Social Distance Detection Using YOLO

Ashok V

Post Graduate Student in Computer Science

Government Model Engineering College Kochi, Kerala, India

***Abstract*—The world has been facing a unique situation for the past two years, COVID-19. It reached every corner of the world. There are different methods to prevent the spread. Usage of masks, avoiding social gatherings, washing your hands, maintaining social distancing, etc. Maintaining social distance is the best strategy to prevent the spread of the virus. But people do not care about social distancing. It causes the further spread of the virus. The goal of social distancing is to keep healthy people and those who are sick from coming into close contact with one another. It slows down the spread of the virus. It is very difficult to monitor social distancing in public places. But by using technologies like machine learning, deep learning, and computer vision, it is possible to monitor social distancing automatically. This paper proposes a method in which the video frame from the camera is used as input and the open source object detection pre-trained model based on the YOLO-v4 algorithm is employed for detection of people. Then the distance between people is estimated. The pair that violates the social distancing will be highlighted red. With the help of YOLO, real-time detection is possible.**

***Index Terms*—COVID-19, YOLO, deep learning**

I. Introduction

COVID-19 has rapidly affected our daily life, global economy, business, etc. The virus is spreading rapidly across the world. Most of the companies are stopped their manufacturing processes, software companies are started work from home culture , classes in schools and colleges are happening in online mode the world is changed in different way. All these changes are due to a small virus called Corona virus.

COVID-19 also known as corona virus disease 2019 is a disease which is caused by the severe acute respiratory syndrome conronavirus 2 (SARS-CoV-2). The virus is first reported at Wuhan, China in the month of December 2019. It is spread worldwide, leading to a pandemic. The most common symtoms of the COVID-19 disease are fever, cough, headache, fatigue, breathing difficulties, loss of smell, and loss of taste. After exposed to virus, the symptoms may begin within fourteen days. In some cases symptoms may not be visible. But those people can also spread the virus. The virus is transmitted when people breathe in air contaminated by droplets and small airborne particles containing the virus. Since the beginning of the COVID-19 pandemic, the SARS-CoV-2 coronavirus that causes COVID-19 has changed (mutated) resulting in different variants of the virus. Variants are given names according to the Greek alphabets. One of these is called the Delta variant [15], it is considered to be a dangerous variant because it appears to be more easily transmitted from one person to another. Another dangerous variant is Omicron [14]. There is an evidence suggesting the Omicron variant is more infectious than the delta variant. There is a chance of spreading these viruses through the air.

To prevent the spreading of virus many countries implemented the lockdown system [16]. The doctors and health professionals stated that the effective way to prevent the spreading of virus is by avoiding the close contact. In order to avoid close contact between persons, they must follow social distancing or physical distancing. Social distancing is a process of keeping a minimum space between two persons. For an effective social distancing, people must keep at least 6 feet between each other.

Meeting, gatherings, travels, workshops, praying, etc are banned by government in order to implement the social distancing. Instead of these, people can conduct meetings, workshops etc in virtual mode. But it is not possible to depend virtual meeting and gatherings all the time. To overcome this problem, governments reduced the restrictions once the number of new covid-19 has dropped below a certain level. But after removing restrictions there will be a chance of spreading the virus again. To reduce the possibility infection, people should avoid activities like shaking hands, hugging, talking in very close, etc. Everyone should maintain at least a distance of 1m from each other.

In many countries governments recommended several prevention measures in workplace, public places, colleges. These measures include implementing social distancing measures, increasing physical space between people, etc. But people sometimes tend to forget or neglect the restrictions. So this work is focused to provide automated detection of social distance violations in public places using a deep learning model. The deep learning model can be used effectively to detect social distance vioilations in public places. Deep learning methods have gained more attention for human detection purposes [17]. As a result, developing an automatic social distance detection tool will be extremely beneficial to the government in monitoring violations. This tool classifies the people by evaluating the real-time video streams from the camera.

II. Related Work

Object detection is a computer technique which is related to deep learning, computer vision, image processing, etc. Object detection involves locating objects in digital images or videos.

Pedestrian detection, face detections, are the example of object detection. Object detection consists of two things,

- Image classification
- Object Localization

Image classification deals with the classification of different objects in an image or video frame. Classification is divided into two. One is binary classification and other is multiple classification. If the requirement is to classify the objects into multiple classes, then it will come under multiple classification. Wheras the requirement is only knowing the presence of a single object, then it will come under the binary classification. In order to classify an image, first step is to locate the object. So the task of locating the object in an image is known as object localisation. In an application like face detection, the application is focused on finding a face in the image, whereas in number plate detection, the application will focus on detecting vehicle number plates. So the object detection will help to identify the image segment that the application needs to focus on. By using object detection it is possible to reduce the dimension of the image to only capture the object of interest it will leads to improve the overall performance of a system also it increases the accuracy as well. Object detection is achieved by using traditional methods, machine learning-based techniques, or deep learning techniques.

Traditional object detection methods were built to support handcrafter features. And it uses some classifier methods like SVM, so as to classify the item, it involves a substantial amount of calculation. So using these methods for a real time application would be difficult. Papers like Motion Based Recognition of Pedestrians [1] discuss the traditional object detection methods. Motion-based recognition proposes an algorithm for detecting pedestrians in colour images taken from a moving camera. It detects pedestrians by checking the characterestic motion of the legs of a pedestrian walking parallel to image plane. Every image is segmented into region-like pictures parts by clustering pixels in a combined color/position feature space. This is an example of traditional object detection technique. This section is more focusing on other two object detection techniques, that is machine learning based object detection and deep learning based object detection techniques.

Machine learning based object detection methods are achieved more improvements when compared to traditional object detection methods. In the machine learning based object detection approach, user defines the fearures and then using these features train the classifier. The major papers which discuss the machine learning based object detection methods are scale invarient feature transform, Viola Jones face detector [3] and Object detection using Histogram of Oriented Gradients features [5].

The scale invarient feature transform [2] was created in 1999 by David Lowe from the University British Columbia. This approach is used for image feature generation. It takes an image and transforms it into a large collection of local feature vectors. These feature vectors are invarient to any scaling, rotation and translation of the image. The Viola Jones face detector also known as VJ detector [3] is the first efficient face detecting algorithm. In this approach the user is hardcoded all the features of the face and then it is trained on Support Vector Machine classifier. The features are known as Haar Cascades. But the probem with this algorithm is, this algorithm can not detect the faces in other orientations. The object detection using Histogram of Oriented gradients (HOG) [5] features was created in 2005 by Navneet Dalal and Bill Triggs.In this approach the local object appearence and shape within an image can be described by the distribution of intensity gradients or edge directions. Now the image is represented by using these HOG features. The training and validation of classifiers such as support vector machine happens using this descriptor.

The deep learning methods are based on convolutional neural networks(CNN). These methods are capable of doing end-end object detection without defining the features. These methods are able to learn robust and high level feature representation of an image. AlexNet [7], Region based convolutional neural network [9],Single Shot Detector [12], YOLO, are some of the examples of deep learning methods.

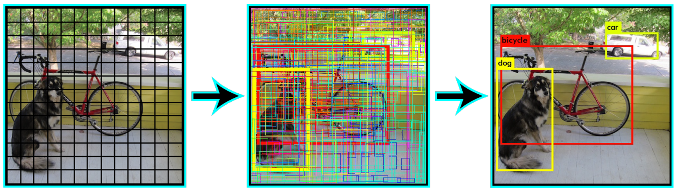


Fig. 1: The illustration of YOLO

The Alexnet was created in 2012 by Alex Krizhevesky in collaboration with Ilya Sutskever and Geoffrey Hinton. AlexNet won the ImageNet Large Scale Visual Recognition Challenge(ILSVRC) by outperforming all prior competitors. ImageNet [18] is a large database of quality controlled, human-annotated images that help test algorithms that are built to store, retrieve, or annotate multimedia data. Convolutional Neural Networks became the gold standard for image classification after Kriszhevsky's CNN's performance during ImageNet. The convolutional neural networks are slow and expensive in terms of computation. R-CNN solved this problem by reducing the number of bounding boxes that are fed to the classifier [4]. In R-CNN nearly 2000 bounding boxes are fed into the classifier. This selective search method replaces the exhaustive search method which is used to capture object location. After extracting the region proposals [10] it compute the CNN features. Then it classify the regions. The single shot detector(SSD) [12] is published in 2015 by Wei Liu et al. The SSD predicts offset of predefined anchor boxes for every location of the feature map. Feature maps at different levels have different receptive field sizes. Here only one feature is responsible for objects at one particular scale. The YOLO (You Only Looks Once) [11] is an example of single stage detector. Because it directly predicts the bounding boxes and probability of each class with a single network in a single evaluation. This will improve the detecton speed. Today different versions of

YOLO are present. On every upgrade its accuracy and speed is improving.

III. Methodology

Social distancing is the best method to prevent the spread of COVID-19. Social distancing is a process of keeping a minimum space between two persons. This will slows down the spread of COVID-19. But people sometimes tend to forget or neglect the restrictions. So this work is focused to provide automated detection of social distance violations in public places using a deep learning model. The deep learning model can be used effectively to detect social distance vioilations in public places.

Deep learning and computer vision techniques are used in the proposed system. An open sourced object detection network based on YOLOv4 [20] algorithm is used to detect persons from the images/videos. Basically YOLO is a multi-class classification algorithm, which means YOLO is capable of classify more than two classes. But here the system need to detect person class only. So here the YOLO is used to detect all persons [13] from an image. After detecting the results, the coordinate values of the bounding boxes can be used to calculate the distance.

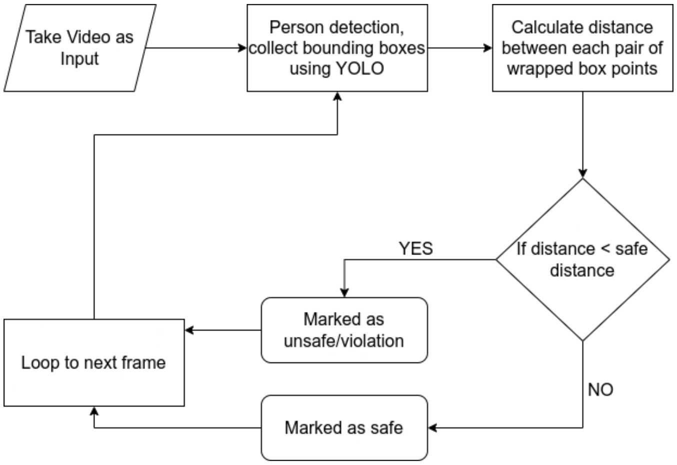
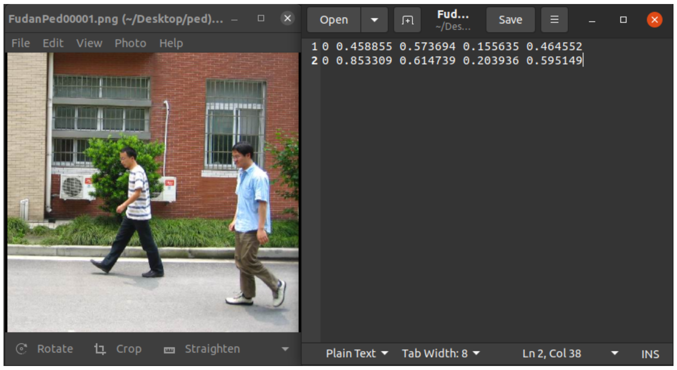

Take Video as Input
Person detection, collect bounding boxes using YOLO
Calculate distance between each pair of wrapped box points
If distance < safe distance
YES
NO
Marked as unsafe/violation
Marked as safe
Loop to next frame

Fig. 2: Flowchart of the system

This system assumes that the pedestrians walks in a flat plane. it is possible to measure the accurate result if the top down view is considered. Before the start of detection, the camera must be calibrated according to the present position. Depending on the distance between two pedestrians the system classifies pedestrians into two classes. One is the class those who are not violating the minimum distance, and the other class is the pair of pedestrians those are violating the minimum distance. The violations are marked in red frame and safe people are marked in green frame. The total number of violations are also calculated. The work was done in Python programming language. The flowchart of this social distance detection system is shown in Fig 2

A. Pedestrian Detection

Compared to traditional and machine learning based object detection methods, deep learning based methods reduced the computational complexity issues by formulating the detection with a single regression problem. When considering the the deep learning methods, the YOLO model is considered the best object detectors which can detect objects in real time with high accuracy [19]. In this work YOLO model is used for pedestrian detection. The YOLO model trained on COCO dataset can able to detect 80 classes including persons. After detecting objects the YOLO model output the object's box coordinates values and corresponding class label. In this work only box coordinate values corresponding to human/pedestrian is used.



```
0 0.458855 0.573694 0.155635 0.464552
0 0.853309 0.614739 0.203936 0.595149
```

Fig. 3: Example of dataset

B. Distance Measurement

After the detection of pedestrians from the image, YOLO return the bounding box coordinate values(x,y,w,h). Using these coordinate values, the centre point is calculated. Then the distance between every pedestrian is calculated based on the center point. The distance is scaled by the scaling factor that is estimated from the cameral caliberation. Consider the center point of two pedestrian are (x_1, y_1) and (x_2, y_2) respectively, then the distance (d) between them is calculated using the Euclidean method as

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

C. Classification

After computing the distance, the next step is classification. This system classify every pedestrians in a frame into two classes. One is people who are following the social distancing and others who are violating the social distancing. In order to decide the violations, a threshold is defined. The threshold is determined according to the cameral caliberation. In this project, the value of the threshold is given as 50 pixels. If the distance between two people is less than this threshold, then these people will classify in unsafe class, otherwise in safe class. The unsafe pair will be marked in red color and the safe pairs will be marked in green color.

$$class = \left\{ \begin{array}{ll} safe & d \geq t \\ unsafe & d < t \end{array} \right\} \tag{2}$$

IV. Result and Discussion

The image dataset used is the COCO (Common Object in Context) dataset. The COCO dataset is one of the most popular open source object recogintion database used to train deep learning programs. This database includes thousands of images with millions of already labeled objects for training An example of a dataset used for this project is shown in Fig 3. In that figure the left side is the image and the corresponding label is in the right side.

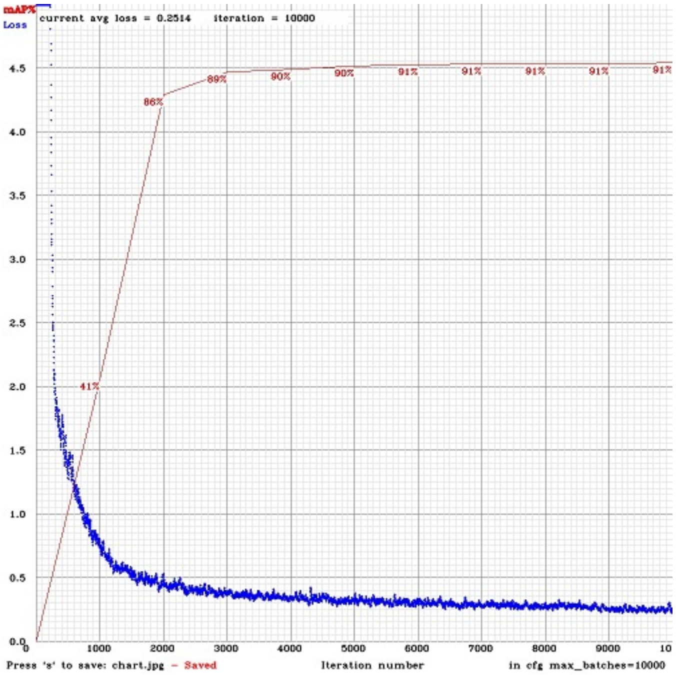


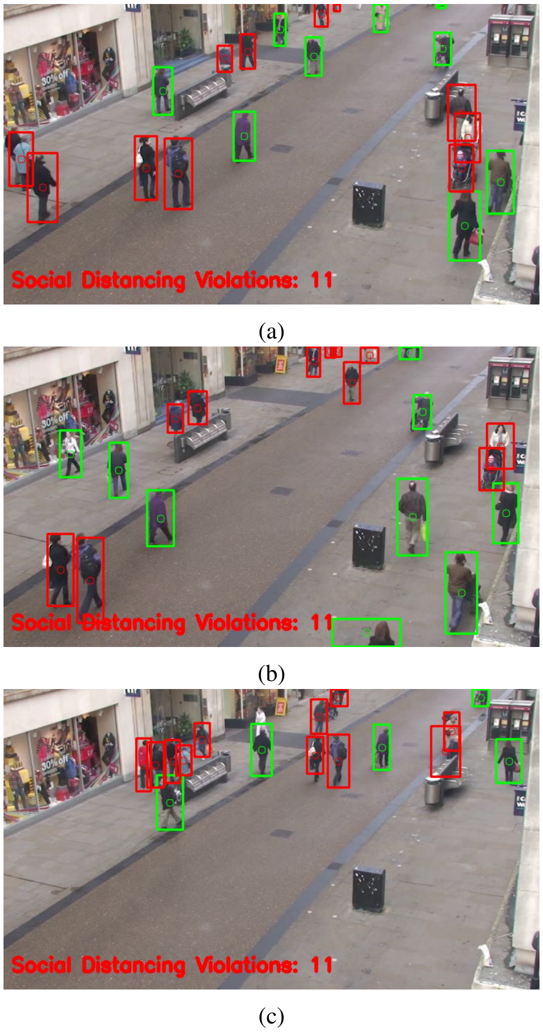
Fig. 4: Loss curve

Here a video of pedestrians walking on a public street is given as input. The camera is fixed at a specified angle. The distance is defined according to the position of camera.

Fig 5 shows the social distancing detection in a video frame.The violations are marked in red bounding boxes and safe people are marked in green bounding boxes. If any two bounding boxes overlap each other, then that will also be considered a violation.

YOLO v4 achieves a good results at a real time speed on COCO dataset [21] with 43.5% AP running at 65 FPS on Tesla V100. Fig 6 shows the comparison of YOLOv4 with other models.

Fig 4 is the loss curve of the model. The blue curve is the training loss or the error on the training dataset. The red line the mean average precision at 50% Intersection-over-Union threshold (mAP@0.5). There are 600 ground truth and prediction bounding boxes number is 485 and the number of correct predictions is 481. The precision and recall is as follows.


Social Distancing Violations: 11
Social Distancing Violations: 11
Social Distancing Violations: 11

(a)

(b)

(c)

Fig. 5: Social distance detection in video frame

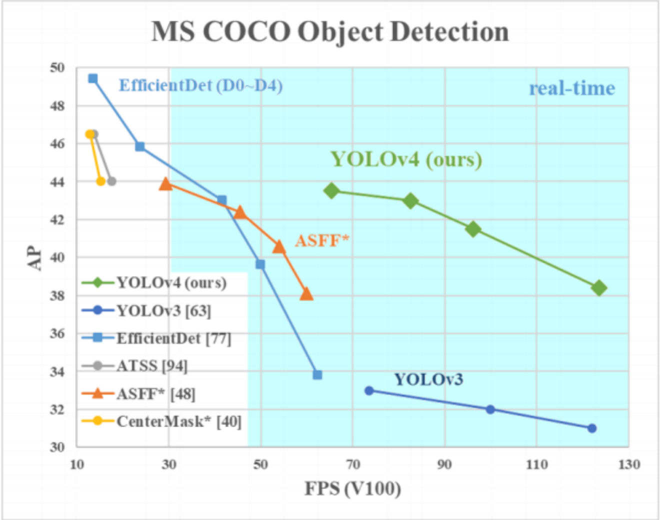


Fig. 6: FPS measured on the Tesla V100 GPU

Precision	Recall	False Positive Rate	Miss Rate
99.37%	80.48%	0.63%	19.52%

TABLE I: Validation Scores

V. Conclusion And Future Works

A methodology for social distancing detection tools using a deep learning model is proposed. By using computer vision and deep learning, the distance between people can be estimated, and any pair of people who are violating the social distancing will be indicated with a red frame. The proposed method was validated using a video showing pedestrians walking down a street. The visualisation results showed that the proposed method is capable of determining the social distancing measures between people, which can be further developed for use in other environments such as offices, restaurants, and schools.

Furthermore, the work can be further improved by optimising 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 automatic calibration of the camera.

References

[1] Heisele, Bernd, and Christian Woehler. Motion-based Recognition Of Pedestrians. *Fourteenth International Conference on Pattern Recognition*, Cat. No. 98EX170, May 1998.

[2] Lowe, David G Object recognition from local scale-invariant features *Proceedings of the seventh IEEE international conference on computer vision. Vol. 2. Ieee, 1999*

[3] Viola, Paul, and Michael Jones. Rapid Object Detection Using A Boosted Cascade Of Simple Features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*, CVPR 2001. Vol. 1. IEEE, 2001.

[4] Wu, Bo, et al. *Fast Rotation Invariant Multi-view Face Detection Based On Real Adaboost.* Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings IEEE, 2004

[5] Dalal, Navneet, and Bill Triggs. Histograms Of Oriented Gradients For Human Detection. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Vol. 1. Ieee, 2005.

[6] Felzenszwalb, Pedro F., et al. Object Detection With Discriminatively Trained Part-based Models. *IEEE transactions on pattern analysis and machine intelligence 32.9 (2009).*

[7] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks *Advances in neural information processing systems 25 (2012)*

[8] Girshick, Ross, et al. Rich Feature Hierarchies For Accurate Object Detection And Semantic Segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.*

[9] Girshick, Ross. Fast R-Cnn. *Proceedings of the IEEE international conference on computer vision*, 2015

[10] Ren, Shaoqing, et al. Faster R-cnn: Towards Real-Time Object Detection With Region Proposal Networks. *arXiv preprint arXiv:1506.01497 (2015).*

[11] Szegedy, Christian, et al. Going deeper with convolutions *Proceedings of the IEEE conference on computer vision and pattern recognition. 2015*

[12] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector *In European conference on computer vision, pp. 21-37. Springer, Cham, 2016*

[13] Lan, Wenbo, et al. Pedestrian detection based on YOLO network model. *2018 IEEE international conference on mechatronics and automation (ICMA)*, IEEE, 2018

[14] Callaway, E., Ledford, H. How bad is Omicron? What scientists know so far. *Nature, 600(7888), 197-199 2021*

[15] Shiehzadegan, S., Alaghemand, N., Fox, M., and Venketaraman, V *Analysis of the delta variant B. 1.617. 2 COVID-19. Clinics and Practice, 11(4) 2021.*

[16] 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)*

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

[18] 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.

[19] 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*

[20] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao Yolov4: Optimal speed and accuracy of object detection *arXiv preprint arXiv:2004.10934 (2020)*

[21] Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context *In European conference on computer vision, pp. 740-755. Springer, Cham, 2014*