



Deep learning-based bird eye view social distancing monitoring using surveillance video for curbing the COVID-19 spread

Raghav Magoo¹ · Harpreet Singh¹ · Neeru Jindal¹ · Nishtha Hooda² · Prashant Singh Rana¹

Received: 4 August 2020 / Accepted: 7 June 2021 / Published online: 2 July 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

The escalating transmission intensity of COVID-19 pandemic is straining the healthcare systems worldwide. Due to the unavailability of effective pharmaceutical treatment and vaccines, monitoring social distancing is the only viable tool to strive against asymptomatic transmission. Pertaining to the need of monitoring the social distancing at populated areas, a novel bird eye view computer vision-based framework implementing deep learning and utilizing surveillance video is proposed. This proposed method employs YOLO v3 object detection model and uses key point regressor to detect the key feature points. Additionally, as the massive crowd is detected, the bounding boxes on objects are received, and red boxes are also visible if social distancing is violated. When empirically tested over real-time data, proposed method is established to be efficacious than the existing approaches in terms of inference time and frame rate.

Keywords Social distancing · Real-time · COVID-19 · Bounding boxes · Deep learning

1 Introduction

The escalating active cases of COVID-19 outbreak, highly burdened medical care units, and substantial mortality are straining the healthcare systems worldwide. The major issue is the asymptomatic transmission of the virus in which people with no symptoms are spreading the virus unaware of the fact that they were infected. The potential spread of COVID-19 virus in 30 days without social distancing is presented in the Fig. 1.

World health care organization is continuously advising social distancing, isolation of cases, and intensive testing as the only effective tools to enable to control of virus spread in the highly populated countries. Researchers are developing models to help WHO in this direction [1–3]. Prompt technical innovative video surveillance monitoring framework using artificial intelligence methods for curbing social distancing can drastically control the spread of virus

in highly dense areas. Recently, the emergence of deep learning in the domain of video- surveillance has provided many promising solutions. The rationale is to utilize computer vision and deep learning model for implementing the video surveillance monitoring framework which acts as a viable tool to strive against the escalating transmission intensity of COVID-19 pandemic. The goal of proposing the innovative surveillance monitoring framework is to minimize physical contact among probably infected people and healthy people.

According to the norms of WHO [4–6], people must manage a minimum of six feet space between each individual to maintain social distancing. According to the current study, social distancing is a significant containment measure and important to avoid SARSCoV-2, since people with no or slight symptoms might fortuitously transfer corona epidemic and affect others [7–9]. Epidemiology is a research of various motives for the transmission of contagious diseases. To examine the phenomena of epidemiology, analytical models are the utmost desired choice.

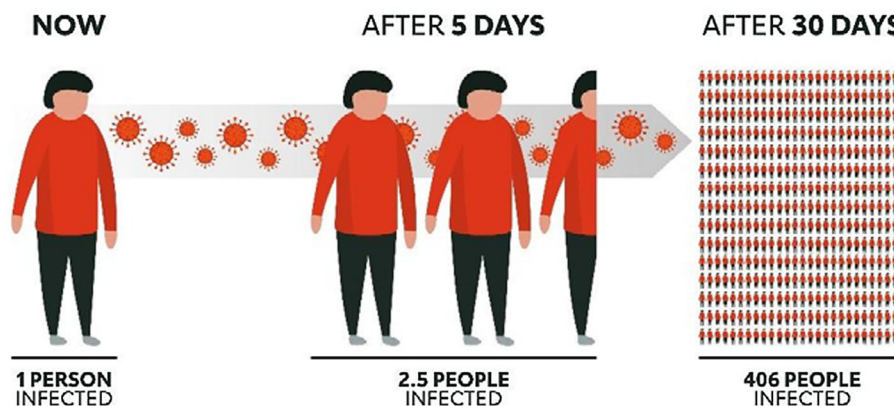
In April 2020 [10, 11], it was declared that the foundation of Artificial intelligence (AI) device can observe social distancing at workstations. Landing AI Company appealed that the forthcoming devices can identify whether the folks are sustaining secure physical distance by

✉ Harpreet Singh
akalharpreet@gmail.com

¹ Thapar Institute of Engineering and Technology, Patiala, Punjab, India

² School of Computing, Indian Institute of Information Technology, Una, Himachal Pradesh, India

Fig. 1 Potential spread of virus without social distancing



scrutinizing live videos from the camera. Moreover, this device can effortlessly combine with existing surveillance cameras present at various workstations to sustain the distance between the workers. The proposed work aims to compare and examine the work of person recognition and tracking procedures for observing social distancing. The major contributions of the proposed work are observed as:

- A novel idea of integrating bird eye view with deep learning algorithm is implemented for monitoring the social distancing at populated areas.
- Bird eye view helps our idea to improve the scalability of a system to calculate the social distancing as the camera will not be needed to be set up in a specific manner i.e., the height of the camera need not be fixed nor the angle of inclination.
- A person-specific social-distance evaluation metric is employed utilizing surveillance video to directly calculate the distance between people without any camera parameters involvement.
- The simulation results demonstrate that the proposed work achieves superior performance than other existing emerging methods like Faster RCNN, SSD, and YOLO V3.

The remaining paper is structured as follows: the work related to this field is given in Sect. 2. Further, Sect. 3 provides detail of the proposed bird eye view-based method to observe social distancing. Later, Sect. 4 gives a conclusion and the future scope.

2 Related work

The corona-virus epidemic is the chief instantaneous challenge to public health at present. Throughout the research, approximately 3 million individuals have recognized positive tests for this severe disease and about 150,000 have passed away. The most reliable procedure to stop the transmission of this contagious disease is social

distancing. Consequently, it was chosen as an unparalleled measure in January 2020, when COVID-19 developed in Wuhan, China [12]. Initially, in the first week of February, the epidemic in China achieved a peak of approx. 2,000 new cases each day. Further, an indication of relief was noticed with no new cases by sustaining distancing [13]. It is obvious that initially social distancing methods were achieved in China, and further accepted globally to handle COVID-19.

The consequences of social distancing procedures on the transmission of COVID-19 endemic were studied by Prem et al. [14]. The authors employed fake area-specific interaction outlines to pretend the current path of the epidemic via susceptible-exposed-infected-removed (SEIR) models. Adolph et al. [15] emphasized the USA's situation, where owing to the deficiency of common agreement between all politicians it was not able to embrace primarily, which caused damage to the public fitness. Further, Kylie et al. [16] studied the severity of social distancing, and it specified that few actions can be allowed through evading an enormous epidemic. As the corona-virus disease commenced, several nations have taken the benefit of technology to hold the epidemic [17–19].

Several countries have utilized GPS for recording the activities of doubted or diseased people, to observe any probability of their contact with fit people. Arogya Setup App is used by the government of India, which utilizes Bluetooth and GPS for localizing the existence of COVID-19 patients in the neighborhood [20]. In contrast, drones and other surveillance cameras have been used by law enforcement sectors for detecting a massive crowd of public, and actions are taken to scatter the mass [21, 22]. Researchers are employing machine learning algorithms for vision-based monitoring system for monitoring crowded areas [23, 24] but very less work has been done in implementation of efficient framework using deep learning-based algorithms.

This physical involvement in such life-threatening circumstances may help to smooth the curve, however, it

carries an exclusive fear to the community and is thought-provoking to workers. All discovered literature undoubtedly creates an image that utilization of human exposure can simply extend to numerous applications to provide the condition that rises currently for example to check suggested standards for sanitization, work practices, social distancing, etc.

3 Experimental investigation

This section presents the data used in the research work along with explaining the experimental setup.

3.1 Dataset

The data are collected from Oxford Town Center data [25] which contains the CCTV video of pedestrians in a crowded market area at Oxford, England. This data are available purely for research purpose in the domain of activity recognition. The video is generated using surveillance camera which includes 2200 people.

3.2 Experimental setup

The training is performed using Python programming. Tensor Flow-based version of YOLOv3 for object detection.

3.3 Computer vision deep learning models

Recently, the emergence of deep learning in the domain of computer vision has provided many promising solutions. Numerous important tasks of computer vision like human activity recognition, object detection, human pose detection, etc., can be improvised using deep learning algorithms. Object detection is the domain of recognizing semantic objects in digital image and videos. The major research contributions of detecting the objects leveraging deep learning algorithm are the variants of convolution neural network (CNN). It usually creates a large set of windows known as bounded boxes that are in the sequel and classified by convolution neural network features.

3.4 Camera resolution

The system can be integrated and applied on all types of CCTV surveillance cameras with any resolution from VGA to Full-HD, with real-time performance. The proposed system used video data with a resolution of 1920×1080 sampled at 25 FPS. The video was down sampled to a standardized resolution of 1280×720 before it was fed to the object detection models.

The models implemented for object detection are:

1. *YOLO V3*: It is fully convolutional network. This object detection model forms a deep neural network level from residual network structure [26]. The output is generated eventually by applying a 1×1 kernel on a feature map. It predicts the objectness score of each bounding box using logistic regression as shown in the Fig. 2.
2. *Faster RCNN*: This model focuses on calculating regional of interest (ROI) region for finding convolution. YOLO v3 performs better than Faster RCNN as it performs fully end to end training along with employing classification and detection at the same time [27].
3. *Single Shot Multibox Detector (SSD)*: This model was initially developed for object detection which employed deep neural network for solving the problems of computer vision and object detection. It is different in the terms that it does not resample features for bounding box hypothesis [28].

4 Proposed framework

The proposed framework is explained well in detail in this section:

The abstract view of the proposed framework is presented in Fig. 3. It shows the three major implemented modules in the proposed work and explained in detail.

1. *Calibration*: This the first step, we will have to calibrate the camera such that all the people walking on the road can be assumed to be on the same plain road, such that the distance calculated between each person will be w.r.t the bird eye view. For the calibration, we need to find the bird eye view of the perspective that is seen by the camera. That bird eye view will be converted using four-point perspective. Finally, we also find the scale factor to convert the Euclidean distance into real life distance.
2. *Detection*: The second step contains the detection of people walking on the road. For this, an efficient computer vision YOLOv3 object detection model is implemented.
3. *Distance calculation*: Once the persons are detected, bounding box of the persons is received. After the detection, Euclidean distance is calculated. The calculated distance is multiplied with the scaling factor to get the distance in real life metrics such as feet. If d is the Euclidean distance which is defined as 6 feet for real-time social distancing is defined as:

$$D \text{ (Euclidean distance)} = 6 \text{ feet (real life metric)}$$

$$1 \text{ (Euclidean distance)} = 6/D \text{ feet}$$

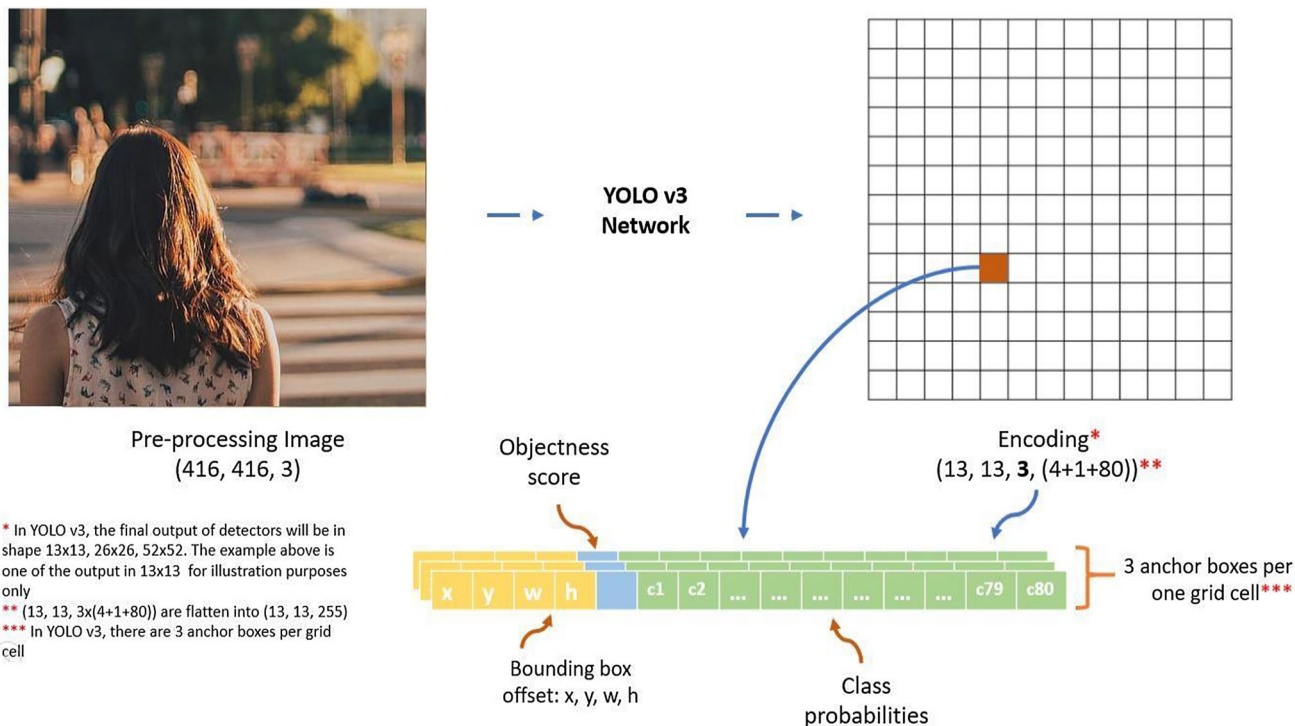


Fig. 2 Architecture of YOLO v3

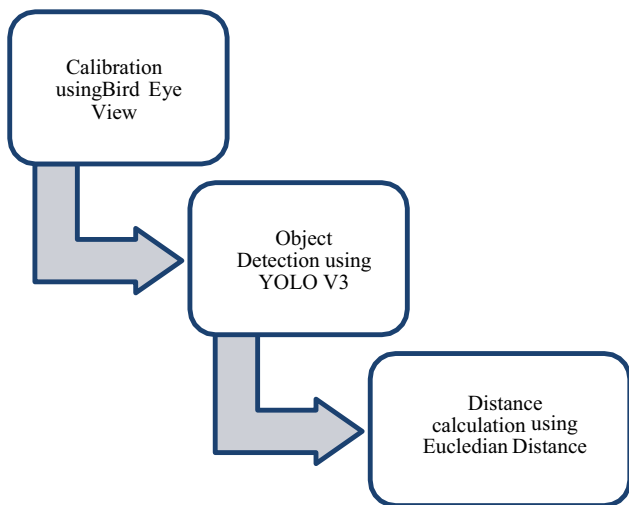


Fig. 3 Abstract view of Proposed Framework

Then, X is used for curbing social distancing as:

$$X \text{ (Euclidean distance)} = (6/D) * X \text{ feet}$$

If $[(6/D)*X]$ is less than 6, this implies the social distancing is not being maintained, and red signal will be given.

The detailed view of the proposed framework is presented in this Section.

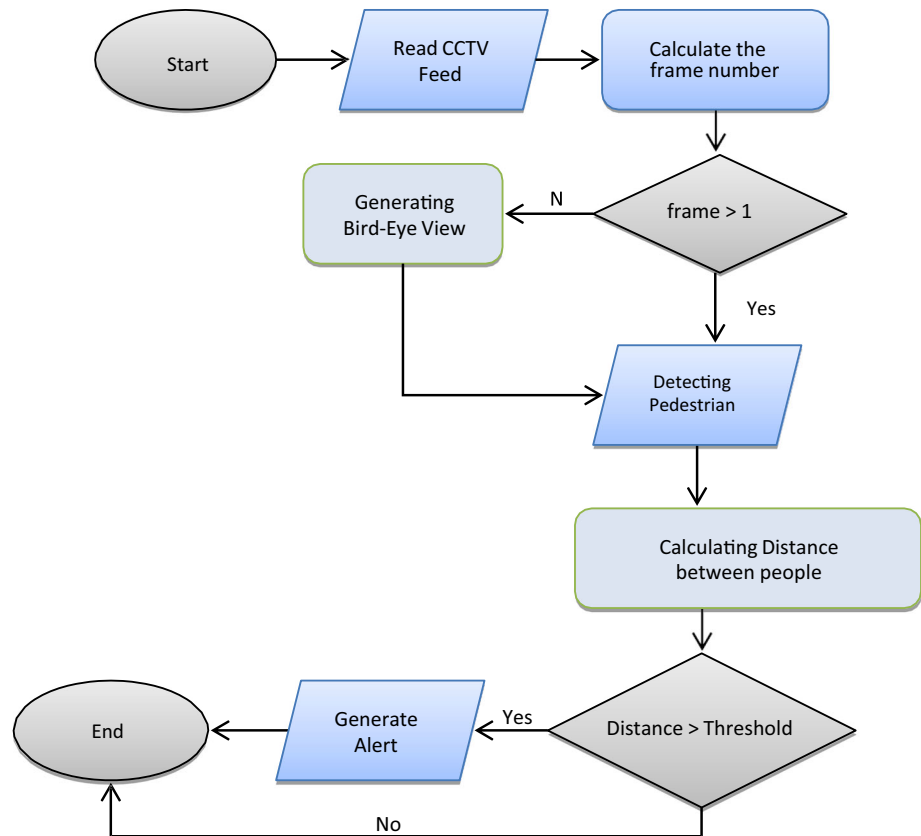
Figure 4 presents the proposed workflow diagram. The system starts by capturing the video feed from the CCTV camera and calculate the frame number if the frame

number is first then that means that the system is being calibrated and the bird eye view of the frame will be calculated and then pedestrians will be detected but if the frame number is not one then the system directly starts detecting pedestrians once the pedestrians are detected, and bounding boxes are obtained the next step is to find the distance between people, for that the mid-point of the bottom of the bounding boxes are taken, and distance between those points are taken in the bird eye view, if the respective distance is less than six feet than alert will be generated else not.

The novelty in our approach is that we have calculated social-distance after finding the bird eye view of the frame. This bird eye computer vision is useful in the ways that it brings the people walking on the road on the same plane so that the distance that we calculate between them does not depend on their relative position to the camera.

Bird eye view helps our idea to improve the scalability of a system to calculate the social distancing as the camera will not be needed to be set up in a specific manner i.e., the height of the camera need not be fixed nor the angle of inclination. Our first step i.e., the calibration step is used to make it easy for the people to calibrate the system they just need to click on 4 points the system is calibrated. If this system is turned to a product it can be very easily adopted.

Fig. 4 Proposed workflow



5 Results and discussion

This section presents the discussion of results and comparison of the proposed framework with state-of-the-art methods. An efficient computer vision YOLOv3 is implemented for object detection model. Simulation results are presented in Fig. 5.

The performance of the proposed version of YOLO v3 using bird eye view is using three important evaluation parameters i.e., framerate, mAP, and inference time are explained below:

1. *Framerate*: This evaluation metric is used in the computer vision for object detection to record frames per seconds processing speed (FPS). It is another way to test the processing speed of model. The results of the proposed YOLO v3 are compared with two state-of-the-art models i.e., Faster RCNN and SSD using Framerate are presented in Fig. 6. It can be observed that SSD is performing very well, and the proposed YOLO v3 is not leading much in the speed. For testing further, mean Average Precision is compared.
2. *Mean Average Accuracy (mAP)*: It is a very popular parameter to evaluate the accuracy of object detection models. The rationale is to find the average precision value for recall value over 0 to 1. Precision helps in

testing how accurate the model is performing the predictions and recall focuses on how well the trained model is predicting the positive samples.

3. *Inference time*: This evaluation metric helps measuring the predictive performance of an already training model. The inference time for the trained model should be as low as possible.

The results of the proposed YOLO v3 are compared with two state-of-the-art models i.e., Faster RCNN and SSD using Framerate and mAP are depicted in Fig. 6.

Observing Fig. 6, both SSD and YOLOv3 can be used to detect pedestrians on the same Oxford Town Center Dataset and the model which performs better on the dataset. In order to avoid confusion, Inference time is also compared. Figure 7 shows the results of the proposed YOLO v3 with two Faster RCNN and SSD models using Inference time. The inference should be as low as possible and as we can observe the inference time of improved YOLOv3 is the lea stand SSD is not that far behind but faster RCNN is way too slow. In order to check the comprehensive performance of improved version of YOLO V3 with bird eye vision, it can be observed that the proposed model is far better than the SSD and Faster RCNN and hence, proposed method is established to be efficacious than the existing state-of-the-art models.



Fig. 5 Simulation results on a dataset using YOLO v3

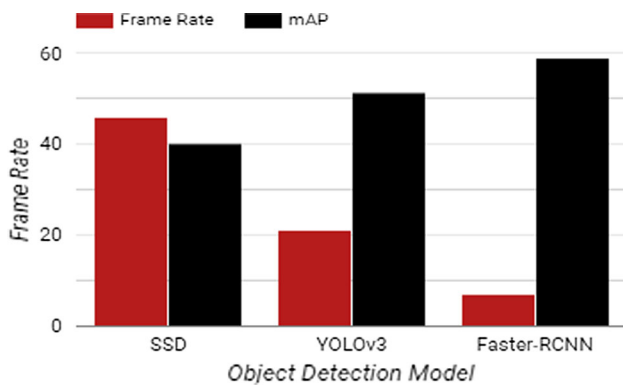


Fig. 6 Frame rate comparison of proposed YOLO v3 with faster RCNN and SSD model

Figure 8a and b presents the performance of model in terms of mAP on standard benchmark dataset with normal and dim light conditions. Light conditions have been checked on six sets of different populations, and experimental results show that the proposed model exhibits good performance with balanced mean average precision (mAP) score. Frame rate is also varied w.r.t. populations, but no

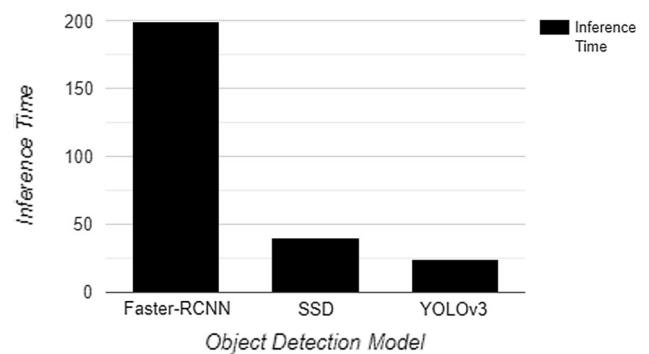


Fig. 7 Inference time comparison of proposed YOLO v3 with faster RCNN and SSD

major change has been observed with normal and dim light conditions. The inference time gives minimal change with normal and dim light conditions, and hence proves robustness of the proposed model.

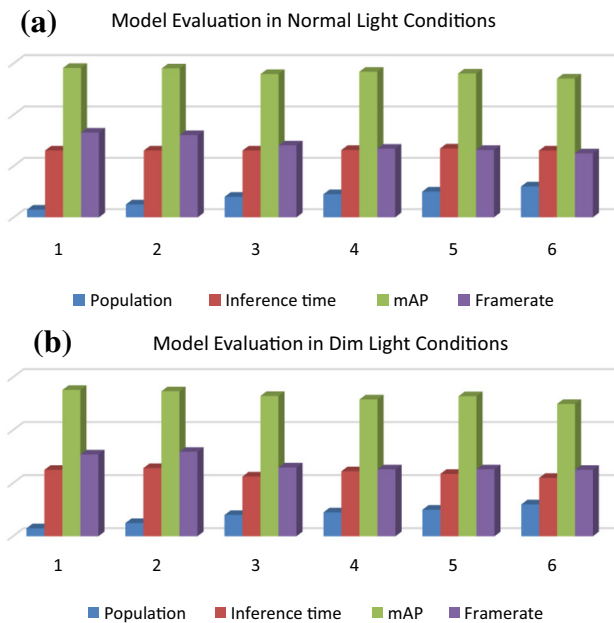


Fig. 8 **a** Model Evaluation in Normal Light Conditions at Home Environment with 6 sets of different populations **b** Model Evaluation in Dim Light Conditions at Home Environment with 6 sets of different populations

6 Conclusion

COVID-19 is not only limited to cause the burden on sufferers and healthcare services, it also advances the world to a significant economic loss. This paper presents a fast and efficient framework for monitoring social distancing through object recognition and deep learning models. The proposed framework first performs calibration using bird eye view. In the second step, YOLOv3 convolutional neural network-based object detection model is trained. The third step employs bounding boxes to detect individuals violating social distancing norms. The comprehensive performance of the proposed framework is validated using real-time data of surveillance video with different evaluation metrics like mAP, FPS score, and inference time and is established to be outperformed than the existing object detection models like SSD and Faster RCNN. As the proposed approach is extremely sensitive to the spatial position of the camera, it can also be modified to adjust better with an equivalent field of view.

For future work, an entire system will be created to curb the spread of coronavirus in the containment zones to keep an eye on the massive crowd by integrating a crowd density detection module for monitoring approximate people in the public and checking whether an individual is wearing a mask or not.

Funding This research did not receive any grant from funding agencies.

Declarations

Conflict of interest Authors have no conflict of interest.

References

- Rahimi I, Chen F, Gandomi AH (2021) A review on COVID-19 forecasting models. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-020-05626-8>
- Choi YJ, Park MJ, Park SJ et al (2021) Types of COVID-19 clusters and their relationship with social distancing in Seoul Metropolitan area in South Korea. *Int J Infect Dis*. <https://doi.org/10.1016/j.ijid.2021.02.058>
- Li S, Lin Y, Zhu T et al (2021) Development and external evaluation of prediction models for mortality of COVID-19 patients using machine learning method. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-020-05592-1>
- World Health Organization (2020) WHO corona-viruses (COVID-19). <https://www.who.int/emergencies/diseases/novel-corona-virus-2019>. Available at. Accessed 24 April 2020
- World Health Organization (2020) WHO director-general's opening remarks at the media briefing on covid-19—11 March 2020. Available at: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>. Accessed 16 March 2020
- Hensley L (2020) Social distancing is out, physical distancing is in. Available: <https://globalnews.ca/news/6717166/what-is-physical-distancing/> Accessed 3 May 2020
- European Centre for Disease Prevention Control (2020) Considerations relating to social distancing measures in response to COVID-19 second update. Available at: <https://www.ecdc.europa.eu/en/publications-ta/considerations>. Accessed 17 May 2020
- Fong MW, Gao H, Wong JY, Xiao J, Shiu EY, Ryu S, Cowling BJ (2020) Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures. *Emerg Infect Dis*. <https://doi.org/10.3201/eid2605.190995>
- Ahmed F, Zviedrite N, Uzicanin A (2018) Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. *BMC Public Health*. <https://doi.org/10.1186/s12889-018-5446-1>
- Landing AI (2020) Landing AI Named an April 2020 Cool Vendor in the Gartner Cool Vendors in AI Core Technologies. Available at: <https://landing.ai/landing-ai-named-an-april-2020-cool-vendor-in-the-gartner-cool-vendors-in-ai-core-technologies/> Accessed 6 May 2020
- Ng AY Curriculum Vitae Available at: <https://ai.stanford.edu/~ang/curriculum-vitae.pdf>. Accessed 10 July 2020
- BBC News (2020) China coronavirus: Lockdown measures rise across Hubei province. Available at: <https://www.bbc.co.uk/news/world-asia-china51217455>. Accessed March 31, 2020
- National Health Commission of the People's Republic of China (2020) Daily briefing on novel coronavirus cases in China. Available at: <http://en.nhc.gov.cn/2020-03/20/c78006>. Accessed: 30 March 2020
- Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a

- modelling study. *The Lancet Public Health*. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6)
15. Adolph C, Amano K, Bang-Jensen B, Fullman N, Wilkerson J (2020) Pandemic politics: Timing state-level social distancing responses to COVID-19. *J Health Polit Policy Law*. <https://doi.org/10.1215/03616878-8802162>
 16. Ainslie KE, Walters CE, Fu H et al (2020) Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment. *Wellcome Open Res*. <https://doi.org/10.12688/wellcomeopenres.15843.2>
 17. Sonbhadra SK, Agarwal S, Nagabhushan P (2020) Target specific mining of COVID-19 scholarly articles using one-class approach. *Chaos Soliton Fract*. <https://doi.org/10.1016/j.chaos.2020.110155>
 18. Punn NS, Agarwal S (2020) Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. *Appl Intell*. <https://doi.org/10.1007/s10489-020-01900-3>
 19. Punn NS, Sonbhadra SK, Agarwal S (2020) COVID-19 epidemic analysis using machine learning and deep learning algorithms. *MedRxiv*. <https://doi.org/10.1101/2020.04.08.20057679>
 20. O. website of Indian Government (2020) Distribution of the novel coronavirus-infected pneumoni Aarogya Setu Mobile App. Available at: <https://www.mygov.in/aarogya-setu-app/> Accessed 26 May 2020
 21. Robakowska M, Tyranska-Fobke A, Nowak J, Slezak D, Zuratynski P, Robakowski P, Nadolny K, Ładny JR (2017) The use of drones during mass events. *Disaster Emerg Med J*. <https://doi.org/10.5603/DEMJ.2017.0028>
 22. Harvey A, LaPlace J (2019) Megapixels: Origins, ethics, and privacy implications of publicly available face recognition image datasets. Available at: <https://megapixels.cc/> Accessed 21 January 2020
 23. Zeroual A, Harrou F, Dairi A, Sun Y (2020) Deep learning methods for forecasting COVID-19 time-series data: a comparative study. *Chaos Soliton Fract*. <https://doi.org/10.1016/j.chaos.2020.110121>
 24. Ahmed I, Ahmad M, Rodrigues JJ, Jeon G, Din S (2020) A deep learning-based social distance monitoring framework for COVID-19. *Sustain Cities Soc*. <https://doi.org/10.1016/j.scs.2020.102571>
 25. Harvey A, LaPlace J (2020) Megapixels: Oxford Town Centre. Available at: https://megapixels.cc/oxford_town_centre/ Accessed 21 July 2020.
 26. Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. *arXiv preprint*. Available at: <https://arxiv.org/pdf/1804.02767.pdf>
 27. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. <https://doi.org/10.1109/TPAMI.2016.2577031>
 28. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016*. *ECCV 2016. Lecture Notes in Computer Science*, vol. 9905, Springer, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.