COVID-19: Machine learning for safe transportation

Subbiah Sankari | Subramaniam Sankaran Varshini | Savvas Mohamed Aafia Shifana

Department of Information Technology, Sri Sai Ram Engineering College, Chennai, India

Correspondence
Sankari Subbiah, Department of Information Technology, Sri Sai Ram Engineering College, Chennai, India.
Email: sankari2705@gmail.com

Abstract
Entire world has been affected by Covid-19 pandemic. In fighting against the Covid-19, social distancing and face mask have a paramount role in freezing the spread of the disease. People are asked to limit their interactions with each other, to reduce the spread of the disease. Here an alert system has to be maintained to caution people traveling in vehicles. Our proposed solution will work primarily on computer vision. The video stream is captured using a camera. Footage is processed using single shot detector algorithm for face mask detection. Second, YOLOv3 object detection algorithm is used to detect if social distancing is maintained or not inside the vehicle. If passengers do not follow the safety rules such as wearing a mask at any point of the time in the whole journey, alarm/alert is given via buzzer/speaker. This ensures that people abide by the safety rules without affecting their daily norms of transportation. It also helps the government to keep the situation under control.

KEYWORDS
computer vision, COVID-19, face mask, single shot detector, social distancing, YOLOv3

1 | INTRODUCTION

The pandemic COVID-19 has touched all parts of the world. Social distancing & face mask have a paramount role in containing the spread of the disease. A recent newspaper report shows that most of the passengers traveling in assorted RTC buses and trains were heedless in maintaining social distancing while they were returning to Visakhapatnam after Dussehra. Many passengers revealed that maintaining social distancing was an incubus in their return journey to the city after the festival though it had been made mandatory by the Indian Government. While the number of corona positive cases has begun to decrease in many countries, India is still battling with COVID-19. Also, a new variant of the virus has begun to spread in about 18 states in India. In order to keep the situation under control, Government of India have taken many positive measures which include social distancing, wearing of face masks, and redesigning of public buses.

According to the official reporting’s from WHO, the COVID-19 virus is spreading majorly and sprightly from one bloke to another bloke via tiny droplets from the snoot and mouth of corona positive patients. People traveling in public transport are more prone to getting affected by corona virus as there is no sufficient space between any two people. Figure 1 shows overcrowding of public bus where people are at high risk due to lack of social distancing.

Social distancing and wearing face masks are the superlative practices where people can curtail physical contact with viable corona virus hauliers by maintaining the safety range of not less than a meter apart from each other. Figures 2 and 3 show the precautionary measures taken in public buses, with social distancing and wearing face mask, that preclude people from taking a seat next to each other.

This study is intended to mitigate the spread of COVID-19 by:

• Detection of face using single shot detector (SSD) algorithm.
• Classification of each detected face into two classes, that is, masked face or unmasked face using MobileNetv2 trained model.
• Detection of people using YOLOv3.
FIGURE 1  Overcrowding in public bus

FIGURE 2  Social distancing in Metro Manila

FIGURE 3  Social distancing in New Delhi

- Calculation of distance between centroids of two adjacent people.
- Providing an alert if rules are violated.

2  LITERATURE SURVEY

At abstract level this study puts forward the perception of persons in public transports chiefly in buses to abate the spread of corona virus, where people can be monitored for maintaining social distancing and wearing face masks. Therefore, some related research works have been studied based
on this initial purpose. The reverberations are utilitarian in allusion to the proposed objectives since both the studies lie on the same groundwork, that is, based on person detection.

### 2.1 Object detection using YOLO (real-time)

You only look once (YOLO) is a regression-based object detector. There are many other object detectors such as SSD. These two detectors, that is, YOLO and SSD have evinced to be exceptionally faster than other territory based object detectors. YOLO takes an entire image as the input promptly while territory based detectors extrapolate regions which are then given to the classifier. They prognosticate object classes and bounding boxes for the entire image in single run of the algorithm. Thus, YOLO is very much faster than the other object detectors and hence it has been the most popular choice. The approach involved in this model (YOLO) is a single deep convolutional neural network. The input is an RGB image that is split input into a mesh of tiny squares $S \times S$. Each square is solely responsible for prognosticating the bounding boxes. The result is a number of similar bounding boxes that are amalgamated into a single prediction by a post-processing step. They are habitually used for real-time object detection as, by and large, they set off a minuscule amount of accuracy for colossal upswing in speed. Figure 4 shows the YOLO object detection algorithm pipeline.

### 2.2 Person detection using segmented ROI

This study focuses on utilizing the MobileNet Single Shot Multibox Detector (SSD) object pursuing model for espying bloke in territory of interest and OpenCV library for image processing. The interspace in the middle of any two persons detected in the live camera footage is estimated. Then it is juxtaposed with a threshold pixel value. The interspace in the middle of the people is calculated using their central points and the imbricating boundary between blokes in the segmented tracking area. If unsafe distances are detected between people, an alert or warning can be given to keep the distance safe. Another prominent aspect of the system is to detect if any person is present in the controlled areas. If the system detects any person in the controlled areas, a warning is triggered. Effectiveness test of the program for both purposes was performed using some analysis techniques. The sequels of the analysis show that the social distance tracking system achieved an accuracy in the middle of 56.5% to 68% for tests performed on outdoor and exacting input videos, while the system showed much greater accuracy when the testing was performed in controlled indoor environment. For the safety violation alert attribute which is based on segmented ROI, it manifested better accuracy, for all tested input videos.

### 2.3 Object detection in manga images using CNN

Japanese comics are known as manga. It has wide reach all around the world. This study is made mainly to procure the metadata from Japanese comics manga images. Manga objects that are used in this study include character face, panel layout, text, and speech balloon. For recognizing manga images content, self-executing techniques were studied. A study was made on manga images in which convolutional neural network (CNN)}
was used to detect the objects present in them. Region of interest was generated using selective search in R-CNN and also in fast R-CNN. While CNN layers called region proposal network (RPN) were used to generate the territory of interest in faster R-CNN. Single shot multibox detector (SSD) is one of the contemporary object detection methods. This method is effective for performing object detection in images with smaller region of interest because it brings about object categorization and box adjustment for small territory in an image. But it is not definitely known whether this method will work satisfactorily to manga images or not, because those images have traits that are incompatible with natural images. In this study, the effectiveness of manga object detection is measured by correlating fast R-CNN, faster R-CNN, and SSD. Preliminary results show that fast R-CNN is effectual for speech balloon and panel layout, whereas faster R-CNN is effectual for text and character face.

2.4 Multiple object detection

In this study real time objects are detected. Detection of real time objects is one among the tedious tasks as it involves rapid computational power in recognizing the object at that particular time. Data produced from any present time system are huge and unlabeled. So, this frequently needs enormous labeled dataset for the purpose of effective training. This study proposes a high-speed detection method for present time object detection which is dependent on CNN model called single shot mult-box detection. This is faster because the feature resampling stage is eliminated. It combines all estimated results as a single final output. Yet there are places which lack computational power such as mobile phones, laptops and so forth. Hence there is a need for a much more lightweight network model. Thus, a lightweight network model called MobileNet is used in this study. It makes use of depth-wise separable convolution. Exploratory results show that use of MobileNet in conjunction with SSD model increases the precision level in recognizing the real time objects.

3 PROPOSED SYSTEM

In the proposed system, it can be detected whether social distancing and wearing of face masks are maintained or not inside the public transports. This system will work primarily based on computer vision. The video stream is live recorded using a camera. The recorded footage is processed with the help of SSD algorithm. SSD is used to detect each face and takes only one shot to detect multiple faces present in a video stream. SSD has high detection accuracy and this is because of using multiple boxes with different sizes and aspect ratio for detecting objects. After face detection, each detected face is classified based on two classes using MobileNetV2 trained model. The MobileNetV2 is used for feature extraction and semantic segmentation. The MobileNetV2 is trained by ImageNet dataset. The trained model has two classifiers, namely masked face and unmasked face. This pretrained model is applied on each detected face and classification of two classes is performed.

The next part is monitoring social distancing and this is achieved using YOLOv3 Object detection algorithm. Each passenger inside the vehicle is detected by using YOLOv3. The pixel distance between any two people is estimated using the Euclidean distance formula. If the inter space in the middle of any two persons is not less than the minimum distance, it will be indicated with a green frame and with the text “SAFE.” If any noncompliant pair of people is detected, then it will be specified as red frame and with the text “ALERT”.

The final part is the alert system. For passengers who do not follow the safety rules, that is, either not maintaining social distancing or not wearing face masks or both, the proposed system triggers the alert system which gives an alarm/warning for a maximum of three times failing which necessary action can be taken upon the violating individual by the bus authorities or any associated authorities who have the right to impose fines or take necessary action upon the erring passenger. Figure 5 shows the architecture diagram of the proposed system.

4 EXPERIMENTAL SET-UP

The proposed idea in this system is developed based on Python3 and OpenCV. To make use of the image processing methods OpenCV library is used. This face mask and social distancing detecting gizmo was invented to detect whether safety measures are followed in public transports. If rules are violated by any of the individuals, an alert or warning will be given to bus authorities or any associated authorities. This system comprises three modules which have been described in this section.

4.1 Face mask detection using SSD and MobileNetv2

In this module, the live stream video will be captured using a camera. Dataset is trained using MobileNetV2. The dataset used comprises 1400 images with facial observation categorized into two classes: with_mask faces and without_mask faces. Face detection is performed using SSD algorithm. SSD has two components, namely backbone model and head model. Backbone model is a pretrained image classification network as
a feature extractor. The head model is the one or more convolutional layers added to the backbone model. The output is the list of bounding boxes around each detected face. The detected face is classified based on two classes (with mask/without mask) using MobileNetV2 trained model. It accepts our input dataset and fine-tunes MobileNetV2 upon it to create a face mask classifier model. During training data augmentation is performed where the random rotation, zoom, shear, and shift are applied on our images. This pretrained model is applied on each detected face and classification of two classes is performed. Figure 6 shows the self-explanatory flowchart of face mask detection from web camera.

(A) Face detection using SSD

As depicted pictorially in the above images, face detection using SSD can be done in five basic steps. First, the image is loaded from the video frame as shown in Figure 7A. Next the loaded image frame is split into grids of size $S \times S$ as shown in Figure 7B. As shown in Figure 7C, faces at different scales are detected using receptive field. Anchor boxes are used to detect multiple faces in grids as in Figure 7D. Finally, the faces detected are displayed as output with bounding boxes as shown in Figure 7E. Zoom Levels and aspect ratios can be used to detect faces in the image at a longer distance.

(B) Social distancing monitoring

In this module, social distance monitoring is performed using you only look once object detection (YOLOv3) algorithm. The real time video stream is recorded using a camera. The recorded video is processed using YOLOv3 to detect passengers inside the vehicle. It takes the whole image promptly and prognosticates the coordinates of bounding box and probability of a particular class. The greatest lead of using YOLOv3 is its fastest speed. It can process up to 45 frames per second and it’s a prevalent object representation. A given image is taken as input in YOLO algorithm and
FIGURE 7  Face detection using SSD. (A) Load image from video frame, (B) divide the image using grids, (C) detect face at different scales using receptive field, (D) use anchor box to detect multiple faces in grids, and (E) display final output.

concurrently learning bounding box coordinates. After performing human detection, the pixel distance between any two people is estimated using the Euclidean distance formula as shown in Equation (1).

\[ d = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2} \]  

where \( d \) is the calculated distance between the centroids of people, \((x_1, y_1)\) are the position coordinates of person1, \((x_2, y_2)\) are the position coordinates of person2.

If the distance between any two persons is not less than the fixed threshold distance, ‘\( t \)’ is the threshold value and will be compared with the calculated distance ‘\( d \)’. If ‘\( d \)’ value is greater than or equal to ‘\( t \)’, then ‘\( i \)’ will be indicated with a green frame and alert with a text "SAFE." If any noncompliant pair of people is detected, then it will be specified with a red frame and notified through the message with a text "ALERT" as in Equation (2).

\[ i = \{ \text{red if } d < t, \text{ green if } d \geq t \} \]
In Reference 8, YOLO is a regression-based object detector. It gives much better performance on all the parameters including high frames per second for real-time usage. Instead of predicting the region of interest, it draws bounding boxes across different classes of objects in the whole image in a single run.

There are five main parameters for each bounding box, namely

- Center of the bounding box \((x, y)\),
- Width \((w)\),
- Height \((h)\),
- Class value \((c)\),
- Probability \((p_c)\).

Initially the image is split into grids of size \(S \times S\). In each cell, \(K\) bounding boxes are predicted. A human is said to be present in a cell if and only if center coordinates of the anchor box lie in that respective cell. The center coordinates are estimated always with respect to the cell while the height \((h)\) and width \((w)\) are calculated with respect to the whole image size. During a single pass of forward propagation, the class probability is calculated as in Equation (3).

\[
score_{ij} = p_c \times c_i
\]

Succeeding step in the process is non-max suppression. It is done in order to help the algorithm to do away with extraneous anchor boxes. Non-max suppression uses intersection over union (IoU) to eliminate the unwanted anchor boxes. Thus, the human is detected in the image using YOLO. Figure 8 shows the pictorial representation of IoU operation.

(D) Alert system

Using the above two modules, the person violating the rules will be identified. Once the distance violation is encountered, then an alarm is given. Also, if the passenger does not wear a mask at any point of the time in the whole journey, alarm/alert is given. If any noncompliant pair of passengers is detected, then it will be specified as a red frame and automatically a warning or buzzer sound is triggered for a maximum of three times to alert passengers inside the vehicle failing which necessary action can be taken upon the violating individual by the bus authorities or any associated authorities who have the right to impose fines or take necessary action upon the erring passenger.

5 RESULTS

Figure 9 depicts the accuracy curve and training loss for the model trained using the above said experimental set-up. Training loss and accuracy graph is plotted by having epoch in x-axis and loss/accuracy in the y-axis. An epoch represents the number of passes, the machine learning algorithm has made through the entire training dataset. Typically, datasets are organized into batches. More than one epoch is used to create training models. The relationship between dataset size \((D)\), number of epochs \((E)\), batch size \((B)\) and number of iterations \((I)\) is shown in the Equation (4). In this equation, \(D, E\) are directly proportional to \(B, I\).

\[
D \times E = B \times I
\]

In machine learning, accuracy can be calculated as in Equation (5) from the predicted and expected value. Let us consider the following:

People wearing mask and maintaining social distancing—true positive (TP).
People wearing mask and not maintaining social distancing—true negative (TN).
People not wearing mask and maintaining social distancing—false negative (FN).
People not wearing mask and not maintaining social distancing—false positive (FP).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  

The validation accuracy in our model is 1.0, indicating that our system performs well when compared with the training accuracy. Similarly, the validation loss is less when compared with training loss. Thereby, our proposed system showcases the accurate prediction of the person wearing mask with social distance.

Figure 10 shows the training and validation accuracy model graph, in which the red line indicates Validation accuracy and blue line indicates the training accuracy of the person wearing mask and maintaining social distancing.

Figure 11 shows the training and validation loss model graph, in which the red line indicates Validation loss and blue line indicates the training loss of the person wearing mask and maintains social distancing.
FIGURE 12 People not wearing mask

FIGURE 13 People who do not follow social distancing

FIGURE 14 People following safety measures

FIGURE 12 shows red box and also alert signal for people not wearing mask.
Figure 13 depicts the outcome in which individuals violated the social distancing rule, resulting in an immediate alert.
Figure 14 shows the output where social distancing and wearing of face mask are followed.

6 CONCLUSION AND FUTURE ENHANCEMENT

Wearing a face mask and following social distancing is one of the prominent precautionary measures to reduce the spread of corona virus. Face mask detector and social distance monitoring tool that have been developed can possibly contribute to public healthcare where a person traveling in public transport will find it safe and secure. The three classes that are concurrently perceived are face with mask, face without mask as well as entire people. Using the coordinates produced by the detection of the class “person”, the relative interspace in the middle of any two individuals
is estimated and generating the alert system that can be handled and monitored. This system works very effectively and efficiently in identifying
the social distancing between the people in public transports, there by safety measures are achieved inside the vehicle which ensures Covid free
traveling without affecting their daily norms of transportation. This also helps government to keep situations in control.

The work can be further enhanced by advancing the passenger detection algorithm. Amalgamating other algorithms like human body tempera-
ture detection which can add more safety to the public and also improving the computational power.

DATA AVAILABILITY STATEMENT
Data availability statement not needed and no new data generated

ORCID
Subbiah Sankari https://orcid.org/0000-0003-1067-1177

REFERENCES
1. THE HINDU. Social distancing goes for a toss at bus station; 2020. www.thehindu.com/news/cities/Visakhapatnam/social-distancing-goes-for-a-toss-at-bus-station/article31653287.ece
2. TOI. New Covid variant found in 18 states in India; 2021. https://timesofindia.indiatimes.com/india/coronavirus-live-news-updates-march-23-2021/liveblog/81642011.cms
3. TOI. RTC buses undergo changes to maintain social distancing; 2020. https://timesofindia.indiatimes.com/city/visakhapatnam/rtc-buses-undergo-changes-to-maintain-social-distancing/articleshow/75745336.cms
4. The HINDU. Inter-district bus, rail services to resume in Tamil Nadu; 2020 https://www.thehindu.com/news/national/tamil-nadu/bus-rail-services-between-districts-to-resume-in-tamil-nadu-from-september-7/article32503303.ece
5. Inkl. ‘Social distance’ cuts down public transport in Metro Manila; 2020. https://www.inkl.com/news/watch-social-distance-cuts-down-public-transport-in-metro-manila
6. Mint. Passengers maintain social distancing while sitting in a bus; 2020.https://www.livemint.com/news/world/six-feet-distancing-not-enough-to-stop-virus-transmission-in-light-winds-study-11589952990283.html
7. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S. SSD: single shot multibox detector; 2015. arXiv:1512.02325.
8. Lu Y, Zhang L, Xie W. YOLO-compact: an efficient YOLO network for single category real-time object detection. Proceedings of the Chinese Control and Decision Conference (CCDC); 2020. doi:10.1109/CCDC49329.2020.9164580
9. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014:580-587; Columbus, OH. doi: 10.1109/CVPR.2014.81
10. MDPI. YOLO-TOMATO: a robust algorithm for tomato detection based on YOLOv3; 2020. https://www.mdpi.com/1424-8220/20/7/2145/htm
11. Ahamad AH, Zaini N, Latip MF. Person detection for social distancing and safety violation alert based on segmented ROI. Proceedings of the 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE); August 21, 2020:113-118; IEEE. doi:10.1109/ICCSCE50387.2020.9204934
12. Yanagisawa H, Watanabe TYH. A study on object detection method from manga images using CNN. Proceedings of the International Workshop on Advanced Image Technology, IWAIT; 2018. doi: 10.1109/IWAIT.2018.8369633
13. Kanimozi S, Gayathri G, Mala T. Multiple real-time object identification using single shot multi-box detection. Proceedings of the International Conference on Computational Intelligence in Data Science, ICCIDS; 2019. doi: 10.1109/ICCIDS.2019.8862041, 2019.
14. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks; 2018. arXiv:1801.04381.
15. Towards Data Science. OLO — You only look once; 2020. https://towardsdatascience.com/yolo-you-only-look-once-3dbbb608ec4

How to cite this article: Sankari S, Varshini SS, Aafia Shifana SM. COVID-19: Machine learning for safe transportation. Concurrency Computat Pract Exper. 2022;34(19):e7041. doi: 10.1002/cpe.7041