was no militant in the population of 154, where infection occurred after the enforcement of social distances. Landing AI [14] has suggested a social distance sensor to highlight people whose physical separation is below the specified threshold using a monitoring camera. A related method [15] has been implemented to observe employee operations including instantaneous speech warnings at the industrial plant. *Venkateswaran et al.* [21] suggested the Covid-19 pandemic model based on a system dynamic (SD) model. The model is a compartmental approach focused on age, which detects different modes of disease transmission and dramatically increases the SEIR approach. The prototype was modified as per the right population ladder, touch-level matrixes, and external arrivals for India. They also directly monitored the effects of models such as touch-recording, isolation of COVID-positive patients, quarantine, and use of masks, enhanced grooming procedures, social distance, and touch-level in different places in the home, college, school, and other different sites. Outcomes of the analysis indicate that a significant number of the pathogen will remain even after a prolonged lockout and the pandemic will resurface. Social distance monitoring is also called a Visual Social Distancing (VSD) issue in [23]. The approach implemented the detection of skeleton centered on interpersonal distance measurement. It also discussed the impact of the social context on the social distance of people and posed privacy concerns. Discussions are promising, but they are again not yielding good results for tracking social distances and opening up the issue.

More recently, numerous prototypes were developed to track social distances using machine learning and sensing techniques. In order to prevent the air transport of the COVID-19 virus, the "social distance" approach of about 6 feet was suggested in the proposed method by *Feng et al.* [30]. Winds also have a dynamic, highly dependent, and located effect on droplet transport and wake flow patterns, secondary flow intensities between the two artificial entities, and calm currents. High RH=99.5 percent will increase the condensation effect, and during transport in air, the size of the cough droplets continues to rise until partial pressure at the droplet is equal to the water vapor saturation pressure. RH=40 percent, on the other hand, causes the coughing gout to evaporate and leads to a decrease in the droplet, which may contribute to a long duration of suspension in the air. Using deep learning techniques, a drone was proposed by *Ramadass et al.* [24] for the inspection of social distances and also to check whether or not an individual is wearing a mask. The qualified Yolov3 algorithm is designed into a drone camera with custom data collection. That uses the YOLOv3 to determine whether or not social space is conserved and whether individuals wearing masks are in the crowd. The drone is designed to operate automatically. *Reluga et al.* [12] suggested a conditional test to determine whether people will take advantage of social distances and associated self-protective behaviors. The conditional test is used as a mitigation method to examine the potential utility of social distancing by calculating the actions of balance under a set of cost functions. Once an epidemic has been detected, the accumulated cost of infection under balance procedures was calculated using statistical approaches in the time before mass immunization.

3. REQUIREMENT ANALYSIS

3.1 TensorFlow

It is a library that is developed to be used for research fields as well as in the development and production areas to perform fast numeric computing. It was published under Apache 2.0 open source license and currently is maintained by Google. It was nominally developed for the Python programming language, but there is access to the underlying C++ API to be run on single CPU systems, GPUs as well as mobile devices, and large-scale distributed systems of hundreds of monitors.

3.2 OpenCV

OpenCV is a Python linking as well as a cross-platform library that can be used to develop real-time computer vision applications and to solve computer vision problems. It was initially developed in C++ language with the addition of Python and Java language bindings. It can run on a variety of operating systems such as windows, OSx, Net BSD, Linux, etc.

3.3 Python

Python programming language was developed by Guido van Rossum during 1985-1990 which is a high-level, general-purpose, and interactive, object-oriented programming language designed to be highly readable and is easy to learn. The language and purpose-based (object oriented) methodology were designed to help programmers create a dynamically typed and straight forward logical code for large and small projects.

Python is also defined as an efficient language because of its robust standard library and memory management techniques including dynamic typing and a combination of reference counting and a loop detection garbage collection. It also connects process and variable names during program execution using the dynamic name resolution method.

3.4 Artificial Intelligence

Artificial Intelligence (AI), is the intelligence expressed by computers and is also known as machine intelligence. It is a study of "intelligent agents" a mechanism that perceives its environment and acts to exploit its opportunities to effectively achieve its goals as defined by Landing AI.

This term is sometimes used to describe the systems or machines or computers that imitate the "cognitive functions that humans associate with human brains, such as "learning" and "problem-solving". Contemporary computer abilities commonly known as AI include understanding human voice, autonomously driving vehicles, intelligent content delivery network routing, and military simulations.

3.5 Deep Learning

It is a subfield of machine learning and also a crucial function of artificial intelligence (AI) that has the ability to learn without human supervision and has also evolved hand-in-hand with the expansion of the digital era to mimics the working nature of the human brain in pattern creation and data processing by drawing both unstructured and unlabeled data. It has various purposes that include object detection, language translation, speech recognition, and decision making.

3.6 YOLOv3

YOLO stands for "*You Only Look Once*" which is one of the fastest object detection algorithms that use a variant of Darknet originally trained on Imagenet consisting of 53 layers. It was developed by Joseph Redmon for the task of detection and computer vision. The YOLOv3 is the latest version of this single deep convolutional neural network that can recognize 80 different objects in images and videos. The architecture of this third version is a 106 fully convolutional layer that boasts of residual skip connections and upsampling that makes predictions at three different scales which are precisely given by downsampling the dimension of the input image. The whole YOLO framework works on several steps that include taking an input image and then dividing that particular input image into grids of several desired dimensions. Further, each of those grids is subjected to image classification and localization through which the YOLO predicts the bounding boxes and their corresponding probabilities for a particular class of various objects.

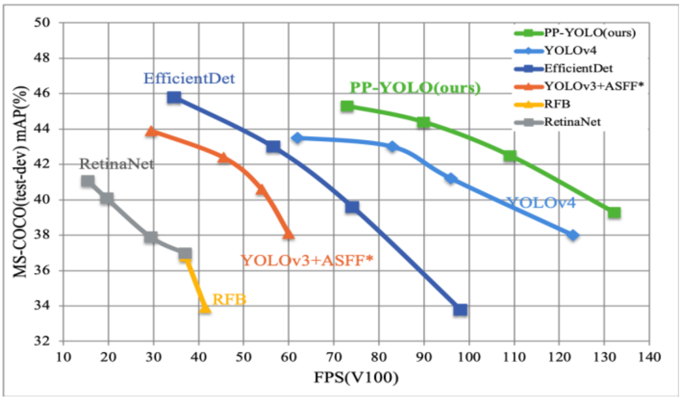

Figure-2:-Comparison of different object detection models

4. PROPOSED METHODOLOGY

The proposed system is a three-level module that focuses on human identification or detection, their tracking, and their inter-distance assessment with computer vision and deep learning with an OpenCV library as a full solution for social distance tracking and zone-based threat analysis for image/video streams. This arrangement can be combined with all forms of CCTV monitoring cameras with the real-time display from VGA to Full-HD.

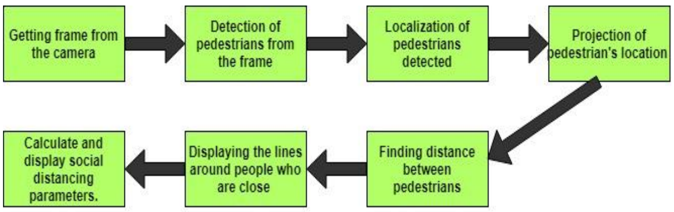

Figure-3: Workflow of the proposed system

Approach

The idea is to develop a real-time system that will be certain about the following criteria:-

1. Detect the object in the frame
2. Track the object's movement
3. Measure the distance
4. Generate the output by distinguishing the objects being in threat or safe

4.1 Camera calibration

Calibration is the process in which the transformation is measured by converting the perspective view into a top-down (bird-eye) view. The input frames that are taken from a single camera i.e. monocular frames and from an arbitrary viewpoint involve the calibration by mapping the four selected perspective viewpoints into the corners of a rectangle in the top-down view. This process is carried out in order to adopt the fact that each person detected lies on the same flat ground plane. This gives us the technique that can be applied to the entire perspective view for deriving the transformation by mapping. During the calibration, the number of pixels representing the 6-feet distance in real-life e.g. scale factor is also estimated in bird's eye view.

4.2 Region of Interest

It is a proposed region extracted from an original image that should be tested by the feature extraction techniques to be put further onto the feature maps. Putting those on a feature maps and applying quantization to them helps maps to get placed on the matrix hence making pooling easier.

ROI pooling is done after we map extracted ROI onto a single feature map for all RPN-generated plans in a single pass.

Later, fixed-size function maps from non-uniform inputs are generated by applying max-pooling on the given inputs and it also results in having an equal number of input and output channels on that particular layer. The pooling layer of the region of interest has two inputs:

A feature map was obtained from the Convolutional Neural Network (CNN) after several convolutions and pooling layers.

Each proposal generated from this consists of total of five values, the first being the index, and the other four are the proposal coordinates. Usually, it represents the upper left and lower right corners of the proposal.

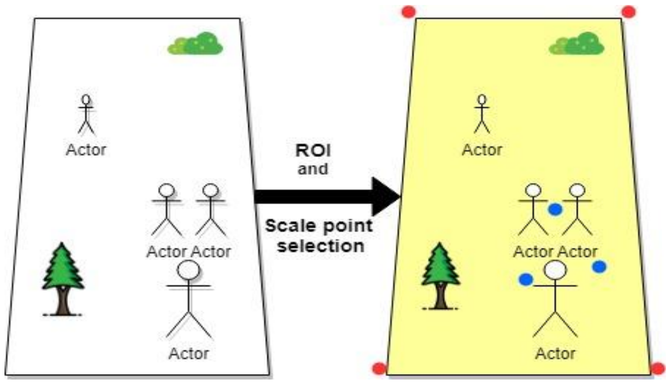

Figure-4: Region of Interest and scale point selection

4.3 People Detection

The next step to be performed is pedestrian detection and outlining them with the bounding boxes. For precise allocation of bounding boxes and to avoid the chances of overfitting, techniques such as minimal post-processing, such as non-max suppression (NMS), and various rule-based heuristics can be applied.

Many object detection algorithms such as YOLO, SSD, Mask RCNN, and RetinaNet are applied and verified by many people in recent years that has benefitted the area of the object or people detection.

Here's an object detection method that uses features learned by a deep, convoluted neural network to detect an object named YOLO stands for You Only Look Once.

YOLO only uses convolutionary layers to make it a fully interconnected network (FCN). It has 75 confounding layers, skip connections, and sampling layers. No type of bundling is used to evaluate the characteristics of the maps and a convolutionary step 2 layer is used. This helps to prevent the sometimes pooling loss of low-level features. The video input sequences are obtained by a CCTV camera and passed into our Deep Network model. The model's output is the individuals with their particular bounding boxes detected in the scene.

4.4 Distance Measurement

(i) Determination of a person's location

For this purpose, the ROI range is compared with each of the ground plane point as it is more suitable than centre point value in order to determine the bounding box of a person as well as segment involved.

(ii) Calculation of centroid of a bounding box

The centroid, *C(x,y)*, of the detected person's bounding box is calculated by midpoint equation.

$$C(x,y)=((x_{min}+x_{max})/2, (y_{min}+y_{max})/2)$$

The centre point of the bounding boxes is calculated from each of the minimum and maximum values for the corresponding width, *Xmin* and *Xmax* and height, *Ymin* and *Ymax* of the bounding boxes.

(iii) Calculation of distance between two bounding boxes

For this, the centroid of the bounding boxes is used to determine the two different locations of the persons. The formula used to measure the distance, $C1(x_{min}, y_{min})$ and $C2(x_{max}, y_{max})$, between two person's bounding box is

$$d(C1,C2)=((x_{min}+x_{max})^2+(y_{min}+y_{max})^2)^{1/2}$$

After getting those centroids, the algorithm will calculate the distance with the minimum pixel range defined in the code.

Implementation

- Calculation of Euclidean distance between two points
- Converting the center coordinates into rectangular coordinates
- Filtering the person in the frame by detecting them and getting a bounding box around the centroid for each human detected.
- Checking which person bounding boxes are close to each other
- Displaying the extent of the threat.

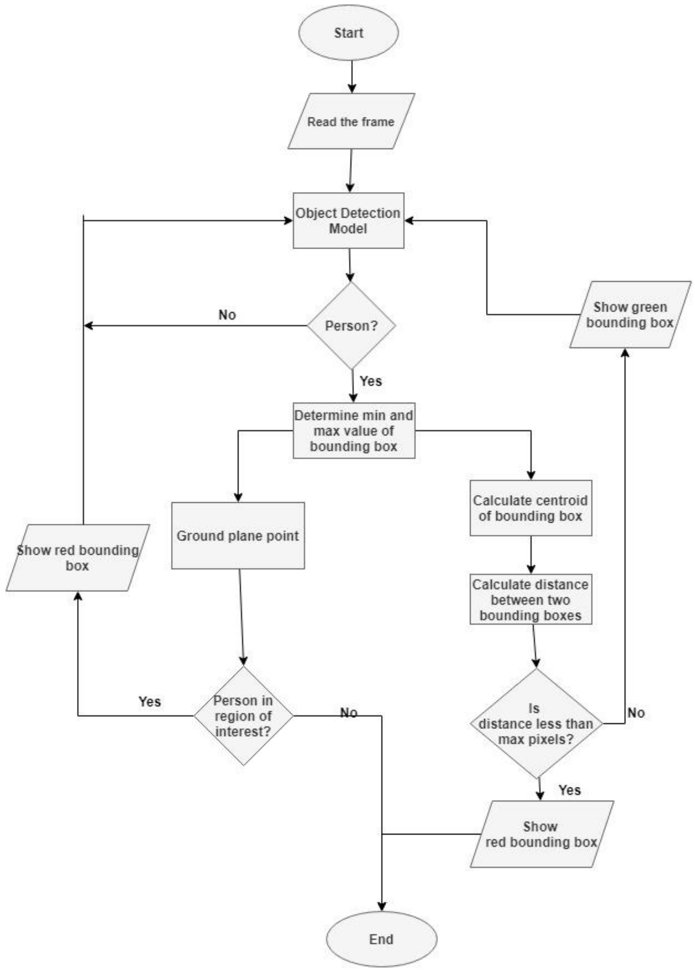

Figure-5: Flowchart of the entire system

Running the program will give us an initial frame on which the region of interest is drawn and distance scales are primed. The whole viewpoint is converted into a bird's eye view by which the horizontal and vertical distance around 180cm is measured in order to get the region of interest and distance scales in the initial frame.

Next, we have a two-step process: Selection of the region of interest and scale points for the initial frame and to identify the person to get the specifics of the bounding box after drawing them around each distinguished person in the frame.

After that, the position of each person in the frame is estimated by taking the pivot point at the lower base of the bounding box referring to it as the person's location, and then applying the transformation to that pivot point to get the (x,y) coordinates of person's location into the bird's eye view. Hence, calculating the pivot point at the base of boxes will help us to project certain points in Bird's eye view.

Finally, the bird's eye view distance will be determined between individual pairs (position) to compare distances by calibrating the scales horizontally and vertically.

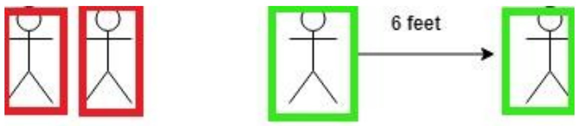

Figure-6: Encircling based on the 6-feet violation

5. CONCLUSION

Emerging developments and the availability of smart technology allow us to create new models that help meet the needs of the emerging world. So here I have created a social distance detector that could potentially apply to public health care. The model proposes an effective, synchronous scheme using object detection and tracking mechanism to automatically monitor social distancing by defining each human with the help of bounding boxes. Identifying gatherings of people who satisfy the proximity property computed with the aid of the Bird's eye view method. The number of violations is confirmed by the measurement of the number of groups created and the infringement index term measured the ratio between the number of people and the number of groups.

This system works very effectively and efficiently in identifying the social distancing between the people and generating the alert that can be handled and monitored. Additionally, it also displays labels as per 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. The screening is possible in mass places such as train stops, bus stations, markets, highways, mall entrances, schools, colleges, etc. By observing 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.

REFERENCES

[1] D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock,O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drostenet al., "The continuing 2019-ncov epidemic threat of novel coronaviruses to global health the latest 2019 novel coronavirus outbreak in wuhan, china," International Journal of Infectious Diseases, vol. 91, pp. 264–266, 2020

[2] C. T. Nguyen, Y. M. Saputra, N. Van Huynh, N.-T. Nguyen,T. V. Khoa, B. M. Tuan, D. N. Nguyen, D. T. Hoang, T. X.Vu, E. Dutkiewicz et al., "Enabling and emerging technologies for social distancing: A comprehensive survey," arXiv preprint arXiv:2005.02816, 2020.

[3] S. Yadav, Deep learning based safe social distancing and face mask detection in public areas for covid-19 safety guidelines adherence

[4] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212–3232, 2019

[5] M. Greenstone and V. Nigam, "Does social distancing matter?" University of Chicago, Becker Friedman Institute for Economics Working Paper, no. 2020-26, 2020

[6] 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)

[7] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in Advances in neural information processing systems, 2016, pp. 379–387

[8] F. Sener, N. Ikizler-Cinbis, Two-person interaction recognition via spatial multiple instance embedding, Journal of Visual Communication and Image Representation 32, 63 (2015)

[9] M. Bielecki, R. Z¨ust, 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)

[10] 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)

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: To-wards real-time object detection with region proposal net-works," in Advances in neural information processing systems, 2015, pp. 91–99

[12] T.C. Reluga, Game theory of social distancing in response to an epidemic, PLoS Compute Biol 6(5), e1000793 (2010)

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pat tern recognition, 2016, pp. 779–788.

[14] LandingAI, Landing AI Creates an AI Tool to Help Customers Monitor Social Distancing in the Workplace, 2020 (accessed May 3, 2020). [Online].

[15] P. Khandelwal, A. Khandelwal, and S. Agarwal, "Using computer vision to enhance safety of workforce in manufacturing in a post covid world," arXiv preprintarXiv:2005.05287, 2020

[16] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263–7271.

[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y.Fu, and A. C. Berg, "Ssd: Single shot multi box detector," in European conference on computer vision. Springer, 2016,pp. 21–37.

[18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll ́ar, "Focal loss for dense object detection," in Proceedings of the IEEE international conference on computer vision, 2017, pp.2980–2988

[19] M. Tan, R. Pang, and Q. V. Le, "Efficient det: Scalable and efficient object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020, pp. 10 781–10790.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587

[21] 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)

[22] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448

[23] M. Cristani, A. Del Bue, V. Murino, F. Setti, and A. Vincia-relli, "The visual social distancing problem," arXiv preprintarXiv:2005.04813, 2020

[24] 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)

[25] ——, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[26] Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4:Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020

[27] K. Chen, C. C. Loy, S. Gong, and T. Xiang, "Feature mining for localised crowd counting." in BMVC, vol. 1, no. 2, 2012,p. 3.

[28] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, pp. 2871–2878.9

[29] Charles Courtmanche, Joseph Garuccio, Anh Le, Joshua Pinkston, Aaron Yelowitz. "Strong Social Distancing Measures In The United States Reduced the COVID-19 Growth Rate". Project HOPE- The People-to-People Health Foundation, Inc., DOI:10.1377/hlthaff.2020.00608 HEALTH AFFAIRS 39, NO. 7 (2020) .

[30] 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)