# Social Distancing Detection with Deep Learning Model

Yew Cheong Hou[1], Mohd Zafri Baharuddin[2], Salman Yussof[1], Sumayyah Dzulkifly[1]
[1]Institute of Informatics and Computing in Energy
[2]College of Engineering
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
email: {ychou, zafri, salman, sumayyah}@uniten.edu.my

***Abstract*—The paper presents a methodology for social distancing detection using deep learning to evaluate the distance between people to mitigate the impact of this coronavirus pandemic. The detection tool was developed to alert people to maintain a safe distance with each other by evaluating a video feed. The video frame from the camera was used as input, and the open-source object detection pre-trained model based on the YOLOv3 algorithm was employed for pedestrian detection. Later, the video frame was transformed into top-down view for distance measurement from the 2D plane. The distance between people can be estimated and any noncompliant pair of people in the display will be indicated with a red frame and red line. The proposed method was validated on a pre-recorded video of pedestrians walking on the street. The result shows that the proposed method is able to determine the social distancing measures between multiple people in the video. The developed technique can be further developed as a detection tool in real-time application.**

***Keywords—social distancing, pedestrian detection, deep learning, convolutional neural network.***

## I. Introduction

When the novel coronavirus (Covid-19) pandemic emerges, the spread of the virus has left public keep anxiety if they do not have any effective cure. The World Health Organization (WHO) has declared Covid-19 as a pandemic due to the increase in the number of cases reported around the world [1]. To contain the pandemic, many countries have implemented a lockdown where the government enforced that the citizens to stay at home during this critical period. The public health bodies such as the Centers for Disease Control and Prevention (CDC) had to make it clear that the most effective way to slow down the spread of Covid-19 is by avoiding close contact with other people [2]. To flatten the curve on the Covid-19 pandemic, the citizens around the world are practicing physical distancing.

To implement social distancing, group activities and congregations such as travel, meetings, gatherings, workshops, praying had been banned during the quarantine period. The people are encouraged to use phone and email to manage and conduct events as much as possible to minimize the person-to-person contact. To further contain the spread of the virus, people are also informed to perform hygiene measures such as frequently washing hands, wearing mask and avoiding close contact with people who are ill. However, there is a difference between knowing what to do to reduce the transmission of the virus and putting them into practice.

The world has not yet fully recover from this pandemic and the vaccine that can effectively treat Covid-19 is yet to be discovered. However, to reduce the impact of the pandemic on the country's economy, several governments have allowed a limited number of economic activities to be resumed once the number of new cases of Covid-10 has dropped below a certain level. As these countries cautiously restarting their economic activities, concerns have emerged regarding workplace safety in the new post-Covid-19 environment. To reduce the possibility of infection, it is advised that people should avoid any person-to-person contact such as shaking hands and they should maintain a distance of at least 1 meter from each other.

In Malaysia, the Ministry of Health Malaysia (MOHM) has recommended several disease prevention measures for workplaces, individuals, and families at home, schools, childcare centres, and senior living facilities [3]. These measures include implementing social distancing measures, increasing physical space between workers at the workplace, staggering work schedules, decreasing social contacts in the workplace, limiting large work-related gatherings, limiting non-essential work travel, performing regular health checks of staff and visitors entering buildings, reducing physical activities especially for organizations that have staff in the high-risk category, and conducting company events or activities online.

Individuals, communities, businesses, and healthcare organizations are all part of a community with their responsibility to mitigate the spread of the Covid-19 disease. In reducing the impact of this coronavirus pandemic, practicing social distancing and self-isolation have been deemed as the most effective ways to break the chain of infections after restarting the economic activities. In fact, it has been observed that there are many people who are ignoring public health measures, especially with respect to social distancing. It is understandable that given the people's excitement to start working again, they sometimes tend to forget or neglect the implementation of social distancing. Hence, this work aims to facilitate the enforcement of social distancing by providing automated detection of social distance violation in workplaces and public areas using a deep learning model. In the area of machine learning and computer vision, there are different methods that can be used for object detection. These methods can also be applied to detect the social distance between people. The following points summarizes the main components of this approach:

a. Deep learning has gained more attention in object detection was used for human detection purposes.
b. Develop a social distancing detection tool that can detect the distance between people to keep safe.
c. Evaluation of the classification results by analyzing real-time video streams from the camera.

## II. Related Work

This section highlights some of the related works about human detection using deep learning. A bulk of recent works on object classification and detection involve deep learning are also discussed. The state-of-the-art review mainly focuses on the current research works on object detection using machine learning. Human detection can be considered as an object detection in the computer vision task for classification and localization of its shape in video imagery. Deep learning has shown a research trend in multi-class object recognition and detection in artificial intelligence and has achieved outstanding performance on challenging datasets. Nguyen et al. presented a comprehensive analysis of state-of-the-art on recent development and challenges of human detection [4]. The survey mainly focuses on human descriptors, machine learning algorithms, occlusion, and real-time detection. For visual recognition, techniques using deep convolutional neural network (CNN) have been shown to achieve superior performance on many image recognition benchmarks [5].

Deep CNN is a deep learning algorithm with multilayer perceptron neural networks which contain several convolutional layers, sub-sampling layers, and fully connected layers. Later, the weight in the whole layers in the networks are trained for each object classification based on its dataset. For object detection in image, the CNN model was one of the categories in deep learning which are supervised feature learning methods robust in detecting the object in different scenarios. CNN has achieved great success in large-scale image classification tasks due to the recent high-performance computing system and large dataset such as ImageNet [6]. Different CNN models for object detection with its object localization had been proposed in terms of network architecture, algorithms, and new ideas. In recent years, CNN models such as AlexNet [5], VGG16 [7], InceptionV3 [8], and ResNet-50 [9] are trained to achieve outstanding results in object recognition. The success of deep learning in object recognition is due to its neural network structure that is capable of self-constructing the object descriptor and learning the high-level features which are not directly provided in the dataset.

The current state-of-the-art object detectors with deep learning had their pros and cons in terms of accuracy and speed. The object might have different spatial locations and aspect ratios within the image. Hence, the real-time algorithms of object detection using the CNN model such as R-CNN [10] and YOLO [11] had further developed to detect multi-classes in a different region in images had been developed. YOLO (You Only Look Once) is the prominent technique for deep CNN-based object detection in terms of both speed and accuracy. The illustration for the YOLO model is shown in Figure 1

Adapting the idea from the work [12], we present a computer vision technique for detecting people via a camera installed at the roadside or workspace. The camera field-of-view covers the people walking in a specified space. The number of people in an image and video with bounding boxes can be detected via these existing deep CNN methods where the YOLO method was employed to detect the video stream taken by the camera. By measuring the Euclidean distance between people, the application will highlight whether there is sufficient social distance between people in the video.

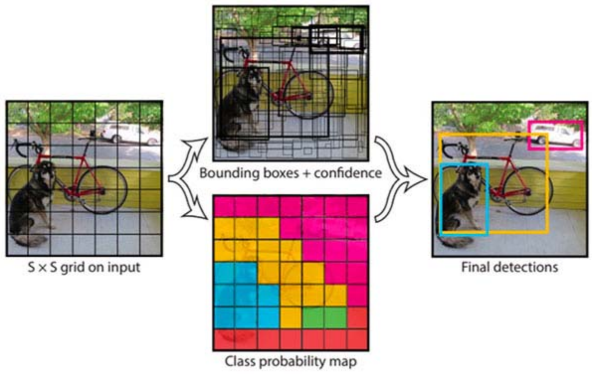


Figure 1 The illustration of YOLO model for object detector pipeline. [11].

## III. Methodology

This social distancing detection tool was developed to detect the safety distance between people in public spaces. The deep CNN method and computer vision techniques are employed in this work. Initially, an open-source object detection network based on the YOLOv3 [13] algorithm was used to detect the pedestrian in the video frame. From the detection result, only pedestrian class was used and other object classes are ignored in this application. Hence, the bounding box best fits for each detected pedestrian can be drawn in the image, and these data of detected pedestrians will be used for the distance measurement.

For camera setup, the camera is captured at fixed angle as the video frame, and the video frame was treated as perspective view are transformed into a two-dimensional top-down view for more accurate estimation of distance measurement. In this methodology, it is assumed that the pedestrians in the video frame are walking on the same flat plane. Four filmed plane points are selected from frame and then transformed into the top-down view. The location for each pedestrian can be estimated based on the top-down view. The distance between pedestrians can be measured and scaled. Depending on the preset minimum distance, any distance less than the acceptable distance between any two individuals will be indicated with red lines that serve as precautionary warnings. The work was implemented using the Python programming language. The pipeline of the methodology for the social distancing detection tool is shown in Figure 2.

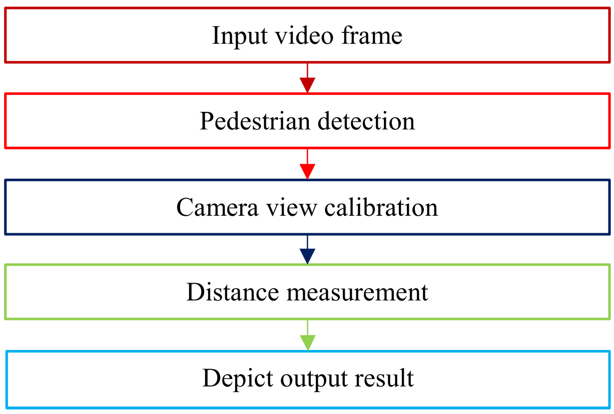


Figure 2 Pipeline for social distancing detection.

### A. Pedestrian detection

Deep CNN model was the object detection approach was proposed that mitigated the computational complexity issues by formulating the detection with a single regression problem [11]. When it comes to deep learning-based object detection, the YOLO model is considered one of the state-of-the art object detectors which can be demonstrated to provide significant speed advantages will suitable for real-time application. In this work, the YOLO model was adopted for pedestrian detection is shown in Figure 3. The YOLO algorithm was considered as an object detection taking a given input image and simultaneously learning bounding box coordinates ($t_x$, $t_y$, $t_w$, $t_h$), object confidence and corresponding class label probabilities ($P_1$, $P_2$, …, $P_c$). The YOLO trained on the COCO dataset which consists of 80 labels including human or pedestrian classs. In this work, the only box coordinates, object confidence and pedestrain object class from detection result in the YOLO model were used for pedestrian detection.

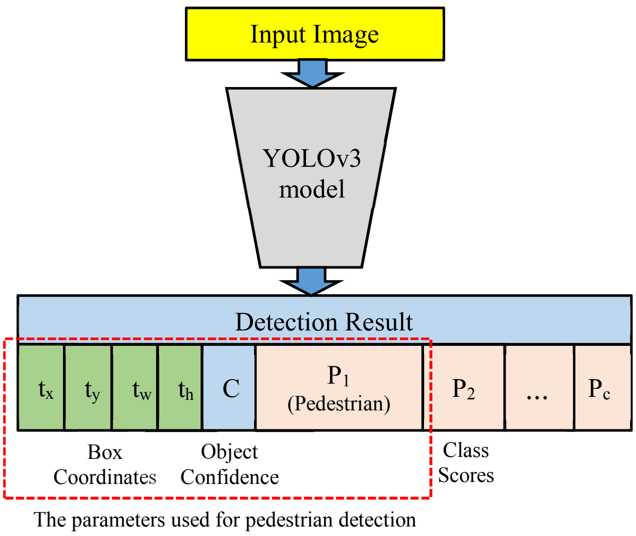


Figure 3 YOLO model applied for pedestrian detection.

### B. Camera view calibration

The region of interest (ROI) of an image focuses on the pedestrian walking street was transformed into a top-down 2D view that contains 480×480 pixels as shown in Figure 4. Camera view calibration is applied which works by computing the transformation of the perspective view into a top-down view. In OpenCV, the perspective transformation is a simple camera calibration method which involves selecting four points in the perspective view and mapping them to the corners of a rectangle in the 2D image view. Hence, every person is assumed to be standing on the same level flat plane. The actual distance between pedestrians corresponds to the number of pixels in the top-down view can be estimated.

### C. Distance measurement

In this step of the pipeline, the location of the bounding box for each person (*x, y, w, h*) in the perspective view is detected and transformed into a top-down view. For each pedestrian, the position in the top-down view is estimated based on the bottom-center point of the bounding box. The

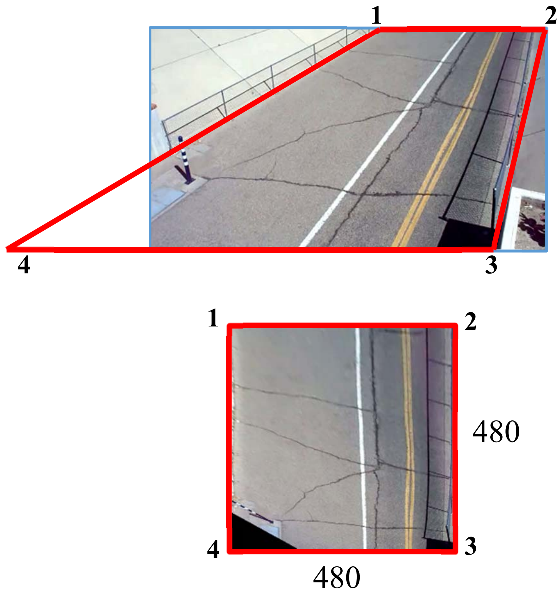


Figure 4 Sample of original perspective view (top) and top down view after calibration (bottom).

distance between every pedestrian pair can be computed from the top-down view and the distances is scaled by the scaling factor estimated from camera view calibration. Given the position of two pedestrians in an image as ($x_1$, $y_1$) and ($x_2$, $y_2$) respectively, the distance between the two pedestrians, *d*, can be computed as:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

The pair of pedestrians whose distance is below the minimum acceptable distance, *t*, is marked in red, and the rest is marked in green. A red line is also drawn between the pair of individuals whose distance is below the pre-defined threshold. The bounding box's color threshold operation, *c*, can be defined as:

$$c = \begin{cases} red & d < t \\ green & d \geq t \end{cases} \quad (2)$$

Overall, the flowchart of social distancing detection is depicted in Figure 5.

## IV. Result and Discussion

The video shows the pedestrian walking on a public street. In this work, the video frame is fixed at a specified angle to the street. The perspectives view of the video frame is transformed into a top-down view for more accurate estimation of distance measurement. Figure 6 shows the social distancing detection in a video frame and the results of the top-down view. The sequences are depicted from top to bottom. The points represent each pedestrian for social distancing detection. The red points represent the pedestrians whose distance with another pedestrian is below the acceptable threshold and the green points represent the pedestrians who keep a safe distance from other pedestrians. However, there are also a number of detection errors are shown in Figure 7. These errors are possibly due to the pedestrians walking too near to another pedestrian until they are overlaid on the camera view. The precision of the distance measurement between pedestrians is also affected by the

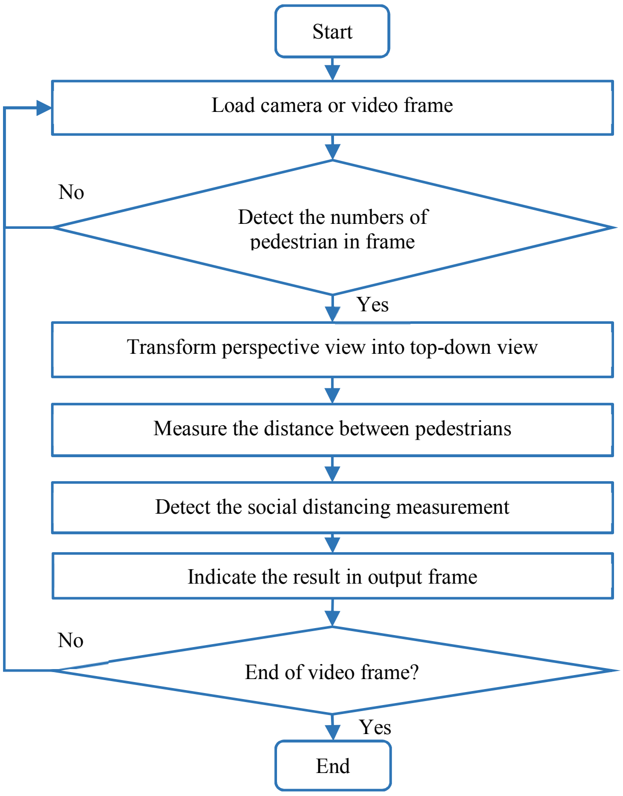


Figure 5 Program flowchart of social distancing detection for each video frame.

pedestrian detection algorithm. The YOLO algorithm is also able to detect the half body of the pedestrian as an object by showing the bounding box, the position of the pedestrian corresponds the middle-point of bottom line is estimated based on the bounding box will less precise. To overcome the detection errors, the proposed methodology had been improved by adding a quadrilateral box to observe the appointed region in an image as shown in Figure 8. Hence, only the pedestrians walking within the specified space will be counted for people density measurement.

## V. Conclusion and Future Works

A methodology of social distancing detection tool using a deep learning model is proposed. By using computer vision, the distance between people can be estimated and any noncompliant pair of people will be indicated with a red frame and a red line. The proposed method was validated using a video showing pedestrians walking on a street. The visualization results showed that the proposed method is capable to determine the social distancing measures between people which can be further developed for use in other environment such as office, restaurant, and school. Furthermore, the work can be further improved by optimizing the pedestrian detection algorithm, integrating other detection algorithms such as mask detection and human body temperature detection, improving the computing power of the hardware, and calibrating the camera perspective view.

## Acknowledgment

This work was funded by Yayasan Canselor Uniten (201901001YCU/17). The authors would like to acknowledge our research collaborators, Itxotic Sdn. Bhd., for their contribution, expertise and equipment.

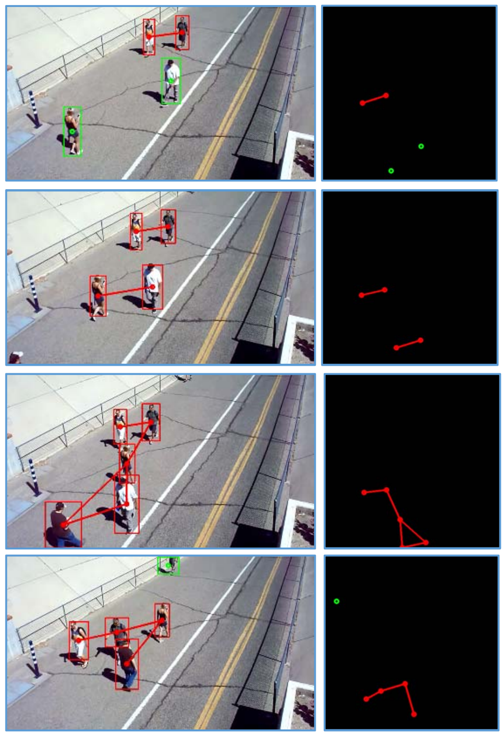

Figure 6 Social distancing detection in video frame (left) and in top-down view (right). The sequences are depicted from top to bottom.

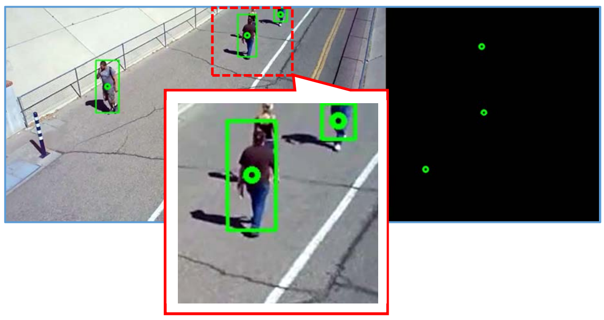

Figure 7 Detection errors due to overlay between pedestrians and half-body detection respectively.

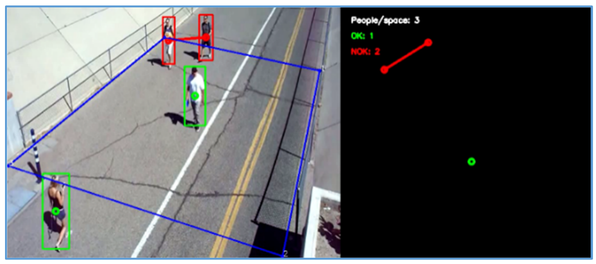


Figure 8 An example of a blue quadrilateral box placed on the image for people density calculation in specified region.

## References

[1] Centers for Disease Control (CDC). Implementation of Mitigation Strategies for Communities with Local COVID-19 [Online]. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (Accessed 8 May 2020).

[2] Centers for Disease Control (CDC). Implementation of Mitigation Strategies for Communities with Local COVID-19 Transmission [Online]. Available at https://www.cdc.gov/coronavirus/2019-ncov/downloads/community-mitigation-strategy.pdf (Accessed 8 May 2020).

[3] Ministry of Health Malaysia (MOHM) Official Portal. COVID-19 (Guidelines) [Online]. Available at https://www.moh.gov.my/index.php/pages/view/2019-ncov-wuhan-guidelines (Accessed 8 May 2020).

[4] D.T. Nguyen, W. Li, P.O. Ogunbona, "Human detection from images and videos: A survey", Pattern Recognition, 51:148-75, 2016.

[5] A. Krizhevsky, I. Sutskever, G.E. Hinton, "Imagenet classification with deep convolutional neural networks", In Advances in neural information processing systems, pp. 1097-1105, 2012.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database", In Computer Vision and Pattern Recognition, 2009.

[7] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826, 2016.

[9] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.

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

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

[12] Landing AI Creates an AI Tool to Help Customers Monitor Social Distancing in the Workplace [Onlive]. Available at https://landing.ai/landing-ai-creates-an-ai-tool-to-help-customers-monitor-social-distancing-in-the-workplace/ (Access on 4 May 2020).

[13] J. Redmon, A. Farhadi, "Yolov3: An incremental improvement", arXiv preprint arXiv:1804.02767, 2018.