

A Real-Time Social Distancing and Face Mask Detection System Using Deep Learning



Suet Nam Wai, Sew Sun Tiang, Wei Hong Lim, and Koon Meng Ang

Abstract It has been more than two years since the transmission of COVID-19 virus has affected the public health globally. Due to its natural characteristic, the virus is very likely to undergo mutation over time and consistently changes to a new variant with higher severity and transmission rate. The pandemic is expected to prolong with the increment in number of daily cases which leads to why preventive measures like practising distance apart rule and wearing facemask are still mandatory in the long run. This paper is prepared to develop a social distancing model using deep learning for COVID-19 pandemic. The tracking accuracy of the proposed model is discussed in the paper and compared with other deep learning methods as well. The efficiency of the detection model is observed and evaluated by performing quantitative metrics. The monitoring model is trained by implementing YOLOv4 algorithm and has achieved an accuracy of 93.79% with F1-score of 0.87 in detecting person and facemask. The model is applicable for real-time and video detection to monitor social distance violation as an effort to flatten the curve and slow down the transmission rate in the community.

Keywords Deep learning · Social distancing · Mask detection · YOLOv4

1 Introduction

Even though vaccines for COVID-19 are now available worldwide to fight against the pandemic, the fundamentals of preventive measures are still highly anticipated. Vaccination is just an additional step in reducing the severity effect of the disease and death. The extend of how much it can protect a person from the infection and transmitting the virus to others is still unknown [1]. The term, social distancing, can be described as a public health practice that limits any in-person contact with anyone by staying at home and away from public spaces to reduce the airborne transmission

S. N. Wai · S. S. Tiang (✉) · W. H. Lim · K. M. Ang
Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia
e-mail: tiangss@ucsiuniversity.edu.my

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
M. A. Abdullah et al. (eds.), *Advances in Intelligent Manufacturing and Mechatronics*,
Lecture Notes in Electrical Engineering 988,
https://doi.org/10.1007/978-981-19-8703-8_2

[2, 3]. Ainslie et al. revealed that the number of new cases dropped significantly during the imposition of strict social distancing and movement restrictions towards mainland China and Hong Kong SAR from late January to early February 2020 [4]. Prem et al. had studied the effectiveness of physical distancing in Wuhan whereby it decreased the median number of infections by more than 92% in middle of 2020 and 24% at the end of 2020 [5]. Fong et al. and Kahalé had also verified that social distancing is an effective preventive measure in combatting the pandemic [6, 7]. On the other side, deep learning is a universal learning approach that can perform in almost all application domains in cases where humans do not have to be present in the scene to conduct the specific task. It can be defined as a subset of machine learning that uses neural networks with many layers and is introduced to mimic the function of the human brain in data processing [8], object detection [9, 10], and fault detection [11, 12]. It has been evolving for the past decades with improvised algorithms to produce higher accuracy percentage and generate data concurrent with the present situations. Developing a social distancing monitoring model using deep learning can contribute to slowing down the virus transmission rate that is affecting the public health by identifying social distance violation through person detection.

2 Related Work

A summary of other similar works in using deep learning for object detection to monitor the practice of preventive measures is shown in Table 1.

Uddin et al. [13] used ResNet50 as the CNN architecture to develop an intelligent model that categorized people based on body temperature which resulted in person tracking accuracy at 84%. Saponara et al. [14] applied YOLOv2 to monitor social distance and body temperature through thermal camera using two different datasets and achieved accuracy detection of 95.6% and 94.5%, respectively. Pun et al. [15] utilized YOLOv3 framework with the addition of Deepsort approach that can track the identified people by assigning them with unique IDs. The proposed model had

Table 1 Comparison of quantitative analysis data based on different social distancing models

Paper	Methods	Dataset size	Accuracy (%)	Precision	Recall	F1-Score
[13]	CNN (ResNet50)	11 880	84	0.84	0.82	0.82
[14]	YOLOv2	775	95.6	0.95	0.96	0.95
		800	94.5	0.94	0.95	0.94
[15]	YOLOv3 and deepsort	800	84.6	–	–	–
[16]	YOLOv3 with transfer learning	–	95	0.86	0.83	0.84
[17]	YOLOv4	7 363	97.84	0.85	0.97	0.91
[18]	YOLOv4	3.76M	99.8	0.998	0.976	0.99

84.6% accuracy. Ahmed et al. [16] proposed his model to detect human from overhead perspective by implementing YOLOv3 adopted with transfer learning which in return achieving 95% accuracy. Rahim et al. [17] developed a social distancing monitoring model specifically for low-light environment targeting night-time using YOLOv4 algorithm. Despite the limitation of having the proposed model to focus in the environment temporarily before monitoring, the accuracy result was 97.84%. Razaeei and Azarmi [18] aimed to have a viewpoint-independent human classification algorithm to monitor social distancing that can overcome limitation of light condition and challenging environment without needing to consider the angle and position of the camera. Their proposed model was built on YOLOv4 algorithm and obtained an accuracy of 99.8%.

3 Methodology

3.1 Dataset Preparation

A total of 530 images are collected randomly from various online sources shown in Google Images as well as selectively from raw images published by X. zhangyang's GitHub [19] and Prajnash's GitHub [20]. These images are taken with people from all ages and gender in different situations like walking, standing, sitting, and other possible body positions to maximize the stimulated conditions for detecting person with and without facemask. The dataset consists of both closed-up and distant images with 200 images focusing on single person with mask only, 160 images focussing on single person without mask only, and 170 images mix with a group of people with and without mask. They are pre-processed by resizing and orienting to establish a base size and orientation to be fed into the framework. This helps in improving the quality and consistency of the data for feature extraction as shown in Fig. 1.

3.2 Model Training

The model training process is conducted via Google Colab to utilize their GPU acceleration for extra computational power. The reason behind choosing YOLOv4 algorithm rather than other deep learning methods is that it is the only framework that can run in a conventional GPU that is easily accessible with minimal cost from home. Besides, its performance in speed and accuracy has been proven with astonishing outcomes, and it suits the real-time application for the proposed model [21]. The network size for model training is 416×416 . The hyperparameters, which cannot be inferred by the model, are set such that momentum is configured to 0.949, weight decay is configured to 0.0005, and the learning rate is at 0.001. The classification model is trained to predict three classes namely person, with mask, and without mask.

3.3 Performance Evaluation

Quantitative metrics are the measurements of how robust the model is and act as a form of feedback to determine which aspects of the model can be improved. Since the proposed model focuses on classification performance, the metrics used for performance evaluation are precision, recall, F1-score, mean average precision (mAP), and intersection over union (IoU).

Precision is used to measure the ratio of true positives (TP) to the total positives predicted as expressed in Eq. (1). It is more on how many predictions did the model capture correctly.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Recall, also known as sensitivity, is used to measure the ratio of TP to the actual number of positives as expressed in Eq. (2). It is more on how many predictions did the model miss.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Both precision and recall can be represented in a single score called F1-score. It takes the harmonic mean of those two metrics as expressed in Eq. (3).

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Meanwhile, average precision (AP) is the result of the area under the precision–recall curve and can be calculated using Eq. (4). This is where mean average precision (mAP) comes into the picture to calculate the average of AP for all the classes as shown in Eq. (5).

$$\text{AP} = \frac{1}{11} \sum_{\text{Recall}_i} \text{Precision}(\text{Recall}_i) \quad (4)$$

$$\text{mAP} = \frac{1}{N} \times \sum_{i=1}^N \text{AP}_i \quad (5)$$

3.4 Deployment of Classifier Model

The model deployment is conducted in PyCharm Community Edition 2021.2.1. The overview workflow of the classifier model can be seen in Fig. 2. The model is

begun by reading the input video and converting it into frames. The ability of object detector is then applied to classify three classes based on the confidence value. If the predicted object is a person, the model would proceed with evaluating the inter-distance measurement. If the predicted object is with mask, purple bounding box is generated with “mask” text labelled on top of it. If the predicted object is without mask, red bounding box is generated with “no mask” text labelled on top of it. A mini dashboard is updated at the top left corner of the output video to show the monitoring status according to the number of bounding boxes generated per frame.

The inter-distance calculation is performed by measuring the distance between the centre point of every bounding box of predicted person. The centroid coordinate of the bounding box can be obtained by adding the lowest and highest value of the same axis and divide them by two as expressed in Eq. (6). C_i , which is also equivalent to (X_i, Y_i) , represents the centroid coordinate. X_{\min} and X_{\max} are the lowest and highest x-coordination of the bounding box, respectively. Likewise, Y_{\min} and Y_{\max} are the lowest and highest y-coordination of the bounding box, respectively.

$$C_i = (X_i, Y_i) = \left(\frac{X_{\min} + X_{\max}}{2}, \frac{Y_{\min} + Y_{\max}}{2} \right) \quad (6)$$

After that, Euclidean distance criterion is applied here to translate the distance between the pixels in the input frame to metric distance format. The equation of Euclidean formula is shown in Eq. (7). The distance between two centroid points of the bounding boxes is represented as $D(C_1, C_2)$. X_{\max} and Y_{\max} represent the coordinates from either one of the centroid points that has the largest value. X_{\min} and Y_{\min} represent the coordinates from the other the centroid point that has the smallest value.

$$D(C_1, C_2) = \sqrt{(X_{\max} - X_{\min})^2 + (Y_{\max} - Y_{\min})^2} \quad (7)$$



Fig. 1 Samples of images for dataset preparation



Fig. 2 Flowchart of social distancing monitoring model

Initially, the bounding boxes will not be drawn first when they are detected. Once the inter-distance calculation is computed, the model will decide whether the bounding boxes will be in green or red. The violation distance is denoted as the violation threshold value in this case. If $D(C_1, C_2)$ is more than or equal to the violation threshold value, then the bounding boxes will be drawn in green with the text “safe” as the label on top of them. If $D(C_1, C_2)$ is smaller than the violation threshold value, the bounding boxes be drawn in green with the text “at risk” as the

Table 2 Comparison analysis between three different training models

	YOLOv4	YOLOV3	YOLOv2
mAP@0.50 (%)	93.79	93.07	92.29
Precision	0.82	0.85	0.79
Recall	0.94	0.91	0.93
F1-score	0.87	0.88	0.85

label on top of them. This process is repeated in loop for every frame in real-time video.

4 Results and Discussion

4.1 Quantitative Analysis of Deep Learning Methods

In this work, three different deep learning models are pre-trained with the same dataset and hyperparameters for comparisons. The labelled images are split into 80% of training set and 20% of testing set to measure the robustness of the models.

Based on the quantitative metrics tabulated in Table 2, it is analysed that YOLOv2 model has the lowest performance out of the three training models. On the contrary, the overall robustness of both YOLOv4 and YOLOv3 models are quite similar as their precedence is the other's flaw and vice versa. YOLOv4 model has the upper hand in terms of accuracy and recall whereas YOLOv3 model has the upper hand in terms of precision and F1-score. After some considerations, YOLOv4 model is selected to be deployed as the classifier in the proposed social distancing monitoring model due to having the highest accuracy detection of 93.79% when compared to the other two models. It also has the best sensitivity in not missing out any true positives with recall value at 0.94 and a fair F1-score at 0.87.

4.2 Performance of Social Distancing Monitoring Model

The experiment is done at public areas that have potential widespread of COVID-19 transmission. Hence, the videos are captured from three different cases. Figure 3a, b shows the results at one of the rest stops beside Lebuhraya Utara-Selatan in Perak as an example of open space area. The second case is aimed at enclosed space, for example, like Mid Valley Megamall, and the results are shown in Fig. 3c, d. The third case is aimed at public semi-enclosed place like KL Sentral Transit Hub as shown in Fig. 3e, f.

By referring to the output frames in Fig. 3, it is observed that the overall result of object classification and localization is executed well towards detecting objects that

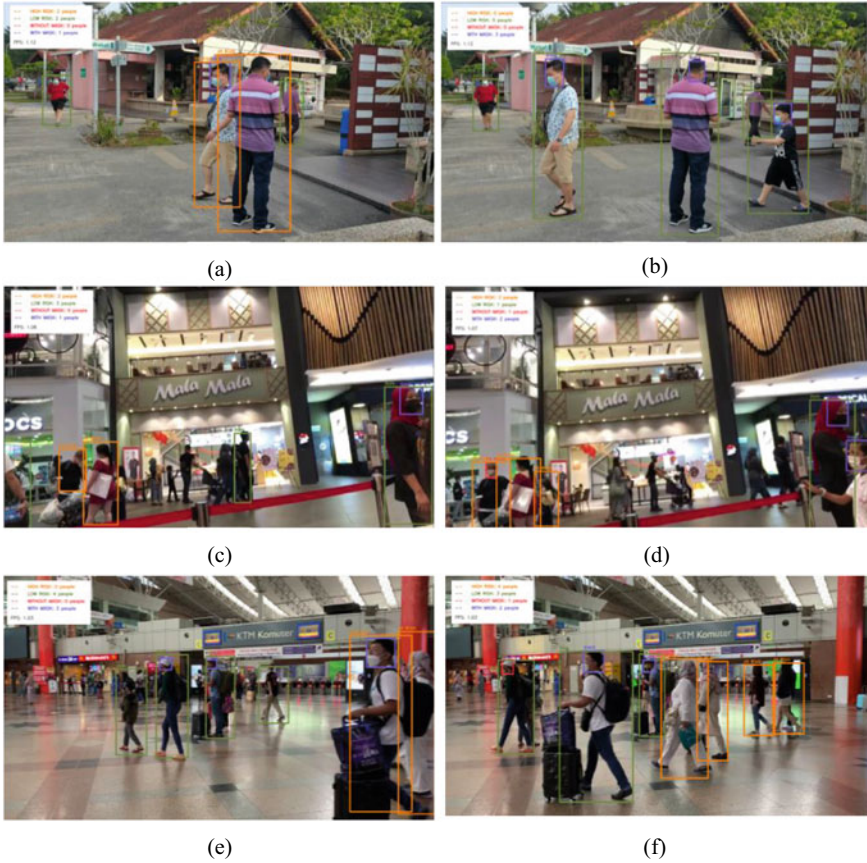


Fig. 3 Visualization of classification and localization as well as monitoring social distancing

are close to the camera. Besides, the monitoring of social distance violation is well performed as expected and the mini dashboard is updated correctly for every single frame according to the number of bounding boxes generated. It can be interpreted that camera position should be taken into considerations as the monitoring performance is able to execute better when the camera is positioned at eye level rather than at lower angle assuming at sitting position level. The performance of social-distance monitoring model, in terms of number of high risks, number of low risks, number of individuals without mask, and number of individuals with mask, is summarized in Table 3.

Table 3 Performance of social-distance monitoring model at Perak Rest Stop, Mid Valley Megamall, and KL Sentral Transit Hub

Venues	Perak Rest Stop		Mid Valley Megamall		KL Sentral Transit Hub	
	Figure 3a	Figure 3b	Figure 3c	Figure 3d	Figure 3e	Figure 3f
# High risks	2	0	2	3	2	4
# Low risks	2	5	3	1	4	3
# Without mask	0	0	0	1	0	1
# With mask	1	3	1	2	3	2

5 Conclusion

The development of social distancing monitoring model using deep learning and the analysis of the model performance are covered in this paper. The effectiveness of social distancing is studied before building the model to understand better in relation to the objective of the project. The process of model training using YOLOv4 method is discussed so that the proposed model can work with real-time and video detection. As a result, the model has achieved accuracy detection of 93.79% and F1-score of 0.87. In terms of deployment performance, it is shown that the object classification and localization as well as the evaluation of social distance violation are executed well towards predicted objects that are close to the camera at eye level position. The outcome of the social distancing monitoring model can be implemented in situations where public health is emphasized corresponds to the practice of preventive measures during COVID-19 pandemic. An additional feature of facemask detection is included too in an effort to mitigate the transmission rate of airborne virus in public places. Nevertheless, improvements can be made in future work to detect a wider range of the crowds since the proposed model only works with objects that are close to the camera.

Acknowledgements This work was supported by the Ministry of Higher Education Malaysia under the Fundamental Research Schemes with project codes of Proj-FRGS/1/2019/TK04/UCSI/02/1 and the UCSI University Research Excellence & Innovation Grant (REIG) with project code of REIG-FETBE-2022/038.

References

1. COVID-19 Vaccines Advice. <http://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines/advice>. Accessed 22 July 2021
2. What is social distancing and how can it slow the spread of COVID-19? | Hub, Mar. <http://hub.jhu.edu/2020/03/13/what-is-social-distancing/>. Accessed 27 July 2021
3. Coronavirus, Social and Physical Distancing and Self-Quarantine | Johns Hopkins Medicine. <http://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/coronavirus-social-distancing-and-self-quarantine>. Accessed 27 July 2021

4. Ainslie KEC et al (2020) Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment. *Wellcome Open Res* 2020 5(5):81
5. 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. *Lancet Public Health* 5(5):e261–e270
6. Fong MW, Gao H, Wong JY, Xiao J, Shiu EYC, Ryu S, Cowling BJ (2020) Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures. *Emerg Infect Dis* 26(5):976
7. Kahlé N (2020) On the economic impact of social distancing measures. SSRN Electron J
8. Jdid B, Lim WH, Dayoub I, Hassan Kais, Rizon M (2021) Robust automatic modulation recognition through joint contribution of Hand-crafted and conceptual features. *IEEE Access* (9):104530–104546
9. Voon YN, Ang KM, Chong YH, Lim WH, Tiang SS (2022) Computer-vision-based integrated circuit recognition using deep learning. In: Zain MZ et al (eds) *Proceedings of the 6th international conference on electrical, control and computer engineering, LNEE*, vol 842. Springer, Singapore, pp 913–925
10. Low JW, Tiang SS, Lim WH, Chong YH, Voon YN (2022) Tomato leaf health monitoring system with SSD and MobileNet. In: Zain MZ et al (eds) *Proceedings of the 6th international conference on electrical, control and computer engineering, LNEE*, vol 842. Springer, Singapore, pp 795–804
11. Alrifay M, Lim WH, Ang CK, Natarajan E, Solihin MI, Rizon M, Tiang SS (2022) Hybrid deep learning model for fault detection and classification of grid-connected photovoltaic system. *IEEE Access* 10:13852–13869
12. Alrifay M, Lim WH, Ang CK (2021) A novel deep learning framework based RNN-SAE for fault detection of electrical gas generator. *IEEE Access* 9:21433–21442
13. Uddin MI, Shah SAA, Al-Khasawneh MA (2020) A novel deep convolutional neural network model to monitor people following guidelines to avoid COVID-19. *J Sens*
14. Saponara S, Elhanashi A, Gagliardi A (2021) Implementing a real-time, AI-based, people detection and social distancing measuring system for Covid-19. *J Real-Time Image Process* 1–11
15. COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques, <https://arxiv.org/abs/2005.01385v4>. Accessed 27 July 2021
16. Ahmed I, Ahmad M, Rodrigues JJPC, Jeon G, Din S (2021) A deep learning-based social distance monitoring framework for COVID-19. *Sustain Cities Soc* 65:102571
17. Rahim A, Maqbool A, Rana T (2021) Monitoring social distancing under various low light conditions with deep learning and a single motionless time of flight camera. *PLoS ONE* 16(2):e0247440
18. Rezaei M, Azarmi M (2020) DeepSOCIAL: social distancing monitoring and infection risk assessment in COVID-19 pandemic. *Appl Sci* 10(21):7514
19. GitHub—X-zhangyang/Real-World-Masked-Face-Dataset: real-World Masked Face Dataset, 口罩人脸数据集, <http://github.com/X-zhangyang/Real-World-Masked-Face-Dataset>. Accessed 24 Feb 2022
20. GitHub—Prajnasb/observations. <http://github.com/prajnasb/observations>. Accessed 24 Feb 2022
21. YOLOv4: Optimal Speed and Accuracy of Object Detection. <http://arxiv.org/abs/2004.10934v1>. Accessed 24 Feb 2022