



# Automatic Face Mask Detection and Violation of Social Distancing Application

Jordan Pang Min Han, Chuan-Chin Teo<sup>(✉)</sup>, and Han-Foon Neo

Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia  
{ccteo, hfneo}@mmu.edu.my

**Abstract.** Ever since worldwide COVID-19 epidemic, social distancing has been prioritised and advised to practice everywhere on a daily basis. It has been and will continue to be one of the most efficient and effective measures to fight the pandemic and saving lives. In Malaysia, social distancing is a proactive action to minimize the spread of COVID-19 among human beings. No matter we are staying indoor or outdoor, a tolerable distance among each other will save ourselves and others. However, there are a lot of constraints to govern the violation of social distancing manually. In this research, an automatic detection of the violation of social distancing is proposed to create a healthier environment. The proposed research could be conducted in public places or confined places. Particularly, a security camera combined with another camera is used to monitor the social distancing in a confined area. Social distancing violation will be detected using Euclidean distance analysis.

**Keywords:** Social distancing violation · mask and head count detection · Euclidean analysis

## 1 Introduction

Pandemics refers to widespread outbreaks of infectious disease affecting a large area and substantial number of people and has a great impact to society and politic worldwide [1]. In 2003, Severe Acute Respiratory Syndromes (SARS) outbreak has caused the death of 800 people in 17 countries and costs the world's economic loss of USD40 billion. Since then, there has been other epidemic outbreaks such as bird flu, swine flu, Middle East Respiratory Syndrome, Ebola and the most recent COVID-19.

According to World Health Organization (WHO), the spread of COVID-19 can be reduced if certain control measures are observed such as keeping a physical social distance of at least one meter with another person and avoiding overcrowded places or crowded groups [2]. Obviously, both the government and everyone is playing a big role in reducing the cases of COVID-19. However, there is a lack of artificial intelligent components available that can contribute to ensure the physical social distancing is well practiced by all the stakeholders especially in a confined area.

Social distancing should be practiced everywhere on a daily basis during COVID pandemic. Practicing social distancing is one of the most effective ways to fight the

pandemic and saving lives. It is unclear how social separation should be performed or controlled, particularly when it comes to limiting everyone's social activities [3]. On the other hand, [4] has reported that a proper physical social distancing policies and residents' compliance can minimise the spread of virus.

With the introduction of computer vision and deep learning in artificial intelligence (AI), there are many applications which helps to detect automatically whether a person or individual is wearing a mask and if one is maintaining social distancing with other individuals in the public. In this research, a prototype of social distancing violation detection system is proposed. The system conducts an automatic head count and mask detection in the targeted area and an alert is notified if there is any violation of social distancing.

## 2 Related Works

### 2.1 Object Detection

#### 2.1.1 Region Proposal Network (RPN) and Fast R-Convolutional Neural Network (R-CNN)

Reference [14] has proposed Region Proposal Network (RPN) for efficient and accurate region generation and enhance it as fast R-CNN later on. By merging comparable pixels and textures into multiple rectangular boxes, the faster R-CNN seeks to locate areas that may represent an object. The basic R-CNN utilised search selection to generate for example 2000 suggested regions (rectangular boxes). Fast R-CNN, on the other hand, is a substantial improvement since it sends the original picture to a pre-trained CNN model once, rather than applying CNN 2000 times to the specified regions. Faster R-CNN is more robust than Fast R-CNN in terms of performance.

#### 2.1.2 Face Mask Detection

Reference [5] has proposed face mask detection using "You Only Look Once" (YOLO) and faster R-CNN models. The convolutional neural network (CNN) in YOLO is widely used in image and video processing applications. The main reason is it has mature feature extraction technique and it consumes lower computational cost. To extract higher-level features, the source pictures or feature maps is convolved with convolution kernels in CNN.

The Residual Network (ResNet) was designed to train deeper neural networks by learning the identity mapping from the preceding layer. It allows the layers to suit a residual mapping than expecting that each of the stacked layers directly match to a desired mapping. The networks are built by stacking the residual blocks on top of one other.

Mobile Network (MobileNet) was developed because object detectors are typically installed on gadgets that are mobile or integrated with low computing resources. MobileNet's computational cost is considerably cheaper than networks that utilise ordinary convolutions because it uses channel-wise convolutions to change channel numbers and depth-wise convolutions to extract features. In CNN, a kernel or a filter is passed over the picture and altered according to the filter's values.

In YOLO, a single Convolutional Neural Networks (CNNs) predicts several bounding boxes as well as the class probabilities for those boxes all at a fixed similar time. YOLOv3's architecture includes 53 ImageNet's trained convolutional layers. For recognition, another 53 layers are layered on top of it, giving YOLOv3 a 106 layer fully convolutional underlying architecture. Detection in YOLOv3 is accomplished by using detection kernels on feature maps of three different sizes at three distinct locations throughout the network.

The detection kernel has the size of  $1 \times 1 \times (B \times (5 + C))$ .  $B$  refers to the number of bounding boxes a feature map cell can predict; The coefficient "5" refers to the four bounding box characteristics plus one object confidence; and  $C$  refers to the number of classes. The kernel size in YOLOv3 is  $1 \times 1 \times 255$  since  $B = 3$  and  $C = 80$ . The 82nd layer is the first to detect something. The picture is down-sampled by the network for the first 81 levels, with the 81st layer having a stride of 32.

## 2.2 Crowd Management

People's flow to a building or area can be controlled using an AI system. Reference [6] use Convolutional Neural Network and a Deep Learning algorithm for capacity control to help the stake holders to comply with the COVID-19 standard operating procedure (SOP) by counting the number of individuals entering and exiting a location, providing physical separation, measuring body temperature, and alerting visitors and management to infractions. Similarly, [7] has proposed an affordable IoT-enabled COVID-19 SOP compliance application which can perform head counts. For example, it could detect the number of people going into or leaving a premises, monitoring social distancing and measuring body temperature. In addition, it is able to give warnings to the attendees and managers of violations. Meanwhile, [8] has developed a similar surveillance application which upholds the importance of privacy by using adaptive social distance estimation in addition to crowd monitoring. In [9], big data analysis is conducted using graph for better pedestrian group mining. Lastly, [10] used a macroscopic method to analysis crowd density and a microscopic method to compute the distance between them. The mentioned research works are the corner stone for our proposed project.

## 2.3 Case Study

### 2.3.1 Aarogya Setu

This application is a smartphone application developed as part of the India's COVID-19 contact tracing, mapping of syndromes and self-assessment digital service by the National Informatics Centre under the Ministry of Electronics and Information Technology. There are several features of this application, namely it supports 11 universal languages, bluetooth and global positioning system capabilities. If someone came within 6-feet of a positive COVID-19 instance, it offers information. IN addition, the software suggests several measurements, self-assessment tools, and do's and don'ts in the event of a global pandemic on how to maintain social distancing. Finally, this software might also serve as an e-pass, allowing users to go from one area to another with ease [11].

### 2.3.2 Safe Space

SafeSpace is an application designed to keep people safe in indoor spaces by detecting whether the person is wearing a mask, based on AI. This software is based on machine learning technique. The application has a high face mask detection accuracy of 99.33 percent. It consists of personalized screen messages and audio alerts [12].

### 2.3.3 Capacity Tracker

Most business facilities have lower capacity limits to control the crowd as an effort in preventing the spread of COVID-19. Capacity Tracker was created to facilitate businesses, places of worship, restaurants, schools, and other establishments to keep track of indoor gatherings to ensure everyone's safety. The application is able to update the count, and individuals can come and go as they like. When a facility's capacity is reached, it will be notified. It also keeps track of all accesses and exits by syncing the count across many devices [13].

## 3 Methodology

### 3.1 Tools

To implement the prototype, python code and python compatible software are used to program such as Pycharm and Anaconda (Miniconda)'s jupyter notebook. Anaconda terminal is served as a location to download necessary packages such as NumPy, Tensorflow, Keras, and OpenCV. Tensorflow-GPU package is used for the purpose of model output. In order to use this package, Nvidia's CUDA toolkit and Cudnn need to be installed. Furthermore, a built-in camera will be used for demonstration purpose. There can be multiple cameras for output but configuration has to be done. Yolov3 weights and Darknet as well as pretrained models are used in this research to train the mask dataset.

### 3.2 System Features

The prototype design in this research contains several features such as the ability to trace the number of people in the targeted confined area, the number of people who violate the SOP prescribed as total serious violation and total abnormal violation, set red flag on users who did not adhere to maintaining social distancing, and users who did not wear a face mask. This prototype is able to calculate social distancing violation via the Euclidean distance formula to calculate the centroids.

### 3.3 Face Mask Detection

In this research, a social distancing and face mask detection system prototype is created. First, YOLOv3 weights are pretrained where the pretrained YOLOv3 is used in people detection. YOLOv3 is chosen as the method because its accuracy is higher compared to previous versions. Second, a pretrained single-shot detector (SSD) is used to detect the human faces, while a pretrained MobileNetv2 is used to classify people into groups of

wearing mask and not wearing mask. In this prototype, Euclidean distance is used for distance estimation between people. This is to show whether a person has violated the social distancing rules in a specific premise.

For customisation or configuration of YOLOv3, transfer learning is used. When training a new model, transfer learning is a strategy for reusing the weights in one or more layers from a previously learned network model by retaining the weights, fine adjusting the weights, and completely changing the weights. In this research, the YOLOv3 network model reuse the pre-trained weights of a neural network from an open source neural network framework which is Darknet.

### 3.4 Social Distancing Analysis

To start the video capture of built-in camera is to be set to value 0 and to connect with other out source of camera like cctv and webcam set the value as 1, 2 meaning the number of connected cameras.

Every targets are included with a rectangle and text for detecting the faces with their result that they wear mask or not, if wear mask then both rectangle and text will display Green color or else it will display red colour indicating alert or violation.

Blob are constructed, face detection and initialize list, and those functions set to return. These lists include the region of interest of faces. Inside the loop, weak detections are filtered out and extract bounding boxes to make sure the box coordinates do not the image bound. After extracting region of interest in the face and pre-processing, it will expand the shape of an array.

The video capture load real-time images to detect the face and check the social distancing. To detect and calculate total persons in a frame and those maintain social distancing at a time there are at least two people detections and centroids will be extracted from the results and compute Euclidean distances between all pairs of the centroids. For social distance to be in a safe position, centroid of each person will be used to judge the social distance whether safe or violated.

## 4 Experimental Results

### 4.1 Face Mask Detection

Figure 1 shows the training loss and accuracy after trained for face mask detection, the results show initially train is 78%+— accuracy but it increase continuously until 100% accuracy as epoch increased and thus training loss also decreases. On the other hand, Table 1 shows the accuracy result of different architecture used.

### 4.2 Human Capacity Control

The proposed system has detected the face and check the social distancing. Figure 2 shows a sample of safe social distancing. The are two head count detected and the reasonable Euclidean distance calculation. In contrast, Fig. 3 demonstrates a sample of the violation of social distancing with automatic calculation of the head count as 9 given shorter Euclidean distance.

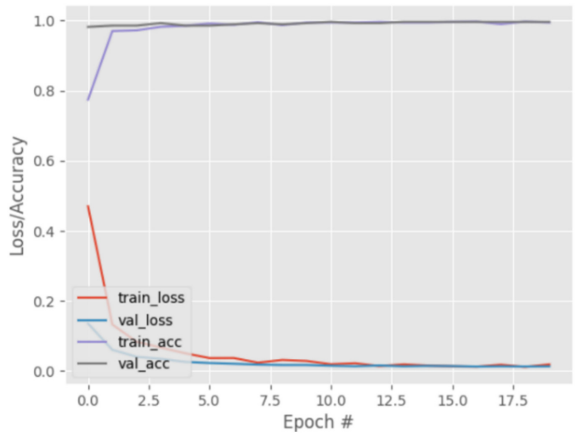


Fig. 1. Training and Loss Analysis

Table 1. Comparison between accuracy of different architectures used to detect mask

Architecture	Accuracy
SSDMNV2	92.64%
Yolov3 with Mobilenetv2	99.99%



Fig. 2. Sample safe social distancing

5 Conclusion

Malaysia had been hit by waves of coronavirus disease (COVID-19) since early 2020 and government-imposed movement control order (MCO) for the citizen to restrict them from moving further than a specific area to help reduce the spread of the coronavirus. Covid-19 cases have increased rapidly every day after MCO is lightened and some people



**Fig. 3.** Sample violation of social distancing's output

tend to break the rules and protocol by not wearing a face mask or did not follow social distancing impose by World Health Organization (WHO). The increase of cases also meant that the frontliners that are working hard in general hospitals or the healthcare field are facing a very big challenge to overcome as the healthcare system is in critical condition with almost all of the intensive care unit (ICU) in the country occupied by COVID-19 patients. This is will cause the healthcare system to collapse and patient with other sickness or disease could not seek immediate treatment from doctors.

The government and its citizens play a crucial part in reducing the cases of COVID-19 by implementing appropriate solutions and obeying the rules respectively. Thus, Information Technology and Artificial Intelligence field will play a part in this challenging period in helping the public and ensure the specific protocol such as wearing a mask and maintain social distance are obeyed by the public. With the introduction of computer vision and deep learning in Artificial Intelligence field will help in detecting whether a person or individual is wearing a mask and detecting if one is maintaining social distance with other individual in the public. Therefore, a prototype of social distancing violation and mask detection system is developed as an alternative in fighting the pandemic.

In this research, a prototype for Automatic Detection of the Violation of Social Distancing is proposed. This project is beneficial to the entire populations to reduce the impact of pandemic COVID19. The proposed system can estimate the physical distance among target user in confined area. It is especially an initiative to reduce the number of positive cases in Malaysia and also help to boost the local economy.

With the introduction of this prototype, government bodies can ensure that business operators and the public follow the Standard Operation Procedure (SOP) the was introduce to keep the Covid-19 pandemic away from Malaysian. Besides, the alert functions built in the prototype are used to ensure that business operators and users in a confined area followed the SOP of wearing a mask in the public area as well as maintaining social distancing. Business operators can use this prototype to make sure consumers in a confined area such as in a shopping complex, a room, etc., maintain a distance from the other consumer to cope with the spread of the virus. The mask detection function allows business operators to detect and keep track of consumers that do not wear a mask and flag

an alert to the employees in charge. The implementation of this prototype makes users who are in that particular confined area feel more secure and safe from the spreading of the virus. This system can be used as long as pandemic is still there till the future, when there are better algorithm developed, this system can be enhance by just retrain the data with the better algorithm for greater and more accurate detection of mask and social distancing considering complex angle too.

**Acknowledgments.** This work is supported by funding of IR Fund, from Multimedia University (MMU/RMC/GRPROP/IR FUND/2020/38612 – MMUI/210035).

**Authors' Contributions.** Jordan Pang Min Han has contributed in application development and original draft preparation. C.C. Teo is the corresponding author and is the supervisor for this research. He has contributed to reviewing and editing the research paper. H.-F. Neo is responsible for providing overall research guidance and proofreading the research paper.

## References

1. N. Madhav, B. Oppenheim and M. Gallivan, Pandemics: Risks, Impacts, and Mitigation, In: Jamison DT, Gelband H, Horton S, et al., editors. Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank, Chapter 17. Retrieved from: <https://www.ncbi.nlm.nih.gov/books/NBK525302/>.
2. World Health Organization (WHO), COVID-19: physical distancing, Retrieved from <https://www.who.int/westernpacific/emergencies/covid-19/information/physical-distancing>.
3. Z. Wu, Social distancing is a social dilemma game played by every individual against his/her population, PLoS ONE, 2021, vol. 16, no. 8, e0255543. <https://doi.org/10.1371/JOURNAL.PONE.0255543>.
4. P.-C.Chung and T.-C. Chan, Impact of physical distancing policy on reducing transmission of SARS-CoV-2 globally: Perspective from government's response and residents' compliance, PLoS ONE, 2021, vol. 16, no. 8, e0255873. <https://doi.org/10.1371/journal.pone.0255873>.
5. S. Singh, U. Ahuja and M. Kumar, Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment, Multimedia Tools Application 80, 2021, pp. 19753–19768.
6. C. R. Rodriguez, D. Luque, C. Rosa, D. Esenarro and B. Pandey, Deep Learning Applied to Capacity Control in Commercial Establishments in Times of COVID-19, 12th International Conference on Computational Intelligence and Communication Networks, 2020, pp. 423–428.
7. A. Bashir, U. Izhar and C. Jones, IoT-Based COVID-19 SOP Compliance and Monitoring System for Businesses and Public Offices, Engineering Proceedings 2020, 2020, vol. 2, no. 1, p. 14.
8. M. Al-Sa'd, S. Kiranyaz, I. Ahmad, C. Sundell, M. Vakkuri and M. Gabbouj, A Social Distance Estimation and Crowd Monitoring System for Surveillance Cameras, Sensors, 2022, vol. 22, no. 2, p. 418. <https://doi.org/10.3390/s22020418>.
9. A. Corbetta, J.A. Meeusen, C. Lee, R. Benzi and F. Toschi, Physics-based modeling and data representation of pairwise interactions among pedestrians, Phys Rev E., 2018, vol. 98, no. 6, pp. 1-18. <https://doi.org/10.1103/PhysRevE.98.062310>.



10. F. Bouhlel, H. Mliki and M. Hammami, Crowd Behavior Analysis based on Convolutional Neural Network: Social Distancing Control COVID-19, In Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2021, vol. 5, pp. 273–280, ISBN 978-989-758-488-6; ISSN 2184-4321.
11. Arogya Setu App: Learn more about how it Works & Benefits, Retrieved from <https://www.godigit.com/health-insurance/diseases/novel-coronavirus/arogya-setu-app>.
12. The first face mask detection app for keeping you safe, Retrieved from <https://safer.work/safe-space/#section-4>.
13. Capacity Tracker—Stay within the Guidelines., Retrieved from <https://capacitytrackerapp.com/>.
14. S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, 2015, Neural Information Processing Systems, vol. 1, pp. 91–99.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

